Forecasting Hotel Daily Occupancy for High-Frequency and Complex Seasonality Data

Mr. Phoom Ungtrakul

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering in Industrial Engineering
Department of Industrial Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2018

การพยากรณ์จำนวนห้องพักรายวันของโรงแรมสำหรับข้อมูลแบบฤดูกาลเชิงซับซ้อนและความถี่สูง

นายภูมิ อึ่งตระกูล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมอุตสาหการ ภาควิชาวิศวกรรมอุตสาหการ
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2561

Thesis Title	Forecasting Hotel Daily Occupancy for High-Frequency and Complex Seasonality Data
By	Mr. Phoom Ungtrakul
Field of Study	Industrial Engineering
Thesis Advisor	Assistant Professor NARAGAIN PHUMCHUSRI, Ph.D.

Accepted by the Faculty of Engineering, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Engineering

.... Dean of the Faculty of Engineering

(Professor SUPOT TEACHAVORASINSKUN, D.Eng.)

THESIS COMMITTEE

.... Chairman

(Associate Professor WIPAWEE THARMMAPHORNPHILAS, Ph.D.)

.... Thesis Advisor

(Assistant Professor NARAGAIN PHUMCHUSRI, Ph.D.)

.... Examiner

(Amonsiri Vilasdaechanont, Ph.D.)

.... External Examiner

(Assistant Professor Nantachai Kantanantha, Ph.D.)

ภูมิ อึ่งตระกูล : การพยากรณ์จำนวนห้องพักรายวันของโรงแรมสำหรับข้อมูลแบบฤดูกาลเชิงซับซ้อนและความถี่สูง. (Forecasting Hotel Daily Occupancy for High-Frequency and Complex Seasonality Data) อ.ที่ปรึกษาหลัก : ผศ. ดร.นระเกณฑ์ พุ่มชูศรี

การพยากรณ์จำนวนห้องพักโรงแรมแบบรายวันที่แม่นยำเป็นข้อมูลที่สำคัญสำหรับการบริหารรายได้โรงแรม งานวิจัยนี้นำเสนอการพยากรณ์ทั้งตัวแบบอนุกรมเวลาและตัวแบบที่ใช้ปัจจัยภายนอกสำหรับโรงแรมกรณีศึกษาระดับ 4 ดาวในจังหวัดภูเก็ต ประเทศไทย ตัวแบบที่ใช้ประกอบด้วย Holt-Winters, Box-Jenkins, Box-Cox transformation, ARMA errors, trend and multiple seasonal patterns (BATS), Trigonometric BATS (TBATS), โครงข่ายประสาทเทียมแบบใช้ขั้นตอนการส่งค่าย้อนกลับ (BPNN), และซัพพอร์ตเวคเตอร์รีเกรสชัน (SVR) ปัจจัยที่ใช้ในตัวแบบที่ใช้ปัจจัยภายนอกในการพยากรณ์ คือ ข้อมูลแปลงค่าของโรงแรมในอดีต จำนวนนักท่องเที่ยวในภูเก็ตจากกลุ่มประเทศหลัก ราคาน้ำมัน อัตราแลกเปลี่ยน เป็นต้น การเปรียบเทียบตัวแบบพยากรณ์สำหรับข้อมูลรวมทุกห้อง (aggregated) ใช้การประเมินค่าเฉลี่ยร้อยละสัมบูรณ์ (MAPE) และการประเมินค่าเฉลี่ยร้อยละ (MAE) ผลวิจัยพบว่าตัวแบบพยากรณ์โครงข่ายประสาทเทียมแบบใช้ขั้นตอนการส่งค่าย้อนกลับให้ผลที่แม่นยำมากที่สุดที่ MAPE เท่ากับ 8.96% อย่างไรก็ตามตัวแบบพยากรณ์ซัพพอร์ตเวคเตอร์รีเกรสชันให้ผลที่แม่นยำรองลงมาที่ MAPE เท่ากับ 12.30% ไม่ได้มีความแตกต่างอย่างมีนัยสำคัญทางสถิติเมื่อเทียบกับโครงข่ายประสาทเทียมแบบใช้ขั้นตอนการส่งค่าย้อนกลับ ในขณะที่มีความแตกต่างอย่างมีนัยสำคัญทางสถิติกับตัวแบบพยากรณ์อื่นๆ แสดงให้เห็นว่าตัวแบบการเรียนรู้ของเครื่อง (Machine Learning) ในงานวิจัยนี้มีประสิทธิภาพสูงกว่าตัวแบบอนุกรมเวลาที่ออกแบบสำหรับข้อมูลแบบฤดูกาลเชิงซับซ้อนอย่าง BATS และ TBATS ทั้งนี้โรงแรมกรณีศึกษาสามารถนำผลวิจัยไปใช้วางแผนจัดการจำนวนห้องพักที่ว่างเพื่อกระจายจำนวนห้องเหลือให้แก่ตัวแทนบริษัทนำเที่ยวที่ขายราคาถูกได้อย่างมีประสิทธิผลต่อไปในอนาคต

สาขาวิชา	วิศวกรรมอุตสาหการ	ลายมือชื่อนิสิต	....
ปีการศึกษา	2561	ลายมือชื่อ อ.ที่ปรึกษาหลัก	....

6071303521 : MAJOR INDUSTRIAL ENGINEERING
KEYWORD: Forecasting, Revenue Management, High-Frequency Seasonality, BATS/TBATS, BPNN, SVR

Phoom Ungtrakul : Forecasting Hotel Daily Occupancy for High-Frequency and Complex Seasonality Data. Advisor: Asst. Prof. NARAGAIN PHUMCHUSRI, Ph.D.

Accurate hotel daily occupancy forecasting is an important input for hotel revenue management. This research presents forecasting models both time series and causal methods for a case study 4-star hotel in Phuket, Thailand. Holt-Winters, Box-Jenkins, Box-Cox transformation, ARMA errors, trend and multiple seasonal patterns (BATS), Trigonometric BATS (TBATS), back-propagation Neural Network (BPNN), and Support Vector Regression (SVR) are explored in this work, where Same Day Last Year is considered as benchmark forecast. For causal method, independent variables used as regressor inputs are transformed data observed in the past periods, the number of tourist arrivals from main countries to Phuket, Oil prices, exchange rate, etc. Accuracy for Aggregated data are tested using Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE). Findings suggested that back-propagation Neural Network (BPNN) outperforms other models with the lowest MAPE of 8.96%. However, SVR which gives second least error of 12.30% are not statistically different compared to BPNN, while all others are significantly different. It shows that Machine Learning techniques studied in this research outperform the sophisticated time series methods designed for complex seasonality data like BATS and TBATS. The results obtained can be applied to the case study hotel's future planning about the forecasted number of left-over rooms so that they effectively allocate to their discounted online travel agent (OTA) more effectively.

Field of Study:	Industrial Engineering	Student's Signature ....
Academic Year:	2018	Advisor's Signature ....

ACKNOWLEDGEMENTS

I would like to express my utmost gratitude to my dissertation major advisor who has guided, assisted and encouraged me throughout each stage of my research. Her critical and insightful comments have provided thoughtful ideas that have greatly improved the quality of the thesis. I would also like to thank Dr. Nantachai Kantanantha in the Faculty of industrial Engineerig, Kasetsart University for his helpful assistance and care given to me. Furthermore, I would like to thank you, my lovely family, particularly my friends, who have encouraged and supported my ambitions over the years.

Phoom Ungtrakul

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

Chapter I: Introduction

1.1 Hotel Industry Overview

Tourism industry is considered very important for Thailand's economy. Tourism contributes a significant amount of income to Thai GDP. In 2016 alone, tourism in Thailand generates 2.53 trillion baht in revenue, which is 17.7 percent of the country's GDP (Council). For the past few decades, Thai government has agenda to sponsor tourism by means of appointing an organization named Tourism Authority of Thailand (TAT) to oversee the entire tourism enterprise.

Phuket, an island province situated in the southern region of Thailand, in Andaman sea, dubbed 'Pearl of Andaman' has served the country as the finest tourist destination for many years. In 2005, people vote Phuket as one of the world's top five travel destinations. And, in 2017, 10 million foreign visitors traveled to Phuket and generated an estimation of 385-billion-baht accounting for nearly 14 percent of the year's total GDP (Council).

In Phuket island, there are as many as 60,000 hotel rooms scattering throughout the province. Most of them locate alongside famous beaches, namely Patong, Kata, Karon, Rawai, Nai Yang, Mai Kao, etc. Hotels that are situated next to beaches are favorite location for Russian tourists, though Chinese tourists are indifferent whether the hotel is next to the beach or not.

Hotels in Phuket welcome most of the guests from two channels: tourists that come from tour agencies, and free independent traveler (FIT). There are many tour agencies that bring tourists to Phuket. For Russian market, Pegas Tourist, Biblio Globus, Anex Tourism, etc. bridge the gap between travelers and the hotel. For Chinese market, Siam Impression Phuket, Thai Morning Sun Tour, Sunny Sea Holidays, etc. bring Chinese travelers to the hotel. For other market H.I.S. Tours brings Japanese tourists; NS Tour brings Korean tourists; Oy Aurinkomattkat Suntours brings Finnish tourists.

Free independent travelers (FIT) are more various. Chinese, Russian, Scandinavian, Australian, British, Singaporean, Korean, and Malaysian tourists can be found on the island. They usually book hotel rooms through booking.com, Agoda, Expedia, etc. These FIT tourists usually have experienced in Phuket and can speak English, so they are confident to travel by themselves.

Figure 1 shows data from Tourism Authority of Thailand, exhibiting number of total tourist arrivals in Phuket. The graph reveals strong upwards trend and seasonality, expressing long term growth of the number of tourists. In consequence, there are opportunities for hotel business growth. At the same time, the competition in this industry is also likely to be higher.

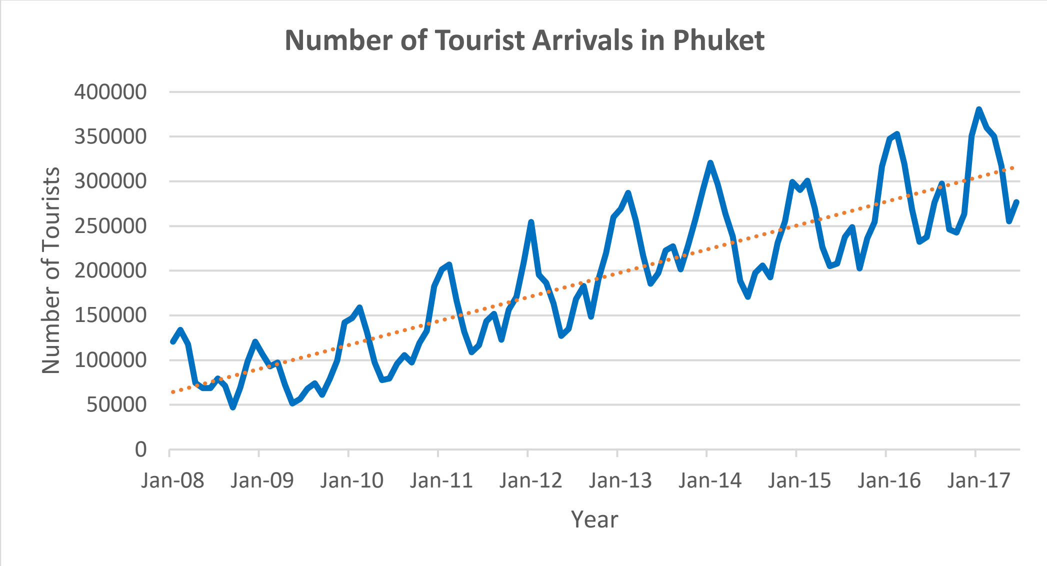

Figure 1: The number of tourist arrivals in Phuket from 2008-2017

1.2 Case-study Hotel Information

The case-study hotel has 152 rooms in total, categorized into 4 types: Deluxe; Seaview; Pool Access; and Villa. Deluxe room type has 48 rooms, constituting 32% of the total rooms. Seaview type (the largest room type) has 75 rooms, counting into 49% of total rooms. Pool Access type has 15 rooms and Villa room type has 14 rooms, counting for 10% and 9%, respectively. Thus, this thesis focuses on Deluxe and Seaview room type because the sum of these two types are 123 rooms, which are 81% of total rooms. Table 1 summarizes the number of rooms for each category. Seaview room type of the case-study hotel is shown in figure 2.

Table 1: Composition of room types of the case-study hotel

Room Type	No. of room	Percentage
Deluxe	48	32%
Seaview	75	49%
Pool Access	15	10%
Villa	14	9%
Total	152	100%

Figure 2: Seaview room type of the case-study hotel

1.3 Hotel Revenue Management Overview

In tourism industry, room rate varies across the year due to seasonality, so does in Phuket. There are 3 main seasons on the island: Low season, High season, and Peak season. Low season starts from May 1^{st} to October 31^{st}; High season from November 1^{st} to March 15^{th}; Peak season from December 20^{th} to January 20^{th}. For example, A collectible average room rate of Deluxe room type of the case-study hotel in Low, High, Chinese New Year period, and Peak season are 1,500, 2,500, 3,300, and 3,600 THB respectively, yielding in yearly average room rate at 2,051 THB per night. In these 3 seasons, variation in room rate appears so greatly. Data from the hotel shows that the highest and the lowest room rate differs more than twice. Hence, marketing team supposes to contemplate in this fact to be able to proficiently manage the hotel's revenue. Moreover, occupancy can fluctuate across seasons and within season. Though high demand persists in High and Peak season, room occupancy sometimes is not at maximum, causing some loss in net revenue.

Nowadays, management team deals with revenue management by executive insight and common practice. Room availability is filled few months in advanced. When the actual date is approaching, sometimes there are unoccupied rooms even in the high season.

1.4 Problem Statement

Though hotel management software has provided many essential functions in hotel business, forecasts provided by the program are only calculate Same Date Last Year Naïve, delivering little substantial benefits to management team. Same Date Last Year Naïve is calculated by aligning days of the latest year to the preceding year. For example, with existing data ended at year 2018, the first Monday of January 2018 was aligned with the first Monday of January 2017. Then, values in year 2017 become forecast for year 2018 and so on. Table 2 shows the forecasting results computed by the existing program. Forecasting errors are excessive in some day, yielding final error measurement (MAPE) of 73.15%. Hence, this forecasting method is not applicable to business practice.

Because of the inventory nature of hotel rooms is perishable, meaning when the room is unoccupied the hotel will lose sales opportunities when the day has passed. If the number of daily room occupancy is accurately forecasted in advance, hotel management team will be able to exercise marketing strategy such as last day promotion to attract different groups of customers via different sales channels before such total occupancy is realized.

Thus, it is beneficial to the hotel if management team can understand the pattern and accurately forecast the future daily occupancy. In addition, expected revenue in near future obtained from the forecasts enable hotel management team to plan many things, e.g., marketing expenditure in promotion for periods with low demand to attract more customers. This can provide opportunities to obtain more revenue.

Table 2: Error from hotel management software forecasting module

Date	Actual	Forecast	|Error|
1/01/2018	149	148	0.67%
2//01/2018	130	150	15.38%
3//01/2018	149	122	18.12%
…	…	…	…
21/11/2018	15	100	566.67%
22/11/2018	18	102	466.67%
23/11/2018	38	119	213.16%
…	…	…	…
29/12/2018	141	144	2.13%
30//12/2018	146	152	4.11%
31//12/2018	141	152	7.80%
		MAPE	73.15%

Furthermore, the case-study hotel situates in Phuket, amassing high competition due to nature of attractive location and government supports. Figure 3 shows graph of numbers of hotel supply in Phuket from 2016 to 2021 (Bill Barnett). The numbers of hotel supply in Phuket are expected to raise in the future.

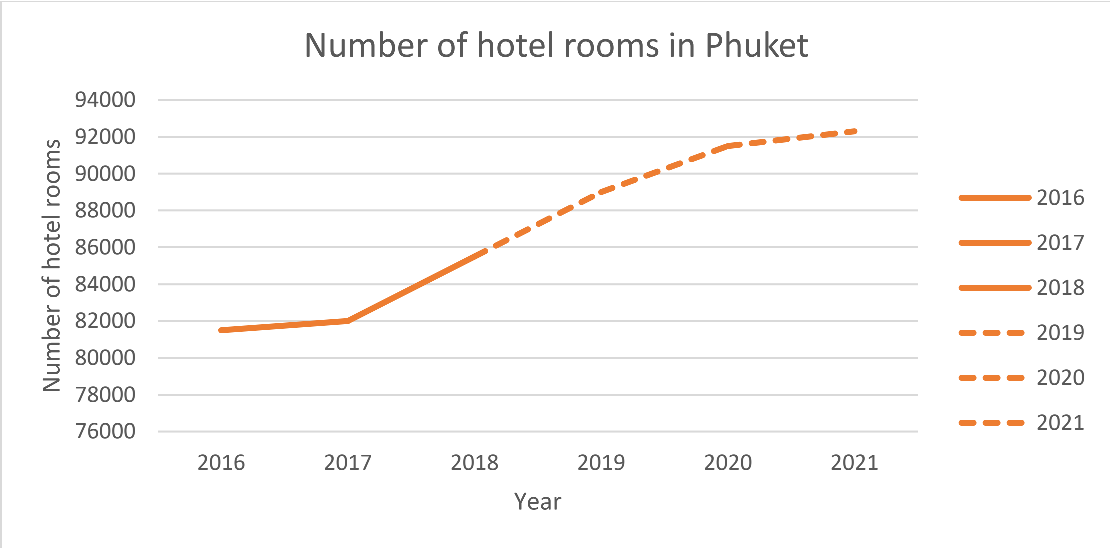

Figure 3: The number of hotel supply in Phuket (2019-2021 are expected number)

In most case, hotel daily time series comprised of high frequency and complex seasonal properties that period of seasons can also be non-integer. Daily data constituted high frequency due to high number of days in one period (365 days in a year). Besides, daily data could propagate complex seasonal pattern, having weekly seasonal pattern with a period of 7 and an annual seasonal pattern with a period of 365.25. In contrast to monthly, quarterly, and yearly data that had smaller period within single seasonal pattern, and such period can only be integer (4 for quarterly data or 12 for monthly data). The existing researches about daily hotel occupancy focused only on accuracy comparison among time series methods and did not yet compare them with other methods. In addition, the comparison of aggregated and disaggregated forecasts of hotel daily demand has not yet been explored.

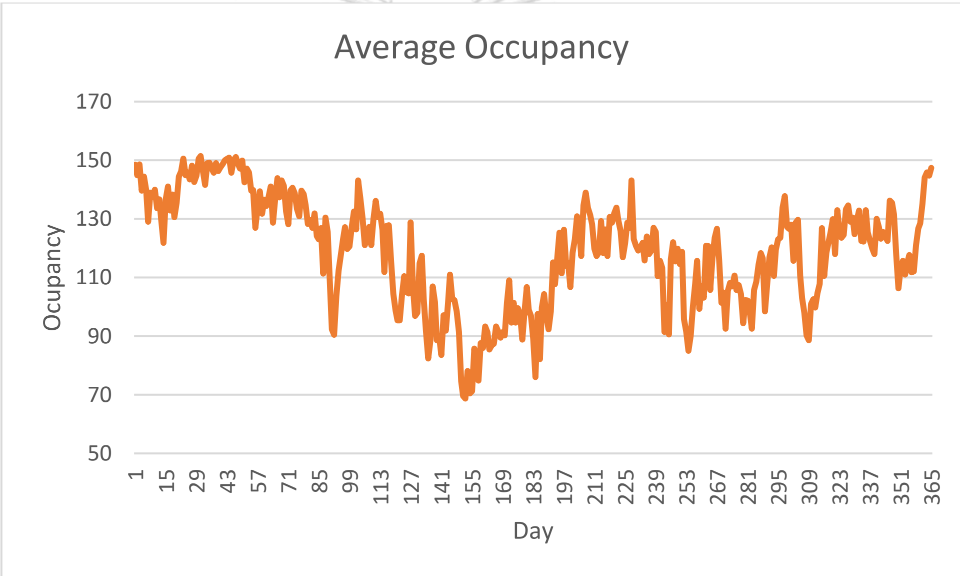

Figure 4: Graph shows daily average occupancy of case-study hotel

Graph in figure 4 shows daily average occupancy of case-study hotel. Graph peaks at the beginning and in the end, representing high occupancy around November, December, January, February, and March of every year. The lowest point of the year is around April, May, and June. This behavior corresponds to tourism seasonality of Phuket and confirms high frequency characteristic of daily data. Hence, the graph exhibits that high frequency and seasonal properties of daily demand data of hotel does exist.

So, it is interesting to compare both time series models and explanatory models for daily hotel room occupancy with high frequency and complex seasonality, and to find insights on how to obtain the accurate forecast results.

1.5 Objectives

The objective of this thesis is to find forecasting models, both time series and explanatory methods, for daily hotel room occupancy having high frequency and complex seasonality pattern. These models can provide lower errors compared to the existing method currently used and can capture informative insights for the case-study hotel management team for their future planning.

1.6 Scopes

1. This thesis focuses on both time series models, i.e. Holt-Winters' exponential smoothing, Seasonal autoregressive integrated moving average (SARIMA), Box-Cox transformation, ARMA errors, trend and multiple seasonal patterns (BATS), and Trigonometric BATS (TBATS), as well as causal models i.e. back-propagation Neural Network (BPNN), and Support Vector Regression (SVR). The examples of factor used in explanatory models are the number of tourist arrivals in Phuket,

exchange rates, Customer Price Index in Southern Region, global incidents i.e. sanction on Russia, etc.

2. The comparison between data aggregation (forecasting all room types) and disaggregation (forecasting each room type separately) are performed to explore which one provides more accurate result for the case-study hotel.

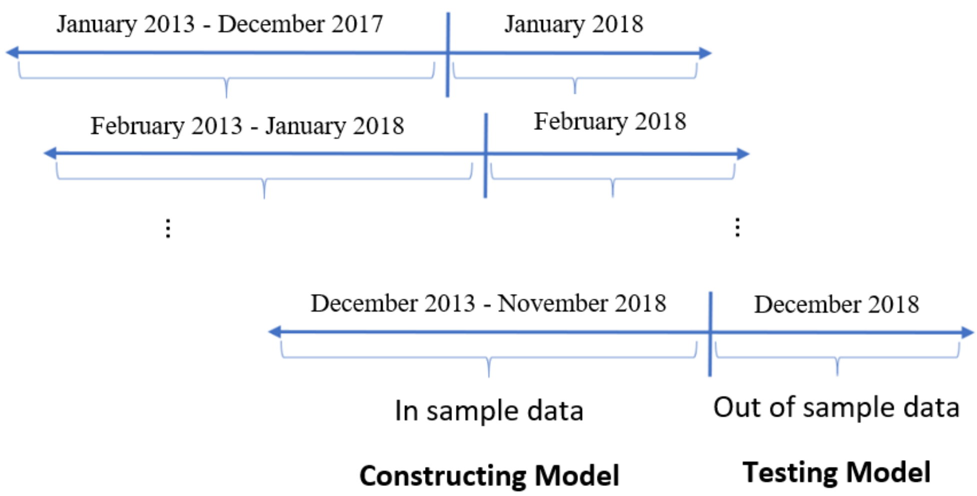

Figure 5: Data separation for model constructing and testing of forecasting horizon model

3. One-month forecasting horizon of both aggregated and disaggregated data of the most appropriate model is conducted to simulate business activity. Figure 5 shows data separation for constructing model and testing of forecasting horizon model.

4. The accuracy performance of forecasting models was evaluated in term of Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE). They are calculated by the following equations.

$$\boldsymbol{MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|} \quad (1)$$

$$\boldsymbol{MAE = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t|} \quad (2)$$

1.7 Thesis Outcomes

The outcomes of this thesis are:

1. Forecasting model with improved accuracy for daily hotel room occupancy with high frequency and complex seasonality for the case-study hotel.
2. Comparison of results when data is aggregated versus disaggregated by category.
3. One-month forecasting horizon of both aggregated and disaggregated data.

1.8 Benefits of this Thesis

Several benefits can be acquired from this thesis, as the following:

1. The result can provide management team with reliable forecasting method on daily hotel room occupancy both in total and in each category.

2. The insights can assist management team in comprehension of trend, seasonality pattern, and affecting variables that explain the data pattern.

.

Chapter II: Literature Review

2.1 Related Theory

In revenue management system, there are many issues to consider in forecasting. (Weatherford and Kimes 2003) discussed these issues in detail. First, we considered what to forecast; arrivals or occupancy. Arrivals are future demand which may not be fulfilled due to cancellation, but occupancy is demand in the past that already occurred.

Second, we investigated level of aggregation. In hotel industry, mostly there are different types of room, for example: deluxe, sea view, mountain view, city view, pool access, family suite, etc. These types are categorized for marketing purpose and target segmentation. Otherwise, types of room can also be classified into rate categories, which assign each of the type by monetary value. Plus, level of aggregation can be defined with length of stay, or the total night spent in the hotel for each booking.

Thus, data can be disaggregated into 'Rate Categories' (RC), or 'Length of stay' (LoS), and even can be fully disaggregated by RC and LoS. In the aggregated end, data can be fully aggregated by combining all data in every rate category into one single number. Or, aggregated forecast is executed, then it is disaggregated by probability distribution.

Third, we explored selection of data included in the forecasting model, which are number of periods, types of data, and outlier. Firstly, number of periods were considered whether we include all data points available, or just some selected number of periods. Secondly, we considered types of data, which are whether we include only completed stay nights, or completed and incomplete stay nights. Lastly, outlier was determined whether to be included or not included in the model.

2.1.1 What to forecast?

(Zakhary et al. 2011) introduced Monte Carlo simulation for approximation of arrivals by occupancy data to forecast hotel room demand in Egypt. Historical data or occupancy was used to quantify simulation model's parameters. Then, arrivals, cancellation, length of stay, trend, and seasonality were computed from the parameter estimation.

(Chen and Kachani 2007) proposed optimization model by network flow formulation. Authors used both arrivals and occupancy data on forecasting side, employing, classical and advanced pickup method, simple exponential smoothing, and combined methods (pickup and exponential smoothing).

(Weatherford and Kimes 2003) presented the use of daily arrivals data to forecast. Arrivals data from Choice hotels was used to form a suitable forecasting method. Then, Arrivals data from Marriot hotels was used to verify the findings.

(Pereira 2016) introducing new useful forecasting method for high frequency daily data, used occupancy data from 300-room Portuguese four-star hotel.

2.1.2 Level of Aggregation

(Baker and Collier 1999) showed that most of hotels nowadays collect customer spending at different rate category for marketing purposes. For example, Marriot and Hyatt kept customer profile according to amount of money spent on amenity, food, and meeting room usage.

(Weatherford, Kimes, and Scott 2001), studying forecast accuracy on 4 different levels of aggregation with 2-year daily arrival data by rate category and length of stay, reported that disaggregated forecast highly outperforms every aggregated forecast. Therefore, (Weatherford, Kimes, and Scott 2001) suggested that hotel should gather arrivals data by rate category and length of stay. But, if disaggregation on both rate category and length of stay are not possible, the hotel should at least collect data by rate category or the type of room.

(Pereira 2016) gathered fully disaggregated data on room type from Portuguese four-star hotel including classic rooms, deluxe rooms, and suites. However, the author chose to use aggregated data to show new forecasting technique on high frequency data in large amount (up to 1500 days).

2.1.3 Data Selection

(Rajopadhye et al. 2001) introduced forecasting technique on unconstrained room demand. This meant that cancellation of the booking is also reflected in this model. Thus, it was imperative to use actual booking activity that includes both completed and incomplete stay nights. Weekly data was collected from both occupancy and arrivals.

(Chen and Kachani 2007) recommended that using very outdated data could influence dynamicity of the forecasting model, and thus rendered model unresponsive to predictions. (Chen and Kachani 2007) selected 4, 8, and 12 weeks of processed data for the testing set, which constitutes both partial and full data history.

(Weatherford and Kimes 2003) speculated that “Hotels using revenue management typically update forecasts on a daily basis for occupancy dates in the near future (1–2 weeks) and update on a weekly basis for dates farther away (2–8 weeks)”. (Koupriouchina, van der Rest, and Schwartz 2014) also supported this fact, stating that revenue management system recalculates the same date forecasts many times as the actual date approaches. And, (Koupriouchina, van der Rest, and Schwartz 2014) affirmed that it is more beneficial to management team to have advance knowledge about hotel room occupancy because there is time to implement strategies to correct the situation. On the other hand, if the time of such insight is only few days in advance, nothing much but inventory control could be implemented.

2.2 Related Researches

2.2.1 Researches in Thailand

In tourism industry, there were many attempts on forecasting different type of time series data. Many past researches in Thailand focused on forecasting number of tourists. (Keerativibool 2013) examined appropriate forecasting model for number of international tourist arrivals to Thailand. Utilizing 72 monthly data from Tourism Authority of Thailand (TAT), author employed Box-Jenkins' ARIMA, Multiplicative Winters' exponential smoothing, decomposition, and combined method.

(Saothayanun 2014) compared two forecasting models on number of international tourist arrivals to Thailand. Those proposing models were Box-Jenkins' ARIMA, and Winters' exponential smoothing. Root Mean Square Error (RMSE) and Mean Absolute Percent Error (MAPE) were used to measure forecast accuracy. Results showed that forecast accuracy reaches around 10% of the error, in which Winters' exponential smoothing yields lower RMSE and MAPE in every case.

(Rungjindarat 2016) predicted number of Russian tourist arrivals to Thailand at Suvarnabhumi airport with SARIMA model. Sixty-six-month data from

January 2010 to June 2015 were used to estimate the forecast. Accuracy of the forecasts was measured by Root Mean Square Error (RMSE) and Mean Absolute Percent Error (MAPE). Result revealed that number of Russian tourist arrivals to Thailand is affected by both trend and seasonality.

In conclusion, researchers applied Moving Average, decomposition, Holt-Winters' exponential smoothing, Box-Jenkins' ARIMA, SARIMA, and combined method. There was no single dominant forecasting model for all cases that yields highest accuracy, rather it depended on characteristics of time series data such as trend, seasonality, etc.

2.2.2 Hotel Demand Forecasting

2.2.2.1 Time Series

In hotel industry side, (Rajopadhye et al. 2001) simulated booking curve for unconstrained room demand with daily data by combined forecast method. The forecasting model combined short term forecast (STF), and long term forecast (LTF). STF was carried out using booking profile, and LTF was carried out using Holt-Winters' exponential smoothing.

(El Gayar et al. 2011) proposed incorporation of group reservation bookings and integrality constraints into forecasting model using decomposition method to extract components such as reservations, cancellations, duration of stay, no shows, seasonality, trend, arrivals, occupancy rates, etc. Then, the model generated actual process of reservations to receive forecasts of each future components. Benefits of such model were density of the forecasts and confidence intervals can be produced, allowance for scenario analysis such as effect of overbooking, and estimation of sensitivity of the arrival forecast due to changes of some control variables.

Many researches on forecasting often review monthly, quarterly, or annual data. (Mukma 2018) studied monthly, quarterly, and annual data of number of Chinese tourist arrivals to Thailand during 1985 to 2016. Authors employed three forecasting methods: Box-Jenkins' ARIMA, Brown's double exponential smoothing, and regression analysis. Accuracy of the forecast was measured by Root Mean Square Error (RMSE) and Mean Absolute Percent Error (MAPE). As a result, Brown's double exponential smoothing outperformed every method in every case (monthly, quarterly, and annual data).

For more specific studies using hotel room data, (Lim, Chang, and McAleer 2009) used hotel-motel monthly occupancy in New Zealand from January 1997 to December 2006 to calculate 1-month ex post forecast for 12 periods. Then this forecast was compared with occupancy data in year 2007 to quantify forecasting accuracy.

Furthermore, (Haensel and Koole 2011) studied forecasting model on combination of regions, day of week arrivals, and length of stays. With weekly data, authors used Additive Holt-Winters, and Multivariate Vector Autoregression to forecast accrued booking curve, plus expected number of total reservations in each day. Then, forecasts were adjusted to the earlier observations with penalized least squares and historical proportion methods, which was one of the updating procedures. Consequently, average accuracy of Vector Autoregression bettered Additive Holt-Winters method, meaning that correlation between base vectors (early and late bookings) must be deliberated to be able to form an accurate forecasting model.

In case of daily data, it is critical to consider this separately. (Pereira 2016) advised that daily time series are dissimilar to monthly, quarterly, or annual data because daily time series present high frequency and complex seasonal patterns. In most case, daily time series comprised of double seasonal effect that period of seasons can be non-integer. Usually, daily data had weekly seasonal pattern with a period of 7 and an annual seasonal pattern with a period of 365.25. Therefore, conventional time series forecasting method, such as exponential smoothing or ARIMA/SARIMA, were not appropriate for forecasting high-frequency and complexed time series. Those conventional methods were modeled for smaller period within single seasonal pattern that such period can only be integer (4 for quarterly data or 12 for monthly data).

Hence, (Pereira 2016) introduced a novel forecasting method on daily hotel data called **B**ox-Cox Transformation, **A**RMA Errors, **T**rend and Multiple **S**easonal Patterns (BATS), **T**rigonometric BATS (TBATS), and Double Seasonal Holt-Winters, which are extended from Holt-Winters' exponential smoothing. (Pereira 2016) used these mathematical techniques on total 1517 points of daily room demand. As a result, MAPE accuracy measurement ranged around 10%.

BATS and TBATS were proposed by (De Livera 2010b) and (De Livera, Hyndman, and Snyder 2011) in 2010 and 2011, applying these forecasting methods to weekly number of barrels of motor gasoline product supply in the United States from February 1991 to July 2005. This series had seasonality length of 52.179 (365.25/7: yearly/weekly) with 745 observations in total, 484 observations for training period and 261 for testing set. In the same study, (De Livera, Hyndman, and Snyder 2011) also tested this forecast on number of call arrivals handled on weekdays between 7am and 9:05pm from March 3^{rd} to May 23^{rd}. Time series data for this test included 10,140 observations or 12 weeks of data, containing daily seasonal pattern with period of 169 and weekly seasonal pattern with period of 845 (169*5). Finally, (De Livera, Hyndman, and Snyder 2011) performed this method on Turkey electricity demand. This time series exhibited dual-calendar effect due to the Seker (also known as Eid ul-Fitr) and

Kurban (also known as Eid al-Adha) following Hijri calendar, and to national holidays which follows Gregorian calendar.

For other examples use of BATS and TBATS formulations, (Ha, Bianchi, and R. 2017) experimented Short-Term Load Forecasting (STLF) with 40 electricity usage time series collected from US and Norway at different levels of aggregation. (Ha, Bianchi, and R. 2017) chose five suitable models to perform forecasting on up to 4 years of hourly electricity load, presenting a strong seasonal pattern of high demand for electricity during winter and low demand in summer. These five models consisted of ARIMA, Holt-Winters exponential smoothing, Nonlinear Non-Autoregressive Regression, TBATS, and Semi-Parametric Additive model.

(Huang et al. 2017) studied forecast accuracy on monthly tourism demand on several European countries i.e. Austria, Cyprus, Germany, Greece, Netherlands, Portugal, Spain, Sweden and the United Kingdom. Authors employed nine alternative parametric and non-parametric models on data from January 2000 until December 2013. These models were Moving Average, Weighted Moving Average, ARIMA, Exponential Smoothing, Neural Networks, TBATS, ARFIMA, recurrent SSA, and vector SSA.

2.2.2.2 Machine Learning

(Hong, Thong, and Tam 2006) studied monthly revenue per unit (RPU) forecasts for DTAG car rental company in United States. Data from different airport locations dated from January 2002 to September 2004. Training data set was 20 months, validation data set was 5 months, and testing data set was 8 months. Authors implemented immune algorithm that selected parameters for Support Vectors Regression (SVR) yielding the lowest Normalized Root Mean Square Error (NRMSE).

(Claveria, Monte, and Torra 2015) investigated monthly number of international tourist arrivals of 5 origins (France, the United Kingdom, Belgium, the Netherlands, and Germany) to Catalonia from January 2009 to July 2012. This study evaluated the forecasting performance of three different Artificial Neural Network (ANN) models in a multivariate setting based on multiple-input multiple-output (MIMO) structures. Multi-layer perceptron neural network (MLP), Radial basis function neural network (RBF), and Elman Neural Network were ANN models in competition. Several unit roots tests such as ADF test (Dickey and Fuller), PP test (Phillips and Perron), and the KPSS test (Kwiatkowski et al.) were conducted, leading to a fact that tourist arrivals were multicointegrated. Thus, multivariate multiple-output neural network approach was deemed appropriated. Root Mean Squared Error (RMSE) was used to measure forecasting accuracy of the models.

As a result, MLP and RBF networks showed lower RMSE values than Elman networks.

(Martinez-de Pison et al. 2016) compared Grid search (GS) and Genetic Algorithm (GA) on nine regression models which are Model Tree-based on Quinlan's M5 algorithm (M5P), Instance-based learning (IBL) method, Multilayer perceptron (MLP) neural network, Support vector regression (SVR), Extreme machine learning (ELM) algorithm, Locally weighted learning for linear regression (LWLLIN), Random forest (RF) algorithm, Extreme Gradient Boosting (XGB) algorithm, and linear ridge regression (LIN) as the reference prediction. Forecasting performance of daily historical booking records of the Spanish hotel for six years (2004–2010) were observed. Regressors also included a list of local, regional, and national festivities, and other additional information such as the day of the week, the local weather conditions, some sociological information, several indicators comprising the macro-economic situation of the area, and independent factors affecting the hotel room demand provided by the Spanish National Institute of Statistics. Root Mean Squared Error (RMSE) and Mean Squared Error (MAE) were computed to compare forecasting accuracy, plus computation time for each regression model with Grid search and Genetic Algorithm was shown in comparisons.

(Urraca et al. 2015) optimized a Knowledge Discovery in Databases (KDD) scheme using genetic algorithms. Booking data were obtained from a hotel located in a small village of La Rioja region in northern Spain Six years of data were used to train and validate the models, while a test set was created from the latest year during January and July. Also, 119 attributes of macro-economic indicators, temporal situation, social patterns, meteorological data, and local and regional holidays were selected by experts. Then, these indicators were decreased to 22 by detecting dependencies via scatterplots and matrix correlation. Forecasting accuracies were computed by Root Mean Squared Error (RMSE) and Generalized Degrees of Freedom (GDF) was used for the selection of parsimonious models. Results presented the three best performing models were those that included feature selection (FS).

(Cao and Tay 2001) applied Support Vector Regression to daily stock price forecasts in comparison with back-propagation (BP) neural network and the regularized radial basis function (RBF) neural network. Standard & Poor 500 stock index futures (CME-SP), United Sates 30-year government bond (CBOT-US), Unite States 10-year government bond (CBOT-BO), German 10-year government bond (EUREX-BUND), and French government stock index futures (MATIF-CAC40) were used in this research. Five-day lagged periods (RDP-5, RDP-10, RDP-15, and RDP-20) and one transformed closing price (EMA100) were calculated as regressor variables based on Relative Difference in Percentage of Price (RDP) and Exponential Moving Average (EMA). This research compared forecasting results of multiple machine learning algorithms such as Support Vector Regression (SVR), Back Propagation Neural Network (BP-ANN), and Regularized Redial Basis Function (RBF) Neural Network. Accuracy of the forecasts are measured by Normalized Mean Squared Error (NMSE), Mean Absolute Error (MAE), and directional symmetry (DS), in which DS provides an indication of the correctness of the predicted direction.

(Thomason 1999) suggested that forecasting horizon for daily stock price should be long enough to compensate the over-trading resulting in excessive transaction costs, but in forecasting aspect, forecasting horizon should be short because information hidden in financial time series existed in limited duration. Therefore, (Thomason 1999) recommended that regressors for daily stock price should be converted into 4 lagged Relative Difference in Percentage of Price (RDP), that this transformed data will become more symmetrical and will follow more closely to a normal distribution. To illustrate, RDP-5 is calculated by $RDP_5 = \frac{A_t - A_{t-5}}{A_{t-5}} * 100$, where A_t is the actual observation at time t.

(Thomason 1999) also proposed that Exponential Moving Average (EMA) should be included in the regressor variables as well because RDP transformation of time series data could remove some useful information hidden in the data.

Thus, EMA is added to the input variables to maintain such information. (Thomason 1999) emphasized that exponential moving average is implemented not normal moving average because exponent functions would highlight more weight on recent observations. It is imperative for daily stock price prediction that most of insight in the financial data existed in the most recent closing prices. EMA100 is calculated by subtracting a 100-day exponential moving average from the closing price. The subtraction is performed to eliminate the trend in price as the maximum value and the minimum value is in the ratio of about 2:1 in data sets. However, this speculation could be indifferent to other types of time series.

2.2.3 Forecasting Accuracy Measures

(Koupriouchina, van der Rest, and Schwartz 2014) presented enhanced view of forecasting accuracy measurement. Authors described in detail about each measurement properties. Accuracy measurers were defined into two main categories: scale-dependent forecasting accuracy measures and scale-independent forecasting accuracy measures.

Scale-dependent measurers comprised of Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Error (ME), Mean Absolute Error (MAE), and Median Absolute Error (MdAE). These measures are dependent because scale of themselves depend on scale of the data used in forecasts, that is the larger the value of elements in forecasting, the larger the quantity of error is possible. For that reason, scale-dependent measurers are not fitted for comparing across different set of data with different scales.

Scale-independent measurers are more resilient because they do not depend on the scale of the data in forecast. Thus, scale-independent measurers are suitable for comparing data error across data with different size and across multiple forecasting model. However, one immense drawback of scale-independent measurers is that denominator of the measurers cannot be zero. Since these measures are based on percentage error, which percentage error is given by $\frac{(Y_t - F_t)}{Y_t} * 100$ where Y_t and F_t are actual value and forecast value respectively. To overcome this drawback researcher may disregard those value that has actual value of zero, use scale-dependent measurers, or apply Theil's Inequality Coefficient (denoted by U) instead.

Scale-independent measurers included Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE), Median Absolute Percentage Error (MdAPE), Symmetric Mean Absolute Percentage Error (sMAPE), Symmetric Median Absolute Percentage Error (SMdAPE), Mean Relative Absolute Error (MRAE), Median Relative Absolute Error (MdRAE), Geometric Mean Relative Absolute Error (GMRAE), Mean Absolute Scaled Error (MASE), Root Mean Square Percentage Error (RMSPE), and Root Median Square Percentage Error (RMdSPE).

Most of researchers in Thailand used MAE and MAPE to measure accuracy. (Saothayanun 2014); (Mukma 2018); and (Rungjindarat 2016) also used Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) to measure forecast accuracy. However, (Keerativibool 2013) used Correlation Coefficient (R) to measure forecast accuracy.

For other uses of accuracy measurement, (Weatherford and Kimes 2003) applied Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) to calculate accuracy of the forecast. (Haensel and Koole 2011) used Total Relative Absolute Errors (TRAE), Mean Squared error (MSE), and Mean Relative Absolute Error (MRAE). (Chen and Kachani 2007) implemented Theil's U, and Mean Percentage Error (MPE). (El Gayar et al. 2011) used Mean Squared Error (MSE), and the Mean Relative Absolute Error (MRAE).

For different use of central tendency measurer such as median, (Pereira 2016) applied Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Median Absolute Percentage Error (MdAPE), Symmetric Mean Absolute Percentage Error (sMAPE), Mean Relative Absolute Error (MRAE), and Median Relative Absolute Error (MdRAE). (Weatherford, Kimes, and Scott 2001) used Mean Absolute Deviation (MAD) or also known as Mean Absolute Error (MAE), and Median Relative Absolute Error (RAE).

One outstanding benefit of MAPE accuracy measure is the scale-independency. MAPE is widely used by researchers because it is easy to compare (both within data set and across data sets), and to be interpreted by universal users (Weatherford and Kimes 2003); (Pereira 2016); (Saothayanun 2014); (Mukma 2018); and (Rungjindarat 2016). However, the worst drawback of this measure is the denominator (actual value) cannot be zero since it is undefined. Moreover, (Koupriouchina, van der Rest, and Schwartz 2014) suggested that MAPE cannot be directly compared with Naïve 1 (random walk) or Naïve 2 (de-seasonalized random walk) models. MAPE is calculated by the following equation.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{3}$$

where; A_t is actual observation, F_t is forecasted value, and n is total number of observations

MAE is also popular among researchers due to its easiness to calculate and easiness to understand by non-specialists (Weatherford and Kimes 2003); (Weatherford, Kimes, and Scott 2001); (Saothayanun 2014); (Mukma 2018); and (Rungjindarat 2016). Absolute operation in this measure ensures that error is meaningful because, without absolute operation error can equal zero, which cannot be determined whether either error is zero, or another case that positive and negative error completely offsets each other. MAE is calculated by the following equation

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t| \quad (4)$$

where; A_t is actual observation, F_t is forecasted value, and n is total number of observations

In addition, Randomized Complete Block Design (RCBD) are performed in this research to provide a measure to test differences in means between forecasting models. If p-value of the test is greater than 0.05, the forecasting models in comparison have no statistically significant differences. Every model is test in pair with Tukey LSD test to identify whether models are significantly different.

2.3 Summary of Hotel Revenue Management Forecasting

The summary of literature review regarding hotel revenue management is showed in table 3.

Table 3: Summary of literature review regarding hotel revenue management

Authors	Variables used	Data frequency	Forecasting model
Rajopadhye et al., 2001	Unconstrained hotel room demand	Weekly	Holt-Winters method
Weatherford et al., 2001	Fully aggregated: by length of stay with rate-category Disaggregated (by rate category with length of stay)	Daily	Classical-pickup method, moving averages, linear regression, simple exponential smoothing, and random walk
Baker et al., 2002	Rate class, arrival date, length of stay, product	Daily (simulated)	Forecasting-allocation approach

	class, stay overnight		
Weatherford and Kimes, 2003	Fully aggregated: by length of stay with rate-category Fully disaggregated (by rate category with length of stay)	Weekly	Classical pickup, advanced pickup, multiplicative, and regression
Vu and Turner, 2006	Guest arrivals at accommodation establishments in 9 cities	Monthly	Box-Jenkins SARIMA, and BSM models (decomposition, trend, seasonal, cycle, irregular)
Chen and Kachani, 2007	Demand distribution for each night, for each rate, and for each length of stay	Weekly	Classical & advanced pickup, linear regression, simple exponential smoothing, advanced pickup exponential smoothing, advanced pickup + exponential smoothing
Lim et al., 2009	Guest night demand	Monthly	Holt-Winters exponential smoothing, and Box-Jenkins ARMA Models
El Gayar et al., 2011	Arrivals, reservation data, folio history, occupancy rates	Daily	Advanced room demand forecast model and an optimization mode

Haensel and Koole, 2011	Combinations of region, arrival day of week, and length of stay	Weekly	Additive Holt-Winters, and the vector autoregressive forecasts
Zakhary et al., 2011	Daily hotel arrivals and hotel occupancy	Daily/weekly	Monte Carlo simulation system
Pereira, 2016	Daily hotel room demand time series	Daily	Exponential smoothing approach
Martinez-de Pison et al., 2016	Daily historical booking records of the Spanish hotel and independent variables: list of festivities; day of the week; local weather conditions; sociological information; and local macro-economic indicators	Daily	Compared Grid search (GS) and Genetic Algorithm (GA) on nine regression models
Urraca et al., 2017	Booking data of La Rioja region in northern Spain and independent variables: macro-economic indicators; temporal situation; social patterns; meteorological data; and regional holidays	Monthly	KDD scheme with GA
This Thesis	Aggregated and disaggregated daily hotel room demand time series	Daily	Moving Average, Holt-Winters, SARIMA, BATS, TBATS, Artificial Neural Network, and Support Vector Regression

There were many foreign researches regarding hotel revenue management but for Thailand there was only one (Vu and Turner 2006), studying monthly data on guest arrivals at accommodation establishments in 9 cities for Thailand from 1996 to 2002. Moreover, there were very limited researches in all academic world on forecasting daily hotel occupancy using machine learning. (Urraca et al. 2015) used monthly hotel data and various regressors to study forecasting accuracy. (Martinez-de Pison et al. 2016) composed a study similarly to (Urraca et al. 2015) but with daily hotel data.

The most related research to this paper is (Pereira 2016); and (Martinez-de Pison et al. 2016), examining properties of daily hotel data. (Pereira 2016) applied new time series forecasting approach to tackle with high frequencies and multiple seasonality patterns, that persist in hotel daily data. Still, (Pereira 2016) did not explore forecasts on disaggregated data, in which this paper will provide. (Martinez-de Pison et al. 2016) conducted Machine Learning algorithm on hotel data, employing independent variables such as list of festivities; day of the week; local weather conditions; sociological information; and local macro-economic indicators, 119 of such variables were provided by experts. Then, these indicators were deducted to 22 by detecting dependencies via scatterplots and matrix correlation.

However, it was deemed heavy burdens for practitioners to forecast using numerous independent variables, since workloads should be impinged on practitioners to select and refine those variables. Therefore, this paper aimed to study hotel revenue management on forecasting daily room occupancy with both time series models and regression models, and to investigate the use transformation of time series data as inputs for regression models, comparing with independent variables.

Chapter III: Methodology

3.1 Data in this Research

Research methodology is presented in this section. First, time series data is collected from the hotel management software. These time series data include both aggregated and disaggregated by categories. Second, data is analyzed and used to construct forecasting model. Third, forecasting results are compared across all time series data set. Finally, interpretation and recommendation are provided for management use.

3.1.1 Hotel Occupancy

Comanche program has stored various type of hotel's data, including room occupancy. Room occupancy on this program can be retrieved with ST_26 function (Figure 6).

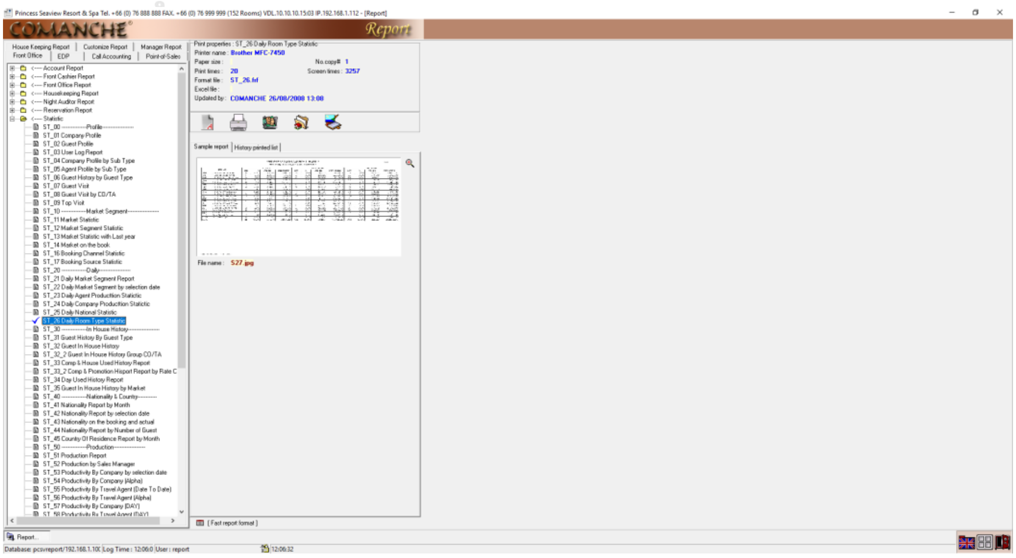

Figure 6: ST_26 function in Comanche program

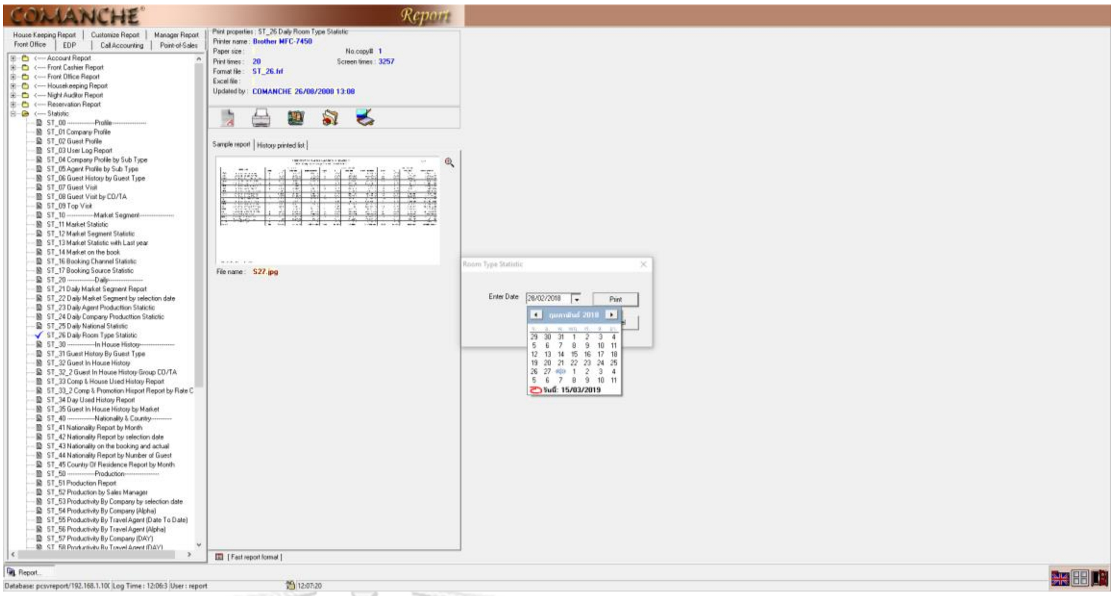

Figure 7: Accessing ST_26 function in Comanche program

ST_26 Daily Room Type Statistic On 01/01/2018 Page #		1											
Printed On : 09/09/2018 Time : 05:45:01													
T O D A Y	Month-To-Date	Year-To-Date											
Room Type	Rms	%	Avg.Rate	Room Revenue	Rms	%	Avg.Rate	Room Revenue	Rms	%	Avg.Rate	Room Revenue	
PDD	PRINCESS DELUXE DOUBLE ROOM	20	13.42%	1,941.38	38,827.56	20	0.13%	1,941.38	38,827.56	20	0.13%	1,941.38	38,827.56
PDS	PRINCESS DELUXE DOUBLE SEAVIEW ROOM	25	16.78%	2,213.05	55,326.27	25	0.17%	2,213.05	55,326.27	25	0.17%	2,213.05	55,326.27
PPD	PRINCESS DELUXE POOL ACCESS DOUBLE ROOM	4	2.68%	2,812.24	11,248.96	4	0.03%	2,812.24	11,248.96	4	0.03%	2,812.24	11,248.96
PDT	PRINCESS DELUXE TWIN ROOM	28	19.00%	2,227.82	62,378.95	28	0.20%	2,227.82	62,378.95	28	0.19%	2,227.82	62,378.95
PTS	PRINCESS DELUXE TWIN SEAVIEW ROOM	50	33.56%	2,163.19	108,159.64	50	0.34%	2,163.19	108,159.64	50	0.34%	2,163.19	108,159.64
PFD	PRINCESS FAMILY SUITE DOUBLE ROOM	2	0.00%	1,699.24	3,398.48	2	0.00%	1,699.24	3,398.48	2	0.01%	1,699.24	3,398.48
PFT	PRINCESS FAMILY SUITE TWIN ROOM	12	8.05%	1,916.74	23,000.90	12	0.08%	1,916.74	23,000.90	12	0.08%	1,916.74	23,000.90
PPT	PRINCRESS DELUXE POOL ACCESS TWIN ROOM	8	5.37%	2,901.21	23,209.64	8	0.05%	2,901.21	23,209.64	8	0.05%	2,901.21	23,209.64
DMF	DUMMY (PERMANENT MASTER)	0	0.00%	0.00	0.00	0	0.00%	0.00	0.00	0	0.00%	0.00	0.00
TOTAL	**149**	**98.03%**	**2,184.90**	**325,550.40**	**149**	**98.03%**	**2,184.90**	**325,550.40**	**149**	**98.03%**	**2,184.90**	**325,550.40**	

Figure 8: Example of .html file exported from Comanche program

01/02/2018	Page #	1											
Printed On : 09/09/2018 Time : 06:09:55													
T O D A Y	Month-To-Date	To-											
Room Type	Rms	%	e	Revenue	Rms	%	e	Revenue	Rms	%	e	Revenue	
PDD	PRINCESS DELUXE DOUBLE ROOM	20	13.16%	1,359.38	27,187.70	20	0.13%	1,359.38	27,187.70	598	0.13%	1,561.79	933,951.87
PDS	PRINCESS DELUXE DOUBLE SEAVIEW ROOM	25	16.45%	1,598.81	39,970.22	25	0.16%	1,598.81	39,970.22	782	0.17%	1,835.07	1,435,022.46
PPD	PRINCESS DELUXE POOL ACCESS DOUBLE ROOM	5	3.29%	1,778.42	8,892.11	5	0.03%	1,778.42	8,892.11	152	0.03%	2,189.15	332,751.43
PDT	PRINCESS DELUXE TWIN ROOM	28	18.42%	1,463.52	40,978.65	28	0.18%	1,463.52	40,978.65	870	0.19%	1,736.71	1,510,936.16
PTS	PRINCESS DELUXE TWIN SEAVIEW ROOM	50	32.89%	1,544.44	77,221.76	50	0.33%	1,544.44	77,221.76	1578	0.34%	1,764.12	2,783,782.54
PFD	PRINCESS FAMILY SUITE DOUBLE ROOM	2	1.32%	1,189.46	2,378.92	2	0.01%	1,189.46	2,378.92	56	0.01%	1,545.24	86,533.61
PFT	PRINCESS FAMILY SUITE TWIN ROOM	12	7.89%	1,366.47	16,397.64	12	0.08%	1,366.47	16,397.64	352	0.07%	1,586.07	558,297.64
PPT	PRINCRESS DELUXE POOL ACCESS TWIN ROOM	10	6.58%	1,469.84	14,698.40	10	0.07%	1,469.84	14,698.40	308	0.07%	2,018.30	621,636.22
DMF	DUMMY (PERMANENT MASTER)	0	0.00%	0	1,784.20	0	0.00%	0	1,784.20	0	0.00%	0	5,097.71
TOTAL	**152**	**100.00%**	**1,509.93**	**229,509.60**	**152**	**######**	**1,509.93**	**229,509.60**	**4696**	**######**	**1,760.65**	**8,268,009.64**	

Figure 9: Example of Microsoft Excel table for daily occupancy

Comanche stores room occupancy data by date. Daily data can be retrieved by accessing ST_26 function, inputting the desired date to this box (Figure 7) to retrieve daily room occupancy data. The most convenient method to take such data is to export this ST_26 function into .html file. In Today section, total room occupancy is displayed within red circle (Figure 8). Also, disaggregated occupancies in each room category are also displayed in this function. Then, it can be easily copied to Microsoft Excel (Figure 9).

3.1.2 Outlier

Room occupancy data storage started since the opening of the hotel on September 2011. However, data in year 2011 was considered outliers because it did not have full year of data.

Thus, after excluding outlier, monthly data has 84 instances (from January 2012-December 2018). And, daily data has 2557 points (from January, 1^{st} 2012-December, 31^{st}2018).

3.1.3 Data Transformation

Time series models forecasting uses original data retrieved from hotel's database, but data for machine learning models includes data transformed into Relative Difference in Percentage (RDP), and Moving Average (MA). Relative Difference in Percentage (RDP) is calculated by obtaining percentage increase from the period before. To illustrate, RDP-7 is calculated by $RDP_7 = \frac{A_t - A_{t-7}}{A_{t-7}} *$

100, where A_t is the actual observation at time t. Four lags of RDP are computed for use in machine learning, which are RDP-7, RDP-30, RDP-52, RDP-365. Another transformed variable is MA-360, generated by subtracting 360-day exponential moving averages from the occupancy ($MA_{360} = y_t - \frac{y_{t-1}+\cdots+y_{t-360+1}}{360}$). Table 4 shows transformation values of daily hotel occupancy time series.

Table 4: Example of transformed data used in forecasting models

	Occupancy	MA360	RDB-7	RDB-14	RDB-21	RDB-28
1/1/2014	149	21.57	-81.71	-119.12	1.97	-23.14
1/2/2014	148	20.61	-23.33	-174.07	-1.37	-12.12
1/3/2014	149	21.47	1.32	-181.13	-21.14	1.97
1/4/2014	149	21.16	1.97	-238.64	-46.08	1.32
1/5/2014	134	5.76	6.29	-243.59	-42.55	8.84
1/6/2014	151	22.51	-4.14	-164.91	-75.58	-9.42
1/7/2014	149	20.49	0.00	-132.81	-56.84	-12.88
1/8/2014	152	23.21	-2.70	-85.37	-123.53	0.00
1/9/2014	134	5.24	10.07	-11.67	-148.15	8.22
1/10/2014	150	21.24	-0.67	0.66	-183.02	-21.95
1/11/2014	144	15.22	-7.46	5.26	-227.27	-41.18
1/12/2014	142	13.26	5.96	0.70	-264.10	-51.06
1/13/2014	133	4.36	10.74	8.28	-133.33	-54.65

1/14/2014	144	15.02	5.26	3.36	-125.00	-51.58
1/15/2014	150	20.87	-11.94	-1.35	-82.93	-120.59
1/16/2014	152	22.77	-1.33	-2.01	-26.67	-181.48
1/17/2014	150	20.77	-4.17	-0.67	0.66	-183.02
1/18/2014	150	20.79	-5.63	-11.94	1.32	-240.91
1/19/2014	149	19.31	-12.03	1.32	-4.20	-282.05
1/20/2014	133	3.26	7.64	10.74	8.28	-133.33
1/21/2014	150	20.08	0.00	1.32	-0.67	-134.38
1/22/2014	151	21.08	0.66	-12.69	-2.03	-84.15
1/23/2014	152	22.04	-1.33	-1.33	-2.01	-26.67
1/24/2014	152	21.96	-1.33	-5.56	-2.01	-0.66
1/25/2014	152	21.94	-2.01	-7.04	-13.43	0.00
1/26/2014	152	21.94	-14.29	-14.29	-0.66	-6.29
1/27/2014	150	19.76	0.00	-4.17	-0.67	-3.45
1/28/2014	152	21.73	-0.66	-1.33	0.00	-2.01
1/29/2014	141	10.86	7.24	7.24	-5.22	4.73
1/30/2014	149	18.53	1.97	0.67	0.67	0.00
1/31/2014	152	21.19	0.00	-1.33	-5.56	-2.01
1/8/2014	152	23.21	-2.70	-85.37	-123.53	0.00
2/1/2014	151	19.74	0.66	-1.34	-6.34	-12.69

2/2/2014	150	18.38	0.00	-12.78	-12.78	0.66
2/3/2014	152	20.38	0.00	-1.33	-5.56	-2.01
2/4/2014	152	19.98	-7.80	-0.66	-1.33	0.00
2/5/2014	151	18.58	-1.34	0.66	0.66	-12.69
2/6/2014	152	18.96	0.00	0.00	-1.33	-1.33
2/7/2014	151	17.36	0.00	0.66	-0.67	-4.86
2/8/2014	152	17.79	-1.33	0.00	-2.01	-7.04
2/9/2014	152	17.68	0.00	-1.33	-14.29	-14.29
2/10/2014	149	14.71	1.97	1.97	0.67	-3.47
2/11/2014	147	12.21	2.65	-4.26	2.65	2.00
2/12/2014	152	16.57	0.00	-2.01	0.00	0.00
2/13/2014	152	16.30	-0.66	0.00	0.00	-1.33
2/14/2014	152	16.30	0.00	-0.66	0.00	-1.33
2/15/2014	151	15.23	0.66	-0.67	0.66	-1.34
2/16/2014	151	15.21	-1.34	0.66	-0.67	-13.53
2/17/2014	152	16.17	-3.40	0.00	0.00	-1.33
2/18/2014	135	-0.71	11.18	10.60	4.26	10.60
2/19/2014	152	16.29	0.00	0.00	-2.01	0.00
2/20/2014	150	14.31	1.32	0.66	1.32	1.32
2/21/2014	150	14.32	0.66	1.32	0.66	1.32

2/22/2014	142	6.16	5.96	6.58	5.33	6.58
2/23/2014	136	0.33	10.53	8.72	10.53	9.33
2/24/2014	127	-8.40	5.93	13.61	16.45	16.45
2/25/2014	110	-24.94	27.63	27.63	27.15	21.99
2/26/2014	134	-0.74	10.67	11.84	11.84	10.07
2/27/2014	146	11.32	2.67	3.95	3.31	3.95
...	...	...	...	...	...	...

3.1.4 Independent Variables

Various attributes of macro-economic indicators, temporal situation, number of tourist Arrivals in Phuket, days of the week, month, season, weather conditions are used as regressor variables in machine learning. Table 5 shows all variables used in this study.

To analyze, Number of Tourist Arrivals in Phuket are selected because, they would have direct effect on number of guests staying in the hotel. Economic indicators such as Customer Price Index (CPI), oil price, exchange rates can indicate purchasing power of tourists. For example, if CPI increases, living cost of tourists in the region would then increase, thus lowering the number of guests and occupancy of the hotel. When oil price increases, it would influence transportation expense, thus lowering the number of guests and occupancy of the hotel. Average room rate is selected according to demand-supply law: the higher the price the lower the demand. Exchange rates also dictate purchasing power of tourists. If exchange rates over Thai Baht of any country devalues, purchasing power of such country would plummet and thus lowering the number of guests and occupancy of the hotel. Dummy variables such as days of the week, month of the year, season, and crisis would help explaining data patter in daily time series, that other independent variables might not be able to.

Table 5: Independent variables used in machine learning models

Category	Sub-category	Frequency	Remarks
Number of Tourist Arrivals in Phuket by Country of Residence	China Russia Australia Korea Malaysia Singapore Sweden United Kingdom	Monthly	- Comprise of 80% of total number of tourist arrivals. - Tourism Authority of Thailand
Customer Price Index of Southern Region of Thailand (CPISouth)		Monthly	Bureau of Trade and Economic Indices

Oil Price		Daily	Crude Oil WTI Futures (CLM9) in USD
Exchange Rates over Thai Baht	United Stated (USD) China (CHY) Russia (RUB) Australia (AUD) Korea (KRW) Malaysia (MYR) Singapore (SGD) Sweden (SEK) United Kingdom (GBP)	Daily	Daily quotes from Bank of Thailand
Average Room Rate (ARR)		Daily	Hotel Management Program
Days of the Week	Monday Tuesday Thursday Friday Saturday	Dummy	- Exclude Wednesday - Wednesday is expected to be the lowest day of the week

	Sunday		
Month of the Year	January February March April June July August September October November December	Dummy	- Exclude May - May is the lowest month of the year
Monsoon		Dummy	Monsoon season: March - October
Season	Peak High	Dummy	Exclude Low season Peak: December 20^{th} to January 20^{th} High: November 1^{st} to March 15^{th} Low: May 1^{st} to

			October 31^{st}
RusSanc		Dummy	Sanction on Russia: March 2014 - present
RDP	RDP-7 RDP-30 RDP-52 RDP-365	Daily	Transformed data
MA-360		Daily	Transformed data
			Total: 46 Variables

3.2 Forecasting Model

3.2.1 Exponential Smoothing (Holt-Winters)

(Holt 1957) and (Winters 1960) incorporated seasonality fraction into Holt's method. The Holt-Winters seasonal method included one forecast equation, and three smoothing equations. These three smoothing equations are level (m_t), trend (b_t), and seasonal (c_t) with $\alpha, \beta,$ and γ as smoothing parameters. Holt-Winters' exponential smoothing consist of two types: additive and multiplicative. Holt-Winters' multiplicative exponential smoothing is selected for use in this research because amplitudes of the seasonal variations are dependent of the data level.

Equations 5 to 9 show Holt-Winters' multiplicative exponential smoothing equations (Holt 1957) and (Winters 1960).

$$y_t \ = \ (m_t + b_t t)c_t + \epsilon_t, \tag{5}$$

$$\widehat{m}_t \ = \ \alpha \frac{y_t}{\hat{c}_{t-s}} + (1-\alpha)\big(\widehat{m}_{t-1} + \hat{b}_{t-1}\big), \tag{6}$$

$$\hat{b}_t \ = \ \beta(\widehat{m}_t - \widehat{m}_{t-1}) + (1-\beta)\hat{b}_{t-1}, \tag{7}$$

$$\hat{c}_t \ = \ \gamma \frac{y_t}{\widehat{m}_t} + (1-\gamma)\hat{c}_{t-s}, \tag{8}$$

$$\hat{y}_{t+r} = \big(\widehat{m}_t - \hat{b}_t \tau\big)\hat{c}_{t+\tau-s}. \tag{9}$$

where

m_t is mean component that gives the level of the time series at time t,

b_t is trend component that indicates direction in which the series is evolving,

c_t is the seasonal component that indicates the periodic variations in the level of the series,

ϵ_t is random error,

$\alpha, \beta, and\, \gamma$ are smoothing constants for the base, trend, and seasonal components respectively,

$\hat{y}_{t+r}$ is forecast for future time τ.

The following figure shows code in Rstudio computing Holt-Winters exponential smoothing (Figure 10).

```
library(forecast)
library(TSPred)
library(ggplot2)

#Importing the data
dailyData <- read.csv("C:\\users\\ph156\\Desktop\\Forecast2018\\DailyForecast\\DailyOcc2015.csv")
dailyData
dim(as.matrix(dailyData))

ts.data = ts(dailyData$Occupancy, frequency = 365, start = c(2015,1), end = c(2018,365))
plot(ts.data, ylim = c(0,160))
dim(as.matrix(ts.data))
summary(ts.data)

#Training and testing data set
data.train = window(ts.data, frequency = 365, start = c(2015,1), end = c(2018,60))
plot(data.train, ylim = c(0,160))
dim(as.matrix(data.train))

data.test = window(ts.data, frequency = 365, start = c(2018,61), end = c(2018,365))
plot(data.test, ylim = c(0,160))
dim(as.matrix(data.test))

Y4DailyHW1 <- HoltWinters(data.train) # default additive
Y4DailyHW2 <- HoltWinters(data.train, seasonal = c("multiplicative"))

#Plot fitted values
plot.ts(ts.data ,ylab="Occupancy", ylim=c(0,195))
lines(Y4DailyHW1$fitted[,1], lty=2, col="blue")
lines(Y4DailyHW2$fitted[,1], lty=2, col="red")

#Predict
Y4DailyHW1.pred <- predict(Y4DailyHW1, 305, prediction.interval = TRUE)
Y4DailyHW2.pred <- predict(Y4DailyHW2, 305, prediction.interval = TRUE)

#Plot prediction
plot.ts(ts.data, ylab="Occupancy", xlim=c(2015,2019), ylim=c(0,225))
lines(Y4DailyHW1$fitted[,1], lty=2, col="red")
lines(Y4DailyHW1.pred[,1], col="blue")
lines(Y4DailyHW1.pred[,2], col="seagreen",lty=2)
lines(Y4DailyHW1.pred[,3], col="seagreen",lty=2)

plot.ts(ts.data, ylab="Occupancy", xlim=c(2015,2019), ylim=c(0,225))
lines(Y4DailyHW2$fitted[,1], lty=2, col="red")
lines(Y4DailyHW2.pred[,1], col="blue")
lines(Y4DailyHW2.pred[,2], col="seagreen",lty=2)
lines(Y4DailyHW2.pred[,3], col="seagreen",lty=2)

summary(Y4DailyHW1.pred)
Y4DailyHW1.pred
Y4DailyHW2.pred
accuracy(Y4DailyHW1.pred, data.test)
accuracy(Y4DailyHW2.pred, data.test)
```

Figure 10: Holt-Winters exponential smoothing computation code in Rstudio

3.2.1 SARIMA

Autoregressive Integrated Moving Average (ARIMA) forecasting model was proposed by (Box and Jenkins 1970) to predict time series. The method started with identify whether the time series is stationary with autocorrelation function (ACF), and partial autocorrelation function (PACF). If the time series was not stationary by mean, differencing is executed to the time series, but if the time series was not stationary by variance, time series is transformed with square root or natural logarithm. Next, maximum likelihood estimation or non-linear least-squares estimation would be performed on the time series to estimate variables. Then, the predicted model would be tested for stationary univariate process, that is residuals should be constant over time and be independent of each other. This process was check by Ljung–Box test or plotting autocorrelation and partial autocorrelation of the residuals.

SARIMA (Seasonal Autoregressive Integrated Moving Average) is an extended version of ARIMA. As is in the name, this SARIMA model incorporated seasonal effect, in contrast to ARIMA that incorporates only trend. To be more specific, ARIMA required time series that is either not seasonal or has the seasonal components eliminated. A time series $\{X_t | t = 1,2, \ldots, N\}$ is estimated by SARIMA(p,d,q)(P,D,Q) by the following equation (Box and Jenkins 1970).

$$\phi_p(B)\, \Phi_P(B^s)(1 - B)^d(1 - B^s)^D Y_t = \delta + \theta_q(B)\, \Theta_Q(B^s)\varepsilon_t \tag{10}$$

where N is the number of observations, p, d, q, P, D and Q are integers; s is the seasonal period length,

$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ is the regular autoregressive operator (AR) of order p,

$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}$ is the seasonal autoregressive operator (AR) of order P,

$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_p B^q$ is the regular moving average operator (MA) of order q,

$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_P B^{Qs}$ is the seasonal moving average operator (MA) of order Q,

$\delta = \mu\phi_p(B)\, \Phi_P(B^s)$ is constant, in which μ is average of stationary time series,

B is Backward Operator, in which $B^s Y_t = Y_{t-s}$

d is the number of regular differences,

D is the number of seasonal differences,

ε_t is the estimated residual at time t that is identically and independently distributed as a normal random variable with a mean equal to zero and a constant variance.

Figure 11 shows code in Rstudio computing SARIMA model.

```
library(uroot)
library(astsa)
library(fracdiff)
library(MASS)
library(tseries)
library(forecast)

#Importing the data
dailyData <- read.csv("C:\\users\\ph156\\Desktop\\Forecast2018\\DailyForecast\\DailyOcc2015.csv")
dailyData
dim(as.matrix(dailyData))

ts.data = ts(dailyData$Occupancy, frequency = 365, start = c(2015,1), end = c(2018,365))
plot(ts.data, ylim = c(0,160))
dim(as.matrix(ts.data))
summary(ts.data)

#Training and testing data set
data.train = ts(dailyData$Occupancy, frequency = 365, start = c(2015,1), end = c(2018,60))
plot(data.train, ylim = c(0,160))
dim(as.matrix(data.train))

data.test = ts(dailyData$Occupancy, frequency = 365, start = c(2018,61), end = c(2018,365))
plot(data.test, ylim = c(0,160))
dim(as.matrix(data.test))

arima = auto.arima(data.train, trace=TRUE, d = 0, D = 0, max.p = 5, max.q = 5, max.P = 1, max.Q = 1,
                    max.order = 6, max.d = 1, max.D = 1, start.p = 0, start.q = 0, start.P = 0, start.Q = 0,
                    stepwise = TRUE, nmodels = 10000000, seasonal.test = c("hegy"), test = "kpss", ic = "aic")

library(forecast)
summary(arima)
confint(arima)

#Residual diagnosis
plot.ts(arima$residuals)
Box.test(arima$residuals, lag = 20, type = "Ljung-Box")
acf(arima$residuals, lag.max = 24, main = "ACF of the model")
Box.test(arima$residuals^2, lag = 20, type = "Ljung-Box")

library(tseries)
jarque.bera.test(arima$residuals)

arima.forecast = forecast(arima, h=305)
plot(arima.forecast, xlab = "years", ylab = "occupancy", ylim = c(0,160))

library(TSPred)
library(forecast)
plotarimapred(data.test, arima, xlim = c(2015,2019), range.percent = 0.05)
accuracy(arima.forecast, data.test)

acf(arima$residuals, lag.max = 24, main = "ACF 511,010")
pacf(arima$residuals, lag.max = 24, main = "ACF 511,010")
```

Figure 11: SARIMA computation code in Rstudio

3.2.2 BATS/TBATS

This original forecasting method was presented by (De Livera 2010b) and (De Livera, Hyndman, and Snyder 2011), projecting forecasts on weekly number of barrels of motor gasoline product supply in the United States, number of call arrivals handled on weekdays, and (Pereira 2016) introduced this forecasting technique on 4-year daily hotel data of a 4-star Portuguese hotel.

BATS model was an extended version of double seasonal Holt Winters, integrated with Box-Cox Transformation to handle with non-linear data, and with ARMA model to account for autocorrelation in time series by residuals. (De Livera 2010b) demonstrated that BATS model enhances prediction accuracy compared to simple State Space Models. Nevertheless, BATS model still did not perform satisfactorily when seasonality is complex and with high frequency. Subsequently, (De Livera, Hyndman, and Snyder 2011) proposed TBATS (Trigonometric BATS) model to incorporate trigonometric functions into the BATS model. TBATS model could decrease model parameters and gives flexibility to manage complexity of seasonality with high frequency. Thus, TBATS could manipulate data with non-integer seasonal period, non-nested periods, and high frequency data.

First, double seasonal Holt-Winters (DSHW) exponential smoothing equation with additive trend and additive seasonality is shown below. This model (equation 11 to 15), developed by (J.W. 2003), was an extension of the Holt-Winters exponential smoothing (Holt 1957) and (Winters 1960).

$$l_t = \alpha\left(y_t - s^{(1)}_{t-m_1} - s^{(2)}_{t-m_2}\right) + (1-\alpha)(l_{t-1} + b_{t-1}), \quad (11)$$

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}, \quad (12)$$

$$s^{(1)}_t = \gamma\left(y_t - l_t - s^{(2)}_{t-m2}\right) + (1-\gamma)s^{(1)}_{t-m_1}, \quad (13)$$

$$s^{(2)}_t = \delta\left(y_t - l_t - s^{(1)}_{t-m_1}\right) + (1-\delta)s^{(2)}_{t-m_2}, \quad (14)$$

$$\hat{y}_t(h) = l_t + hb_t + s^{(1)}_{t-m_1+h} + s^{(2)}_{t-m_2+h} + \phi^h[y_t - (l_{t-1} + b_{t-1} + s^{(1)}_{t-m_1} + s^{(2)}_{t-m_2})] \quad (15)$$

where

l_t and b_t are the smoothed level and trend in period t respectively,

$s^{(1)}_t$ is the seasonal component for the short cycle,

$s^{(2)}_t$ t is the seasonal component for the long seasonal cycle,

m_1, m_2 is the lengths of the shorter and longer seasonal cycles respectively,

$\hat{y}_t(h)$ is the h step-ahead forecast made from forecast origin t,

$\phi^h[y_t - (l_{t-1} + b_{t-1} + s^{(1)}_{t-m_1} + s^{(2)}_{t-m_2})]$ is a simple adjustment for first order autocorrelation,

$\alpha, \beta, \gamma, and\ \delta$ are smoothing parameters.

The following equations shows the extension of double seasonal Holt-Winters (DSHW). Box-Cox transformation, ARMA errors, trend and multiple seasonal patterns (BATS) is expressed by equation 16 to 21 (De Livera 2010b).

$$y_t^{(\omega)} = \begin{cases} \frac{y_t^{\omega}-1}{\omega}, \omega \neq 0 \\ log y_t, \omega = 0 \end{cases} \tag{16}$$

$$y_t^{(\omega)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{T} s_{t-m_i}^{(i)} + d_t, \tag{17}$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t, \tag{18}$$

$$b_t = (1-\phi)b + \phi b_{t-1} + \beta d_t, \tag{19}$$

$$s_t^{(i)} = s_{t-m_i}^{(i)} + \gamma_i d_t, \tag{20}$$

$$d_t = \sum_{i=1}^{p} \varphi d_{t-i} + \sum_{i=1}^{p} \theta_i \varepsilon_{t-i} + \varepsilon_t, \tag{21}$$

where

$m_1, \dots, m_T$ is the periods of T seasonal patterns,

l_t is the local level in period t,

b is the long-run trend,

b_t is the short-run trend in period t,

$s_t^{(i)}$ is the i^{th} seasonal component at time t $(t = 1, \dots, n; i = 1, \dots, T)$,

d_t is $ARMA(p, q)$ process,

ε_t is Gaussian white-noise error term with zero mean and constant variance,

ϕ is the damping constant of the trend,

$\alpha, \beta, and\ \gamma_i$ are smoothing parameters.

The following equations shows the extension of BATS model by replacing equation (17), and (20) respectively, with the following expressions. This conclusion is the TBATS model (equation 22 to 25) (De Livera, Hyndman, and Snyder 2011).

$$y_t^{(\omega)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{T} s_{t-1}^{(i)} + d_t, \tag{22}$$

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}, \tag{23}$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + \gamma_1^{(i)} d_t, \tag{24}$$

$$s_{j,t}^{*(i)} = -s_{j,t-1} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t. \tag{25}$$

where

k_i is the number of harmonics required for the i^{th} seasonal component,

$\gamma_1^{(i)}, \gamma_2^{(i)}$ are smoothing parameters,

and $\lambda_j^{(i)} = \frac{2\pi j}{m_i}$.

BATS and TBATS forecasting model were estimated using these following steps.

Step 1: Specification of all available model combinations which are to be considered for each series

In the BATS modeling framework, a total of 24 models was available for consideration of each series. This framework consisted of 16 model combinations considering each B, A, T, S component and 8 additional models considering a damped trend component. For example, $\omega = 1$ is considered as having no Box-Cox transformation, $\phi = 1$ as having no damping component, p = q = 0 as having no ARMA residual adjustment in the model.

TBATS model formulation was relatively straightforward, the seed states of state space models were usually treated as random vectors. Given trial values of the unknown parameters, the joint steady state distributions of stationary states were derived, and then assigned to associated seed states.

Step 2: Estimation of the models

The initial states x_0, the smoothing parameters, the Box-Cox parameter, the damping parameter and the coefficients for the ARMA components were estimated using an appropriate estimation criterion. Three different estimation criteria were considered for non-linear optimization as follows:

(1) Maximize the log likelihood of the estimates (MLE)

(2) Minimize the Root Mean Square Error of the original data (RMSE)

(3) Minimize the Root Mean Square Error of the transformed data (RMSE_T)

Step 3: Selection of the best of the available models

Akaike information criterion (AIC) was used for choosing between the models. The following ARMA fitting approaches are explored.

(1) Setting $\{p = 0, q = 0\}$ assumed that an ARMA residual adjustment is not necessary.

(2) Finding the values for p and q in all possible ARMA combinations up to $p = q = 5$ were considered, and the ARMA(p,q) combination which minimizes the AIC was chosen, or retrieved the values for p and q in a stepwise procedure.

For TBATS model, number of harmonics was selected by constantly adding harmonics, testing the significance of each one using F-tests.

Step 4: Generation of prediction distributions using the best model.

The following figure shows code in Rstudio computing BATS/TBATS (Figure 12).

```r
library(forecast)
library(TSPred)
library(ggplot2)

#Importing the data
dailyData3 <- read.csv("C:\\users\\ph156\\Desktop\\Forecast2018\\DailyForecast\\DailyOcc2015.csv")
dailyData3
dim(as.matrix(dailyData3))

ts.data3 = ts(dailyData3$Occupancy, frequency = 365, start = c(2015,1), end = c(2018,365))
plot(ts.data3, ylim = c(0,160))
dim(as.matrix(ts.data3))
summary(ts.data3)

#Training and testing data set
data.train3 = ts(dailyData3$Occupancy, frequency = 365, start = c(2015,1), end = c(2018,60))
plot(data.train3, ylim = c(0,160))
dim(as.matrix(data.train3))

data.test3 = ts(dailyData3$Occupancy, frequency = 365, start = c(2018,61), end = c(2018,365))
plot(data.test3, ylim = c(0,160))
dim(as.matrix(data.test3))

#Estimate BATS/TBATS
BATS3 <- bats(data.train3, use.parallel = TRUE, biasadj = TRUE, trace = TRUE)
TBATS3 <- tbats(data.train3, use.parallel = TRUE, biasadj = TRUE, trace = TRUE)

#Plot fitted values
plot.ts(ts.data3 ,ylab="Occupancy", ylim=c(0,170), xlim=c(2012,2019))
lines(BATS3$fitted[], lty=2, col="red")
plot.ts(ts.data3 ,ylab="Occupancy", ylim=c(0,170), xlim=c(2012,2019))
lines(TBATS3$fitted[], lty=2, col="red")

library(forecast)
#Predict
BATS3.pred <- forecast(BATS3, h = 305,level = c(80, 95), prediction.interval = TRUE)
TBATS3.pred <- forecast(TBATS3, h = 305,level = c(80, 95), prediction.interval = TRUE)

#Plot prediction
plot.ts(ts.data3, ylab="Occupancy", xlim=c(2015,2019), ylim=c(0,220), lwd = 2)
lines(BATS3$fitted, lty=2, col="red", lwd = 2.5)
lines(BATS3.pred$mean, col="blue")
lines(BATS3.pred$level, col="seagreen", lwd = 5)

plot.ts(ts.data3, ylab="Occupancy", xlim=c(2015,2019), ylim=c(0,180))
lines(TBATS3$fitted, lty=2, col="red")
lines(TBATS3.pred$mean, col="blue", lwd = 5)
lines(TBATS3.pred$level, col="seagreen", lwd = 5)

library(forecast)
accuracy(BATS3.pred, data.test3)
accuracy(TBATS3.pred, data.test3)
```

Figure 12: BATS/TBATS computation code in Rstudio

3.2.3 Linear Regression

Linear Regression is a statistical model discovering the relationship between a response variable (or dependent variable), and one or more explanatory variable (or independent variable). The model fits linear line with the method of least squares, suggesting that the error from the model should be minimum distance measured from a linear line generated from the model. Regression assumptions are restriction of this regression model that relationships between a response and one or more explanatory variable is linear, and random error terms are independent with normal distribution of mean zero and constant variance. Normal probability plot will approve normality assumptions if the plot resembles straight line, sloping upward to the right. Equal variance assumptions are valid if the residual versus fitted value plot should appear as random scatter around zero mean.

The following figure shows code in Rstudio computing Linear Regression (figure 13).

```
library(tseries)

data <- read.csv("C:\\users\\ph156\\Documents\\Regression\\RegressorsAll.csv")
data1 <- read.csv("C:\\users\\ph156\\Documents\\Regression\\OccData7-12-28-30-31-52-360-364-365-366,EMA30-60-90-180-270-36

data.train = cbind(data[1:1826,2:45], data1[1:1826,3:18])
data.test = cbind(data[1827:2191,2:45], data1[1827:2191,3:18])
Naive <- data$Occupancy[1826:2190]

summary(data)

formular = Occupancy ~ China + Russia + Australia + Korea + Malaysia + Singapore + Sweden + UK + CPISouth + OilPrice +
  USD + CNY + RUB + AUD + KRW + MYR + SGD + SEK + GBP + ARR + Thursday + Friday + Saturday + Sunday + Monday + Tuesday +
  June + July + August + September + October + November + December + January + Febuary + March + April + Monsoon +
  Peak + High + RusSanc + ChiBoatSink + ChiFreeVisa + Occ.EMA30 + Occ.EMA60 + Occ.EMA90 + Occ.EMA180 + Occ.EMA270 +
  Occ.EMA360 + RDB.7 + RDB.12 + RDB.28 + RDB.30 + RDB.31 + RDB.52 + RDB.360 + RDB.364 + RDB.365 + RDB.366

#fitting multiple linear regression
lm.model <- lm(formular, data = data.train)
summary(lm.model)
lm.pred <- predict(lm.model, data.test, method = "qr")
lm <- data.matrix(lm.pred)

#predict test data set
predictions <- lm
actualValues <- data.test$Occupancy

MAPE = sum(abs(actualValues-predictions)/actualValues)/nrow(data.test)*100
MAPE

MAE = sum(abs(actualValues-predictions))/nrow(data.test)
MAE

#Naive1
MAENaive = sum(abs(actualValues-Naive))/nrow(data.test)
MAENaive

MASE = MAE/MAENaive
MASE
```

Figure 13: Linear Regression computation code in Rstudio

3.2.4 Neural Network

Artificial Neural Network is one of the non-linear statistical machine learning models. It is used to find complex relationships or patterns in data between inputs and outputs. There are 2 major types of Neural Network, dividing by connecting patterns of the hidden layers: feed-forward networks and recurrent networks. Feed-forward neural network passes values only from input node to output node (only one direction), but the data flow bidirectional in recurrent networks. Many types of Neural Network models were proposed for examination such as Multilayer Perceptron (MLP), Bayesian Neural Network (BNN), Radial Basis Function Neural Network (RBF), Generalized Regression Neural Network (GRNN), Elman Network, etc. (Claveria, Monte, and Torra 2015); and (Ahmed et al. 2009). Figure 14 is an example structure of Neural Network model.

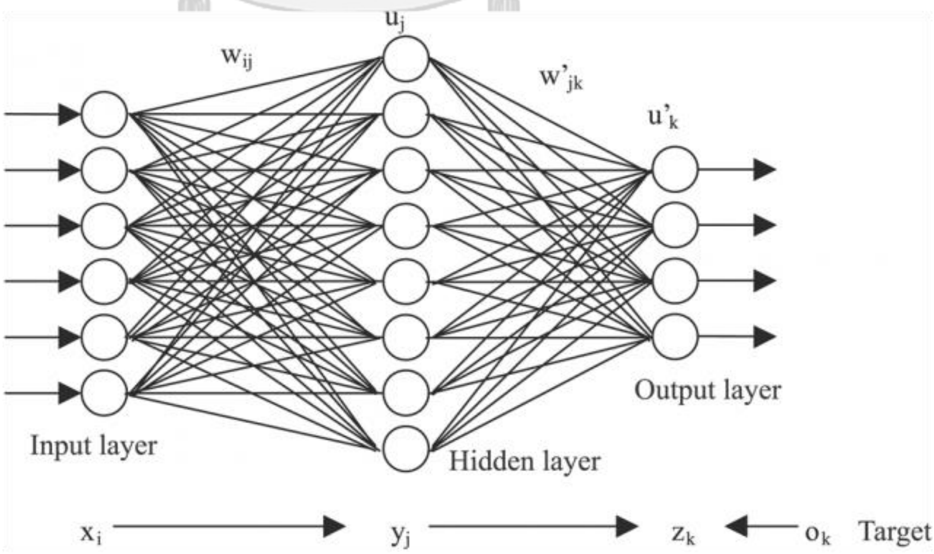

Figure 14: Example structure of Neural Network model

Artificial Neural Network model comprises of input nodes, hidden layers, output nodes, and transfer function. Values of input nodes and output nodes are set with the acquired data. Transfer functions and weights on nodes connection are tested to find lowest model error. One of the problems for Neural Network is how to set up hidden layer nodes structure whether Grid Search (GS) is implemented, or to use the better of Genetic Algorithm (GA) (Martinez-de Pison et al. 2016). The following picture shows example of Neural Network structure, containing 6 input variables, 1 hidden layer with 8 nodes, and 4 output variables.

The following figure shows code in Rstudio computing Neural Network Model (figure 15).

```
formular = Occupancy ~ Occ.EMA360 + RDB.7 + RDB.30 + RDB.52 + RDB.365

data.train <- train.4years
data.test <- test.4years
data.y <- data.matrix(test.4years$Occupancy)
Naive <- Naive.4years

str(data.train)
str(data.test)

neuralModel <- neuralnet(formular, data = data.train, hidden = c(20),
                         threshold = 1, stepmax = 1e+7, rep = 1, startweights = NULL, learningrate.limit = NULL,
                         learningrate.factor = list(minus = 0.5, plus = 1.2), learningrate = NULL, lifesign = "full",
                         lifesign.step = 10000, algorithm = "rprop+", err.fct = "sse", act.fct = "logistic",
                         constant.weights = NULL, likelihood = FALSE)

plot(neuralModel)

#predict test data set
NN.pred <- compute(neuralModel, data.test)
str(NN.pred)

predictions <- NN.pred$net.result
actualValues <- (data.y)

#predictions
#actualValues

MAPE = sum(abs(actualValues-predictions)/actualValues)/nrow(data.y)*100
MAPE

MAE = sum(abs(actualValues-predictions))/nrow(data.y)
MAE

#Naive1

MAENaive = sum(abs(actualValues-Naive))/nrow(data.y)
MAENaive

MASE = MAE/MAENaive
MASE
```

Figure 15: Artificial Neural Network computation code in Rstudio

3.2.5 Support Vector Regression

Support Vector Regression (SVR) is one of the machine learning models that can be used for both classification and regression problems. For regression problem, the specific name Support Vector Regression (SVR) is used instead. SVR can manipulate both linear and non-linear problems by adjusting kernel function. SVR can also operate unsupervised learning approach with unlabeled data by the use of Vapnik's ε-insensitive loss function (Vapnik V.N. 1995).

Given a set of data points $(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)$ ($x_\mathrm{i} \in X \subseteq R^n, y_\mathrm{i} \in Y \subseteq R$, l is the total number of training samples) randomly and independently generated from an unknown function, SVR approximates the function using the following form:

$$f(x) = w \cdot \phi(x) + b \tag{26}$$

where $\phi(x)$ represents the high-dimensional feature spaces which is nonlinearly mapped from the input space x. The coefficient w and b are estimated by minimizing the regularized risk function (27).

$$\text{Minimize} \qquad \frac{1}{2}\|w\|^2 + C\frac{1}{l}\sum_{i=1}^{l} L_\varepsilon(y_i, f(x_i)) \tag{27}$$

$$L_\varepsilon(y_i, f(x_i)) = \begin{cases} |y_i - f(x_i)| - \varepsilon, & |y_i - f(x_i)| \geq \varepsilon \\ 0 \qquad\qquad\qquad\quad , & otherwise. \end{cases} \tag{28}$$

The first term($\|w\|^2$) is called the regularized term. Minimizing $\|w\|^2$ will make a function as flat as possible, thus playing the role of controlling the function capacity. The second term $\frac{1}{l}\sum_{i=1}^{l} L_\varepsilon(y_i, f(x_i))$ is the empirical error measured by the ε-insensitive loss function (28). This loss function provides the advantage of using sparse data points to represent the designed function (26). C is referred to as the regularization constant. ε is called the tube size. They are both user-prescribed parameters and determined empirically.

To get the estimations of w and b, (27) is transformed to the primal objective function (34) by introducing the positive slack variables ξ_i^* ($(*)$ denotes variables with and without $*$)

Minimize $\quad \frac{1}{2}\|w\|^2 + C\frac{1}{l}\sum_{i=1}^{l} L_\varepsilon(\xi_i, \xi_i^*)$

Subjected to $\quad y_i - w \cdot \phi(x) - b \leq \varepsilon + \xi_i$

$$w \cdot \phi(x) + b - y_i \leq \varepsilon + \xi_i^*, \quad i = 1, \ldots l$$

$$\xi_i^{(*)} \geq 0 \tag{29}$$

Finally, by introducing Lagrange multipliers and exploiting the optimality constraints, the decision function (26) has the following explicit form:

$$f(x) = \sum_{i=1}^{l}(a_i - a_i^*)K(x_i, x) + b. \tag{30}$$

In function (35), $a_i^{(*)}$ are the Lagrange multipliers. They satisfy the equalities $a_i \times a_i^* = 0, a_i \geq 0, and\ a_i^* \geq 0$, where $i = 1, \ldots l,$ and they are obtained by maximizing the dual function of (29), which has the following form:

$$W\left(a_i^{(*)}\right) = \sum_{i=1}^{l} y_i(a_i - a_i^*) - \varepsilon\sum_{i=1}^{l}(a_i - a_i^*) - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(a_i - a_i^*)(a_j - a_j^*)\ K(x_i, x_j) \tag{31}$$

with the following constraints:

$$\sum_{i=1}^{l}(a_i - a_i^*) = 0, \qquad 0 \le a_i^{(*)} \le C, \qquad i = 1, \ldots l.$$

$K(x_i, x_j)$ is defined as the kernel function. The value of the kernel is equal to the inner product of two vectors x_i and x_j in the feature space $\phi(x_i)$ and $\phi(x_j)$, that is $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$.

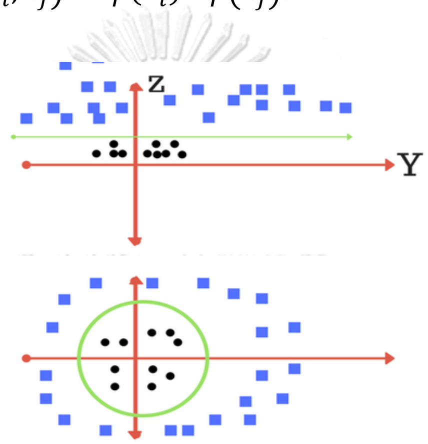

Figure 16: Graphical representation of kernel activation functions in Support Vector Regression

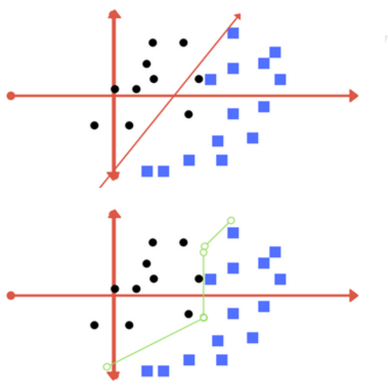

Figure 17: Graphical representation of regularization parameter in Support Vector Regression

Equation 26 to 31 has shown Support Vector Regression Algorithm (Cao and Tay 2001). To simplify, Support Vector Regression can manipulate in higher dimensional separation plane with non-linear error function. SVR configurates three parameters in order to achieve classification results. These three parameters are kernel function, parameter C, and parameter γ. First, SVR model transform data to 2-dimanional separation plane. When data separation is completed, the model utilizes kernel function to transform this separation plane back to original dimensions. Available kernel activation functions are linear, polynomial, RBF, and sigmoid (Figure 16). Also, regularization parameter C are adjusted to find the most suitable separation function. If value of C is large, the optimization will choose a smaller margin hyperplane, and if value of C is small, such hyperplane might misclassify more data points (Figure 17). Finally, γ parameter quantifies how the data points are considered in constructing the separation hyperplane. If γ value is high, only data points located close to the separation plane are considered, but if γ value is low, data points far beyond the separation plane up to the boundary of datapoints are considers (Figure 18). Therefore, the most significant characteristic of Support Vector Regression is the separation hyperplanes are attuned to construct good margin between groups of data points and that hyperplanes, assuring that classification of data points is optimized

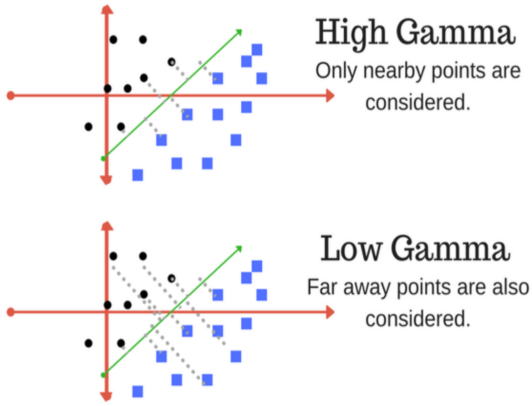

Figure 18: Graphical representation of gamma parameter in Support Vector Regression

The following figure shows code in Rstudio computing Support Vector Regression (figure 19).

```
#install.packages("e1071")
library(e1071)
#install.packages("rpart")
library(rpart)
#install.packages("Matrix")
library(Matrix)
#install.packages("SparseM")
library(SparseM)

#Importing the data
data.occupancy <- read.csv("C:\\users\\ph156\\Documents\\SVM\\OccData7-12-28-30-31-52-360-364-365-366,EMA90-180-270-360.c

train.6years <- data.occupancy[1:1826,2:16]
test.6years <- data.occupancy[1827:2191,2:16]

Naive <- data.occupancy$Occupancy[1826:2190]

formular = Occupancy ~ Occ.EMA360 + RDB.7 + RDB.30 + RDB.52 + RDB.365

data.train <- train.6years
data.test <- test.6years
data.y <- data.matrix(test.6years$Occupancy)

str(data.train)
str(data.test)

#Fitting SVM
svm.model <- svm(formular, data = data.train, scale = TRUE,  type = "eps-regression", kernel = "radial",
                 na.action = na.exclude)

svm.pred <- predict(svm.model, data.test)

svm <- data.matrix(svm.pred)

#predict test data set
predictions <- svm
actualValues <- data.y

MAPE = sum(abs(actualValues-predictions)/actualValues)/nrow(data.y)*100
MAPE

MAE = sum(abs(actualValues-predictions))/nrow(data.y)*100
MAE

#Naive1

MAENaive = sum(abs(actualValues-Naive))/nrow(data.y)*100
MAENaive

MASE = MAE/MAENaive
MASE
```

Figure 19: Support Vector Regression computation code in Rstudio

Chapter IV: Results and Discussion

4.1 Results

Forecasting results are presented in this section. Time series model forecasts are tested with data for model construction. For Machine Learning models, experiments are made to find the most suitable regressor variables and the most accurate model. First, significant effecting variables are discovered with linear regression. Then, forecasting accuracy between independent variables and transformed data are compared to find the most suitable regressor variables. Next, the most accurate model is constructed with the significant effecting variables. Finally, forecasts are generated.

4.1.1 Exponential Smoothing (Holt-Winters)

Figure 20 shows actual values and predicted values of Holt-Winters model. The R Studio software program automatically provides optimized estimates of model parameters (alpha, beta, and gamma) that minimize forecast errors. Holt-Winters model with mean parameter (alpha) = 0.208, trend parameter (beta) = 0, and seasonal parameter (gamma) = 0.498 is selected by the program. Result indicates that Holt-Winters model cannot capture changes in trend component in this hotel daily time series, yielding MAPE 20.69%

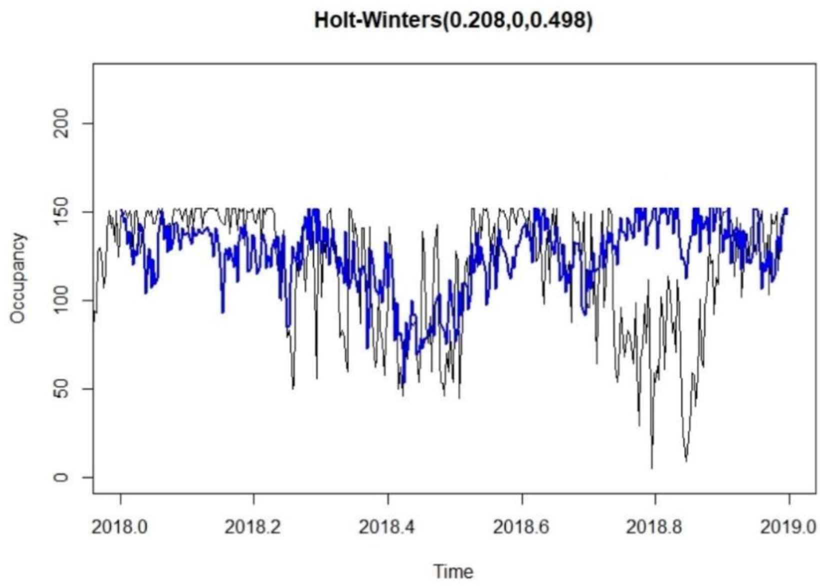

Figure 20: Holt-Winters' Exponential Smoothing forecasting model result

4.1.2 SARIMA

Figure 21 show actual values and predicted values of SARIMA model. Model adequacy is check for each forecasting method. Figure 22, 23 show ACF and PACF plots from SARIMA model. ACF of SARIMA model shows spike at lag 1 though it decays at the end, meaning that trend and seasonality has not been completely removed. Thus, SARIMA model is not reliable and should not be selected for usage. Figure 24 shows residuals versus fitted values of SARIMA model. It can be observed that residuals are randomly distributed around zero mean. However, data points from the plot show downward sloping confinement on the upper right corner. This is anticipated to be a result from nature of hotel time series data that have maximum room capacity.

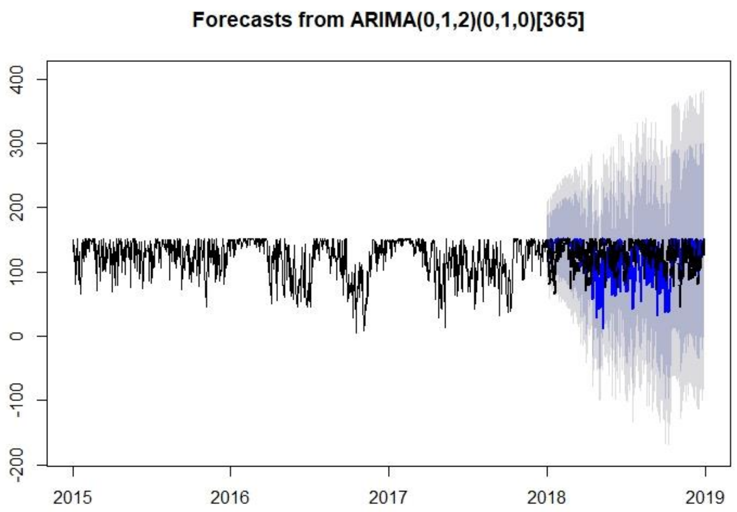

Figure 21: SARIMA forecasting model result

Figure 25 shows flowchart of SARIMA model selection, Autoregressive, moving average, and integrated factor for both trend and seasonal components (p,d,q,P,D,Q) are adjusted to find lowest AIC and lowest MAPE. Model SARIMA(0,1,2)(0,1,0)365 with AIC = 6957.83 , and MAPE = 23.44% is the result of this experiment.

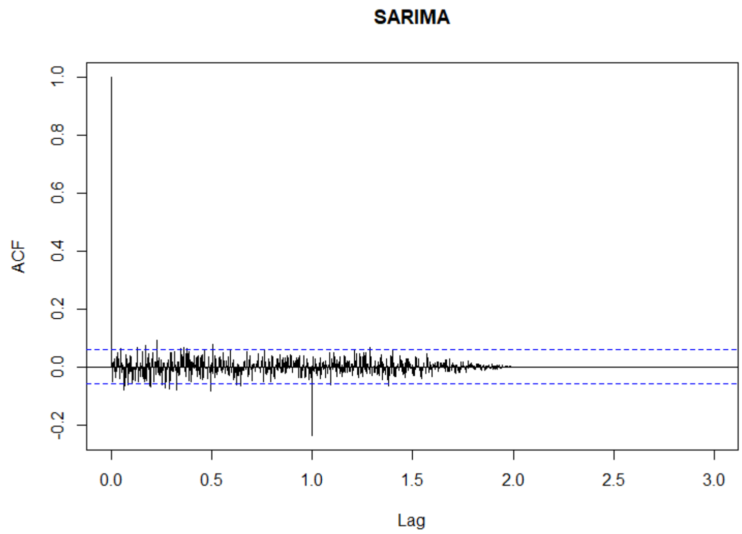

Figure 22: ACF of residuals from SARIMA model

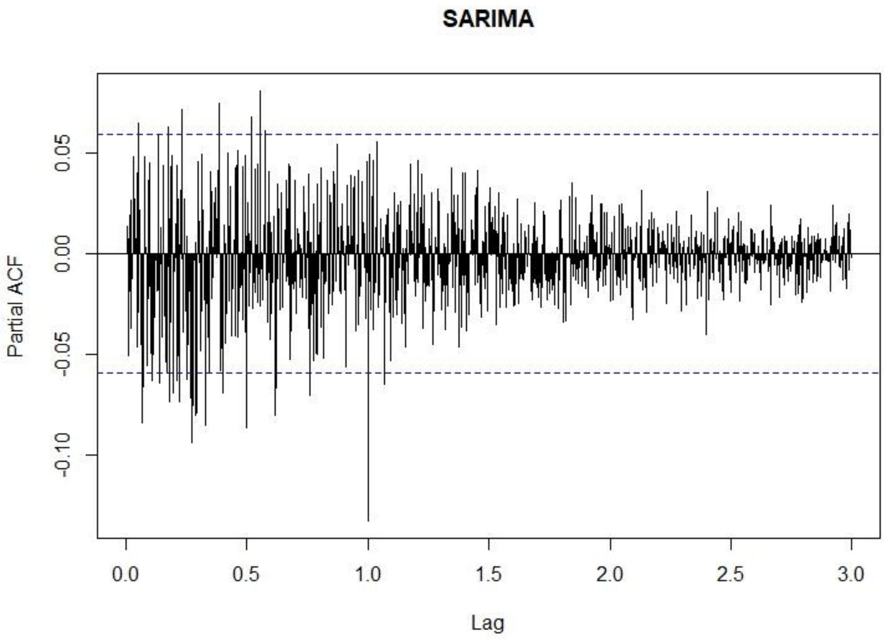

Figure 23: PACF of residuals from SARIMA model

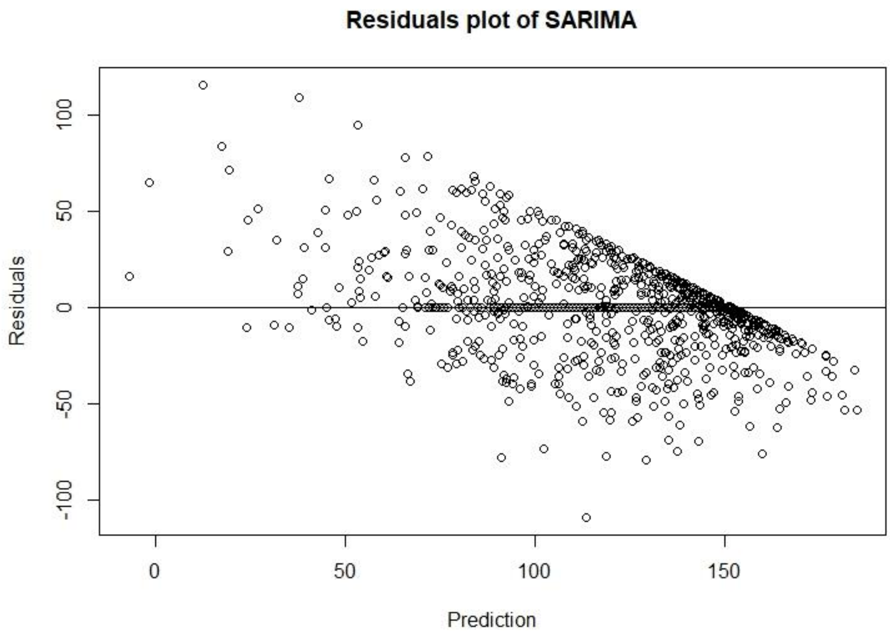

Figure 24: Residuals vs. fitted values plot of SARIMA model

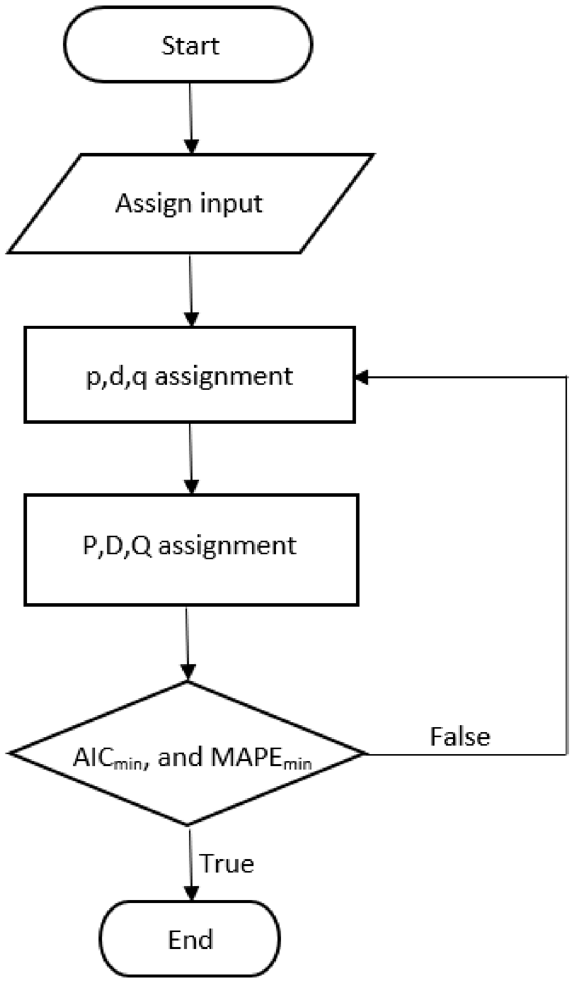

Figure 25: Flowchart showing SARIMA model selection

4.1.3 BATS/TBATS

Figure 26 shows flowchart of BATS and TBATS model construction, smoothing parameters (α, β, γ), box-cox transformation parameter (ω), dampening parameter (ϕ), and autoregressive and moving average component (p,q) are adjusted to find maximum log likelihood of the estimates (MLE), minimum Root Mean Square Error of the original data ($RMSE$) and Root Mean Square Error of the transformed data ($RMSE_T$), minimum Akaike information criterion (AIC), and minimum MAPE) As a result, model BATS$(0.984, \{0,0\}, 0.8, \{365\})$ with MAPE = 19.64% and TBATS$(1, \{2,1\}, -, \{< 365,3 >\})$ with MAPE = 17.24% are selected.

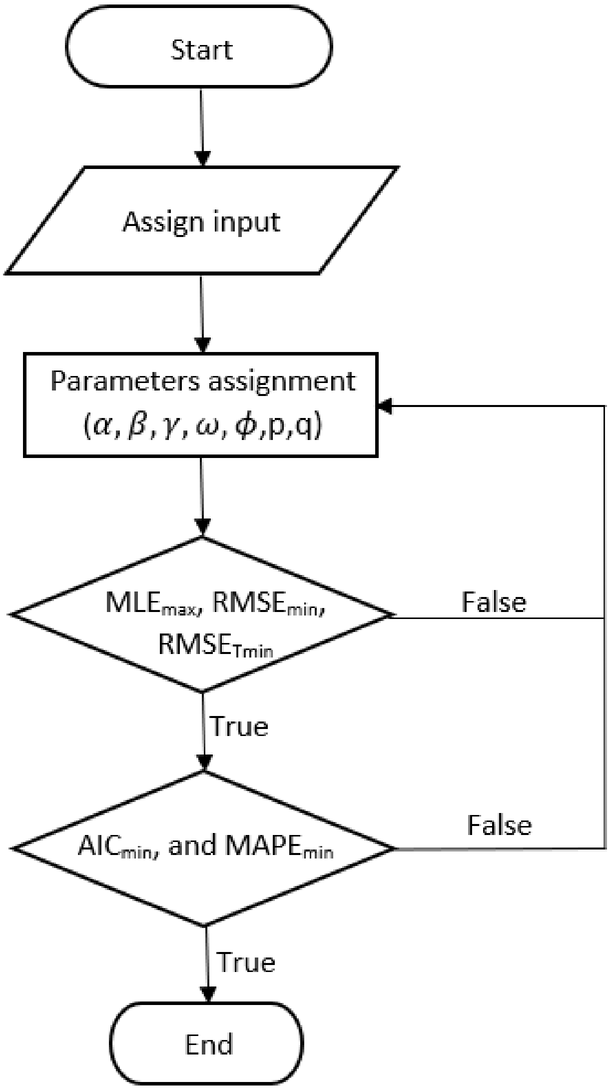

Figure 26: Flowchart showing BATS/TBATS model selection

Figure 27, 28 show actual values and predicted values of BATS and TBATS model. The TBATS model prediction graph shows that seasonality existed in a year that it resembles tourist arrivals graph (Figure 1), peaking at beginning and end of the year and reaching lowest in May and September. This similarity presents that this model can explain daily hotel data. BATS forecasting model generated BATS$(0.984, \{0,0\}, 0.8, \{365\})$, representing the use of Box-Cox transformation, no ARMA error function, disposition of damped trend, and seasonal period of 365. TBATS forecasting model generated TBATS$(1, \{2,1\}, -, \{<365,3>\})$, representing no use of Box-Cox transformation, Autoregressive and Moving Average model of 2 and 1 respectively ARMA(2,1), and seasonal period of 365 with harmonics of 3.

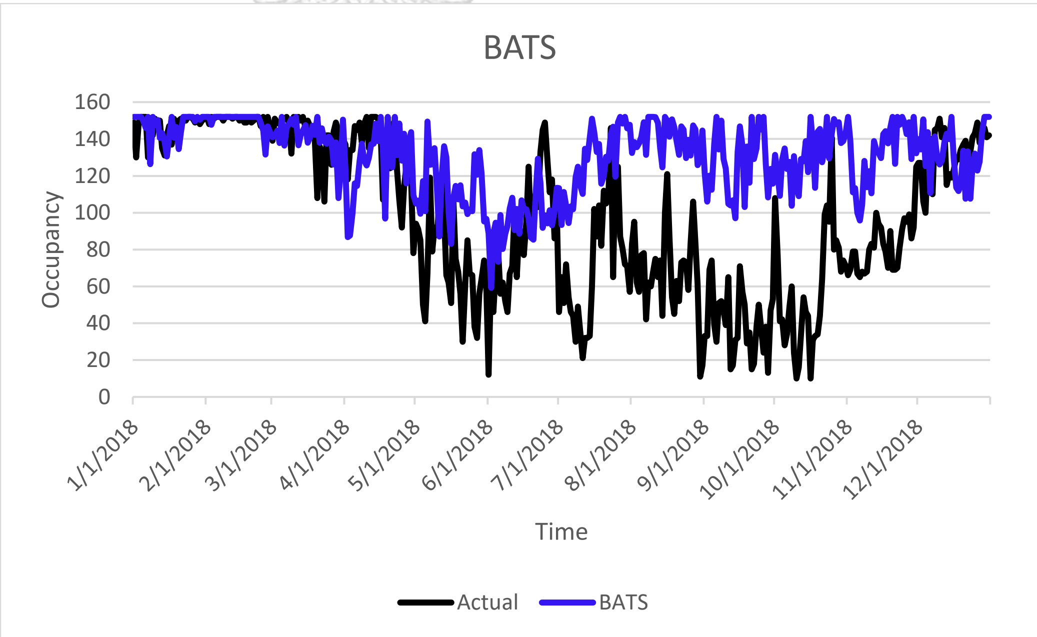

Figure 27: BATS forecasting model result

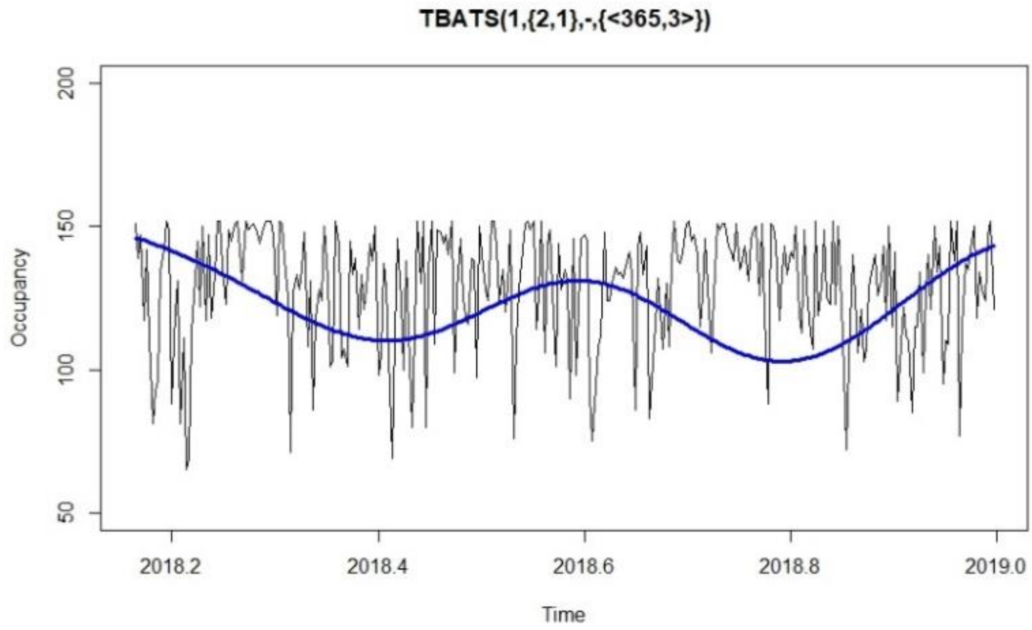

Figure 28: TBATS forecasting model result

Model adequacy is check for each forecasting method. Figure 29, 30, 31, 32 show ACF and PACF plots from TBATS model. Autocorrelation Function (ACF) are calculated at maximum lag of 1095 (3 years of daily data) to check appropriateness of the models. It takes up to 1095 lags because lesser numbers of lag did not decay ACF to zero. Starting from lag one, ACF from TBATS model do not show positive autocorrelation for at high number of lags, meaning that trend and seasonality has been removed. Thus, TBATS model is reliable and could be selected for usage. ACF of BATS model show spikes at lag 1 and 2 though it decays at the end, therefore BATS model is not selected for use.

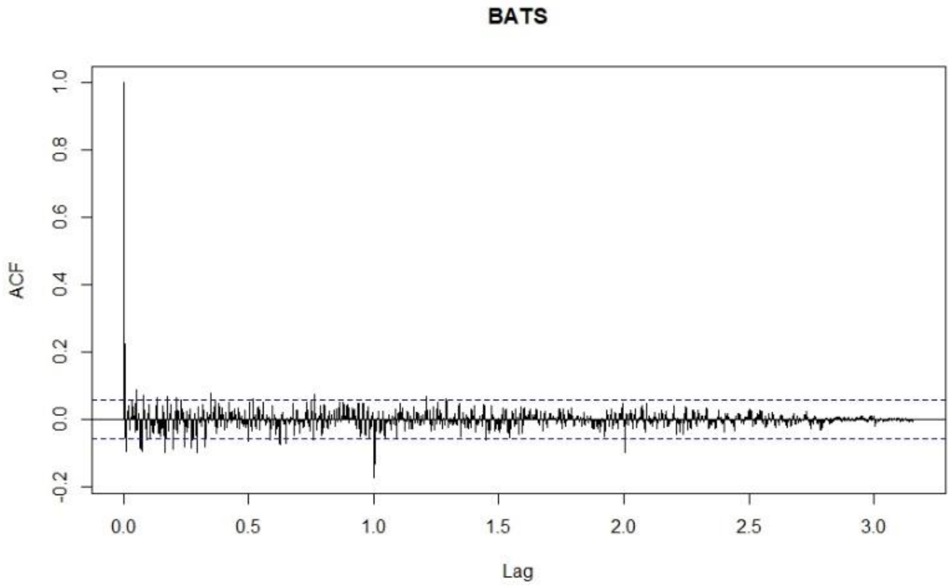

Figure 29: ACF of residuals from BATS model

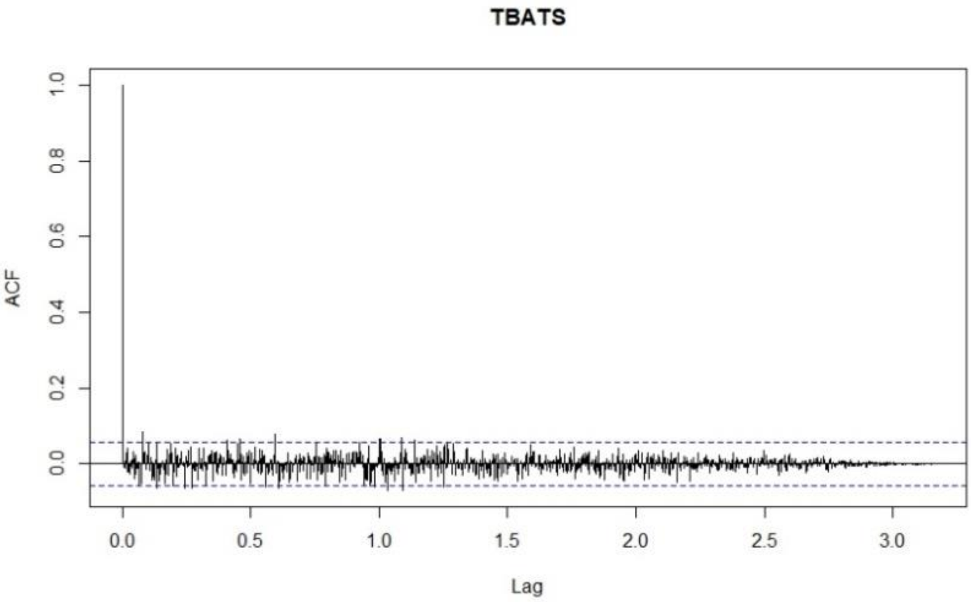

Figure 30: ACF of residuals from TBATS model

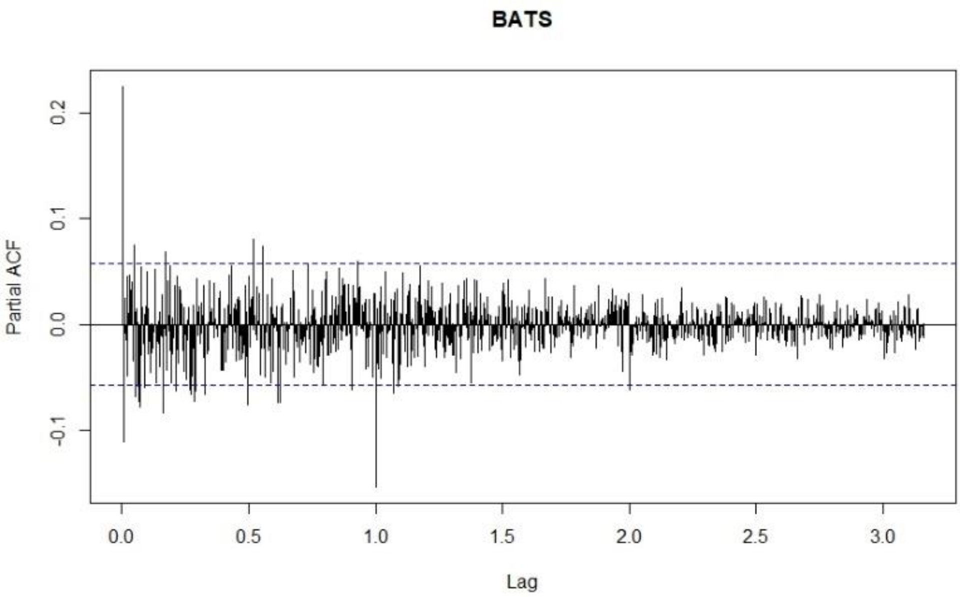

Figure 31: PACF of residuals from BATS model

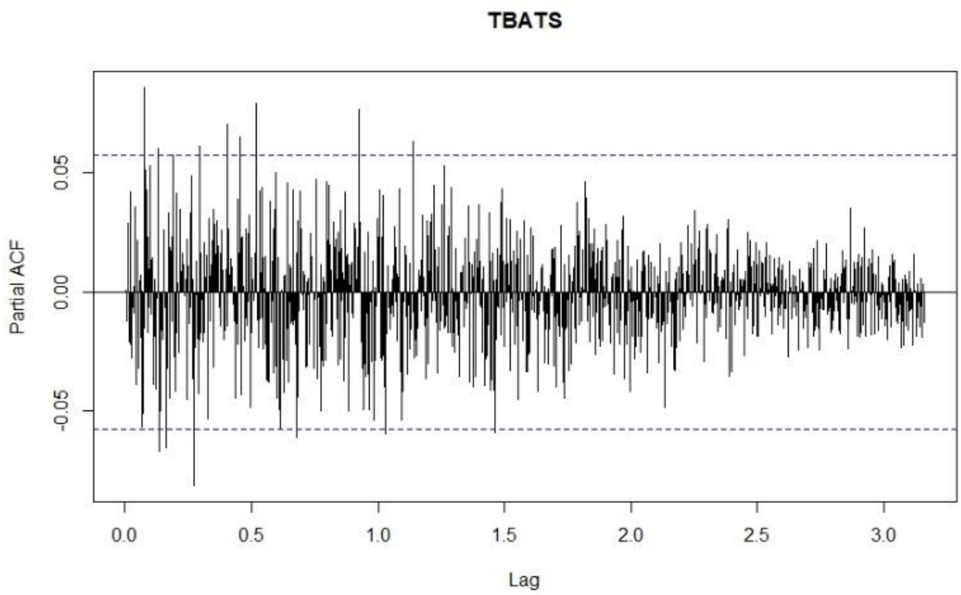

Figure 32: PACF of residuals from TBATS model

4.1.4 Artificial Neural Network

Experiments are conducted to find variables that are significant in Linear Model. Total of 46 independent variables are tested in this section, including number of tourist arrivals in Phuket by country of residence, customer price index of Southern region of Thailand, oil price, exchange rates over Thai Baht, average room rate, days of the week, month of the year, weather, season, sanction on Russia, Chinese boat capsizing, free visa grated. These variables will be used as input for Neural Network model and Support Vector Regression. The test was executed with Minitab program. Linear Regression employed backward elimination of terms with alpha less than 0.05. As a result, there are 30 significant variables in total. Figure 33 shows significant variables from Linear Regression model. The variables that can significantly explain the hotel daily data are

1. MA-360, RDP-7, RDP-30, RDP-52, RDP-365
2. Number of Chinese tourists
3. Number of Russian tourists
4. Number of Australian tourists
5. Number of Korean tourists
6. Number of Malaysian tourists
7. Number of Singaporean tourists

8. Number of Swedish tourists
9. Number of British tourists
10. Customer price index of Southern region of Thailand
11. Oil price
12. Exchange rate of United States Dollar over Thai Baht
13. Exchange rate of Malaysian Ringgit over Thai Baht
14. Exchange rate of Swedish Krona over Thai Baht
15. Exchange rate of British Pound over Thai Baht
16. Average room rate of the hotel
17. Month: July, August, September, October, November, December, and February
18. Peak, and High season dummy variables
19. Sanction on Russia dummy variables

In conclusion, number of tourists from every country is deemed significant by Linear Regression model because these are source of hotel occupancy. Customer price index of Southern region of Thailand and oil price are significant because they determine purchasing power of tourist in the region. Some of the exchange rates are considered significant because they have relation to number of tourists in the island. Only months of the year but not days of the week are significant because seasonal pattern occurs on monthly basis, not on daily basis. Peak season, high season, and sanction on Russia dummy variables are significant because they can explain data pattern in the daily occupancy. Average room rate of the hotel is significant because it is directly disproportionated to the occupancy. Transformed data of the hotel are significant because the data is highly related to daily occupancy.

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	263.1	32.5	8.09	0.000	
China	-0.000219	0.000032	-6.85	0.000	9.73
Russia	0.000844	0.000073	11.63	0.000	12.71
Australia	0.000999	0.000392	2.54	0.011	7.84
Korea	0.000683	0.000207	3.30	0.001	3.86
Malaysia	0.004909	0.000533	9.21	0.000	5.40
Singapore	-0.001369	0.000304	-4.51	0.000	8.47
Sweden	0.000801	0.000284	2.82	0.005	17.23
UK	0.001211	0.000513	2.36	0.018	8.18
CPISouth	0.002693	0.000993	2.71	0.007	11.83
OilPrice	-0.4165	0.0677	-6.15	0.000	8.70
USD	-6.799	0.759	-8.96	0.000	6.36
MYR	-5.97	2.45	-2.44	0.015	15.02
SEK	-16.90	4.59	-3.68	0.000	13.24
GBP	3.210	0.313	10.25	0.000	6.86
ARR	-0.01310	0.00165	-7.92	0.000	4.47
July	20.97	3.05	6.89	0.000	2.61
August	36.67	3.74	9.79	0.000	3.95
September	24.09	2.90	8.31	0.000	2.29
October	7.68	2.55	3.00	0.003	1.84
November	-24.32	4.04	-6.02	0.000	4.46
December	-49.56	4.24	-11.69	0.000	5.05
Febuary	23.39	3.70	6.31	0.000	3.54
Peak	10.02	2.72	3.68	0.000	2.14
High	12.91	3.46	3.73	0.000	10.11
RusSanc	-14.21	2.83	-5.03	0.000	4.51

Figure 33: Significant variables form Linear Regression model

There are two types of regressor variables in this study: independent factors and transformed data. Independent factors are explained thoroughly in table 5. Independent data include variables such as Number of Tourist Arrivals in Phuket by Country of Residence, Customer Price Index of Southern Region of Thailand, Oil Price, Exchange Rates over Thai Baht, Days of the Week, Month, Monsoon, Season, Sanction on Russia, Chinese Boat Capsizing, and Free Visa. Transformed data are RDP-7, RDP-30, RDP-52, RDP-365 (Relative Difference in Percentage), and MA-360 (occupancy subtracted with Moving Average). Table 6 shows Adjusted R-squared of Linear Model with 2 different types of variables.

Table 6: Experiment on types of input variables in Linear Regression model

	Adjusted R-squared		
Variables	Full model	Reduced model	No. of significant variables
Independent factors	43.15%	44.78%	25
Independent factors + Transformed data	99.88%	99.88%	30
Transformed data	97.41%	97.41%	5

All independent variables are also tested for correlation between themselves and dependent variables. Correlation of every pair of dependent and independent variables are shown in Appendix section. Results show that almost every pair of all variables are not considered significant, having coefficient of correlation less than 0.8 or greater than -0.8.

Table 7: Experiment on parameters of Neural Network model

Algorithm	Activation function	MAPE	MAE
rprop-	logistic	11.7975	6.983
rprop-	tanh	12.2517	7.238
rprop+	logistic	11.7974	6.983
rprop+	tanh	12.1301	7.175

Different types of input variables are tested with Neural Network model to find the forecasting accuracy. Table 7 shows parameters selection of Neural Network model including algorithm and activation function. Sum of squared errors (sse) is used as error function, while cross-entropy (ce) error function can be used only for binary response. Only 'rprop-' and 'rprop+' algorithm can achieve threshold of 1 under 10^8 stepmax, the lesser threshold and larger stepmax employ longer computation time which is not practical. Logistic function (logistic) and tangent hyperbolicus (tanh) are used as activation function, both are tested. Results shows that Neural Network model with 'rprop+' as an algorithm and logistic function as activation function performs best. Though all combinations of algorithm do not differ significantly, 'rprop+' and logistic function is selected as parameters.

Tables 8 shows forecasting result on types of input variables from Neural Network model. It can be observed that transformed data as input variables for Neural Network model performs better than both independent factors and independent factors with transformed data in two aspects: MAPE is exceptionally better and number of nodes utilized in the Neural Network structure is lower.

Table 8: Forecasting result on types of input variables from Neural Network model

	Neural Network		
Variables	MAPE	MAE	Nodes
Independent factors	71.39	37.26	100
Independent factors + Transformed data with p-value < 0.05	74.33	37.59	35
Independent factors + Transformed data with lowest p-value	74.34	37.59	55
Transformed data	8.955	5.603	15

Figure 34 shows flowchart of Neural Network model selection, starting with input assignment and error threshold setting at 1. Algorithm and activation function are selected to tested for lowest MAPE, then structure of Neural Network model is tested by adding number of nodes into its structure to find the most accurate model.

An experiment has been made to find the most accurate Neural Network structure, with data up to 1826 points and computation time for Neural Network is of $O(n^5)$, by assuming that gradient descent runs for n iterations, and that there are n layers each with n neurons. The structure of NN model is limited to 1 hidden layer. First, numbers of node vary from 1-25 at lag of 5, then within the range that gives the best model numbers of node are changes by lag of 2 to find all possible models. Table 9 shows result of experiment on number of nodes.

To conclude, the applicable models are 1 hidden layer with 12,14,15,16,18, and 20 nodes, yielding MAPE under 10%. Experiments are conducted to find the most suitable error function and activation function, that will deliver with the lowest MAPE. These Neural Network models employ resilient backpropagation with and weight backtracking as an algorithm in calculating models' weights. Sum of squared error function is used to compute the model error, and logistic function is used as an activation function. Figure 35 shows actual values and predicted values of Neural Network model with backpropagation. Figure 36 shows plot of the most accurate Neural Network model.

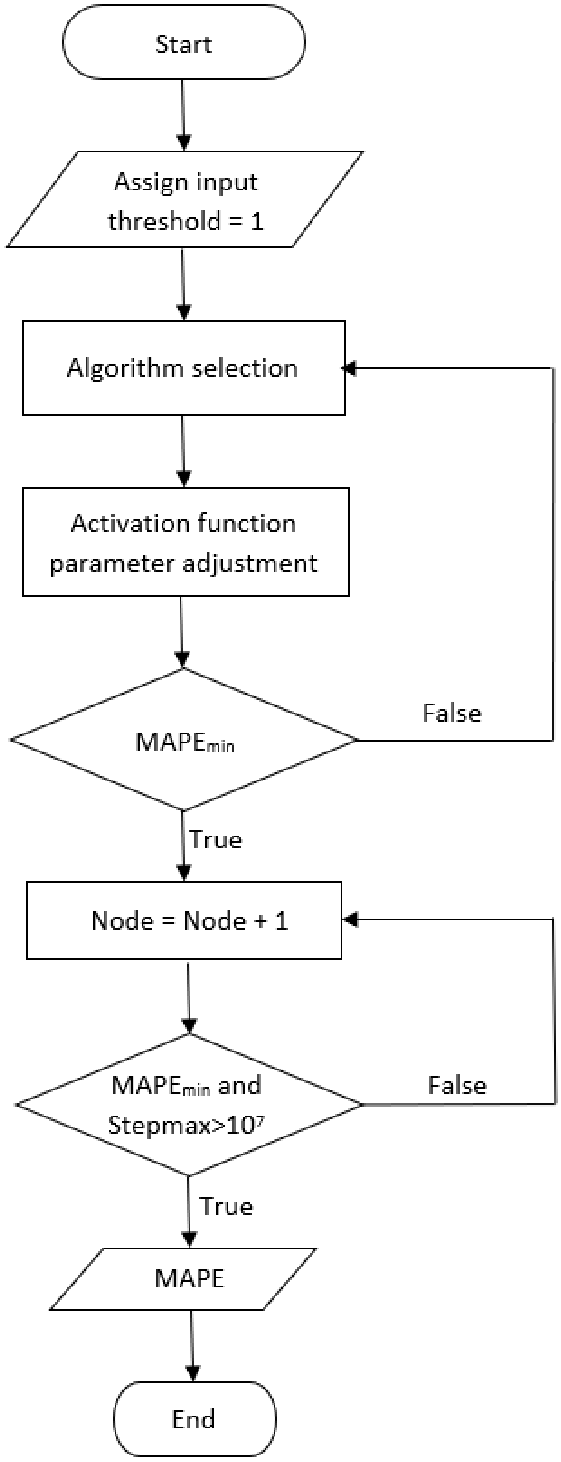

Figure 34: Flowchart showing Neural Network model selection

Table 9: Experiment on number of nodes in Neural Network model

No. of Nodes	Neural Network	
	MAPE	MAE
1	10.72	6.500
5	10.47	6.313
10	10.50	6.579
15	8.955	5.603
20	9.140	5.826
25	10.86	6.683
...	...	...
12	9.334	5.913
14	9.241	5.850
16	9.118	5.835
18	9.660	6.235

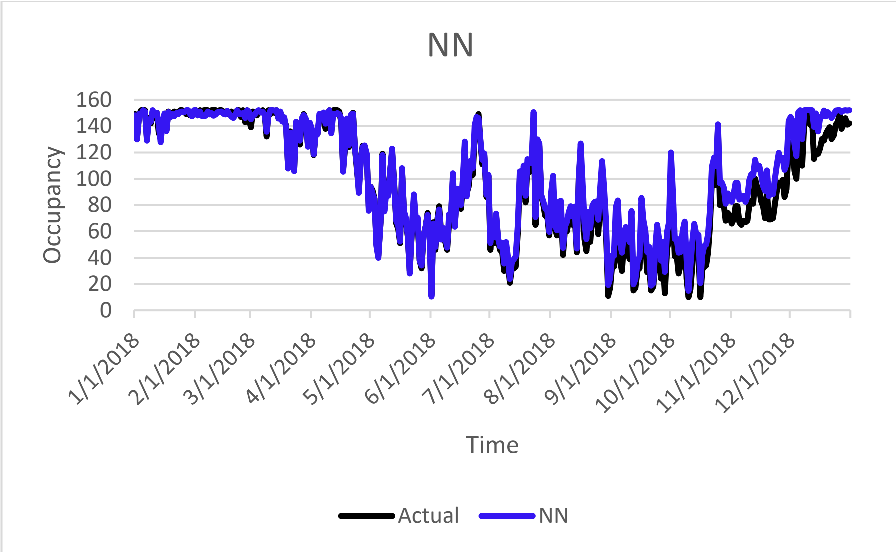

Figure 35: Neural Network forecasting model result

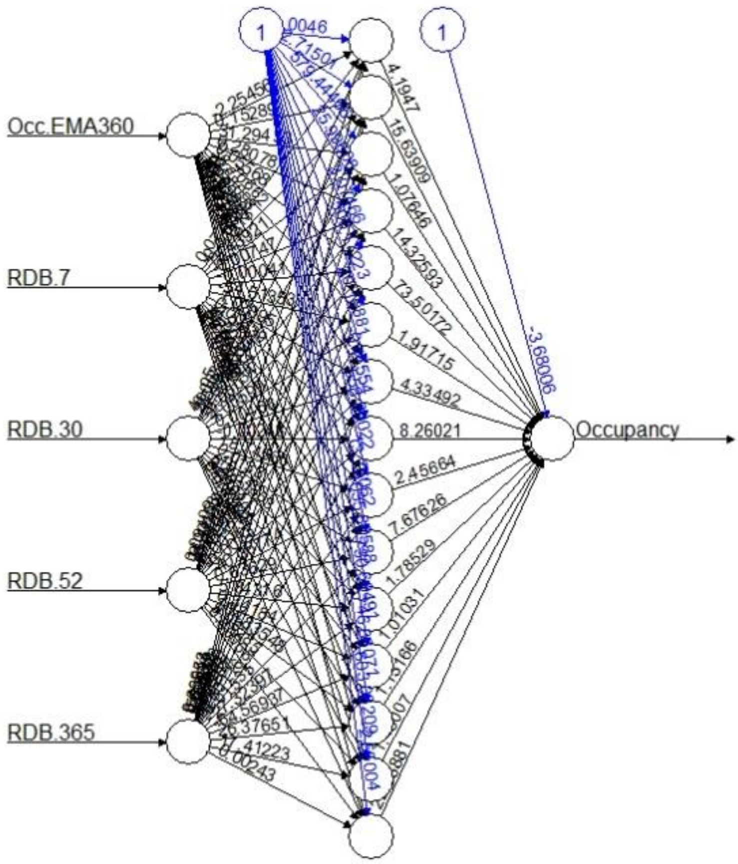

Figure 36: Plot showing the most accurate neural network model

4.1.5 Support Vector Regression

Support Vector Regression model is the most straightforward. It is non-linear regression model without any constraining regression assumption. Experiments are conducted to find the most suitable regression function and kernel function. These Support Vector Regression models utilize epsilon type regression function, and radial basis function as kernel function. The most accurate SVR model gives MAPE at 13.76%.

Table 10: Experiment on parameters of Support Vector Regression

Regression type	Kernel function	MAPE	MAE
eps	Polynomial	63.37	46.17
	Radial basis	13.75	7.779
	Sigmoid	1156.42	434.8
nu	Polynomial	64.29	51.62
	Radial basis	13.82	13.42
	Sigmoid	1080.15	469.2

Experiments were made to adjust parameters for Support Vector Regression (table 10). Gamma and regularization parameter are fixed at first, varying regression type and Kernel function to find the most accurate settings. Later, Gamma and regularization parameter are adjusted but resulted accuracy did not change significantly. Differences of accuracy between eps and nu regression type is negligible. However, eps-regression facilitates best performance by controlling amount of errors in the model, in which solution model could be complex, but nu-regression returns fewer support vectors, requirement for small solution. Figure 37 shows actual values and predicted values of Support Vector Regression model.

Tables 11 shows forecasting result on types of input variables from Support Vector Regression. It can be observed that transformed data as input variables for SVR model performs better than both independent factors and independent factors with transformed data.

Figure 38 shows flowchart of showing Support Vector Regression model selection, starting with input assignment. Kernel function, regularization parameter, and gamma parameter are selected and tested to find the most accurate settings by MAPE. Radial basis as Kernel function with eps-regression type provides the best accuracy, while regularization and gamma parameter are set at default. Then, regularization and gamma parameter are varied to find the best performance.

Table 11: Forecasting result on types of input variables from Support Vector Regression model

	Support Vector Regression	
Variables	MAPE	MAE
Independent factors	84.86	41.56
Independent factors + Transformed data with p-value < 0.05	43.21	22.00
Independent factors + Transformed data with lowest p-value (Russian Tourists)	28.05	14.25
Transformed data	13.75	7.779

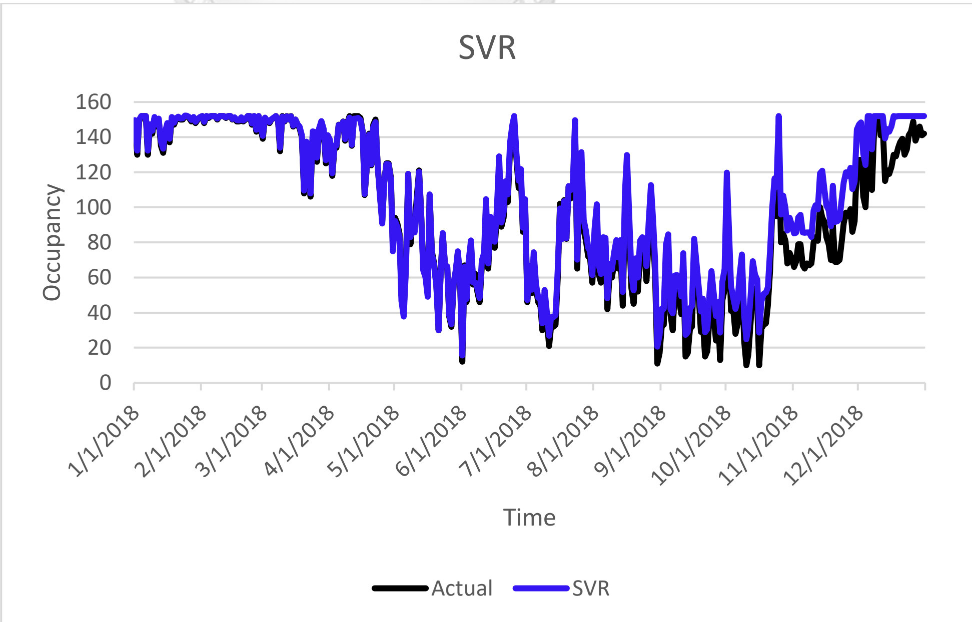

Figure 37: Support Vector Regression forecasting model result

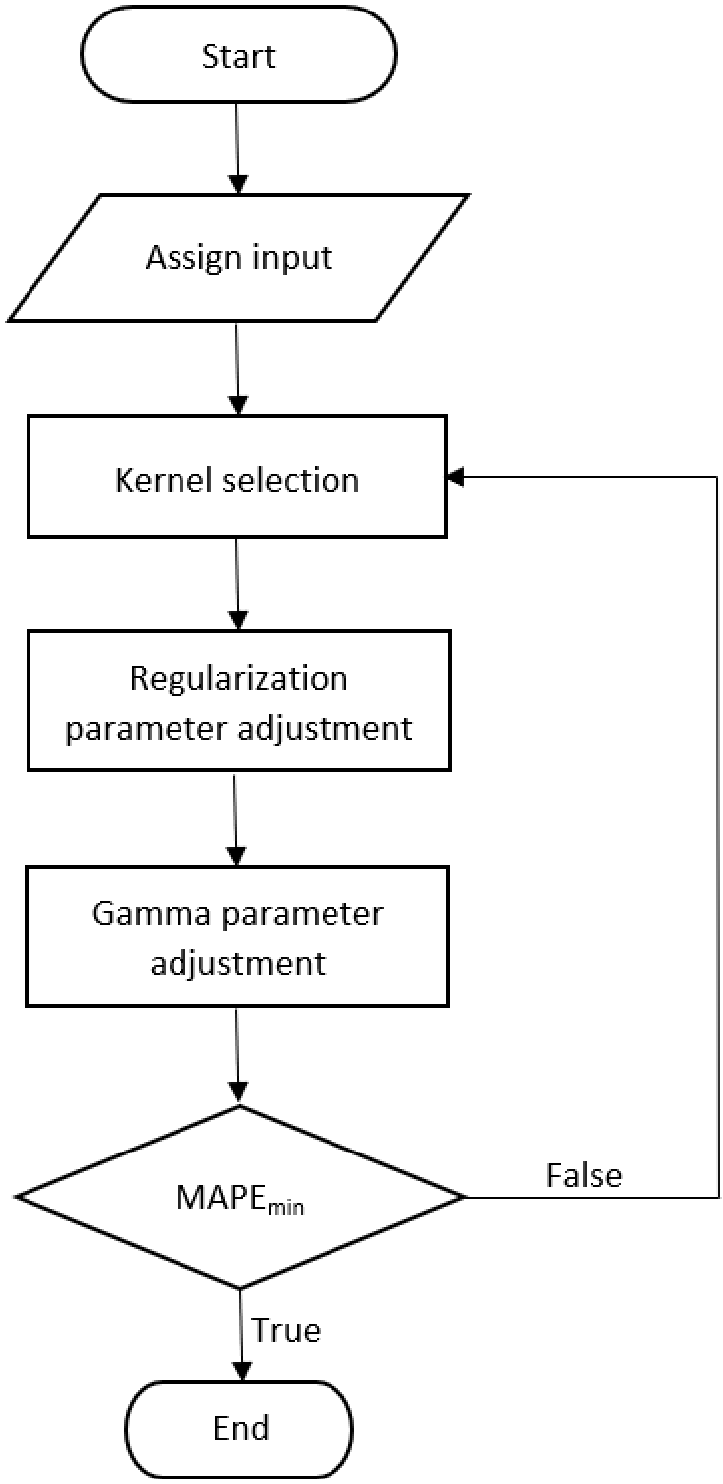

Figure 38: Flowchart showing Support Vector Regression model selection

4.1.6 Benchmark Forecasts

Benchmark forecasts can be computed effortlessly, thus are often used as standards. Benchmark forecasts in this study are Naïve1, Naïve2, Same Date Last Year Naïve, MA(60), MA(90). Benchmark forecasts were calculated to observe lower limit of forecasting performance and effectiveness of forecasting models. If performance of the forecasting models were not better than these benchmarks, one should refer to the benchmarks. Table 12, and 13 show benchmark forecasts according to MAPE, and MAE

Table 12: Benchmark forecasts according to MAPE

		Naïve 1	Naïve 2	SDLY	MA(60)	MA(90)
Fully aggregated:	Occupancy	22.91%	78.36%	73.15%	43.11%	50.09%
Disaggregated:	Deluxe	46.68%	115.19%	117.20%	79.92%	89.28%
	Seaview	33.35%	1.183215	119.09%	62.57%	72.96%
	Pool Access	Inf	Inf	Inf	Inf	Inf
	Villa	Inf	Inf	Inf	Inf	Inf

Table 13: Benchmark forecasts according to MAE

		Naïve 1	Naïve 2	SDLY	MA(60)	MA(90)
Fully aggregated:	Occupancy	13.28	37.98599	34.31	23.94	27.37
Disaggregated:	Deluxe	5.63	11.61127	10.68	8.89	10.07
	Seaview	7.15	18.03324	17.49	11.44	13.07
	Pool Access	2.30	5.68352	5.46	3.78	4.11
	Villa	1.45	6.251604	4.96	3.02	3.59

4.2 Model Comparisons and Selection

In this section, a comparison among proposed models namely, Holt-Winters' exponential smoothing, Seasonal ARIMA, BATS, TBATS, Neural Network with backpropagation, and Support Vector Regression is discussed.

Table 14: Aggregated forecasting results according to MAPE and MAE

Method		MAPE (%)	MAE
Naïve	Last Day	22.91	13.28
	Same Date Last Year	73.15	34.24
Moving Average	30 days	33.90	18.66
	90 days	50.09	27.37
Holt-Winters(0.208,0,0.498)		20.69	16.66
SARIMA(0,1,2)(0,1,0)365		23.44	27.46
BATS$(0.984, \{0,0\}, 0.8, \{365\})$		19.64	21.60
TBATS$(1, \{2,1\}, -, \{< 365,3 >\})$		17.24	19.77
Support Vector Regression		12.30	8.760
Artificial Neural Network		8.955	6.617

Table 14 shows forecasting results of aggregated forecast measured by Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE). In some case, forecast values exceed maximum room capacity. Thus, all forecasts are check whether they exceed the capacity or not. If the predictions surpass room capacity, they are reduced to the maximum number of rooms, and then error measures are calculated.

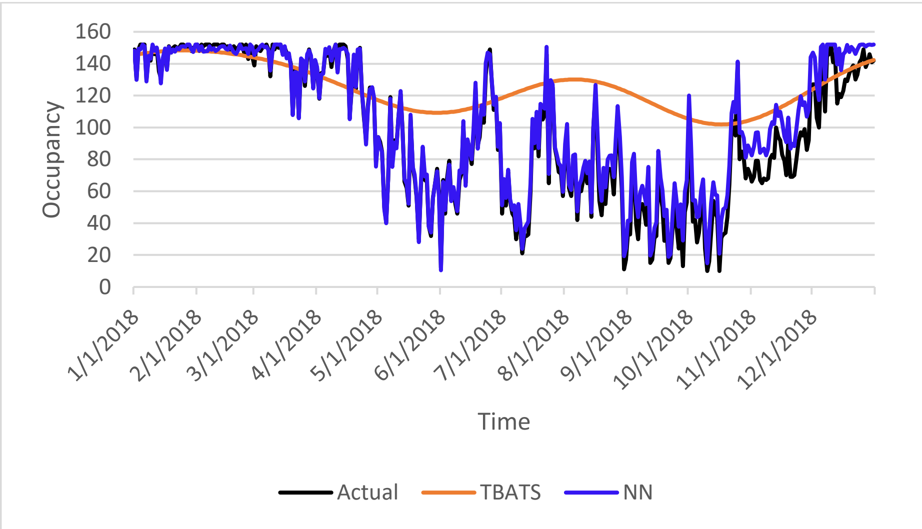

Figure 39: Graphical presentation of forecasting results comparing actual value and forecast value from TBATS and Neural Network model

Figure 39 shows comparisons between actual value and forecast value from TBATS and Neural Network model (the best model from time series and machine learning), depicting that Neural Network model are relatively close to the actual values. As a result, Neural Network model can replicate trend and seasonality of daily hotel occupancy better than TBATS model. The low MAPE indicated that ANN model provided the most accurate forecasting value, followed by SVR, TBATS, BATS, Holt-Winters' exponential smoothing, and SARIMA model. To understand the performance of all methods, the randomized complete block designs (RCBD) and

Tukey Pairwise Comparisons

Grouping Information Using the Tukey Method and 95% Confidence

Factor	N	Mean	Grouping
SDLY	365	34.24	A
SARIMA	365	27.46	B
BATS	365	21.603	C
TBATS	365	19.768	C D
MA(30)	365	18.661	C D
HW	365	16.662	D
SVR	365	8.760	E
NN	365	6.617	E

Means that do not share a letter are significantly different.

Tukey Simultaneous Tests for Differences of Means

Difference of Levels	Difference of Means	SE of Difference	95% CI	T-Value	Adjusted P-Value
MA(30) - SDLY	-15.57	1.35	(-19.68, -11.47)	-11.51	0.000
HW - SDLY	-17.57	1.35	(-21.68, -13.47)	-12.98	0.000
SARIMA - SDLY	-6.78	1.35	(-10.88, -2.67)	-5.01	0.000
BATS - SDLY	-12.63	1.35	(-16.74, -8.53)	-9.33	0.000
TBATS - SDLY	-14.47	1.35	(-18.57, -10.36)	-10.69	0.000
NN - SDLY	-27.62	1.35	(-31.72, -23.51)	-20.40	0.000
SVR - SDLY	-25.48	1.35	(-29.58, -21.37)	-18.82	0.000
HW - MA(30)	-2.00	1.35	(-6.11, 2.11)	-1.48	0.820
SARIMA - MA(30)	8.80	1.35	(4.69, 12.90)	6.50	0.000
BATS - MA(30)	2.94	1.35	(-1.16, 7.05)	2.17	0.368
TBATS - MA(30)	1.11	1.35	(-3.00, 5.21)	0.82	0.992
NN - MA(30)	-12.04	1.35	(-16.15, -7.94)	-8.90	0.000
SVR - MA(30)	-9.90	1.35	(-14.01, -5.80)	-7.31	0.000
SARIMA - HW	10.80	1.35	(6.69, 14.90)	7.98	0.000
BATS - HW	4.94	1.35	(0.83, 9.05)	3.65	0.006
TBATS - HW	3.11	1.35	(-1.00, 7.21)	2.29	0.296
NN - HW	-10.04	1.35	(-14.15, -5.94)	-7.42	0.000
SVR - HW	-7.90	1.35	(-12.01, -3.80)	-5.84	0.000
BATS - SARIMA	-5.86	1.35	(-9.96, -1.75)	-4.33	0.000
TBATS - SARIMA	-7.69	1.35	(-11.80, -3.58)	-5.68	0.000
NN - SARIMA	-20.84	1.35	(-24.95, -16.73)	-15.40	0.000
SVR - SARIMA	-18.70	1.35	(-22.80, -14.59)	-13.81	0.000
TBATS - BATS	-1.83	1.35	(-5.94, 2.27)	-1.36	0.877
NN - BATS	-14.99	1.35	(-19.09, -10.88)	-11.07	0.000
SVR - BATS	-12.84	1.35	(-16.95, -8.74)	-9.49	0.000
NN - TBATS	-13.15	1.35	(-17.26, -9.04)	-9.72	0.000
SVR - TBATS	-11.01	1.35	(-15.11, -6.90)	-8.13	0.000
SVR - NN	2.14	1.35	(-1.96, 6.25)	1.58	0.761

Individual confidence level = 99.76%

Figure 40: Randomized complete block designs for all forecasting models

Tukey pairwise comparisons are tested at 95% confidence interval. Only absolute forecasting error values of different methods are of interest. The example of absolute forecasting errors is shown in table 15. Randomized complete block designs are exhibited in figure 40.

Table 15: Absolute Forecasting Error value of hotel daily occupancy with different method

	Absolute Forecasting Error							
Day	SDLY	MA(30)	HW	SARIMA	BATS	TBATS	NN	SVR
1	1	6.85	2.78	18.48	2.98	0.79	0.95	6.85
2	20	1.11	8.67	20.44	7.63	0.10	2.74	1.11
3	27	1.37	0.67	8.22	0.78	0.41	1.46	1.37
4	4	0.19	32.67	40.78	29.20	3.00	0.57	0.19
5	27	7.91	20.33	7.89	4.31	0.64	0.28	7.91
6	2	9.56	36.67	34.93	35.47	1.00	0.50	9.56
7	17	8.30	43.67	71.35	65.64	1.07	2.96	8.30
8	4	18.75	56.67	33.42	53.82	0.99	1.17	18.75
9	8	19.26	50.67	58.48	50.98	1.19	2.58	19.26
10	10	10.50	15.67	13.53	13.14	3.23	0.09	10.50
...	...	...	...	...	...	...	...	...

4.2.1 Disaggregated Data

The following tables show forecasting results disaggregated forecast (table 16, and 17), measured by Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE) for each category: Deluxe, Seaview, Pool Access, Villa. Results show that Neural Network forecasting model is the most accurate model for disaggregated forecast as well.

Table 16: Disaggregated forecasting results according to MAPE

		SDLY	Last day	TBATS	NN
Fully aggregated:	Occupancy	73.15%	22.91%	17.24%	8.955%
Disaggregated:	Deluxe	117.20%	46.68%	13.93%	13.94%
	Seaview	119.09%	33.35%	16.76%	14.05%
	Pool Access	Inf	Inf	Inf	Inf
	Villa	Inf	Inf	Inf	Inf

Table 17: Disaggregated forecasting results according to MAE

		SDLY	Last day	TBATS	NN
Fully aggregated:	Occupancy	34.31	13.28	20.75	5.603
Disaggregated:	Deluxe	10.68	5.63	4.86	2.742
	Seaview	17.49	7.15	8.50	2.553
	Pool Access	5.46	2.30	3.28	0.6819
	Villa	4.96	1.45	4.70	0.6108

4.2.2 Forecasting Horizons

In this section, three-month and one-month forecasting horizons of both aggregated and disaggregated data of the most appropriate model are investigated to observe performance accuracy. This approach will simulate business activity for both aggregated and disaggregated forecasts. According to (Weatherford and Kimes 2003); and (Koupriouchina, van der Rest, and Schwartz 2014), practitioners usually generated forecast frequently i.e. every few days, to obtain near future forecast (1-3 months). Five years of data are used in this experiment. Four years of data are used as training set, and the last year of available data (year 2018) is used for testing set. Forecasting are made 12 times for one-month forecasting horizons, and 4 times for three-month forecasting horizons. For one-month forecasting horizons, data from January 2013 to December 2017 is used as training set, and data from January 2018 is used as testing set. Next, data from February 2013 to January 2018 is used as training set, and data from February 2018 is used as testing set, and so on (figure 5).

The following tables show three-month and one-month aggregated forecasting horizons forecasting horizons results and one-month disaggregated forecasting horizons with Neural Network model (table 18, 19, 20, 21, 22, and 23). Disaggregated three-month forecasting horizons are not continued because results from aggregated forecasting horizons show that one-month forecasting horizons are more accurate than three-month forecasting horizons.

Table 18: Three-month aggregated forecasting horizon results

	Neural Network	
Quarter	MAPE	MAE
Q1	1.059	1.524
Q2	3.055	2.338
Q3	24.82	10.10
Q4	19.13	10.68
Average	15.67	6.160

Table 19: One-month aggregated forecasting horizon results

	Neural Network	
Month	MAPE	MAE
M1	1.144	1.685
M2	1.241	1.850
M3	1.091	1.449
M4	1.387	1.466
M5	2.195	1.445
M6	3.473	2.790

M7	9.683	5.229
M8	14.80	7.592
M9	26.13	8.582
M10	24.37	10.21
M11	12.30	9.784
M12	7.971	10.02
Average	8.815	5.175

Table 20 : One-month forecasting horizon results of Deluxe category

	Neural Network	
Month	MAPE	MAE
M1	1.605	0.7289
M2	1.169	0.5537
M3	1.302	0.5185
M4	3.437	0.7939
M5	10.81	1.231
M6	12.84	2.105
M7	41.34	4.703
M8	31.39	3.713
M9	33.48	4.454

M10	23.18	5.193
M11	13.59	5.182
M12	9.739	4.340
Average	15.32	2.793

It can be observed from table 18, and 19 that the shorter forecasting horizons period provides the lower MAPE (one-month forecasting horizons is better than three-month forecasting horizons). From table 20, presenting one-month forecasting horizon results of Deluxe category. From July to October, MAPE of the forecast pierces to 20-40% due to Chinese boat incident. In this situation the model is not reliable, but number of room occupancy can be expected by management expertise. Also, this incident influenced one-month forecasting horizon of Seaview category (table 21). MAPE of the forecast spikes to around 30% in September and October.

Table 21: One-month forecasting horizon results of Seaview category

	Neural Network	
Month	MAPE	MAE
M1	0.787	0.576
M2	0.750	0.557
M3	0.726	0.528

M4	1.807	0.942
M5	3.233	1.347
M6	1.598	0.861
M7	3.293	1.119
M8	7.719	2.306
M9	30.02	3.563
M10	34.90	3.494
M11	13.85	5.746
M12	4.658	3.114
Average	8.611	2.013

Table 22: One-month forecasting horizon results of Pool Access category

	Neural Network	
Month	MAPE	MAE
M1	0.8174	0.1193
M2	0.9302	0.1343
M3	1.3519	0.1436
M4	10.670	0.4099
M5	Inf	0.2425
M6	Inf	0.2210

M7	Inf	0.2155
M8	Inf	0.6704
M9	Inf	0.8561
M10	Inf	1.184
M11	Inf	1.666
M12	22.838	2.346
Average	-	0.6841

Table 23: One-month forecasting horizon results of Villa category

	Neural Network	
Month	MAPE	MAE
M1	2.843	0.3117
M2	1.701	0.2282
M3	Inf	0.2479
M4	Inf	0.2514
M5	Inf	0.1654
M6	Inf	0.2259
M7	Inf	0.3713
M8	Inf	0.4337
M9	Inf	0.4691

M10	Inf	0.6384
M11	Inf	1.055
M12	Inf	1.953
Average	-	0.5292

4.3 Analysis

BATS and TBATS model are the most sophisticated time series model, incorporating Box-Cox Transformation to handle with non-linear data, ARMA model to account for autocorrelation in time series by residuals, damped trend, level, and multiple seasonal, plus Fourier transform for TBATS. Still, MAPE is high (17.24%) as exhibited by lower level and declining trend, in which is an effect of Chinese boat incident in July 2018.

Therefore, regression models are introduced to solve this problem with explanatory variables that are not encompassed in time series models. Linear Regression is utilized to explore significant variables. Two types of variables are used in this study: independent factors and transformed data. Independent variables are obtained from various sources such as number of tourists, exchange rates, oil price, etc. Transformed data are calculated from the time series data itself. Then, these variables are used as an input for Neural Network model and Support Vector Regression.

For explanatory model such as Linear Regression, Neural Network model and Support Vector Regression, they need regressor variables as inputs of the model. Independent explainable variables are selected and refined by expertise, in which workloads are impinged on practitioners who select and refine those variables. In contrast, transformed variables can be acquired with little or no cost. Results from Linear Regression presents that model with independent variables can explain only 41.73% of patterns in hotel daily data (Adjusted R-squared), while model with transformed data can explain 97.41% of patterns in hotel daily data (Adjusted R-squared). Additionally, the most important aspect in forecasting is model accuracy. Neural Network model with transformed data is proven to be the most accurate (lowest MAPE). Hence, operating explanatory model with transformed variables is more attractive and efficient than with independent input variables.

Table 24: Comparison on types of variables used in explanatory models

Variables	Advantages	Disadvantages
Independent	- Will be an explanation if significant	- Need priori knowledge - No limit on number of variables
Transformed data	- Easy to obtain	- Cannot be used to explain the pattern in data

Table 24 summarizes advantages and disadvantages of types of variables. Independent variables can be an explanation if significant for trend, level, and seasonality pattern if those variables are significant in the model, while transformed data cannot be used to explain and pattern. However, priori knowledge is needed to be able to obtain independent variables, and there are infinite number of independent variables, while transformed data is much easier to construct.

The best result (lowest MAPE of 8.95%) is acquired from Neural Network model with one hidden layer and 20 nodes. This Neural Network model employs resilient backpropagation with and weight backtracking as an algorithm in calculating models' weights. Sum of squared error function is used to compute the model error, and logistic function is used as an activation function.

In the end, all forecasting methods are tested to check that the selected models are statistically significant. One-way ANOVA (figure 40) or the randomized complete block designs (RCBD) at 95% confident interval examines 365-day Mean Absolute Error (MAE) from each forecasting models. Tukey LSD Method presents that mean of MAE of the models are grouped differently. SVR and NN are in the same group, giving the lowest MAE (NN) and second lowest MAE (SVR). It can be observed that SARIMA and TBATS are in the same group with SDLY. Though SDLY has very high MAPE (73.15%) and TBATS has MAPE under 20% (17.24%), RCBD shows that SDLY and TBATS are not statistically different in mean. Thus, TBATS is not appropriate for model selection. Then, this test indicates that NN and SVR are legible for model selection (both can be employed).

Form papers regarding forecasting, experiments were performed to compare between Neural Network and Support Vector Regression. Most of the researches stated that Support Vector Regression outperformed Neural Network. But in this study, the data pattern is disrupted by unusual incident, resulting in the fact that Support Vector Regression does not outperform Neural Network model. One possible reason behind this could be that Neural Network model exercises resilient backpropagation with and weight backtracking techniques, while Support Vector Regression only maps inputs variables and output of the model with non-linear regression (no corrections on model's weight).

The research is subject to few limitations. The main limitation of this research is availability of daily independent data. Authors collected 42 independent variables from various source, yet data such as average room rate of the region is not available. It is expected by expert judgment that average room rate of the region will significantly explain hotel daily occupancy. Secondly, transformed data in regression models cannot be used to explain seasonal or trend in time series data. However, the findings are reliable and indicate that transformed data perform very in forecasting daily hotel time series.

Chapter V: Conclusion and Future Work

5.1 Conclusions

This thesis compares time series and machine learning forecasting model for hotel daily occupancy and provides one-month forecasting results for business purposes. Empirical results in this study show that Neural Network model with transformed data as regressor variables outperforms all other methods. This thesis was divided into two parts.

The first part was to test hotel daily occupancy with time series model namely, Holt-Winters' exponential smoothing, SARIMA, BATS, and TBATS model. The results show that TBATS model performs the best among time series models with MAPE 17.24%. However, though BATS and TBATS model are the most sophisticated time series model, incorporating Box-Cox Transformation to handle with non-linear data, ARMA model to account for autocorrelation in time series by residuals, damped trend, level, and multiple seasonal, plus Fourier transform for TBATS. Still, MAPE is high (17.24%) as exhibited by lower level and declining trend, in which is an effect of Chinese boat incident in July 2018.

The second part was to apply machine learning model incorporated with transformed data to solve this problem, namely Artificial Neural Network, and Support Vector Regression. Results from Linear Regression presents that model with independent variables can explain only 41.73% of patterns in hotel daily data (Adjusted R-squared), while model with transformed data can explain 97.41% of patterns in hotel daily data (Adjusted R-squared).

Empirical results show that Neural Network model with transformed data is proven to be the most accurate model (lowest MAPE of 8.955%). The use of transformed data has significantly improved accuracy, in contrast to Neural Network model with independent factors. Benefits of using transformed data as inputs for machine learning model are as followings

1. Transformed variables can be acquired with little or no cost by manipulating existing time series data.
2. Transformed variables as regressor variables gives considerably accuracy improvement.
3. Transformed variables in Neural Network model requires a smaller number of nodes in the structure (transformed data 20 nodes, independent factors 100 nodes).

5.2 Practical Contributions

This study makes significant and practical contributions to tourism demand forecasting in the following aspects.

(1) This thesis focused on finding the most suitable model to forecast daily hotel room demand occupancy with substantial changes in level and

trend, while most previous work for hotel room demand had focused on monthly period.

(2) This thesis compared performance of time series models and regression models and recommend the optimal model for predicting daily hotel room demand. Forecasting horizon insights are also provided for future model selection. It was shown that the Neural Network model can produce high levels of forecasting accuracy (with MAPE ≤ 10%) and outperform time series models and explanatory models.

(3) This thesis introduced the use of data transformation in contrast with the use of traditional independent variables. Results showed that Neural Network model using independent variables did not provide acceptable forecasting accuracy, while Neural Network model using transformed data could offer high levels of forecasting accuracy (with MAPE ≤ 10%).

5.3 Recommendations for Future Work

There are several areas that future studies can investigate.

(1) The idea can be extended to cover other hotels in the region or in the country.

(2) Future researchers could consider analysis of different data transformation upon daily hotel room demand occupancy that could be

useful to assess the improvement of hotel revenue management implementation.

(3) Future research may involve forecasting other attributes regarding hotel revenue management such as sale prices, booking profile, cancellation rate, etc. to generate broader insights from forecasts.

(4) Future research may explore an extra hidden layer in Neural Network structure.

Appendix

Pairwise Pearson Correlations

Sample 1	Sample 2	Correlation	95% CI for ρ	P-Value
Russia	Occupancy	0.302	(0.263, 0.339)	0.000
Australia	Occupancy	-0.285	(-0.323, -0.246)	0.000
Korea	Occupancy	0.216	(0.176, 0.256)	0.000
Malaysia	Occupancy	-0.179	(-0.219, -0.138)	0.000
Singapore	Occupancy	-0.329	(-0.366, -0.291)	0.000
Sweden	Occupancy	0.378	(0.342, 0.413)	0.000
UK	Occupancy	0.171	(0.130, 0.212)	0.000
CPISouth	Occupancy	-0.199	(-0.239, -0.159)	0.000
OilPrice	Occupancy	-0.023	(-0.065, 0.019)	0.279
USD	Occupancy	-0.035	(-0.076, 0.007)	0.105
CNY	Occupancy	0.180	(0.139, 0.220)	0.000
RUB	Occupancy	0.034	(-0.008, 0.076)	0.110
AUD	Occupancy	0.176	(0.135, 0.216)	0.000
KRW	Occupancy	0.031	(-0.011, 0.072)	0.152
MYR	Occupancy	0.139	(0.098, 0.180)	0.000
SGD	Occupancy	0.049	(0.007, 0.091)	0.022
SEK	Occupancy	0.168	(0.127, 0.208)	0.000
GBP	Occupancy	0.159	(0.118, 0.199)	0.000
ARR	Occupancy	0.258	(0.219, 0.297)	0.000
Thursday	Occupancy	0.004	(-0.038, 0.046)	0.847
Friday	Occupancy	0.010	(-0.032, 0.052)	0.636
Saturday	Occupancy	-0.016	(-0.058, 0.026)	0.448
Sunday	Occupancy	0.007	(-0.035, 0.049)	0.745
Monday	Occupancy	0.013	(-0.029, 0.054)	0.555
Tuesday	Occupancy	-0.019	(-0.061, 0.023)	0.375
June	Occupancy	-0.217	(-0.257, -0.177)	0.000
July	Occupancy	-0.054	(-0.096, -0.013)	0.011
August	Occupancy	0.021	(-0.021, 0.063)	0.324
September	Occupancy	-0.116	(-0.158, -0.075)	0.000
October	Occupancy	-0.085	(-0.127, -0.044)	0.000
November	Occupancy	-0.042	(-0.084, -0.000)	0.049
December	Occupancy	0.052	(0.010, 0.093)	0.016
January	Occupancy	0.210	(0.170, 0.250)	0.000
Febuary	Occupancy	0.238	(0.198, 0.277)	0.000
March	Occupancy	0.134	(0.093, 0.175)	0.000
April	Occupancy	0.026	(-0.016, 0.068)	0.225
Monsoon	Occupancy	-0.266	(-0.304, -0.226)	0.000
Peak	Occupancy	0.156	(0.115, 0.196)	0.000
High	Occupancy	0.309	(0.271, 0.347)	0.000
RusSanc	Occupancy	-0.168	(-0.209, -0.127)	0.000
ChiBoatSink	Occupancy	-0.344	(-0.381, -0.307)	0.000
ChiFreeVisa	Occupancy	0.052	(0.010, 0.094)	0.015
Occ-EMA360	Occupancy	0.981	(0.979, 0.983)	0.000
RDB-7	Occupancy	0.145	(0.104, 0.186)	0.000
RDB-30	Occupancy	0.206	(0.165, 0.245)	0.000

RDB-52	Occupancy	0.264	(0.224, 0.302)	0.000
RDB-365	Occupancy	0.259	(0.220, 0.298)	0.000
Australia	Russia	-0.420	(-0.454, -0.385)	0.000
Korea	Russia	-0.141	(-0.181, -0.099)	0.000
Malaysia	Russia	-0.237	(-0.276, -0.197)	0.000
Singapore	Russia	0.209	(0.168, 0.248)	0.000
Sweden	Russia	0.786	(0.770, 0.802)	0.000
UK	Russia	0.671	(0.647, 0.693)	0.000
CPISouth	Russia	0.263	(0.223, 0.301)	0.000
OilPrice	Russia	-0.023	(-0.065, 0.019)	0.277
USD	Russia	-0.092	(-0.133, -0.050)	0.000
CNY	Russia	-0.183	(-0.223, -0.142)	0.000
RUB	Russia	0.023	(-0.018, 0.065)	0.272
AUD	Russia	-0.046	(-0.088, -0.004)	0.030
KRW	Russia	-0.008	(-0.050, 0.034)	0.713
MYR	Russia	-0.131	(-0.172, -0.089)	0.000
SGD	Russia	-0.162	(-0.203, -0.121)	0.000
SEK	Russia	-0.042	(-0.083, 0.000)	0.052
GBP	Russia	-0.190	(-0.230, -0.149)	0.000
ARR	Russia	0.665	(0.641, 0.688)	0.000
Thursday	Russia	-0.001	(-0.043, 0.041)	0.954
Friday	Russia	0.001	(-0.041, 0.043)	0.960
Saturday	Russia	0.003	(-0.039, 0.045)	0.889
Sunday	Russia	0.002	(-0.040, 0.044)	0.936
Monday	Russia	0.000	(-0.042, 0.042)	0.988
Tuesday	Russia	-0.003	(-0.045, 0.039)	0.897
June	Russia	-0.297	(-0.334, -0.258)	0.000
July	Russia	-0.288	(-0.326, -0.249)	0.000
August	Russia	-0.275	(-0.313, -0.236)	0.000
September	Russia	-0.281	(-0.319, -0.242)	0.000
October	Russia	-0.053	(-0.095, -0.011)	0.013
November	Russia	0.253	(0.213, 0.292)	0.000
December	Russia	0.439	(0.404, 0.472)	0.000
January	Russia	0.348	(0.310, 0.384)	0.000
Febuary	Russia	0.180	(0.139, 0.220)	0.000
March	Russia	0.228	(0.187, 0.267)	0.000
April	Russia	0.005	(-0.037, 0.047)	0.826
Monsoon	Russia	-0.716	(-0.736, -0.695)	0.000
Peak	Russia	0.389	(0.352, 0.424)	0.000
High	Russia	0.761	(0.742, 0.778)	0.000
RusSanc	Russia	-0.089	(-0.131, -0.048)	0.000
ChiBoatSink	Russia	0.115	(0.074, 0.157)	0.000
ChiFreeVisa	Russia	0.322	(0.284, 0.359)	0.000
Occ-EMA360	Russia	0.349	(0.312, 0.385)	0.000
RDB-7	Russia	-0.029	(-0.071, 0.013)	0.171
RDB-30	Russia	0.068	(0.026, 0.110)	0.001
RDB-52	Russia	0.127	(0.085, 0.168)	0.000
RDB-365	Russia	-0.033	(-0.074, 0.009)	0.128
Korea	Australia	-0.038	(-0.080, 0.003)	0.072
Malaysia	Australia	0.012	(-0.030, 0.054)	0.583
Singapore	Australia	0.079	(0.038, 0.121)	0.000
Sweden	Australia	-0.591	(-0.618, -0.564)	0.000

UK	Australia	-0.133	(-0.174, -0.092)	0.000
CPISouth	Australia	0.044	(0.002, 0.085)	0.041
OilPrice	Australia	0.164	(0.123, 0.204)	0.000
USD	Australia	-0.064	(-0.106, -0.022)	0.003
CNY	Australia	-0.081	(-0.122, -0.039)	0.000
RUB	Australia	-0.056	(-0.097, -0.014)	0.009
AUD	Australia	-0.004	(-0.046, 0.038)	0.854
KRW	Australia	-0.001	(-0.043, 0.041)	0.956
MYR	Australia	0.024	(-0.018, 0.066)	0.262
SGD	Australia	0.070	(0.029, 0.112)	0.001
SEK	Australia	0.045	(0.003, 0.086)	0.037
GBP	Australia	0.000	(-0.042, 0.042)	0.990
ARR	Australia	-0.448	(-0.481, -0.414)	0.000
Thursday	Australia	-0.005	(-0.047, 0.037)	0.801
Friday	Australia	-0.007	(-0.049, 0.035)	0.754
Saturday	Australia	-0.001	(-0.043, 0.041)	0.954
Sunday	Australia	0.006	(-0.036, 0.048)	0.791
Monday	Australia	0.005	(-0.037, 0.047)	0.808
Tuesday	Australia	0.003	(-0.039, 0.045)	0.880
June	Australia	0.146	(0.104, 0.186)	0.000
July	Australia	0.467	(0.433, 0.499)	0.000
August	Australia	-0.057	(-0.098, -0.015)	0.008
September	Australia	0.233	(0.193, 0.272)	0.000
October	Australia	0.280	(0.241, 0.318)	0.000
November	Australia	-0.194	(-0.234, -0.153)	0.000
December	Australia	-0.106	(-0.147, -0.065)	0.000
January	Australia	-0.134	(-0.175, -0.092)	0.000
Febuary	Australia	-0.537	(-0.566, -0.506)	0.000
March	Australia	-0.217	(-0.257, -0.177)	0.000
April	Australia	0.317	(0.279, 0.354)	0.000
Monsoon	Australia	0.560	(0.531, 0.588)	0.000
Peak	Australia	-0.126	(-0.167, -0.084)	0.000
High	Australia	-0.606	(-0.632, -0.579)	0.000
RusSanc	Australia	0.040	(-0.002, 0.082)	0.060
ChiBoatSink	Australia	0.092	(0.050, 0.133)	0.000
ChiFreeVisa	Australia	-0.035	(-0.077, 0.007)	0.103
Occ-EMA360	Australia	-0.287	(-0.325, -0.248)	0.000
RDB-7	Australia	0.062	(0.020, 0.103)	0.004
RDB-30	Australia	-0.011	(-0.053, 0.031)	0.600
RDB-52	Australia	-0.056	(-0.098, -0.014)	0.009
RDB-365	Australia	-0.012	(-0.054, 0.030)	0.570
Malaysia	Korea	-0.116	(-0.157, -0.075)	0.000
Singapore	Korea	-0.272	(-0.310, -0.233)	0.000
Sweden	Korea	0.038	(-0.004, 0.080)	0.075
UK	Korea	-0.103	(-0.144, -0.061)	0.000
CPISouth	Korea	-0.213	(-0.252, -0.172)	0.000
OilPrice	Korea	0.157	(0.116, 0.197)	0.000
USD	Korea	0.022	(-0.020, 0.064)	0.307
CNY	Korea	0.118	(0.076, 0.159)	0.000
RUB	Korea	0.008	(-0.034, 0.049)	0.721
AUD	Korea	0.325	(0.287, 0.362)	0.000
KRW	Korea	0.097	(0.055, 0.138)	0.000

MYR	Korea	0.229	(0.189, 0.268)	0.000
SGD	Korea	0.230	(0.190, 0.270)	0.000
SEK	Korea	0.335	(0.298, 0.372)	0.000
GBP	Korea	0.105	(0.063, 0.146)	0.000
ARR	Korea	0.018	(-0.024, 0.060)	0.405
Thursday	Korea	0.001	(-0.041, 0.043)	0.951
Friday	Korea	0.001	(-0.041, 0.043)	0.956
Saturday	Korea	-0.001	(-0.043, 0.041)	0.968
Sunday	Korea	-0.003	(-0.045, 0.039)	0.902
Monday	Korea	-0.001	(-0.043, 0.040)	0.947
Tuesday	Korea	0.001	(-0.041, 0.043)	0.962
June	Korea	-0.153	(-0.194, -0.112)	0.000
July	Korea	0.171	(0.130, 0.211)	0.000
August	Korea	0.575	(0.546, 0.602)	0.000
September	Korea	-0.104	(-0.145, -0.062)	0.000
October	Korea	-0.046	(-0.087, -0.004)	0.033
November	Korea	-0.163	(-0.203, -0.121)	0.000
December	Korea	0.058	(0.016, 0.099)	0.007
January	Korea	0.148	(0.107, 0.189)	0.000
Febuary	Korea	0.051	(0.009, 0.093)	0.016
March	Korea	-0.230	(-0.269, -0.190)	0.000
April	Korea	-0.177	(-0.218, -0.137)	0.000
Monsoon	Korea	-0.056	(-0.098, -0.014)	0.008
Peak	Korea	0.116	(0.075, 0.157)	0.000
High	Korea	-0.009	(-0.051, 0.032)	0.657
RusSanc	Korea	-0.321	(-0.358, -0.283)	0.000
ChiBoatSink	Korea	-0.172	(-0.212, -0.131)	0.000
ChiFreeVisa	Korea	-0.080	(-0.122, -0.038)	0.000
Occ-EMA360	Korea	0.155	(0.114, 0.196)	0.000
RDB-7	Korea	-0.051	(-0.092, -0.009)	0.018
RDB-30	Korea	0.036	(-0.006, 0.078)	0.092
RDB-52	Korea	0.092	(0.050, 0.133)	0.000
RDB-365	Korea	-0.035	(-0.077, 0.007)	0.100
Singapore	Malaysia	0.481	(0.449, 0.513)	0.000
Sweden	Malaysia	-0.309	(-0.346, -0.270)	0.000
UK	Malaysia	0.080	(0.038, 0.121)	0.000
CPISouth	Malaysia	0.086	(0.044, 0.128)	0.000
OilPrice	Malaysia	-0.153	(-0.193, -0.112)	0.000
USD	Malaysia	0.099	(0.058, 0.141)	0.000
CNY	Malaysia	-0.059	(-0.101, -0.017)	0.006
RUB	Malaysia	-0.051	(-0.092, -0.009)	0.018
AUD	Malaysia	-0.261	(-0.300, -0.222)	0.000
KRW	Malaysia	-0.030	(-0.072, 0.012)	0.159
MYR	Malaysia	-0.215	(-0.254, -0.174)	0.000
SGD	Malaysia	-0.104	(-0.145, -0.063)	0.000
SEK	Malaysia	-0.305	(-0.343, -0.267)	0.000
GBP	Malaysia	-0.186	(-0.226, -0.145)	0.000
ARR	Malaysia	-0.366	(-0.401, -0.329)	0.000
Thursday	Malaysia	0.007	(-0.035, 0.049)	0.740
Friday	Malaysia	0.008	(-0.034, 0.049)	0.725
Saturday	Malaysia	0.001	(-0.041, 0.043)	0.958
Sunday	Malaysia	-0.004	(-0.046, 0.038)	0.840

Monday	Malaysia	-0.005	(-0.047, 0.037)	0.822
Tuesday	Malaysia	-0.004	(-0.046, 0.038)	0.848
June	Malaysia	0.344	(0.306, 0.380)	0.000
July	Malaysia	-0.095	(-0.136, -0.053)	0.000
August	Malaysia	0.064	(0.022, 0.106)	0.003
September	Malaysia	-0.026	(-0.068, 0.016)	0.225
October	Malaysia	-0.126	(-0.167, -0.084)	0.000
November	Malaysia	0.028	(-0.013, 0.070)	0.183
December	Malaysia	0.336	(0.298, 0.372)	0.000
January	Malaysia	-0.496	(-0.527, -0.463)	0.000
Febuary	Malaysia	-0.306	(-0.343, -0.267)	0.000
March	Malaysia	-0.023	(-0.064, 0.019)	0.289
April	Malaysia	-0.008	(-0.050, 0.034)	0.715
Monsoon	Malaysia	0.252	(0.212, 0.291)	0.000
Peak	Malaysia	-0.187	(-0.227, -0.146)	0.000
High	Malaysia	-0.251	(-0.290, -0.212)	0.000
RusSanc	Malaysia	0.176	(0.135, 0.216)	0.000
ChiBoatSink	Malaysia	0.042	(-0.000, 0.083)	0.052
ChiFreeVisa	Malaysia	0.084	(0.043, 0.126)	0.000
Occ-EMA360	Malaysia	-0.166	(-0.206, -0.125)	0.000
RDB-7	Malaysia	0.018	(-0.024, 0.060)	0.395
RDB-30	Malaysia	0.063	(0.021, 0.105)	0.003
RDB-52	Malaysia	0.062	(0.020, 0.103)	0.004
RDB-365	Malaysia	0.093	(0.051, 0.134)	0.000
Sweden	Singapore	-0.013	(-0.055, 0.029)	0.546
UK	Singapore	0.211	(0.170, 0.250)	0.000
CPISouth	Singapore	0.700	(0.678, 0.721)	0.000
OilPrice	Singapore	-0.017	(-0.059, 0.025)	0.416
USD	Singapore	-0.109	(-0.150, -0.067)	0.000
CNY	Singapore	-0.411	(-0.446, -0.376)	0.000
RUB	Singapore	-0.025	(-0.067, 0.017)	0.238
AUD	Singapore	-0.397	(-0.431, -0.361)	0.000
KRW	Singapore	-0.047	(-0.088, -0.005)	0.030
MYR	Singapore	-0.313	(-0.350, -0.275)	0.000
SGD	Singapore	-0.331	(-0.368, -0.293)	0.000
SEK	Singapore	-0.349	(-0.386, -0.312)	0.000
GBP	Singapore	-0.347	(-0.384, -0.310)	0.000
ARR	Singapore	-0.113	(-0.154, -0.071)	0.000
Thursday	Singapore	0.002	(-0.040, 0.044)	0.923
Friday	Singapore	0.001	(-0.041, 0.043)	0.958
Saturday	Singapore	0.002	(-0.040, 0.044)	0.921
Sunday	Singapore	0.001	(-0.041, 0.043)	0.963
Monday	Singapore	-0.000	(-0.042, 0.042)	0.995
Tuesday	Singapore	-0.004	(-0.046, 0.038)	0.850
June	Singapore	0.312	(0.274, 0.349)	0.000
July	Singapore	-0.177	(-0.217, -0.136)	0.000
August	Singapore	0.021	(-0.021, 0.062)	0.335
September	Singapore	0.000	(-0.042, 0.042)	0.992
October	Singapore	0.015	(-0.027, 0.057)	0.477
November	Singapore	0.108	(0.067, 0.150)	0.000
December	Singapore	0.478	(0.445, 0.510)	0.000
January	Singapore	-0.252	(-0.291, -0.212)	0.000

Febuary	Singapore	-0.252	(-0.290, -0.212)	0.000
March	Singapore	0.018	(-0.024, 0.060)	0.407
April	Singapore	-0.194	(-0.234, -0.154)	0.000
Monsoon	Singapore	-0.055	(-0.096, -0.013)	0.011
Peak	Singapore	0.022	(-0.020, 0.064)	0.300
High	Singapore	0.058	(0.016, 0.100)	0.007
RusSanc	Singapore	0.163	(0.122, 0.204)	0.000
ChiBoatSink	Singapore	0.685	(0.662, 0.706)	0.000
ChiFreeVisa	Singapore	0.493	(0.461, 0.524)	0.000
Occ-EMA360	Singapore	-0.227	(-0.267, -0.187)	0.000
RDB-7	Singapore	0.047	(0.005, 0.088)	0.029
RDB-30	Singapore	0.118	(0.076, 0.159)	0.000
RDB-52	Singapore	0.146	(0.105, 0.187)	0.000
RDB-365	Singapore	-0.020	(-0.062, 0.022)	0.354
UK	Sweden	0.510	(0.478, 0.540)	0.000
CPISouth	Sweden	0.012	(-0.030, 0.053)	0.587
OilPrice	Sweden	-0.105	(-0.146, -0.063)	0.000
USD	Sweden	0.018	(-0.024, 0.060)	0.408
CNY	Sweden	0.022	(-0.019, 0.064)	0.294
RUB	Sweden	0.029	(-0.013, 0.071)	0.173
AUD	Sweden	0.033	(-0.009, 0.074)	0.126
KRW	Sweden	0.004	(-0.038, 0.046)	0.853
MYR	Sweden	-0.044	(-0.086, -0.003)	0.038
SGD	Sweden	-0.104	(-0.145, -0.062)	0.000
SEK	Sweden	0.033	(-0.008, 0.075)	0.118
GBP	Sweden	-0.006	(-0.048, 0.036)	0.766
ARR	Sweden	0.775	(0.758, 0.791)	0.000
Thursday	Sweden	0.001	(-0.041, 0.042)	0.979
Friday	Sweden	0.001	(-0.041, 0.043)	0.972
Saturday	Sweden	0.001	(-0.041, 0.043)	0.977
Sunday	Sweden	0.000	(-0.042, 0.042)	0.999
Monday	Sweden	-0.001	(-0.043, 0.041)	0.969
Tuesday	Sweden	0.001	(-0.041, 0.042)	0.980
June	Sweden	-0.238	(-0.277, -0.198)	0.000
July	Sweden	-0.239	(-0.278, -0.199)	0.000
August	Sweden	-0.251	(-0.290, -0.211)	0.000
September	Sweden	-0.244	(-0.283, -0.204)	0.000
October	Sweden	-0.188	(-0.228, -0.147)	0.000
November	Sweden	-0.014	(-0.056, 0.028)	0.509
December	Sweden	0.453	(0.419, 0.485)	0.000
January	Sweden	0.510	(0.478, 0.540)	0.000
Febuary	Sweden	0.398	(0.362, 0.433)	0.000
March	Sweden	0.277	(0.238, 0.316)	0.000
April	Sweden	-0.210	(-0.250, -0.170)	0.000
Monsoon	Sweden	-0.789	(-0.804, -0.772)	0.000
Peak	Sweden	0.497	(0.465, 0.528)	0.000
High	Sweden	0.845	(0.833, 0.857)	0.000
RusSanc	Sweden	-0.126	(-0.167, -0.084)	0.000
ChiBoatSink	Sweden	-0.077	(-0.119, -0.035)	0.000
ChiFreeVisa	Sweden	0.134	(0.093, 0.175)	0.000
Occ-EMA360	Sweden	0.394	(0.358, 0.429)	0.000
RDB-7	Sweden	-0.056	(-0.098, -0.015)	0.008

RDB-30	Sweden	0.039	(-0.003, 0.080)	0.070
RDB-52	Sweden	0.089	(0.048, 0.131)	0.000
RDB-365	Sweden	-0.010	(-0.052, 0.031)	0.625
CPISouth	UK	0.178	(0.138, 0.219)	0.000
OilPrice	UK	-0.381	(-0.416, -0.345)	0.000
USD	UK	0.270	(0.231, 0.309)	0.000
CNY	UK	-0.109	(-0.151, -0.068)	0.000
RUB	UK	-0.016	(-0.058, 0.026)	0.450
AUD	UK	-0.341	(-0.378, -0.304)	0.000
KRW	UK	-0.010	(-0.051, 0.032)	0.656
MYR	UK	-0.466	(-0.498, -0.432)	0.000
SGD	UK	-0.129	(-0.170, -0.087)	0.000
SEK	UK	-0.359	(-0.395, -0.322)	0.000
GBP	UK	-0.315	(-0.352, -0.277)	0.000
ARR	UK	0.306	(0.268, 0.344)	0.000
Thursday	UK	-0.001	(-0.043, 0.041)	0.966
Friday	UK	-0.002	(-0.044, 0.040)	0.921
Saturday	UK	0.002	(-0.039, 0.044)	0.911
Sunday	UK	0.002	(-0.040, 0.044)	0.916
Monday	UK	-0.001	(-0.043, 0.041)	0.952
Tuesday	UK	-0.001	(-0.043, 0.041)	0.965
June	UK	-0.289	(-0.326, -0.250)	0.000
July	UK	-0.057	(-0.098, -0.015)	0.008
August	UK	-0.128	(-0.169, -0.087)	0.000
September	UK	-0.402	(-0.437, -0.366)	0.000
October	UK	-0.052	(-0.093, -0.010)	0.016
November	UK	0.158	(0.117, 0.199)	0.000
December	UK	0.489	(0.456, 0.520)	0.000
January	UK	0.012	(-0.029, 0.054)	0.562
Febuary	UK	0.040	(-0.002, 0.081)	0.063
March	UK	0.250	(0.210, 0.289)	0.000
April	UK	0.266	(0.227, 0.304)	0.000
Monsoon	UK	-0.412	(-0.446, -0.377)	0.000
Peak	UK	0.194	(0.154, 0.234)	0.000
High	UK	0.471	(0.438, 0.503)	0.000
RusSanc	UK	0.301	(0.263, 0.339)	0.000
ChiBoatSink	UK	0.072	(0.030, 0.113)	0.001
ChiFreeVisa	UK	0.208	(0.167, 0.247)	0.000
Occ-EMA360	UK	0.227	(0.187, 0.267)	0.000
RDB-7	UK	0.003	(-0.039, 0.045)	0.890
RDB-30	UK	0.108	(0.066, 0.149)	0.000
RDB-52	UK	0.162	(0.121, 0.202)	0.000
RDB-365	UK	-0.006	(-0.048, 0.036)	0.776
OilPrice	CPISouth	-0.085	(-0.126, -0.043)	0.000
USD	CPISouth	-0.058	(-0.100, -0.016)	0.006
CNY	CPISouth	-0.422	(-0.456, -0.387)	0.000
RUB	CPISouth	-0.035	(-0.077, 0.007)	0.098
AUD	CPISouth	-0.396	(-0.431, -0.360)	0.000
KRW	CPISouth	-0.020	(-0.061, 0.022)	0.358
MYR	CPISouth	-0.297	(-0.335, -0.259)	0.000
SGD	CPISouth	-0.331	(-0.367, -0.293)	0.000
SEK	CPISouth	-0.368	(-0.403, -0.331)	0.000

GBP	CPISouth	-0.344	(-0.380, -0.306)	0.000
ARR	CPISouth	-0.055	(-0.097, -0.013)	0.010
Thursday	CPISouth	-0.002	(-0.043, 0.040)	0.943
Friday	CPISouth	-0.000	(-0.042, 0.042)	0.998
Saturday	CPISouth	0.001	(-0.040, 0.043)	0.945
Sunday	CPISouth	0.003	(-0.039, 0.045)	0.889
Monday	CPISouth	0.005	(-0.037, 0.046)	0.832
Tuesday	CPISouth	-0.004	(-0.046, 0.037)	0.835
June	CPISouth	-0.053	(-0.095, -0.012)	0.012
July	CPISouth	-0.040	(-0.082, 0.002)	0.059
August	CPISouth	-0.018	(-0.060, 0.024)	0.390
September	CPISouth	0.014	(-0.028, 0.056)	0.509
October	CPISouth	0.064	(0.022, 0.106)	0.003
November	CPISouth	0.137	(0.095, 0.177)	0.000
December	CPISouth	0.253	(0.213, 0.292)	0.000
January	CPISouth	-0.079	(-0.120, -0.037)	0.000
Febuary	CPISouth	-0.073	(-0.114, -0.031)	0.001
March	CPISouth	-0.074	(-0.116, -0.032)	0.001
April	CPISouth	-0.069	(-0.110, -0.027)	0.001
Monsoon	CPISouth	-0.142	(-0.183, -0.101)	0.000
Peak	CPISouth	0.063	(0.021, 0.105)	0.003
High	CPISouth	0.117	(0.076, 0.158)	0.000
RusSanc	CPISouth	0.128	(0.086, 0.169)	0.000
ChiBoatSink	CPISouth	0.845	(0.833, 0.857)	0.000
ChiFreeVisa	CPISouth	0.753	(0.734, 0.770)	0.000
Occ-EMA360	CPISouth	-0.064	(-0.106, -0.022)	0.003
RDB-7	CPISouth	0.046	(0.004, 0.088)	0.032
RDB-30	CPISouth	0.147	(0.106, 0.188)	0.000
RDB-52	CPISouth	0.222	(0.181, 0.261)	0.000
RDB-365	CPISouth	-0.093	(-0.135, -0.051)	0.000
USD	OilPrice	-0.756	(-0.774, -0.738)	0.000
CNY	OilPrice	-0.257	(-0.295, -0.217)	0.000
RUB	OilPrice	0.044	(0.002, 0.085)	0.041
AUD	OilPrice	0.741	(0.721, 0.759)	0.000
KRW	OilPrice	-0.043	(-0.085, -0.001)	0.042
MYR	OilPrice	0.757	(0.739, 0.775)	0.000
SGD	OilPrice	0.014	(-0.028, 0.056)	0.504
SEK	OilPrice	0.722	(0.701, 0.741)	0.000
GBP	OilPrice	0.243	(0.203, 0.282)	0.000
ARR	OilPrice	0.152	(0.111, 0.193)	0.000
Thursday	OilPrice	0.000	(-0.042, 0.042)	0.995
Friday	OilPrice	0.000	(-0.042, 0.042)	0.989
Saturday	OilPrice	0.000	(-0.042, 0.042)	0.989
Sunday	OilPrice	0.000	(-0.041, 0.042)	0.985
Monday	OilPrice	-0.001	(-0.043, 0.040)	0.946
Tuesday	OilPrice	0.000	(-0.042, 0.042)	0.999
June	OilPrice	0.050	(0.008, 0.091)	0.020
July	OilPrice	0.049	(0.007, 0.091)	0.021
August	OilPrice	0.018	(-0.024, 0.060)	0.400
September	OilPrice	0.023	(-0.019, 0.065)	0.280
October	OilPrice	0.011	(-0.031, 0.052)	0.620
November	OilPrice	-0.060	(-0.102, -0.018)	0.005

December	OilPrice	-0.101	(-0.142, -0.059)	0.000
January	OilPrice	-0.028	(-0.070, 0.014)	0.188
Febuary	OilPrice	-0.014	(-0.056, 0.027)	0.499
March	OilPrice	-0.015	(-0.056, 0.027)	0.497
April	OilPrice	0.017	(-0.025, 0.059)	0.430
Monsoon	OilPrice	0.120	(0.078, 0.161)	0.000
Peak	OilPrice	-0.063	(-0.104, -0.021)	0.003
High	OilPrice	-0.122	(-0.163, -0.080)	0.000
RusSanc	OilPrice	-0.674	(-0.696, -0.650)	0.000
ChiBoatSink	OilPrice	-0.044	(-0.085, -0.002)	0.042
ChiFreeVisa	OilPrice	-0.092	(-0.134, -0.051)	0.000
Occ-EMA360	OilPrice	-0.063	(-0.105, -0.021)	0.003
RDB-7	OilPrice	-0.018	(-0.060, 0.024)	0.405
RDB-30	OilPrice	-0.024	(-0.066, 0.018)	0.257
RDB-52	OilPrice	-0.038	(-0.080, 0.004)	0.074
RDB-365	OilPrice	-0.024	(-0.066, 0.017)	0.252
CNY	USD	0.616	(0.589, 0.641)	0.000
RUB	USD	-0.073	(-0.114, -0.031)	0.001
AUD	USD	-0.455	(-0.487, -0.421)	0.000
KRW	USD	0.046	(0.004, 0.087)	0.033
MYR	USD	-0.481	(-0.513, -0.449)	0.000
SGD	USD	0.521	(0.490, 0.551)	0.000
SEK	USD	-0.329	(-0.366, -0.291)	0.000
GBP	USD	0.159	(0.118, 0.199)	0.000
ARR	USD	-0.159	(-0.199, -0.118)	0.000
Thursday	USD	-0.001	(-0.043, 0.041)	0.949
Friday	USD	0.000	(-0.042, 0.042)	0.999
Saturday	USD	0.000	(-0.042, 0.042)	0.988
Sunday	USD	0.000	(-0.042, 0.042)	0.988
Monday	USD	0.001	(-0.041, 0.042)	0.979
Tuesday	USD	-0.000	(-0.042, 0.041)	0.982
June	USD	-0.005	(-0.047, 0.037)	0.804
July	USD	0.017	(-0.025, 0.059)	0.422
August	USD	0.031	(-0.011, 0.072)	0.151
September	USD	0.039	(-0.003, 0.081)	0.067
October	USD	0.042	(-0.000, 0.083)	0.052
November	USD	0.068	(0.026, 0.110)	0.001
December	USD	0.100	(0.058, 0.141)	0.000
January	USD	0.009	(-0.033, 0.051)	0.679
Febuary	USD	-0.049	(-0.090, -0.007)	0.023
March	USD	-0.088	(-0.130, -0.047)	0.000
April	USD	-0.121	(-0.162, -0.079)	0.000
Monsoon	USD	-0.076	(-0.118, -0.035)	0.000
Peak	USD	0.062	(0.020, 0.103)	0.004
High	USD	0.054	(0.013, 0.096)	0.011
RusSanc	USD	0.610	(0.583, 0.635)	0.000
ChiBoatSink	USD	-0.052	(-0.094, -0.010)	0.015
ChiFreeVisa	USD	-0.031	(-0.073, 0.011)	0.145
Occ-EMA360	USD	-0.046	(-0.088, -0.004)	0.031
RDB-7	USD	0.034	(-0.008, 0.075)	0.115
RDB-30	USD	0.049	(0.008, 0.091)	0.021
RDB-52	USD	0.052	(0.010, 0.094)	0.015

RDB-365	USD	-0.041	(-0.083, 0.001)	0.056
RUB	CNY	-0.040	(-0.082, 0.002)	0.059
AUD	CNY	0.059	(0.017, 0.101)	0.006
KRW	CNY	0.032	(-0.010, 0.074)	0.133
MYR	CNY	0.249	(0.210, 0.288)	0.000
SGD	CNY	0.752	(0.733, 0.770)	0.000
SEK	CNY	0.292	(0.253, 0.329)	0.000
GBP	CNY	0.798	(0.782, 0.812)	0.000
ARR	CNY	0.001	(-0.041, 0.042)	0.977
Thursday	CNY	-0.001	(-0.042, 0.041)	0.977
Friday	CNY	-0.000	(-0.042, 0.042)	0.997
Saturday	CNY	0.000	(-0.042, 0.042)	0.989
Sunday	CNY	0.000	(-0.042, 0.042)	0.989
Monday	CNY	0.001	(-0.041, 0.043)	0.974
Tuesday	CNY	-0.001	(-0.043, 0.041)	0.958
June	CNY	0.023	(-0.018, 0.065)	0.272
July	CNY	0.004	(-0.038, 0.045)	0.867
August	CNY	0.008	(-0.034, 0.050)	0.703
September	CNY	0.028	(-0.014, 0.069)	0.195
October	CNY	0.007	(-0.035, 0.049)	0.754
November	CNY	0.015	(-0.026, 0.057)	0.470
December	CNY	0.027	(-0.015, 0.068)	0.214
January	CNY	0.049	(0.007, 0.091)	0.021
Febuary	CNY	-0.006	(-0.048, 0.036)	0.778
March	CNY	-0.059	(-0.101, -0.017)	0.006
April	CNY	-0.086	(-0.127, -0.044)	0.000
Monsoon	CNY	-0.051	(-0.092, -0.009)	0.018
Peak	CNY	0.054	(0.012, 0.096)	0.011
High	CNY	0.037	(-0.005, 0.078)	0.087
RusSanc	CNY	0.191	(0.150, 0.231)	0.000
ChiBoatSink	CNY	-0.458	(-0.490, -0.424)	0.000
ChiFreeVisa	CNY	-0.210	(-0.250, -0.170)	0.000
Occ-EMA360	CNY	0.069	(0.027, 0.111)	0.001
RDB-7	CNY	-0.035	(-0.077, 0.007)	0.100
RDB-30	CNY	-0.060	(-0.102, -0.018)	0.005
RDB-52	CNY	-0.079	(-0.121, -0.038)	0.000
RDB-365	CNY	0.023	(-0.019, 0.065)	0.276
AUD	RUB	0.051	(0.009, 0.093)	0.017
KRW	RUB	0.202	(0.161, 0.241)	0.000
MYR	RUB	0.025	(-0.017, 0.067)	0.245
SGD	RUB	-0.031	(-0.073, 0.010)	0.141
SEK	RUB	0.032	(-0.010, 0.073)	0.140
GBP	RUB	-0.005	(-0.046, 0.037)	0.829
ARR	RUB	0.038	(-0.004, 0.080)	0.074
Thursday	RUB	0.008	(-0.034, 0.050)	0.697
Friday	RUB	-0.020	(-0.062, 0.022)	0.348
Saturday	RUB	-0.020	(-0.062, 0.022)	0.348
Sunday	RUB	-0.020	(-0.062, 0.022)	0.348
Monday	RUB	0.056	(0.014, 0.098)	0.009
Tuesday	RUB	0.014	(-0.028, 0.056)	0.511
June	RUB	0.017	(-0.025, 0.058)	0.440
July	RUB	0.003	(-0.038, 0.045)	0.873

August	RUB	0.010	(-0.032, 0.052)	0.639
September	RUB	-0.037	(-0.078, 0.005)	0.087
October	RUB	-0.038	(-0.080, 0.004)	0.077
November	RUB	-0.028	(-0.070, 0.014)	0.188
December	RUB	-0.047	(-0.088, -0.005)	0.029
January	RUB	0.014	(-0.028, 0.056)	0.523
Febuary	RUB	0.097	(0.056, 0.139)	0.000
March	RUB	0.001	(-0.040, 0.043)	0.948
April	RUB	0.003	(-0.039, 0.045)	0.890
Monsoon	RUB	-0.019	(-0.061, 0.023)	0.370
Peak	RUB	-0.005	(-0.047, 0.037)	0.822
High	RUB	0.019	(-0.023, 0.061)	0.377
RusSanc	RUB	0.001	(-0.040, 0.043)	0.949
ChiBoatSink	RUB	-0.040	(-0.082, 0.002)	0.060
ChiFreeVisa	RUB	-0.017	(-0.059, 0.024)	0.415
Occ-EMA360	RUB	0.037	(-0.005, 0.079)	0.081
RDB-7	RUB	-0.001	(-0.043, 0.041)	0.961
RDB-30	RUB	-0.009	(-0.051, 0.033)	0.684
RDB-52	RUB	-0.008	(-0.050, 0.034)	0.710
RDB-365	RUB	0.008	(-0.034, 0.049)	0.724
KRW	AUD	-0.006	(-0.048, 0.036)	0.783
MYR	AUD	0.843	(0.830, 0.854)	0.000
SGD	AUD	0.367	(0.330, 0.403)	0.000
SEK	AUD	0.889	(0.880, 0.898)	0.000
GBP	AUD	0.450	(0.416, 0.482)	0.000
ARR	AUD	0.261	(0.222, 0.300)	0.000
Thursday	AUD	0.001	(-0.041, 0.043)	0.971
Friday	AUD	-0.001	(-0.043, 0.041)	0.963
Saturday	AUD	-0.001	(-0.043, 0.041)	0.967
Sunday	AUD	-0.001	(-0.043, 0.041)	0.967
Monday	AUD	-0.004	(-0.046, 0.038)	0.847
Tuesday	AUD	0.004	(-0.038, 0.046)	0.863
June	AUD	0.013	(-0.029, 0.054)	0.559
July	AUD	0.010	(-0.032, 0.052)	0.644
August	AUD	0.003	(-0.039, 0.045)	0.897
September	AUD	-0.021	(-0.063, 0.021)	0.319
October	AUD	-0.034	(-0.076, 0.008)	0.109
November	AUD	-0.050	(-0.092, -0.008)	0.019
December	AUD	-0.103	(-0.144, -0.061)	0.000
January	AUD	0.080	(0.038, 0.121)	0.000
Febuary	AUD	0.028	(-0.014, 0.070)	0.193
March	AUD	0.030	(-0.012, 0.072)	0.161
April	AUD	0.028	(-0.014, 0.070)	0.186
Monsoon	AUD	0.027	(-0.015, 0.069)	0.208
Peak	AUD	0.015	(-0.027, 0.056)	0.497
High	AUD	-0.020	(-0.062, 0.022)	0.349
RusSanc	AUD	-0.639	(-0.663, -0.613)	0.000
ChiBoatSink	AUD	-0.416	(-0.450, -0.381)	0.000
ChiFreeVisa	AUD	-0.194	(-0.234, -0.153)	0.000
Occ-EMA360	AUD	0.095	(0.053, 0.136)	0.000
RDB-7	AUD	-0.036	(-0.078, 0.006)	0.094
RDB-30	AUD	-0.051	(-0.092, -0.009)	0.018

RDB-52	AUD	-0.063	(-0.104, -0.021)	0.003
RDB-365	AUD	0.013	(-0.029, 0.055)	0.533
MYR	KRW	-0.009	(-0.051, 0.033)	0.672
SGD	KRW	0.057	(0.015, 0.099)	0.008
SEK	KRW	0.004	(-0.038, 0.046)	0.847
GBP	KRW	-0.002	(-0.044, 0.040)	0.923
ARR	KRW	-0.009	(-0.051, 0.033)	0.677
Thursday	KRW	-0.015	(-0.057, 0.026)	0.470
Friday	KRW	-0.019	(-0.060, 0.023)	0.386
Saturday	KRW	-0.019	(-0.060, 0.023)	0.386
Sunday	KRW	-0.019	(-0.060, 0.023)	0.386
Monday	KRW	-0.007	(-0.049, 0.035)	0.742
Tuesday	KRW	-0.014	(-0.056, 0.028)	0.501
June	KRW	-0.023	(-0.065, 0.019)	0.289
July	KRW	0.032	(-0.010, 0.074)	0.137
August	KRW	0.037	(-0.005, 0.078)	0.086
September	KRW	0.028	(-0.014, 0.069)	0.196
October	KRW	-0.013	(-0.055, 0.029)	0.534
November	KRW	-0.013	(-0.055, 0.029)	0.543
December	KRW	-0.013	(-0.055, 0.029)	0.542
January	KRW	0.036	(-0.006, 0.078)	0.089
Febuary	KRW	-0.002	(-0.044, 0.040)	0.939
March	KRW	-0.023	(-0.065, 0.019)	0.285
April	KRW	-0.023	(-0.065, 0.019)	0.281
Monsoon	KRW	-0.005	(-0.047, 0.037)	0.805
Peak	KRW	0.039	(-0.003, 0.081)	0.068
High	KRW	-0.001	(-0.043, 0.041)	0.953
RusSanc	KRW	0.002	(-0.040, 0.044)	0.922
ChiBoatSink	KRW	-0.021	(-0.063, 0.021)	0.327
ChiFreeVisa	KRW	-0.009	(-0.051, 0.033)	0.666
Occ-EMA360	KRW	0.024	(-0.018, 0.066)	0.259
RDB-7	KRW	-0.004	(-0.046, 0.038)	0.858
RDB-30	KRW	0.001	(-0.041, 0.042)	0.978
RDB-52	KRW	0.001	(-0.041, 0.043)	0.959
RDB-365	KRW	-0.002	(-0.044, 0.040)	0.925
SGD	MYR	0.391	(0.355, 0.426)	0.000
SEK	MYR	0.855	(0.843, 0.865)	0.000
GBP	MYR	0.671	(0.647, 0.694)	0.000
ARR	MYR	0.183	(0.142, 0.223)	0.000
Thursday	MYR	-0.000	(-0.042, 0.042)	0.995
Friday	MYR	0.000	(-0.041, 0.042)	0.985
Saturday	MYR	0.001	(-0.041, 0.042)	0.977
Sunday	MYR	0.001	(-0.041, 0.042)	0.977
Monday	MYR	-0.002	(-0.044, 0.040)	0.923
Tuesday	MYR	0.001	(-0.041, 0.043)	0.977
June	MYR	0.046	(0.004, 0.088)	0.031
July	MYR	0.047	(0.005, 0.089)	0.028
August	MYR	-0.001	(-0.043, 0.041)	0.960
September	MYR	-0.037	(-0.078, 0.005)	0.086
October	MYR	-0.040	(-0.082, 0.002)	0.059
November	MYR	-0.061	(-0.103, -0.019)	0.004
December	MYR	-0.078	(-0.119, -0.036)	0.000

January	MYR	0.031	(-0.011, 0.073)	0.143
Febuary	MYR	0.004	(-0.038, 0.046)	0.849
March	MYR	-0.008	(-0.050, 0.034)	0.700
April	MYR	0.027	(-0.015, 0.068)	0.213
Monsoon	MYR	0.061	(0.019, 0.103)	0.004
Peak	MYR	-0.005	(-0.047, 0.036)	0.801
High	MYR	-0.060	(-0.102, -0.018)	0.005
RusSanc	MYR	-0.571	(-0.598, -0.542)	0.000
ChiBoatSink	MYR	-0.309	(-0.347, -0.271)	0.000
ChiFreeVisa	MYR	-0.147	(-0.187, -0.105)	0.000
Occ-EMA360	MYR	0.051	(0.009, 0.092)	0.018
RDB-7	MYR	-0.042	(-0.084, -0.001)	0.047
RDB-30	MYR	-0.061	(-0.103, -0.020)	0.004
RDB-52	MYR	-0.078	(-0.120, -0.036)	0.000
RDB-365	MYR	0.010	(-0.032, 0.052)	0.643
SEK	SGD	0.526	(0.495, 0.556)	0.000
GBP	SGD	0.706	(0.684, 0.726)	0.000
ARR	SGD	-0.015	(-0.057, 0.027)	0.475
Thursday	SGD	-0.000	(-0.042, 0.042)	0.997
Friday	SGD	0.002	(-0.040, 0.044)	0.927
Saturday	SGD	0.002	(-0.040, 0.044)	0.932
Sunday	SGD	0.002	(-0.040, 0.044)	0.932
Monday	SGD	-0.004	(-0.046, 0.038)	0.849
Tuesday	SGD	-0.001	(-0.043, 0.041)	0.964
June	SGD	0.069	(0.027, 0.111)	0.001
July	SGD	0.094	(0.053, 0.136)	0.000
August	SGD	0.069	(0.027, 0.111)	0.001
September	SGD	0.050	(0.008, 0.091)	0.020
October	SGD	0.025	(-0.017, 0.067)	0.235
November	SGD	0.005	(-0.037, 0.047)	0.811
December	SGD	0.008	(-0.034, 0.049)	0.722
January	SGD	-0.008	(-0.049, 0.034)	0.725
Febuary	SGD	-0.089	(-0.130, -0.047)	0.000
March	SGD	-0.139	(-0.180, -0.098)	0.000
April	SGD	-0.097	(-0.138, -0.055)	0.000
Monsoon	SGD	0.047	(0.005, 0.089)	0.027
Peak	SGD	0.020	(-0.022, 0.061)	0.359
High	SGD	-0.083	(-0.124, -0.041)	0.000
RusSanc	SGD	0.109	(0.067, 0.150)	0.000
ChiBoatSink	SGD	-0.348	(-0.384, -0.310)	0.000
ChiFreeVisa	SGD	-0.160	(-0.200, -0.119)	0.000
Occ-EMA360	SGD	-0.046	(-0.088, -0.004)	0.031
RDB-7	SGD	-0.003	(-0.044, 0.039)	0.903
RDB-30	SGD	-0.012	(-0.054, 0.030)	0.583
RDB-52	SGD	-0.027	(-0.069, 0.015)	0.200
RDB-365	SGD	-0.054	(-0.095, -0.012)	0.012
GBP	SEK	0.656	(0.631, 0.679)	0.000
ARR	SEK	0.270	(0.231, 0.309)	0.000
Thursday	SEK	0.001	(-0.040, 0.043)	0.947
Friday	SEK	-0.000	(-0.042, 0.042)	0.993
Saturday	SEK	-0.000	(-0.042, 0.042)	0.998
Sunday	SEK	-0.000	(-0.042, 0.042)	0.998

Monday	SEK	-0.003	(-0.045, 0.039)	0.876
Tuesday	SEK	0.001	(-0.041, 0.043)	0.956
June	SEK	0.015	(-0.027, 0.057)	0.483
July	SEK	0.001	(-0.041, 0.043)	0.951
August	SEK	0.016	(-0.026, 0.058)	0.459
September	SEK	0.025	(-0.016, 0.067)	0.234
October	SEK	-0.002	(-0.044, 0.040)	0.933
November	SEK	-0.067	(-0.109, -0.025)	0.002
December	SEK	-0.059	(-0.100, -0.017)	0.006
January	SEK	0.069	(0.027, 0.111)	0.001
Febuary	SEK	0.037	(-0.005, 0.079)	0.081
March	SEK	0.003	(-0.039, 0.045)	0.899
April	SEK	-0.037	(-0.079, 0.005)	0.085
Monsoon	SEK	0.012	(-0.030, 0.054)	0.573
Peak	SEK	0.028	(-0.014, 0.070)	0.192
High	SEK	-0.010	(-0.052, 0.032)	0.649
RusSanc	SEK	-0.631	(-0.656, -0.605)	0.000
ChiBoatSink	SEK	-0.390	(-0.425, -0.354)	0.000
ChiFreeVisa	SEK	-0.175	(-0.215, -0.134)	0.000
Occ-EMA360	SEK	0.055	(0.013, 0.097)	0.010
RDB-7	SEK	-0.045	(-0.087, -0.003)	0.035
RDB-30	SEK	-0.071	(-0.113, -0.030)	0.001
RDB-52	SEK	-0.081	(-0.122, -0.039)	0.000
RDB-365	SEK	-0.002	(-0.044, 0.040)	0.939
ARR	GBP	0.084	(0.042, 0.125)	0.000
Thursday	GBP	0.000	(-0.042, 0.042)	0.994
Friday	GBP	0.001	(-0.041, 0.043)	0.951
Saturday	GBP	0.001	(-0.041, 0.043)	0.949
Sunday	GBP	0.001	(-0.041, 0.043)	0.949
Monday	GBP	-0.003	(-0.045, 0.039)	0.901
Tuesday	GBP	-0.000	(-0.042, 0.042)	0.994
June	GBP	0.038	(-0.004, 0.079)	0.078
July	GBP	0.009	(-0.033, 0.051)	0.683
August	GBP	0.002	(-0.040, 0.043)	0.942
September	GBP	0.020	(-0.022, 0.061)	0.357
October	GBP	-0.019	(-0.060, 0.023)	0.383
November	GBP	-0.018	(-0.060, 0.023)	0.389
December	GBP	-0.004	(-0.046, 0.038)	0.849
January	GBP	0.038	(-0.004, 0.080)	0.077
Febuary	GBP	0.005	(-0.037, 0.046)	0.830
March	GBP	-0.053	(-0.095, -0.011)	0.013
April	GBP	-0.044	(-0.086, -0.002)	0.040
Monsoon	GBP	-0.012	(-0.054, 0.030)	0.580
Peak	GBP	0.026	(-0.015, 0.068)	0.216
High	GBP	-0.002	(-0.044, 0.040)	0.933
RusSanc	GBP	-0.083	(-0.125, -0.042)	0.000
ChiBoatSink	GBP	-0.356	(-0.392, -0.319)	0.000
ChiFreeVisa	GBP	-0.180	(-0.221, -0.140)	0.000
Occ-EMA360	GBP	0.054	(0.012, 0.096)	0.011
RDB-7	GBP	-0.052	(-0.094, -0.010)	0.014
RDB-30	GBP	-0.060	(-0.102, -0.018)	0.005
RDB-52	GBP	-0.079	(-0.120, -0.037)	0.000

RDB-365	GBP	0.003	(-0.039, 0.045)	0.882
Thursday	ARR	-0.005	(-0.047, 0.037)	0.817
Friday	ARR	-0.007	(-0.049, 0.035)	0.743
Saturday	ARR	0.007	(-0.035, 0.049)	0.741
Sunday	ARR	0.008	(-0.034, 0.050)	0.709
Monday	ARR	-0.006	(-0.048, 0.036)	0.781
Tuesday	ARR	0.011	(-0.031, 0.053)	0.604
June	ARR	-0.209	(-0.249, -0.169)	0.000
July	ARR	-0.197	(-0.237, -0.157)	0.000
August	ARR	-0.204	(-0.244, -0.164)	0.000
September	ARR	-0.212	(-0.252, -0.172)	0.000
October	ARR	-0.201	(-0.241, -0.161)	0.000
November	ARR	0.069	(0.027, 0.110)	0.001
December	ARR	0.282	(0.243, 0.320)	0.000
January	ARR	0.449	(0.415, 0.482)	0.000
Febuary	ARR	0.323	(0.285, 0.360)	0.000
March	ARR	0.179	(0.139, 0.220)	0.000
April	ARR	-0.081	(-0.123, -0.039)	0.000
Monsoon	ARR	-0.657	(-0.680, -0.633)	0.000
Peak	ARR	0.539	(0.508, 0.568)	0.000
High	ARR	0.701	(0.679, 0.722)	0.000
RusSanc	ARR	-0.342	(-0.378, -0.304)	0.000
ChiBoatSink	ARR	-0.123	(-0.164, -0.081)	0.000
ChiFreeVisa	ARR	0.025	(-0.017, 0.067)	0.239
Occ-EMA360	ARR	0.249	(0.209, 0.288)	0.000
RDB-7	ARR	-0.078	(-0.120, -0.037)	0.000
RDB-30	ARR	-0.035	(-0.076, 0.007)	0.106
RDB-52	ARR	0.024	(-0.017, 0.066)	0.253
RDB-365	ARR	-0.047	(-0.089, -0.006)	0.026
Friday	Thursday	-0.167	(-0.207, -0.126)	0.000
Saturday	Thursday	-0.167	(-0.207, -0.126)	0.000
Sunday	Thursday	-0.167	(-0.207, -0.126)	0.000
Monday	Thursday	-0.167	(-0.207, -0.126)	0.000
Tuesday	Thursday	-0.167	(-0.207, -0.126)	0.000
June	Thursday	0.001	(-0.041, 0.043)	0.949
July	Thursday	-0.003	(-0.045, 0.039)	0.900
August	Thursday	0.002	(-0.040, 0.044)	0.925
September	Thursday	-0.003	(-0.045, 0.038)	0.874
October	Thursday	0.002	(-0.040, 0.044)	0.925
November	Thursday	0.001	(-0.041, 0.043)	0.949
December	Thursday	-0.003	(-0.045, 0.039)	0.900
January	Thursday	0.002	(-0.040, 0.044)	0.925
Febuary	Thursday	-0.001	(-0.043, 0.041)	0.974
March	Thursday	0.002	(-0.040, 0.044)	0.925
April	Thursday	-0.003	(-0.045, 0.038)	0.874
Monsoon	Thursday	0.000	(-0.042, 0.042)	1.000
Peak	Thursday	-0.002	(-0.044, 0.040)	0.926
High	Thursday	0.000	(-0.041, 0.042)	0.986
RusSanc	Thursday	-0.001	(-0.043, 0.040)	0.947
ChiBoatSink	Thursday	0.001	(-0.041, 0.043)	0.973
ChiFreeVisa	Thursday	-0.005	(-0.047, 0.037)	0.825
Occ-EMA360	Thursday	0.004	(-0.038, 0.046)	0.846

RDB-7	Thursday	0.002	(-0.040, 0.044)	0.926
RDB-30	Thursday	-0.005	(-0.047, 0.037)	0.807
RDB-52	Thursday	0.009	(-0.033, 0.050)	0.691
RDB-365	Thursday	0.002	(-0.040, 0.044)	0.917
Saturday	Friday	-0.167	(-0.207, -0.126)	0.000
Sunday	Friday	-0.167	(-0.207, -0.126)	0.000
Monday	Friday	-0.167	(-0.207, -0.126)	0.000
Tuesday	Friday	-0.167	(-0.207, -0.126)	0.000
June	Friday	0.001	(-0.041, 0.043)	0.949
July	Friday	-0.003	(-0.045, 0.039)	0.900
August	Friday	0.002	(-0.040, 0.044)	0.925
September	Friday	0.001	(-0.041, 0.043)	0.949
October	Friday	-0.003	(-0.045, 0.039)	0.900
November	Friday	0.001	(-0.041, 0.043)	0.949
December	Friday	-0.003	(-0.045, 0.039)	0.900
January	Friday	0.002	(-0.040, 0.044)	0.925
Febuary	Friday	-0.001	(-0.043, 0.041)	0.974
March	Friday	0.002	(-0.040, 0.044)	0.925
April	Friday	-0.003	(-0.045, 0.038)	0.874
Monsoon	Friday	-0.000	(-0.042, 0.042)	1.000
Peak	Friday	0.003	(-0.039, 0.045)	0.902
High	Friday	0.000	(-0.041, 0.042)	0.986
RusSanc	Friday	-0.001	(-0.043, 0.040)	0.947
ChiBoatSink	Friday	0.001	(-0.041, 0.043)	0.973
ChiFreeVisa	Friday	-0.005	(-0.047, 0.037)	0.825
Occ-EMA360	Friday	0.010	(-0.031, 0.052)	0.625
RDB-7	Friday	0.000	(-0.041, 0.042)	0.985
RDB-30	Friday	0.001	(-0.041, 0.043)	0.976
RDB-52	Friday	-0.000	(-0.042, 0.042)	0.994
RDB-365	Friday	-0.008	(-0.050, 0.034)	0.700
Sunday	Saturday	-0.167	(-0.207, -0.126)	0.000
Monday	Saturday	-0.167	(-0.207, -0.126)	0.000
Tuesday	Saturday	-0.167	(-0.207, -0.126)	0.000
June	Saturday	0.001	(-0.041, 0.043)	0.949
July	Saturday	-0.003	(-0.045, 0.039)	0.900
August	Saturday	0.002	(-0.040, 0.044)	0.925
September	Saturday	0.001	(-0.041, 0.043)	0.949
October	Saturday	-0.003	(-0.045, 0.039)	0.900
November	Saturday	0.001	(-0.041, 0.043)	0.949
December	Saturday	0.002	(-0.040, 0.044)	0.925
January	Saturday	-0.003	(-0.045, 0.039)	0.900
Febuary	Saturday	-0.001	(-0.043, 0.041)	0.974
March	Saturday	0.002	(-0.040, 0.044)	0.925
April	Saturday	0.001	(-0.041, 0.043)	0.949
Monsoon	Saturday	0.000	(-0.042, 0.042)	1.000
Peak	Saturday	0.003	(-0.039, 0.045)	0.902
High	Saturday	0.000	(-0.041, 0.042)	0.986
RusSanc	Saturday	0.002	(-0.040, 0.044)	0.930
ChiBoatSink	Saturday	0.001	(-0.041, 0.043)	0.973
ChiFreeVisa	Saturday	0.006	(-0.036, 0.048)	0.768
Occ-EMA360	Saturday	-0.017	(-0.058, 0.025)	0.437
RDB-7	Saturday	0.008	(-0.034, 0.050)	0.699

RDB-30	Saturday	0.011	(-0.031, 0.053)	0.595
RDB-52	Saturday	-0.003	(-0.045, 0.039)	0.882
RDB-365	Saturday	-0.018	(-0.060, 0.024)	0.401
Monday	Sunday	-0.167	(-0.207, -0.126)	0.000
Tuesday	Sunday	-0.167	(-0.207, -0.126)	0.000
June	Sunday	0.001	(-0.041, 0.043)	0.949
July	Sunday	0.002	(-0.040, 0.044)	0.925
August	Sunday	-0.003	(-0.045, 0.039)	0.900
September	Sunday	0.001	(-0.041, 0.043)	0.949
October	Sunday	-0.003	(-0.045, 0.039)	0.900
November	Sunday	0.001	(-0.041, 0.043)	0.949
December	Sunday	0.002	(-0.040, 0.044)	0.925
January	Sunday	-0.003	(-0.045, 0.039)	0.900
Febuary	Sunday	-0.001	(-0.043, 0.041)	0.974
March	Sunday	0.002	(-0.040, 0.044)	0.925
April	Sunday	0.001	(-0.041, 0.043)	0.949
Monsoon	Sunday	-0.000	(-0.042, 0.042)	1.000
Peak	Sunday	0.003	(-0.039, 0.045)	0.902
High	Sunday	0.000	(-0.041, 0.042)	0.986
RusSanc	Sunday	0.002	(-0.040, 0.044)	0.930
ChiBoatSink	Sunday	0.001	(-0.041, 0.043)	0.973
ChiFreeVisa	Sunday	0.006	(-0.036, 0.048)	0.768
Occ-EMA360	Sunday	0.007	(-0.034, 0.049)	0.729
RDB-7	Sunday	-0.020	(-0.062, 0.022)	0.346
RDB-30	Sunday	0.000	(-0.041, 0.042)	0.983
RDB-52	Sunday	0.014	(-0.028, 0.056)	0.519
RDB-365	Sunday	0.030	(-0.011, 0.072)	0.154
Tuesday	Monday	-0.167	(-0.207, -0.126)	0.000
June	Monday	0.001	(-0.041, 0.043)	0.949
July	Monday	0.002	(-0.040, 0.044)	0.925
August	Monday	-0.003	(-0.045, 0.039)	0.900
September	Monday	0.001	(-0.041, 0.043)	0.949
October	Monday	0.002	(-0.040, 0.044)	0.925
November	Monday	-0.003	(-0.045, 0.038)	0.874
December	Monday	0.002	(-0.040, 0.044)	0.925
January	Monday	-0.003	(-0.045, 0.039)	0.900
Febuary	Monday	0.004	(-0.038, 0.046)	0.845
March	Monday	-0.003	(-0.045, 0.039)	0.900
April	Monday	0.001	(-0.041, 0.043)	0.949
Monsoon	Monday	-0.000	(-0.042, 0.042)	1.000
Peak	Monday	-0.002	(-0.044, 0.040)	0.926
High	Monday	-0.002	(-0.044, 0.040)	0.914
RusSanc	Monday	0.002	(-0.040, 0.044)	0.930
ChiBoatSink	Monday	0.001	(-0.041, 0.043)	0.973
ChiFreeVisa	Monday	0.006	(-0.036, 0.048)	0.768
Occ-EMA360	Monday	0.013	(-0.029, 0.055)	0.533
RDB-7	Monday	0.035	(-0.007, 0.076)	0.105
RDB-30	Monday	0.023	(-0.019, 0.065)	0.286
RDB-52	Monday	0.001	(-0.041, 0.043)	0.959
RDB-365	Monday	-0.004	(-0.046, 0.038)	0.845
June	Tuesday	-0.003	(-0.045, 0.038)	0.874
July	Tuesday	0.002	(-0.040, 0.044)	0.925

August	Tuesday	-0.003	(-0.045, 0.039)	0.900
September	Tuesday	0.001	(-0.041, 0.043)	0.949
October	Tuesday	0.002	(-0.040, 0.044)	0.925
November	Tuesday	-0.003	(-0.045, 0.038)	0.874
December	Tuesday	0.002	(-0.040, 0.044)	0.925
January	Tuesday	0.002	(-0.040, 0.044)	0.925
Febuary	Tuesday	-0.001	(-0.043, 0.041)	0.974
March	Tuesday	-0.003	(-0.045, 0.039)	0.900
April	Tuesday	0.001	(-0.041, 0.043)	0.949
Monsoon	Tuesday	-0.000	(-0.042, 0.042)	1.000
Peak	Tuesday	-0.002	(-0.044, 0.040)	0.926
High	Tuesday	0.000	(-0.041, 0.042)	0.986
RusSanc	Tuesday	-0.001	(-0.043, 0.040)	0.947
ChiBoatSink	Tuesday	-0.004	(-0.046, 0.037)	0.837
ChiFreeVisa	Tuesday	-0.005	(-0.047, 0.037)	0.825
Occ-EMA360	Tuesday	-0.020	(-0.062, 0.022)	0.351
RDB-7	Tuesday	-0.019	(-0.061, 0.023)	0.372
RDB-30	Tuesday	-0.034	(-0.076, 0.008)	0.112
RDB-52	Tuesday	0.008	(-0.034, 0.050)	0.698
RDB-365	Tuesday	-0.001	(-0.043, 0.041)	0.968
July	June	-0.091	(-0.132, -0.049)	0.000
August	June	-0.091	(-0.132, -0.049)	0.000
September	June	-0.090	(-0.131, -0.048)	0.000
October	June	-0.091	(-0.132, -0.049)	0.000
November	June	-0.090	(-0.131, -0.048)	0.000
December	June	-0.091	(-0.132, -0.049)	0.000
January	June	-0.091	(-0.132, -0.049)	0.000
Febuary	June	-0.086	(-0.128, -0.045)	0.000
March	June	-0.091	(-0.132, -0.049)	0.000
April	June	-0.090	(-0.131, -0.048)	0.000
Monsoon	June	0.210	(0.169, 0.249)	0.000
Peak	June	-0.093	(-0.134, -0.051)	0.000
High	June	-0.229	(-0.269, -0.189)	0.000
RusSanc	June	0.020	(-0.022, 0.062)	0.341
ChiBoatSink	June	-0.082	(-0.123, -0.040)	0.000
ChiFreeVisa	June	-0.036	(-0.078, 0.006)	0.093
Occ-EMA360	June	-0.239	(-0.278, -0.199)	0.000
RDB-7	June	0.026	(-0.016, 0.068)	0.229
RDB-30	June	-0.003	(-0.045, 0.039)	0.887
RDB-52	June	-0.057	(-0.099, -0.015)	0.007
RDB-365	June	0.074	(0.033, 0.116)	0.000
August	July	-0.093	(-0.134, -0.051)	0.000
September	July	-0.091	(-0.132, -0.049)	0.000
October	July	-0.093	(-0.134, -0.051)	0.000
November	July	-0.091	(-0.132, -0.049)	0.000
December	July	-0.093	(-0.134, -0.051)	0.000
January	July	-0.093	(-0.134, -0.051)	0.000
Febuary	July	-0.088	(-0.129, -0.046)	0.000
March	July	-0.093	(-0.134, -0.051)	0.000
April	July	-0.091	(-0.132, -0.049)	0.000
Monsoon	July	0.213	(0.173, 0.253)	0.000
Peak	July	-0.094	(-0.136, -0.053)	0.000

High	July	-0.233	(-0.273, -0.194)	0.000
RusSanc	July	0.021	(-0.021, 0.063)	0.333
ChiBoatSink	July	-0.083	(-0.125, -0.042)	0.000
ChiFreeVisa	July	-0.036	(-0.078, 0.005)	0.088
Occ-EMA360	July	-0.067	(-0.109, -0.025)	0.002
RDB-7	July	0.036	(-0.006, 0.078)	0.093
RDB-30	July	0.085	(0.043, 0.126)	0.000
RDB-52	July	0.068	(0.026, 0.109)	0.002
RDB-365	July	-0.034	(-0.075, 0.008)	0.115
September	August	-0.091	(-0.132, -0.049)	0.000
October	August	-0.093	(-0.134, -0.051)	0.000
November	August	-0.091	(-0.132, -0.049)	0.000
December	August	-0.093	(-0.134, -0.051)	0.000
January	August	-0.093	(-0.134, -0.051)	0.000
Febuary	August	-0.088	(-0.129, -0.046)	0.000
March	August	-0.093	(-0.134, -0.051)	0.000
April	August	-0.091	(-0.132, -0.049)	0.000
Monsoon	August	0.213	(0.173, 0.253)	0.000
Peak	August	-0.094	(-0.136, -0.053)	0.000
High	August	-0.233	(-0.273, -0.194)	0.000
RusSanc	August	0.021	(-0.021, 0.063)	0.333
ChiBoatSink	August	0.116	(0.074, 0.157)	0.000
ChiFreeVisa	August	-0.036	(-0.078, 0.005)	0.088
Occ-EMA360	August	0.018	(-0.023, 0.060)	0.389
RDB-7	August	-0.037	(-0.079, 0.005)	0.083
RDB-30	August	0.008	(-0.034, 0.050)	0.717
RDB-52	August	0.074	(0.032, 0.116)	0.001
RDB-365	August	-0.045	(-0.086, -0.003)	0.036
October	September	-0.091	(-0.132, -0.049)	0.000
November	September	-0.090	(-0.131, -0.048)	0.000
December	September	-0.091	(-0.132, -0.049)	0.000
January	September	-0.091	(-0.132, -0.049)	0.000
Febuary	September	-0.086	(-0.128, -0.045)	0.000
March	September	-0.091	(-0.132, -0.049)	0.000
April	September	-0.090	(-0.131, -0.048)	0.000
Monsoon	September	0.210	(0.169, 0.249)	0.000
Peak	September	-0.093	(-0.134, -0.051)	0.000
High	September	-0.229	(-0.269, -0.189)	0.000
RusSanc	September	0.020	(-0.022, 0.062)	0.341
ChiBoatSink	September	0.114	(0.072, 0.155)	0.000
ChiFreeVisa	September	-0.036	(-0.078, 0.006)	0.093
Occ-EMA360	September	-0.116	(-0.157, -0.074)	0.000
RDB-7	September	-0.014	(-0.056, 0.027)	0.499
RDB-30	September	-0.083	(-0.124, -0.041)	0.000
RDB-52	September	-0.082	(-0.123, -0.040)	0.000
RDB-365	September	-0.049	(-0.091, -0.007)	0.022
November	October	-0.091	(-0.132, -0.049)	0.000
December	October	-0.093	(-0.134, -0.051)	0.000
January	October	-0.093	(-0.134, -0.051)	0.000
Febuary	October	-0.088	(-0.129, -0.046)	0.000
March	October	-0.093	(-0.134, -0.051)	0.000
April	October	-0.091	(-0.132, -0.049)	0.000

Monsoon	October	0.213	(0.173, 0.253)	0.000
Peak	October	-0.094	(-0.136, -0.053)	0.000
High	October	-0.233	(-0.273, -0.194)	0.000
RusSanc	October	0.021	(-0.021, 0.063)	0.333
ChiBoatSink	October	0.116	(0.074, 0.157)	0.000
ChiFreeVisa	October	-0.036	(-0.078, 0.005)	0.088
Occ-EMA360	October	-0.071	(-0.112, -0.029)	0.001
RDB-7	October	0.102	(0.061, 0.143)	0.000
RDB-30	October	0.029	(-0.013, 0.071)	0.173
RDB-52	October	-0.036	(-0.078, 0.006)	0.092
RDB-365	October	0.009	(-0.033, 0.051)	0.684
December	November	-0.091	(-0.132, -0.049)	0.000
January	November	-0.091	(-0.132, -0.049)	0.000
Febuary	November	-0.086	(-0.128, -0.045)	0.000
March	November	-0.091	(-0.132, -0.049)	0.000
April	November	-0.090	(-0.131, -0.048)	0.000
Monsoon	November	-0.427	(-0.461, -0.392)	0.000
Peak	November	-0.093	(-0.134, -0.051)	0.000
High	November	0.390	(0.354, 0.425)	0.000
RusSanc	November	0.020	(-0.022, 0.062)	0.341
ChiBoatSink	November	0.114	(0.072, 0.155)	0.000
ChiFreeVisa	November	-0.036	(-0.078, 0.006)	0.093
Occ-EMA360	November	-0.017	(-0.059, 0.025)	0.423
RDB-7	November	0.016	(-0.026, 0.058)	0.451
RDB-30	November	0.086	(0.044, 0.127)	0.000
RDB-52	November	0.094	(0.052, 0.135)	0.000
RDB-365	November	0.012	(-0.030, 0.053)	0.588
January	December	-0.093	(-0.134, -0.051)	0.000
Febuary	December	-0.088	(-0.129, -0.046)	0.000
March	December	-0.093	(-0.134, -0.051)	0.000
April	December	-0.091	(-0.132, -0.049)	0.000
Monsoon	December	-0.435	(-0.468, -0.400)	0.000
Peak	December	0.323	(0.285, 0.360)	0.000
High	December	0.397	(0.361, 0.432)	0.000
RusSanc	December	0.021	(-0.021, 0.063)	0.333
ChiBoatSink	December	0.116	(0.074, 0.157)	0.000
ChiFreeVisa	December	0.393	(0.357, 0.428)	0.000
Occ-EMA360	December	0.084	(0.042, 0.125)	0.000
RDB-7	December	0.002	(-0.040, 0.044)	0.920
RDB-30	December	0.083	(0.041, 0.124)	0.000
RDB-52	December	0.169	(0.128, 0.210)	0.000
RDB-365	December	-0.005	(-0.046, 0.037)	0.832
Febuary	January	-0.088	(-0.129, -0.046)	0.000
March	January	-0.093	(-0.134, -0.051)	0.000
April	January	-0.091	(-0.132, -0.049)	0.000
Monsoon	January	-0.435	(-0.468, -0.400)	0.000
Peak	January	0.601	(0.573, 0.627)	0.000
High	January	0.397	(0.361, 0.432)	0.000
RusSanc	January	-0.108	(-0.149, -0.066)	0.000
ChiBoatSink	January	-0.083	(-0.125, -0.042)	0.000
ChiFreeVisa	January	-0.036	(-0.078, 0.005)	0.088
Occ-EMA360	January	0.211	(0.171, 0.251)	0.000

RDB-7	January	-0.026	(-0.068, 0.015)	0.215
RDB-30	January	0.033	(-0.009, 0.075)	0.119
RDB-52	January	-0.006	(-0.048, 0.036)	0.785
RDB-365	January	-0.008	(-0.050, 0.034)	0.700
March	Febuary	-0.088	(-0.129, -0.046)	0.000
April	Febuary	-0.086	(-0.128, -0.045)	0.000
Monsoon	Febuary	-0.413	(-0.447, -0.377)	0.000
Peak	Febuary	-0.090	(-0.131, -0.048)	0.000
High	Febuary	0.377	(0.341, 0.412)	0.000
RusSanc	Febuary	-0.101	(-0.142, -0.059)	0.000
ChiBoatSink	Febuary	-0.079	(-0.121, -0.037)	0.000
ChiFreeVisa	Febuary	-0.035	(-0.076, 0.007)	0.105
Occ-EMA360	Febuary	0.238	(0.198, 0.277)	0.000
RDB-7	Febuary	-0.043	(-0.085, -0.001)	0.044
RDB-30	Febuary	-0.022	(-0.064, 0.020)	0.298
RDB-52	Febuary	0.014	(-0.028, 0.056)	0.501
RDB-365	Febuary	-0.014	(-0.055, 0.028)	0.527
April	March	-0.091	(-0.132, -0.049)	0.000
Monsoon	March	0.213	(0.173, 0.253)	0.000
Peak	March	-0.094	(-0.136, -0.053)	0.000
High	March	0.072	(0.030, 0.113)	0.001
RusSanc	March	0.021	(-0.021, 0.063)	0.333
ChiBoatSink	March	-0.083	(-0.125, -0.042)	0.000
ChiFreeVisa	March	-0.036	(-0.078, 0.005)	0.088
Occ-EMA360	March	0.128	(0.086, 0.169)	0.000
RDB-7	March	-0.044	(-0.085, -0.002)	0.041
RDB-30	March	-0.075	(-0.117, -0.034)	0.000
RDB-52	March	-0.063	(-0.105, -0.022)	0.003
RDB-365	March	0.001	(-0.041, 0.043)	0.972
Monsoon	April	0.210	(0.169, 0.249)	0.000
Peak	April	-0.093	(-0.134, -0.051)	0.000
High	April	-0.229	(-0.269, -0.189)	0.000
RusSanc	April	0.020	(-0.022, 0.062)	0.341
ChiBoatSink	April	-0.082	(-0.123, -0.040)	0.000
ChiFreeVisa	April	-0.036	(-0.078, 0.006)	0.093
Occ-EMA360	April	0.014	(-0.028, 0.056)	0.514
RDB-7	April	-0.016	(-0.058, 0.026)	0.455
RDB-30	April	-0.066	(-0.107, -0.024)	0.002
RDB-52	April	-0.087	(-0.128, -0.045)	0.000
RDB-365	April	0.011	(-0.031, 0.052)	0.621
Peak	Monsoon	-0.443	(-0.476, -0.408)	0.000
High	Monsoon	-0.914	(-0.920, -0.906)	0.000
RusSanc	Monsoon	0.097	(0.055, 0.138)	0.000
ChiBoatSink	Monsoon	-0.041	(-0.082, 0.001)	0.057
ChiFreeVisa	Monsoon	-0.171	(-0.211, -0.130)	0.000
Occ-EMA360	Monsoon	-0.300	(-0.338, -0.262)	0.000
RDB-7	Monsoon	0.029	(-0.012, 0.071)	0.168
RDB-30	Monsoon	-0.106	(-0.148, -0.065)	0.000
RDB-52	Monsoon	-0.160	(-0.200, -0.119)	0.000
RDB-365	Monsoon	0.008	(-0.033, 0.050)	0.692
High	Peak	0.404	(0.369, 0.439)	0.000
RusSanc	Peak	-0.061	(-0.102, -0.019)	0.005

ChiBoatSink	Peak	-0.009	(-0.051, 0.033)	0.677
ChiFreeVisa	Peak	0.127	(0.085, 0.168)	0.000
Occ-EMA360	Peak	0.169	(0.128, 0.209)	0.000
RDB-7	Peak	0.003	(-0.039, 0.045)	0.894
RDB-30	Peak	0.022	(-0.020, 0.064)	0.295
RDB-52	Peak	0.066	(0.024, 0.107)	0.002
RDB-365	Peak	-0.004	(-0.046, 0.038)	0.857
RusSanc	High	-0.089	(-0.130, -0.047)	0.000
ChiBoatSink	High	0.016	(-0.026, 0.058)	0.449
ChiFreeVisa	High	0.156	(0.115, 0.197)	0.000
Occ-EMA360	High	0.342	(0.305, 0.379)	0.000
RDB-7	High	-0.037	(-0.079, 0.005)	0.082
RDB-30	High	0.083	(0.042, 0.125)	0.000
RDB-52	High	0.144	(0.102, 0.184)	0.000
RDB-365	High	-0.010	(-0.052, 0.032)	0.640
ChiBoatSink	RusSanc	0.134	(0.093, 0.175)	0.000
ChiFreeVisa	RusSanc	0.059	(0.017, 0.100)	0.006
Occ-EMA360	RusSanc	-0.090	(-0.131, -0.048)	0.000
RDB-7	RusSanc	0.035	(-0.007, 0.076)	0.105
RDB-30	RusSanc	0.033	(-0.009, 0.075)	0.121
RDB-52	RusSanc	0.040	(-0.002, 0.082)	0.063
RDB-365	RusSanc	-0.038	(-0.079, 0.004)	0.078
ChiFreeVisa	ChiBoatSink	0.437	(0.403, 0.470)	0.000
Occ-EMA360	ChiBoatSink	-0.220	(-0.259, -0.179)	0.000
RDB-7	ChiBoatSink	0.056	(0.014, 0.097)	0.009
RDB-30	ChiBoatSink	0.125	(0.083, 0.166)	0.000
RDB-52	ChiBoatSink	0.141	(0.099, 0.181)	0.000
RDB-365	ChiBoatSink	-0.129	(-0.170, -0.088)	0.000
Occ-EMA360	ChiFreeVisa	0.141	(0.100, 0.182)	0.000
RDB-7	ChiFreeVisa	0.002	(-0.040, 0.044)	0.933
RDB-30	ChiFreeVisa	0.074	(0.032, 0.115)	0.001
RDB-52	ChiFreeVisa	0.186	(0.145, 0.226)	0.000
RDB-365	ChiFreeVisa	-0.012	(-0.054, 0.030)	0.584
RDB-7	Occ-EMA360	0.166	(0.125, 0.207)	0.000
RDB-30	Occ-EMA360	0.246	(0.206, 0.285)	0.000
RDB-52	Occ-EMA360	0.314	(0.276, 0.351)	0.000
RDB-365	Occ-EMA360	0.252	(0.213, 0.291)	0.000
RDB-30	RDB-7	0.128	(0.087, 0.169)	0.000
RDB-52	RDB-7	0.094	(0.052, 0.135)	0.000
RDB-365	RDB-7	0.098	(0.057, 0.140)	0.000
RDB-52	RDB-30	0.211	(0.171, 0.251)	0.000
RDB-365	RDB-30	0.078	(0.036, 0.119)	0.000

REFERENCES

Ahmed, N., A. Atiya, N. El Gayar, H. El-Shishiny, and E. Giza. 2009. ' An empirical comparison of machine learning models for time series forecasting', *Econometric*, 29: 594–621.

Baker, T.K., and D.A. Collier. 1999. 'A comparative revenue analysis of hotel yield management heuristics', *Decision Sciences*, 30: 239–63.

Bill Barnett. *Phuket Hotel Market Update*.

Box, G.E.P., and G. Jenkins. 1970. 'Time Series Analysis', *Forecasting and Control, Holden-Day, San Francisco, CA*.

Cao, L.J., and F.E.H. Tay. 2001. 'Application of support vector machines in financial time series forecasting', *Omega*, 29: 309–17

Chen, C., and S. Kachani. 2007. 'Forecasting and optimization for hotel revenue management', *Journal of Revenue and Pricing Management*, 6: 163–74.

Claveria, O., E. Monte, and S. Torra. 2015. 'A new forecasting approach for the hospitality industry', *International Journal of Contemporary Hospitality Management*, 27: 1520-38.

Council, Office of the National Economic and Social Development.

De Livera, A.M. 2010b. 'Automatic forecasting with a modified exponential smoothing state space framework', *Department of Econometrics & Business Statistics, Monash University*.

De Livera, A.M., R. Hyndman, and R. Snyder. 2011. 'Forecasting time series with complex seasonal patterns using exponential smoothing', *Journal of the American Statistical Association*, 106: 1513–27.

El Gayar, N., M. Saleh, A. Atiya, H. El-Shishiny, A. Zakhary, and H. Habib. 2011. 'An integrated framework for advanced hotel revenue management', *International Journal of Contemporary Hospitality Management*, 23: 84-98.

Ha, T.D., F.M. Bianchi, and Olsson R. 2017. 'Local short term electricity load forecasting: Automatic approaches, ' *International Joint Conference on Neural Networks*: 4267–74.

Haensel, Alwin, and Ger Koole. 2011. 'Booking horizon forecasting with dynamic updating: A case study of hotel reservation data', *International Journal of Forecasting*, 27: 942-60.

Holt, C.E. 1957. 'Forecasting seasonals and trends by exponentially weighted averages', *O.N.R. Memorandum No. 52, Carnegie Institute of Technology, Pittsburgh USA*.

Hong, SeJoon, James Y. L. Thong, and Kar Yan Tam. 2006. 'Understanding continued information technology usage behavior: A comparison of three models in the context of mobile internet', *Decision Support Systems*, 42: 1819-34.

Huang, Xu, Hossein Hassani, Mansi Ghodsi, Zinnia Mukherjee, and Rangan Gupta. 2017. 'Do trend extraction approaches affect causality detection in climate change studies?', *Physica A: Statistical Mechanics and its Applications*, 469: 604-24.

J.W., Taylor. 2003. 'Short-term electricity demand forecasting using double seasonal exponential smoothing', *J. Oper. Res. Soc.*, 54: 799–805.

Keerativibool. 2013. 'Forecasting Model for the Number of International Tourist Arrivals to Thailand', *Srinakharinwirot Science Journal*, 29.

Koupriouchina, Larissa, Jean-Pierre van der Rest, and Zvi Schwartz. 2014. 'On revenue management and the use of occupancy forecasting error measures', *International Journal of Hospitality Management*, 41: 104-14.

Lim, Christine, Chialin Chang, and Michael McAleer. 2009. 'Forecasting h(m)otel guest nights in New Zealand', *International Journal of Hospitality Management*, 28: 228-35.

Martinez-de Pison, E., J. Fernandez-Ceniceros, A.V. Pernia-Espinoza, F.J. Martinez-de Pison, and A. Sanz-Garcia. 2016. 'Hotel reservation forecasting using flexible soft computing techniques: A case of study in a Spanish hotel', *International Journal of Information Technology & Decision Making*, 15: 1211–34.

Mukma, Chomtee, Payakkapong. 2018. 'Forecasting the Number of Chinese Tourists in Thailand', *Thai Science and Technology Journal*, 26.

Pereira, Luis Nobre. 2016. 'An introduction to helpful forecasting methods for hotel revenue management', *International Journal of Hospitality Management*, 58: 13-23.

Rajopadhye, Mihir, Mounir Ben Ghalia, Paul P. Wang, Timothy Baker, and Craig V. Eister.

2001. 'Forecasting uncertain hotel room demand', *Information Sciences*, 132: 1-11.

Rungjindarat, Phansaita. 2016. 'Forecasting Russian Tourist Arrivals to Thailand using SARIMA Model', *Dusit Thani College Journal*, 10.

Saothayanun, Taweesakulvatchara, Kanjanasakda, Somrang. 2014. 'A Forecasting Methods for the Number of International Tourists in Thailand: Box-Jenkins Method and Winter's Method', *Thai Science and Technology Journal*, 22.

Thomason, M. 1999. 'The practitioner methods and tool', *J. Comput. Intell. in Finance*, 7: 36–45.

Urraca, R., A. Sanz-Garcia, J. Fernandez-Ceniceros, E. Sodupe-Ortega, and F.J. Martinez-de-Pison. 2015. 'Improving hotel room demand forecasting with a hybrid GA-SVR methodology based on skewed data transformation, feature selection and parsimony tuning', *Hybrid Artificial Intelligent Systems,*, 9121: 632–43

Vapnik V.N. 1995. 'The Nature of Statistical Learning Theory', *New York: Springer-Verlag*.

Vu, C.J., and L.W. Turner. 2006. 'Regional data forecasting accuracy: The case of Thailand', *Journal of Travel Research*, 45: 186–93.

Weatherford, Larry R., and Sheryl E. Kimes. 2003. 'A comparison of forecasting methods for hotel revenue management', *International Journal of Forecasting*, 19: 401-15.

Weatherford, Lawrence R., Sheryl E. Kimes, and Darren A. Scott. 2001. 'Forecasting for hotel revenue management: Testing aggregation against disaggregation', *The Cornell Hotel and Restaurant Administration Quarterly*, 42: 53-64.

Winters, P.R. 1960. 'Forecasting sales by exponentially weighted moving averages', *Management Science*, 6: 324–42.

Zakhary, A., A.F. Atiya, H. El-Shishiny, and N.E. Gayar. 2011. 'Forecasting hotel arrivals and occupancy using Monte Carlo simulation', *Journal of Revenue and Pricing Management*, 10: 344-66.

VITA

NAME	Phoom Ungtrakul
DATE OF BIRTH	23 July 1991
PLACE OF BIRTH	Bangkok
INSTITUTIONS ATTENDED	Chulalongkorn University
HOME ADDRESS	138/62 ไอดิโอสาธร-ตากสิน คอนโด ถนนกรุงธนบุรี แขวงบางลำภูล่าง เขตคลองสาน กทม. 10600