National College of Ireland

Project Design

MSc in Data Analytics
Domain Application of Predictive Analytics

Ashish Rawat
X18185801

School of Computing
National College of Ireland

Supervisor: Vikas Sahni

Predictive Analysis of Hotel Booking Cancellations

Ashish Rawat
Master's in Data Analytics
X18185801

Abstract - **Cancellations have adverse effects on revenue management in hospitality. To counter the effect of cancellation hotels implements strategies like overbooking and cancellation policy, which in some cases have negative effects on hotel image and revenue. Therefore, in this research using the data from a resort hotel in Portugal, I will address booking cancellations prediction. Using machine learning classification models like Decision Tree, Logistic Regression and SVM, forecasting models will be built and evaluated on different parameters to identify the best model for forecasting. And the results achieved from the model will help managers and hotel authorities to develop effective strategies.**

I. INTRODUCTION

In hospitality industry, increase in revenue is directly proportional to the sale of hotel rooms. Sheryl et al stated in [1] that increase in revenue can be achieved by allocating a hotel room to the right customer at the right time through the right channel. Hence booking system [2] was introduced which allowed customers to reserve a room and service of a hotel prior to their date of arrival. Although it seems beneficial for the hotel to provision their service in advance it also had its drawback as customers have the right to cancel their booking before the date of arrival.

Nowadays due to availability of multiple channels like goibibo and trivago it has become much easier for the customer to find a hotel which meet their needs. But at the same time, it has also led to an increase in the cancellation rate of bookings. Cancellation rate now vary from 20% and can go as high as 60% of the total bookings [3]. Unlike earlier when cancellations were based on genuine reasons like sickness or bad weather these days the paradigm has shifted. As customers are now more focused on attaining the best deal. Sometimes in the search for the best deal they make multiple bookings in the moment and later ends up cancelling all except one.

In order to counter such situations hospitality industry came up with solutions like overbooking and cancellation policy [4]. Overbooking like the name suggest allows the hotel authorities to accept booking more than their capacity to lodge. Hotels also started implementing strict cancellation policies to compensate for the loss if a cancellation or no show occurs. But sometimes, these approaches backfires and ends up tainting the hotel's image. And may also affect future revenue as the unsatisfied customers might never again book a room with the hotel.

Hence it is necessary to build a model that avoids the consequences of strategies like overbooking and cancellation policy. By analyzing the customer behavior and their patterns in booking it is possible to forecast which customer booking has the possibility to be cancelled. This narrows down the population and certain strategies can be developed to target such customers. Using the historical data of hotel and applying analytical techniques, forecasting models are being build. This domain of analytic is generally termed as Predictive Analytic and is being widely used across various domains to generate insights to make better decisions.

For this research the data is obtained from the Property Management System (PMS) of a hotel in the region of Algarve, Portugal [5]. The dataset contains records from July 2015 to August 2017. It contains 31 variables including the outcome variable; booking confirmed or cancelled. Analyzing these variables which provides information on the characteristics of customers booking patterns can eventually provide insight which can be effectively monopolized.

II. RESEARCH GOALS

In this research I will be analyzing a historical data of a hotel and build a forecasting model. The research will focus on the following:

- Analyzing all the attributes and determining which attributes have a significant impact on cancellations.
- Applying different machine learning techniques on the dataset.

- Assessing the performance of the machine learning techniques on multiple evaluation parameters and determining the best technique for forecasting.

III.. ETHICAL CONCERNS

Each industry is vulnerable to ethical violations, and the hospitality industry is no exception. Since they manage large amounts of people and capital, they always must be on stakeout for issues like fraud transactions, missing items from inventory, unregistered bookings and stealing food. Inside theft are a common occurrence in hotels and loads of transaction cash can tempt some hotel staff to rob. Since hotel staff usually consists of full time and part time workers working long and odd hours. Since many hotels have restaurants within, it is necessary for them to maintain the health and safety regulations set up by the government. Violation of any regulation may lead to termination of the contract and closing of the restaurant. For security purposes CCTV cameras are usually installed in hotels to monitor any suspicious activity or to keep track of their service. But as per hotel security cameras law is illegal to set up cameras in hotel rooms. They should only be installed in public areas and should not breach customers privacy.

The data for this research was made publicly available for educational and research purposes. To protect the identity of the hotel certain features were excluded from the dataset and dataset was named as H2 data. Also, to maintain the ethicality the H2 data does not contain any customers details or information that can lead to identifying personnel.

IV. STRATEGY

For a hotel to thrive in this competitive market it is vital to layout a systematic plan which considers all variables and helps in achieving the objective. Following are the strategies that can eventually help grow the business.

1. Understanding the market.

To have a clear understanding of the market is crucial for any industry to implement a strategy. By understanding the customer needs and their pattern of booking could eventually help in increasing the business [6]. Moreover, it is important to be vary of your competitors before making any strategic decisions like pricing strategies, offers or marketing campaigns.

2. Understanding the customer type

People book hotels for different purposes like business meetings or family vacations. Therefore, identifying the type of customer would provide information on needs of such customers and help increase their satisfaction level [7].

3. Price Optimization

An estimate of Average Daily Rate can be determined by analyzing the market and help in implementing dynamic pricing [6].

4. Understanding distribution channels

Bookings to a hotel usually comes through different sources like travel agent, tour operators or direct bookings. It is necessary to know which sources are profitable for business and targeting segments where more opportunities can be obtained [6].

5. Predictive strategies

With the help of technology and machine learning it is possible to build a system which can help identify bookings that are highly likely to be cancelled. Identification of such bookings can help direct campaign and offers to retain them [7].

V. EXPLORATORY DATA ANALYSIS

By exploring the attributes of H2 data valuable insights are obtained and presented. Therefore, by using Tableau preliminary visualization were made for understanding the attributes and nature of the dataset.

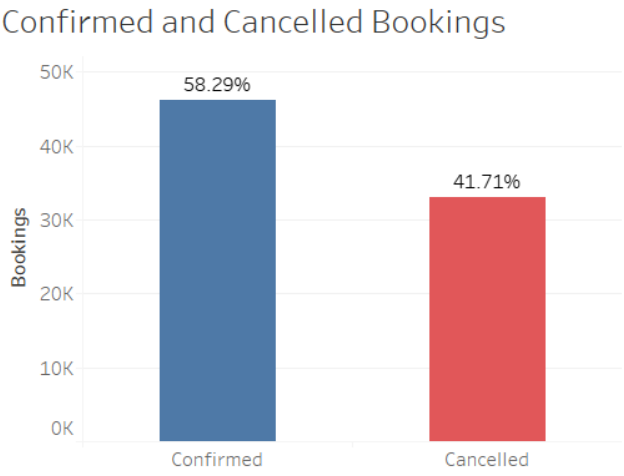

Fig. 1

Total confirmed and cancelled bookings are presented in Figure 1. The number of cancelled booking is 41.7 % of the total bookings which is quite high and not what is desired. Analyzing this data over years is possible and but not preferred because of the lack of data.

The monthly analysis for 2 years' time period is presented in Figure 2. Time period of April to September is when the number of bookings start increasing gradually and August is the month when it reaches its peak. And it can also be observed that with increase in booking, cancellation rate is also increasing.

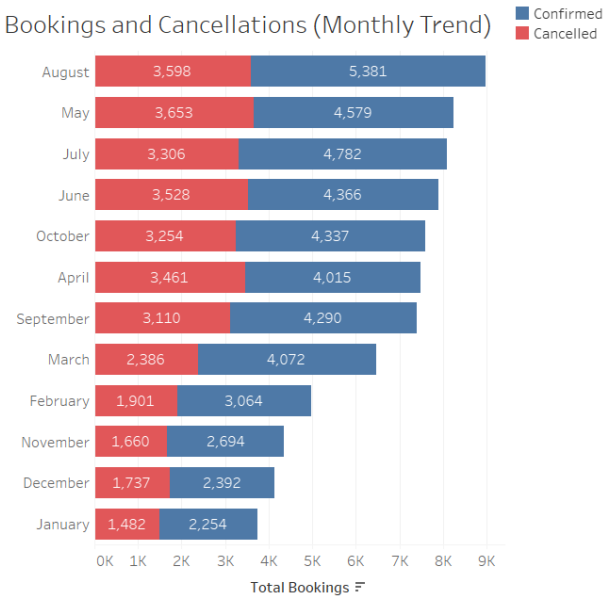

Fig. 2

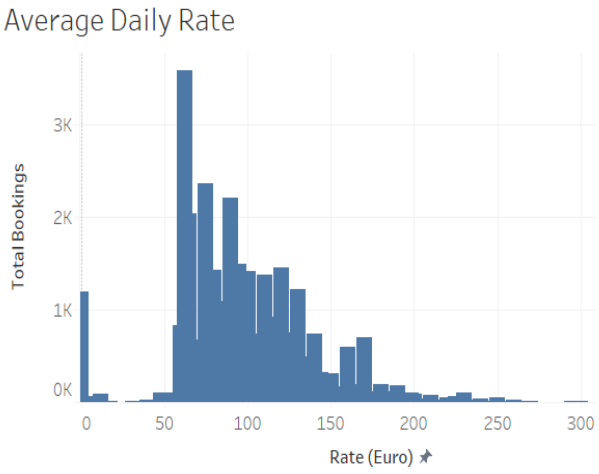

Fig. 3

With increase in demand the price of the hotel rooms also increases. And for majority of the bookings daily rate is between 50€ – 100€. It can also be observed that for 0€, 1200 bookings are shown in figure 3 which should be removed.

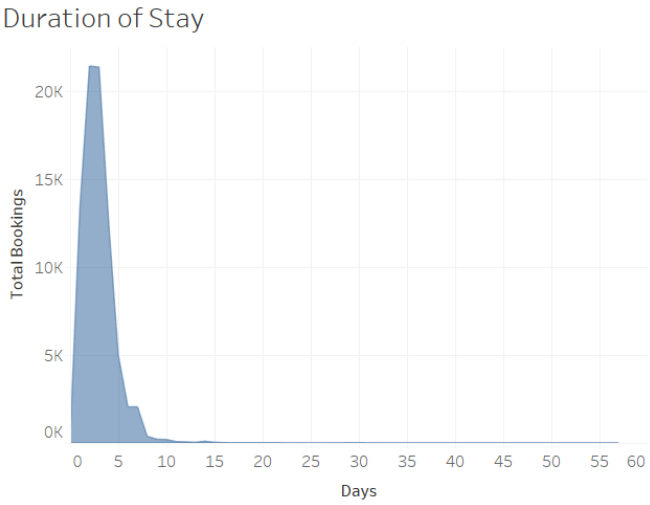

Fig. 4

According to figure 4 majority of the bookings are made for 2 – 4 days. And for very few cases it exceeds 15 days. Therefore, dropping these outliers (days above 15) would aid in analysis.

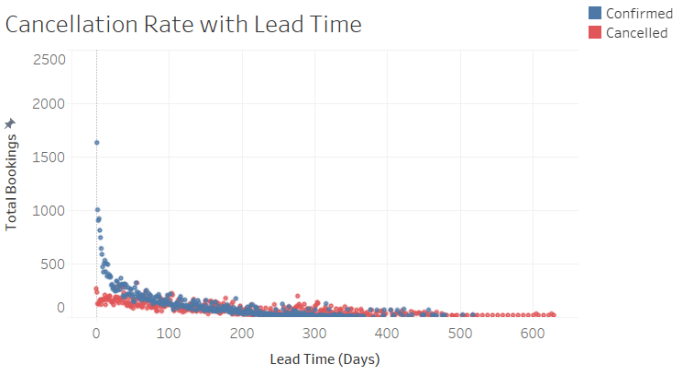

Fig. 5

It is evident from figure 6 that in majority of the records Adults are more likely to book a hotel and only in few cases babies and children accompany them. Hence for the ease of the analysis it would be better to sum up the number of people for which the hotel was booked due to extremely high difference in the number of Adults to that of Children and Babies.

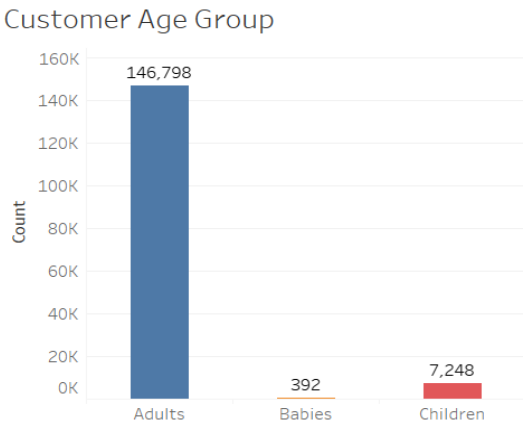

Fig. 6

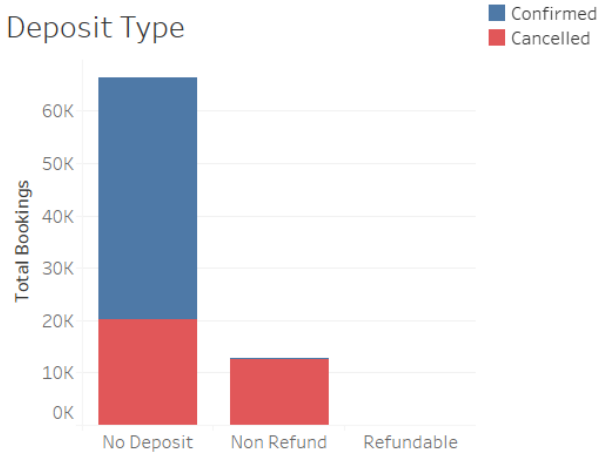

Fig. 7

For majority of the Non-Refund bookings cancellations were observed and for Refundable bookings 100 % success with 6 confirmed bookings was observed. Refundable bookings will be dropped as the data present will bias the predictive model.

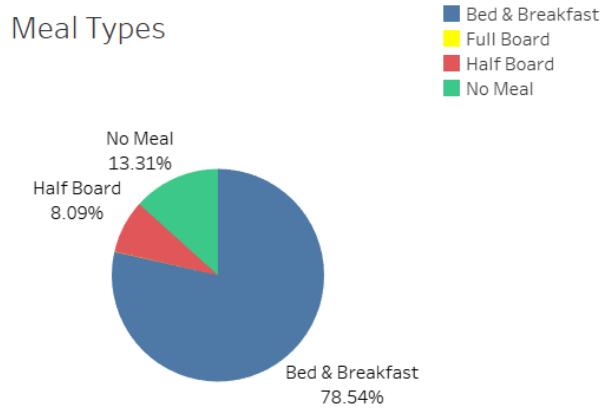

Fig. 8

And of the meal types provided by the hotel Bed & Breakfast is most requested services accounting for 78.54%. The least requested service is Full Board which accounts for 0.06 %. Since it is almost negligible in comparison with Bed & Breakfast it would be much more efficient to combine it with Half Board (8.09%) and create a new category.

VI. APPLICABLE TECHNIQUES

In [8] Nuno, Ana and Luis used classification models like Boosted Decision Tree (BDT), Decision Forest (DF), support vector machine (SVM) and Neural Network (NN) for predicting the cancellations. They used these models because the output variable 'IsCanceled' contains two binary values 0 and 1 which represents confirmed and cancelled bookings respectively. Cross validation techniques like k fold cross-validation was used assessing each model and evaluated the performance of each model on parameters like accuracy, precision, recall, F1 score and AUC. They also took False Positive (FP) as measure to evaluate the model. According to their conclusion DF was slightly better than BDT in terms of accuracy, AUC and predicting least value for FP.

In [9] Han *et al.* used Artificial Neural Network (ANN), specifically back propagation neural network (BPNN) and general regression neural network (GRNN) to build forecasting models. Evaluation parameters like sensitivity and specificity was used which considers the Type 1 and Type 2 errors and a ROC curve was build using them. They concluded that both the models can be utilized to predict cancelations and aid manager in making a better decision.

ARIMA model has been used by Michael in [10]. He approached the issue by analyzing the cancellation on weekly basis. In his research 90 % of the time series data was used for training and rest for testing the model. He achieved mean directional accuracy of 86 % and RMSE value of 274 for the model.

VII. REFERENCES

[1] S. Kimes, J. Wirtz, J. Theng and P. Patterson, "Revenue management: Resolving potential customer conflicts", Journal of Revenue and Pricing Management, vol. 2, no. 3, pp. 216-226, 2003. Available: 10.1057/palgrave.rpm.5170068 [Accessed 24 February 2020].

[2] K. Talluri and G. Ryzin, The theory and practice of revenue management. Boston, MA: Springer-Verlag US, 2005, pp. 1-10. [Accessed 24 February 2020].

[3] D. Romero Morales and J. Wang, "Forecasting cancellation rates for services booking revenue management using data mining", European Journal of Operational Research, vol. 202, no. 2, pp. 554-562, 2010. Available: 10.1016/j.ejor.2009.06.006 [Accessed 25 February 2020].

[4] K. Talluri, I. Karaesmen, G. van Ryzin and G. Vulcano, "Revenue management: Models and methods", Proceedings of the 2009 Winter Simulation Conference (WSC), 2009. Available: 10.1109/wsc.2009.5429322 [Accessed 24 February 2020].

[5] N. Antonio, A. de Almeida and L. Nunes, "Hotel booking demand datasets", Data in Brief, vol. 22, pp. 41-49, 2019. Available: 10.1016/j.dib.2018.11.126 [Accessed 24 February 2020].

[6] "9 Revenue Management Strategies to Grow Your Hotel Business", Revfine.com. [Online]. Available: https://www.revfine.com/revenue-management-strategies-hotel-industry/. [Accessed: 01- Mar- 2020].

[7] "7 Fundamental Hotel Marketing Strategies to Stand Out from Your Competition | RateGain", Your Site NAME Goes HERE, 2017. [Online]. Available: https://rategain.com/blog/fundamental-hotel-marketing-strategies-stand-out/. [Accessed: 06- Mar- 2020].

[8] N. Antonio, A. Almeida and L. Nunes, "Predicting hotel booking cancellations to decrease uncertainty and increase revenue", Tourism & Management Studies, vol. 13, no. 2, pp. 25-39, 2017. Available: 10.18089/tms.2017.13203 [Accessed 1 March 2020].

[9] H. Huang, A. Chang and C. Ho, "Using Artificial Neural Networks to Establish a Customer-cancellation Prediction Model", Semanticscholar.org, 2013. [Online]. Available: https://www.semanticscholar.org/paper/Using-Artificial-Neural-Networks-to-Establish-a-HuangChang/68bfbde1810b422263e13a848d828bea7f996bc1. [Accessed: 03- Mar- 2020].

[10] M. Grogan, "Predicting Hotel Cancellations with ExtraTreesClassifier and Logistic Regression", Medium, 2019. [Online]. Available: https://towardsdatascience.com/predicting-hotel-cancellations-with-extratreesclassifier-and-logistic-regression-fb9229a95d1e. [Accessed: 05- Mar- 2020].