**Members : Huy Sam, Ahmed Mohamed, Raheem Lee, Chau Nguyen,**

**Oluwamayowa Kusemiju**

Dataset: [Hotel booking demand | Kaggle](https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand)

**Table of Contents**

1. Project Description Pg. 1
2. Data Exploration and Reprocessing Pg.1
3. Neural Network Plot Pg 2
4. Model Pg 3
5. Result and Discussion Pg 4
6. Summary Pg 7
7. Reference Pg 8

**Project description**

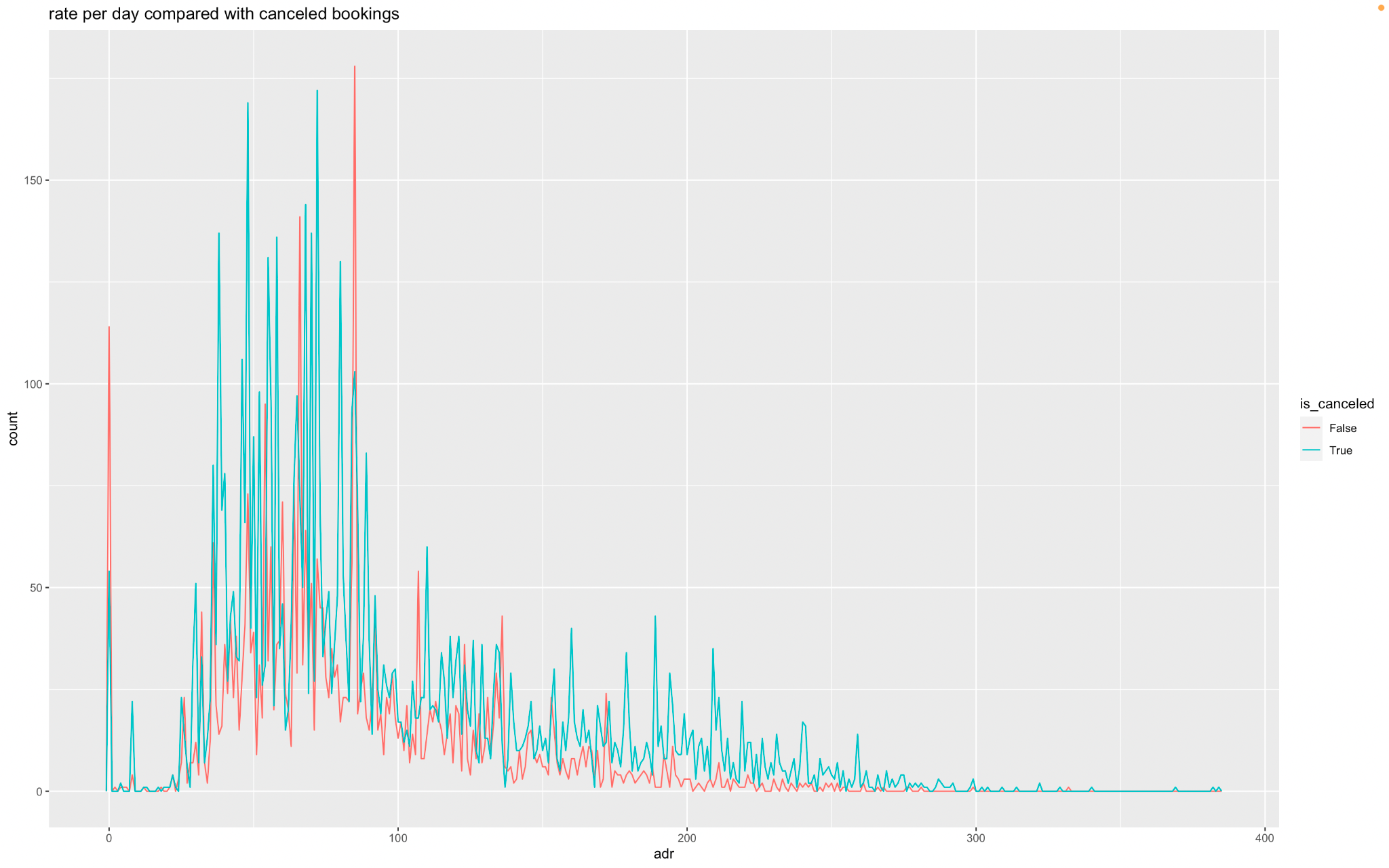
The hoteling industry’s revenue is dependent on demand. Due to several aberrant days related to many yearly holidays, varied events, promotions, and environmental aspects that have evolved, projecting demand in the hotel business is difficult. In the hotel industry, considerable demand seasonality needs to be recognized, by necessitating early planning to avoid revenue reductions. However, as this information is very important to the hotel stakeholders, it can also be translated to become a matter of importance to actual hotel clientele as low demand can equate to lower prices.

This data set: ‘Hotel Booking Demand,’ is a very traditional dataset, with a sample size of 119391, and 32 variables. It comprises booking information for a city hotel and a resort hotel, including dates of booking, duration of stay, number of adults, children, and/or infants, and available parking spaces, among other things. We examined the demand for hotels based on travel data looking at different variables to analyze: country, market segment, reserve room types, etc. Thus, we set a goal of determining the best times to book a hotel by finding when demand is low. Furthermore, we wanted to predict the amount of holiday demand for the future through time series forecasting.

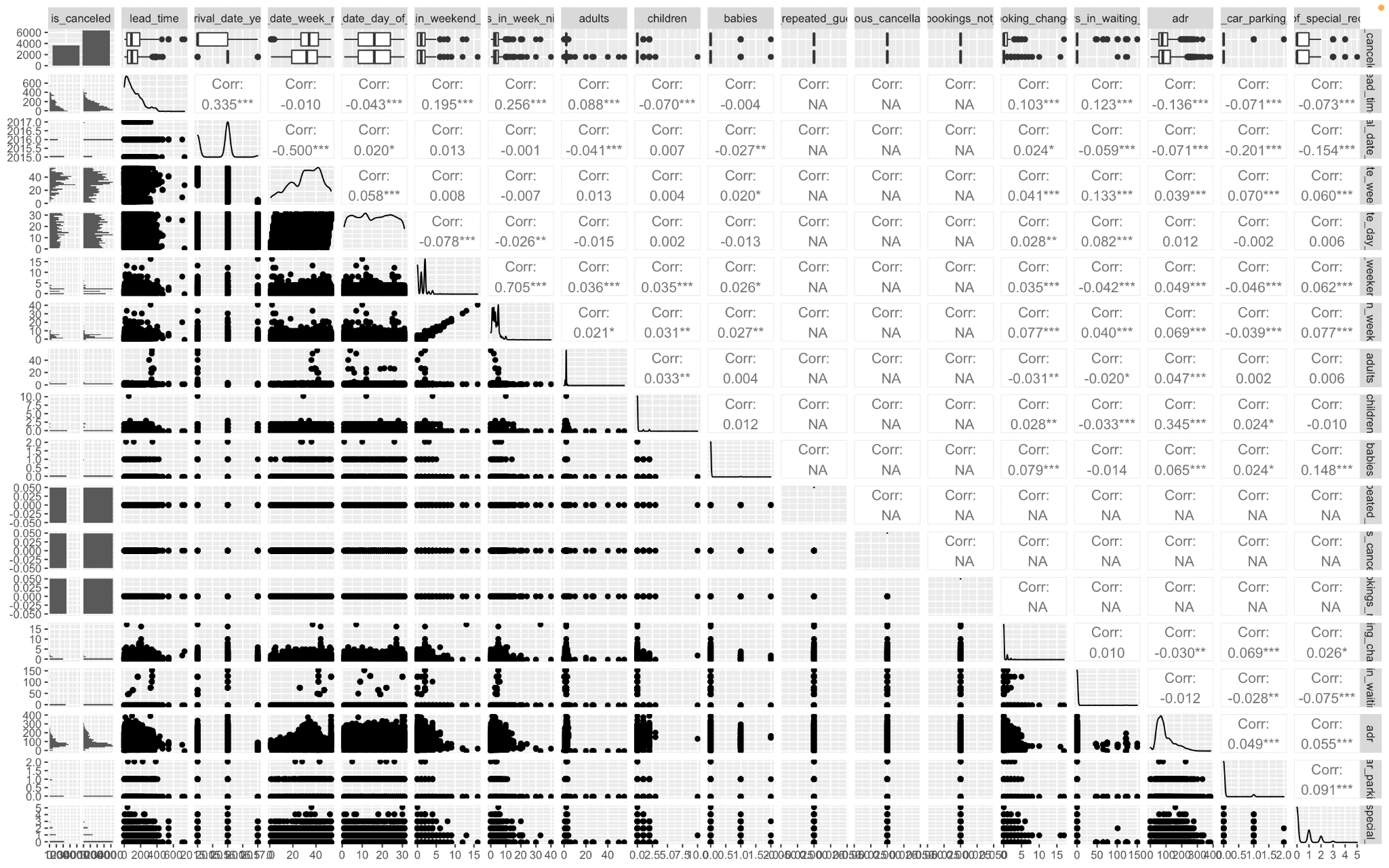
**Data Exploration and Preprocessing.**

Our target variable is is\_canceled and our input variables are different depending on our models because of the different data types.

* TimeSeries
  + Predictors: “reservation\_status\_date” and “total rate”
* NeuralNetwork
* Predictors: Adults, adr and children
* We only used 21250 observations
* Naive Bayes
* Predictors: is\_canceled

****

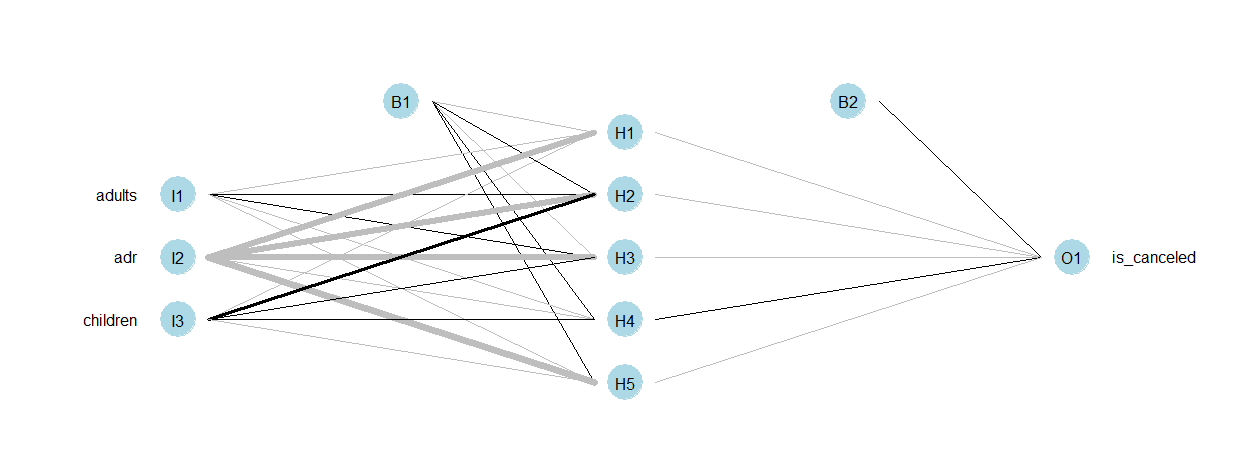
Here we have a comparison of ADR (average rate for rooms) per booking cancellations with the true values equalling a not canceled booking and vice versa.

****

We have a correlation graph we shows each numerical variable the amount they correlate with the predictor variable is\_canceled

**Neural Network Plot**

We used 5 hidden layers for neural network plot



**Point to Note:**

* We have a lot of missing data and errors.
* We used these two method to replace NULL with NA and replace NAs with 0
* There wasn’t any redundant data because not many variables had high correlation, the highest wave was around 33%
* We had a Data transformation, we did have to split the categorical data(chr) and the numerical data in the data frame.

Time series - total\_rate = (total\_nights \* adr)

total\_nights = (stays\_in\_weekend\_nights + stays\_in\_week\_nights)

* We have 60% for train data and 40% for valid data

**Model**

We selected only 21250 observations . Because if we use too many observations, the computer can not process too much data for Neural NetWork

We used a time span between 2016 and the third quarter of 2017 for our Naive TimeSeries prediction. Time series model was formed by taking the “reservation\_status\_date” and “total\_rate” variables and condensing them into its own data frame. The data frame was then used for the Naive seasonal time series prediction.

We used is\_canceled as the independent should be included from a regression equation for all models.

Variable selection techniques used:

i1 <- sapply(hotelBook.df, is.numeric)

hotelBook.df[i1]

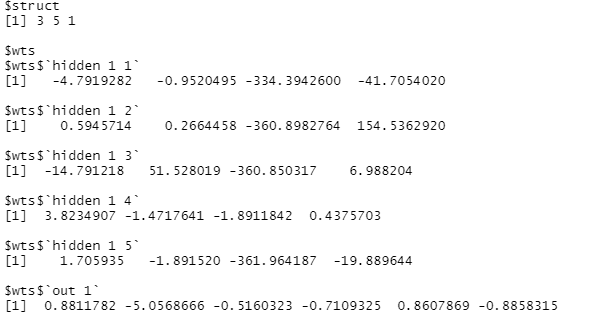
data <- hotelBook.df[i1]

We used this line of code to separate the categorical data (chr) with the numerical data

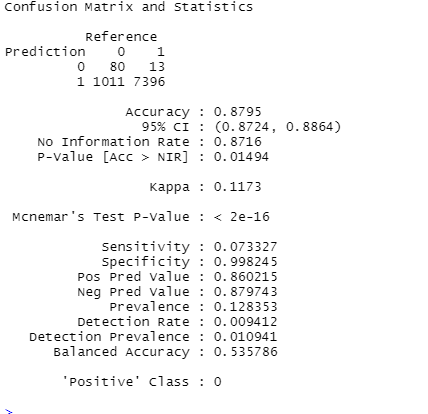
The best model is TimeSeries and we decided to use TimeSeries to forecast the demand of hotels since this will give us best predictions. To predict whether the booking is cancel or not , we use Naive Bayes

**Results and discussion**

**Neural Network**

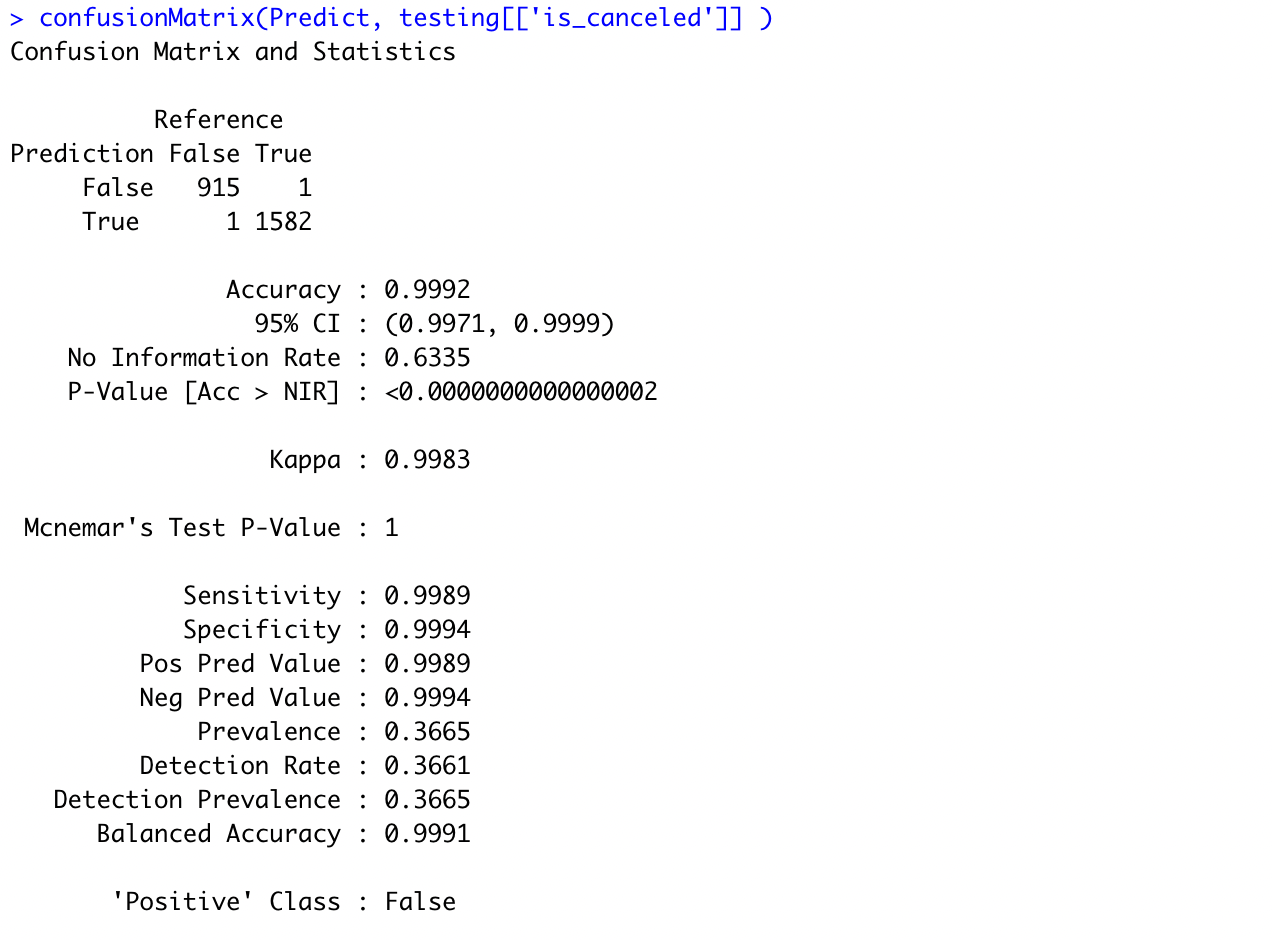


**Description:** Here we have the weights of the hidden layers and for the first layer we see -4.8(is\_canceled)- .952(adults)-334.394(ADR) -41.71(children)



We have 87.95 % accuracy for the Neural Network model.

**Naive Bayes**

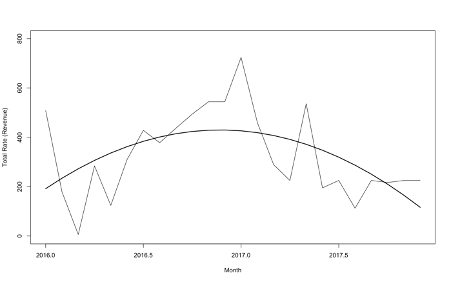
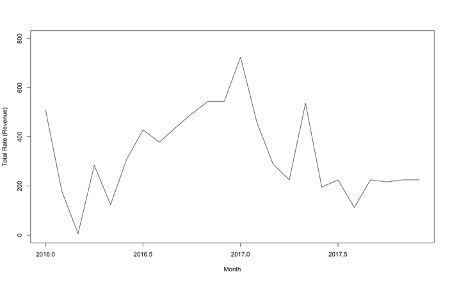


We have the results of The naive bayes with an accuracy of 99.92% of prediction whether the testing data would be canceled.

**TimeSeries**

The operating performance of a hotel or any other lodging business is determined through the Average Daily Rate. The total rate from bookings has always consistently fluctuated. In order to maintain and ensure efficiency, we wanted to take a look at the total rate of bookings per customer in order to predict performance of the last quarter of 2017. The assumption we made is that it’s going to increase due to many people flying out of town and visiting family and friends for the holiday season.

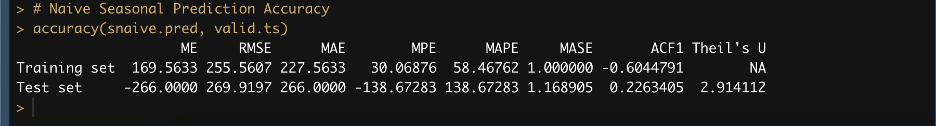
We took the monthly total rate from 2016 through the first three quarters of 2017 and created a time series plot to see if we could find any trends or seasonality.

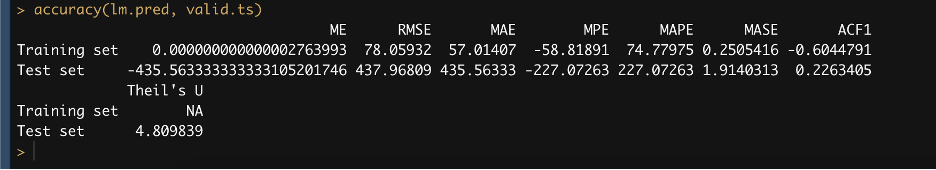
****

Sales overall have generally increased during the year of 2016 but have steadily dropped during most of the year of 2017. You can see that there’s a negative linear trend occurring with some quarterly seasonality trending upward during the summer and winter months. Since we’re calculating future demand, we decided to go with the naïve forecasting method with seasonality.

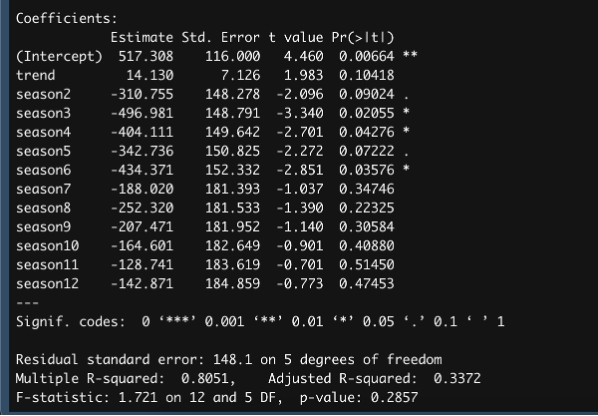
****

Based on our results, as expected, there will be an increase in the total rate per customer during the winter months. What we didn’t expect is that the amount won’t be as high as the previous year. In this case, we suggest coming up with special packages for loyal and new customers during the holiday season. We also suggest offering special events that would create the incentive for a customer to stay longer through the new year.

****

****

Our final Naïve Seasonal Prediction accuracy received a RMSE of 269.92. While our Linear trend seasonality received a RMSE of 437.97 making the Naïve model the most accurate of the two.

****

The linear regression model based on trend and seasonality has a positive coefficient of 14.13 meaning on average, with all things being equal, the average amount of the total rate will increase by $14.13 during the next 6 months.

**6. Summary**

A major pillar supporting a demand forecast is accurate data. A good science-based forecast not only captures valuable historical data, but continuously updates real-time data as well, examining broad factors such as seasonality, market conditions, length of stay, booking pace, and lead time. By incorporating the right data, managers can refine their approach and adjust pricing with greater precision, transforming data into accurate and actionable revenue-enhancing strategies.

A good forecast is crucial to capture the status of market segmentation understanding that each customer segment reflects different preferences, booking trends and patterns, and purchase intentions. A family taking a vacation will have different requirements than a business traveler attending a convention. Predicting demand and requirements for each of these customer segments helps businesses better target their marketing, budget their operational expenses, and achieve an optimal business mix.

**References**

Antonio, Nuno; Almeida, Ana; Nunes, Luis. “Hotel Booking Demand Datasets.” *Data in Brief*, vol 22, February 2022, pp. 41-49, <https://doi.org/10.1016/j.dib.2018.11.126>.