Project Summary

Time Series Forecasting for Archipelago International

The following report is to demonstrate my ability to think on a Masters level and to verify I am ready to work in the professional data science industry using my theoretical knowledge and analytical skills and applying these in the professional business environment by providing strategic solutions to my company and my clients.

I worked for 2 months at Archipelago International, Jakarta as a Data Project Manager. Archipelago International is Indonesia's largest operator of hotels. They have 8 core brands with over 18,500 rooms and apartments under supervision for property owners in Indonesia, the Philippines and Singapore. While they have the largest accommodation portfolio in Indonesia, over 15,000 employees in 136 hotels supports their management. Archipelago International is a true innovator and market leader in its field, bringing condotels to Asia, starting the budget hotel boom in Indonesia and pioneering cloud computing before everyone else in Indonesia. *(Archipelagointernational, 2018)*

I was working closely together with other departments such as Marketing, Sales and Technology. My objective was to develop a forecasting method that is founded on statistical methods and is able to accurately forecast room occupancy for the near future thus the sales department could create a data driven room pricing strategy.

This was done primarily by historical data analysis using statistical modeling, regression and other machine learning models to eventually come to an accurate prediction. The ultimate goal of the project and the company is of course to increase sales volume, optimize pricing strategy, increase profit and most importantly ROI.

My daily routine consisted of internal meetings; data analysis and reporting; data cleansing and model building, analysis of underlying trends and aligning the project strategy and structure. These different tasks and duties gave me diversity providing space for both personal interaction and analytical work.

The report is built on the framework of the **CRISP-DM methodology** (also called CRISP-DM reference model), which provides an overview of the life cycle of my data-mining project. The framework delivers great overview of the different phases of the project as well as the tasks and the relationships in between.
The report consists of 6 main parts namely Business Understanding, Data

Understanding, Data preparation, Modeling, Evaluation and Review.

The first phase was **understanding the Business** itself. Understanding the goals and objectives of Archipelago International and converting this to a data-mining project that can eventually archive the desired objective.

The following phase was **Understanding the Data**, which included the collection of data and familiarizing with the data in order to reveal underlying structure, hidden information and assess data quality.

The **Data Preparation** phase consisted of all the steps that had to be done in order to create the final dataset from the raw data provided by Archipelago. This phase was one of the longest phases due to numerous obstacles I encountered during the process. Besides transforming and cleaning the data, the biggest issue was with the quality of the automatic data, which was eventually unusable and manual data had to be used.

Modeling was one of the most interesting phases for me. I have used various modeling techniques that were applicable for my time series forecast such as **Multiple Linear Regressions, ETS, Decision Tree** and **ARIMA**. It was great to see the accuracy of the different models confirming that the pilot project did eventually reach a positive outcome.

The modeling phase started with a **primary – test design** phase, where I solely used multiple linear regression and started experimenting with training and testing datasets. Strong predictors such as; trend, seasonality and holidays provided the foundation of the time series analysis.

In the **second part** of the modeling phase I selected 86 hotels based on data quality and quantity and wrote the codes in RStudio for ARIMA, ETS and Decision Tree.

Finally, in the **third part** of the modelling phase I created the Shiny Flex board, which is an interactive dashboard that enables easy access to both the explanatory analysis and to the

2
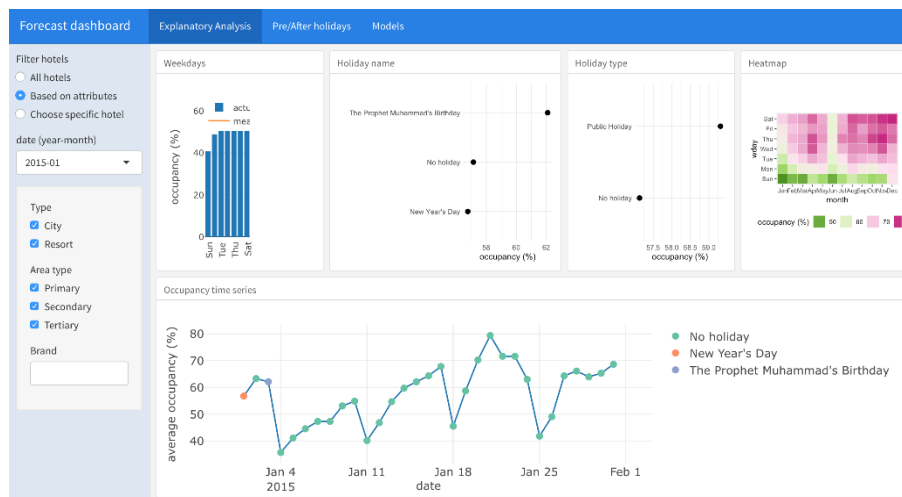
modeling results. In order to provide Archipelago International a functional and valuable asset I had set up a Shiny Server on an AWS instance so staff members of the hotel have access to the forecasting dashboard. As can be seen on the screenshot the dashboard allows easy access to any of the hotels, can select specific date and able to provide joint results to specific hotel types, area types or brands. The first tab shows the explanatory analysis for all or just for the specifically filtered hotels, while the second tab provides analysis on potential holidays.

The third tab on the dashboard is the modeling tab. It allows filtration for any chosen hotel and the model type can be selected. As can be seen from the screenshot this tab provides an interactive plot on model accuracy showing actual data

and prediction. The top right window provides the performance ranking of the models, while the bottom left shows a 60-day forecast, finally the bottom right shows the same in a numerical format.

In the following phase of the report I **evaluated** the models and determined next steps. I have compared their accuracy and identified, which has the highest quality perspective from a data analysis point of view. While in some cases RMSE and MAE scores are relatively low making the prediction more accurate in other cases scores can be high with lower forecasting accuracy. ARIMA and Linear Regression turned out to be the best performing models in nearly all scenarios, while Decision Tree follows them on the third place. ETS did not perform well as it is a simple time series model that does not take into consideration predictors.

In the last phase I **Reviewed** the project reflecting on my own performance and the final product created. I have to say it has been an excellent two months for me, which provided the opportunity to use my skills and knowledge learnt at CEU and I am very happy I was able to create a truly valuable asset for the company.

Due to the time limitation, at the end of my internship the project is still in a pilot phase, regardless the company offered me the opportunity to further proceed with the project and further develop the models.

CEU eTD Collection