"Temporal Insights in Hospitality: ARIMA Forecasting"

A punchline report

Authors:

Suraj Basavaraj Rajolad - 02121154

Deeksha Mallampet - 02120800

Vinayak Gurura Sonter - 02136799

Pavan Kumar Gummuluru - 02121263

ISSUES:

The data is taken from the website called "Boston analysis" which is meant to have 245 different data sets about the city of Boston. We used the data set called "economic indicators" to find some valuable insights from the data. The dataset contains the different interconnected aspects that helps to gauge the economy of the city. The data sets contains the data from 2013-2019.

The data set contains factors like Logan passengers, Logan international flights, hotel occupancy rate, hotel avg daily rate, total jobs, unemployment rate, labor force part rate, pipeline unit, pipeline total development cost, pipeline sqft, pipeline constant jobs, foreclosure pet, foreclosure deeds, med_housing price, housing sales volume, new house construction permit, new affordable housing permit. The total number of logs in this dataset is 85.

We are going to perform some statistical analysis and implement Time series analysis on the hotel occupancy rate.

We address the following questions based on the data present and the analysis we performed.

- What decision can we take to increase the profit of the hotel, based on the population of the passengers arriving to Boston?
- Which is month with the highest average hotel occupancy rate across all years?
- Can we build a machine learning model which can predict the hotel occupancy rate in the near future.

FINDINGS:

- After looking at the data set we were concerned to find the insight on the hotel occupancy and the Logan passenger arriving to Boston. The month with the highest average hotel occupancy rate across all years is 7(July) with an average rate of 0.9.
- So, if the hotel management increase the price of the hotel accommodation by 5% compared to other months in the year, the management can see higher profits.
- We insisted 5% percentage because the drastic increase in the price of accommodation can lead to failure in occupying higher number of people.
- Yes, we can build a Time series model called ARIMA to predict the hotel occupancy rate on the bases of past records.

DISCUSSION:

Firstly, by looking at the dataset, we thought of working on hotel occupancy rate so we plot the bar graphs representing different columns, this shows that the number of people flying to Boston are increasing every year. We selected to work on the analysis of hotel occupancy rate because tourism is one most common and highly accepted. The income that is associated with tourism lends a hand to the economy of a city.

Secondly, we used some of the columns in the data set those are logan passengers, Logan international flights, hotel occupancy rate, hotel avg daily rate. All the other columns in the data set were dropped. we found the correlation between the variable that we have selected. Here we started going deep and analyze monthly occupancy rate of each year from 2013-2019.

We then implemented Time series analysis using ARIMA model. This model can be used to predict the hotel occupancy rate for the near future years. We also calculated the RMSE value for the cross validation.

APPENDIX A Method:

We got the data from 'Boston analysis', this is a website which has a collection of different data sets about the different aspects of Boston city in United States of America. This web site maintains nearly 245 different data sets. We used the data set called "economic indicators" to find some value able insights from the data. The dataset contains the different interconnected aspects that helps to gauge the economy of the city. The data sets contains the data from 2013-2019.

This data set has 17 different columns that are the economic indicators of Boston city. The records contain the information like Number of domestic and international passengers at Logan Airport, Total international flights at Logan Airport, Hotel occupancy for Boston, Hotel average daily rate for Boston, Unemployment rate for Boston, Labor rate for Boston, Number of units approved, Total development cost of approved projects, square feet of approved projects. In these many columns we are going to use only some of useful columns.

We got nearly 85 rows in the dataset to use. We started cleaning the data by looking at the NAN and null values. The data set was cleaned and ready to perform statistical analysis.

We started by visualizing the data using bar horizontal graphs, we also saw the increase in the number of visitors to Boston across the years. Correlation heat map was used to find out the correlation between different variables that can be used to predict the hotel occupancy in the future. The Logan international flight had the highest correlation but that correlation was not much useful.
We started to look after the patterns or trends of the hotel occupancy rate. After visualization of the year wise data, we started work on monthly analysis, the results said that July is the month in every year where the number of visitors are more in Boston.

Since we saw a pattern in the increase and decrease of the occupancy rate of the hotel, we implemented time series analysis on the data present. We plotted ACF and PACF look at the correlation between a time series and its lagged values and the correlation between a time series and its lagged values, removing the effects explained by intermediate lags. Then we implemented fit the

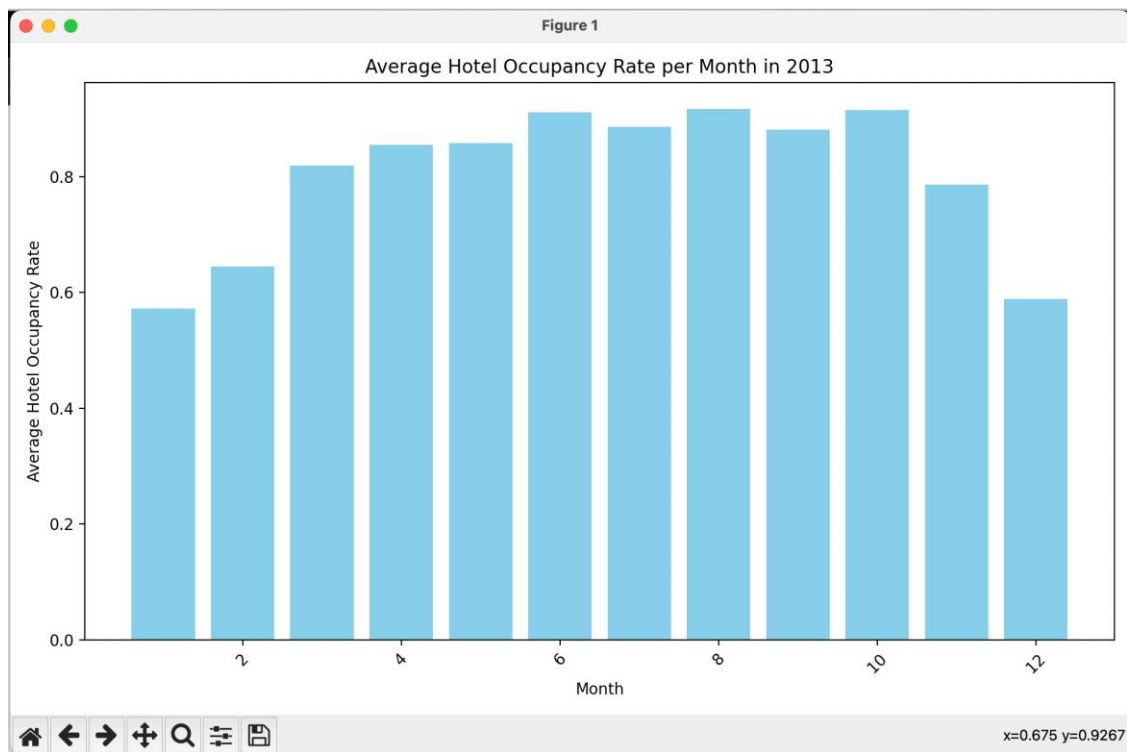model (ARIMA) to forecast the hotel occupancy rate for next coming year.

APPENDIX B Results:
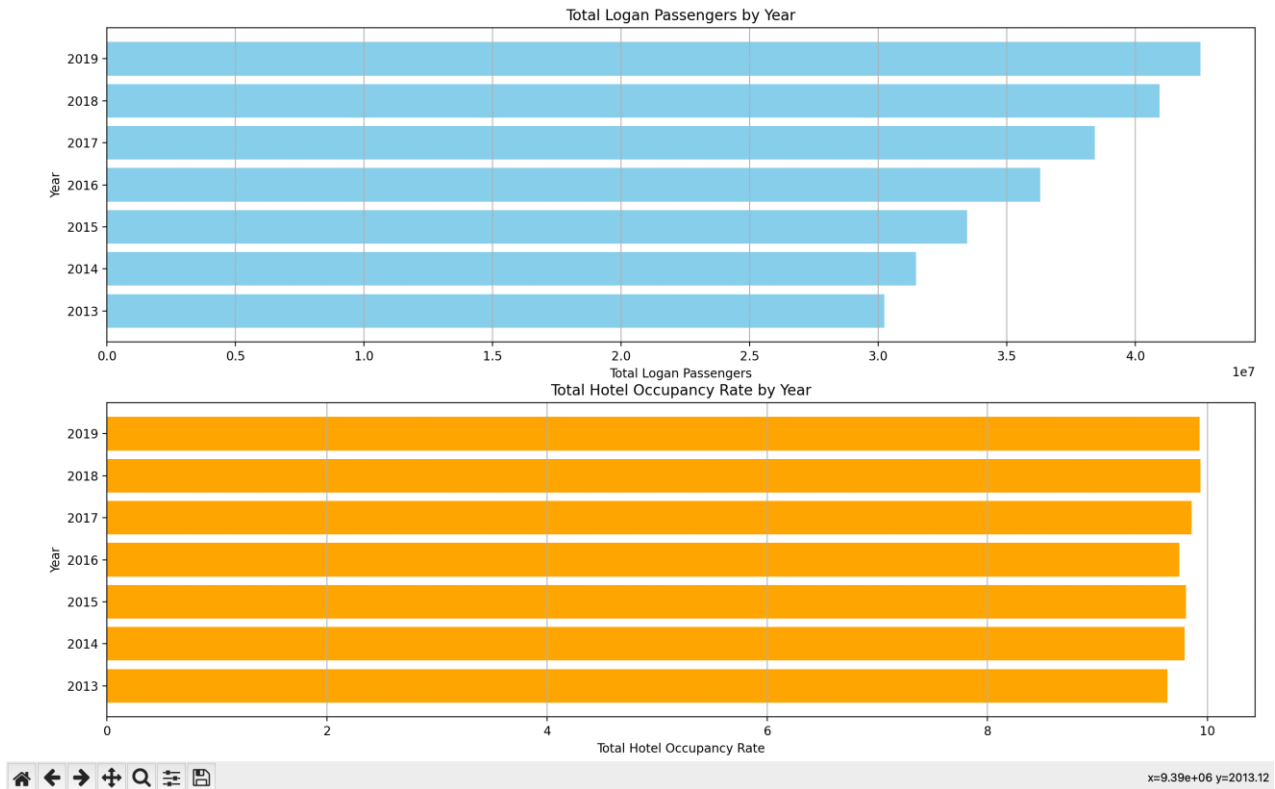
This is the horizontal graph showing the total Logan passengers by years and total hotel occupancy rate by year.
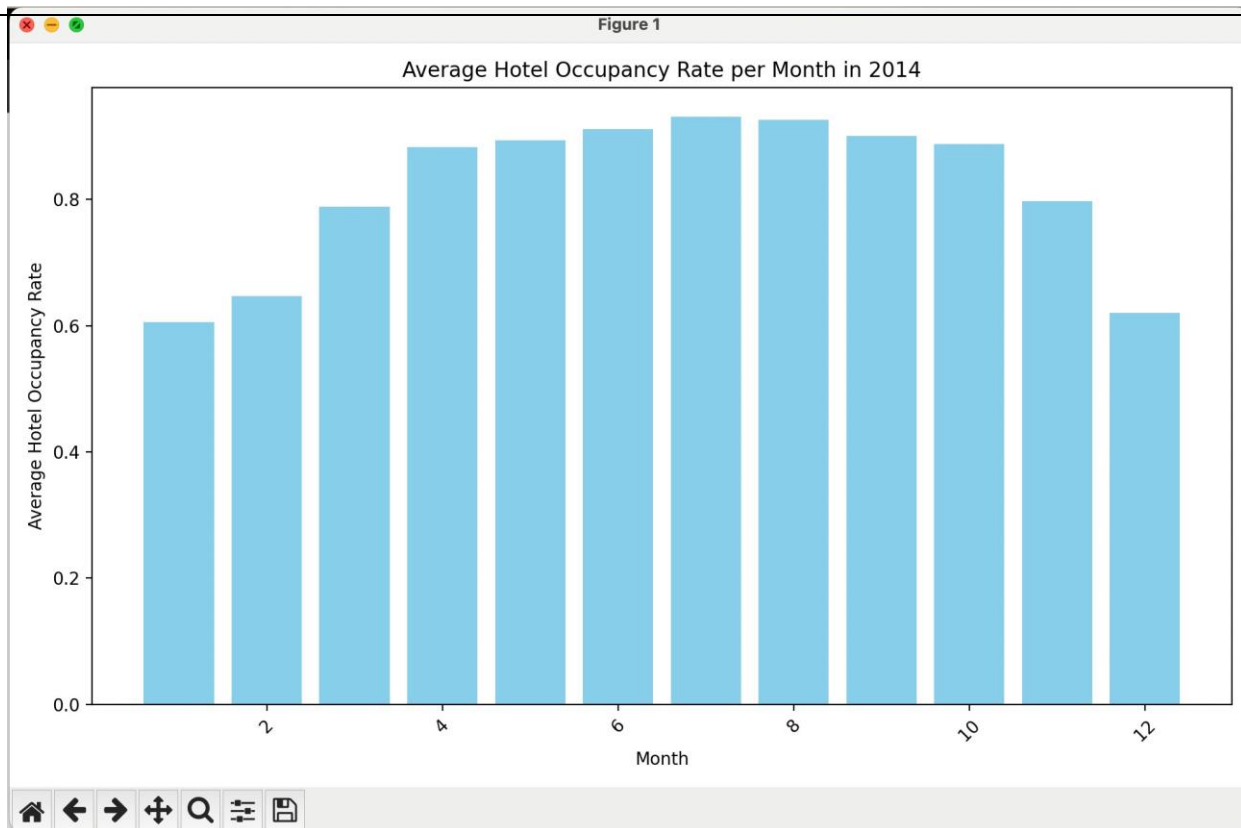
Correlation Heatmap
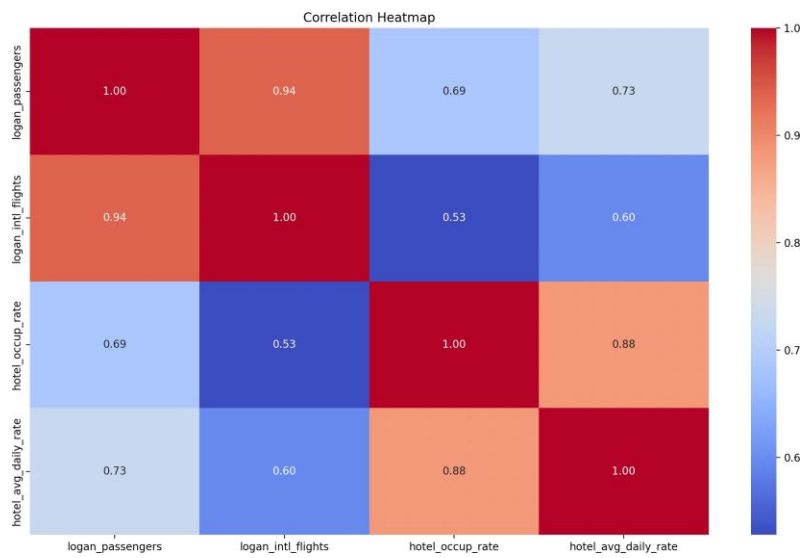
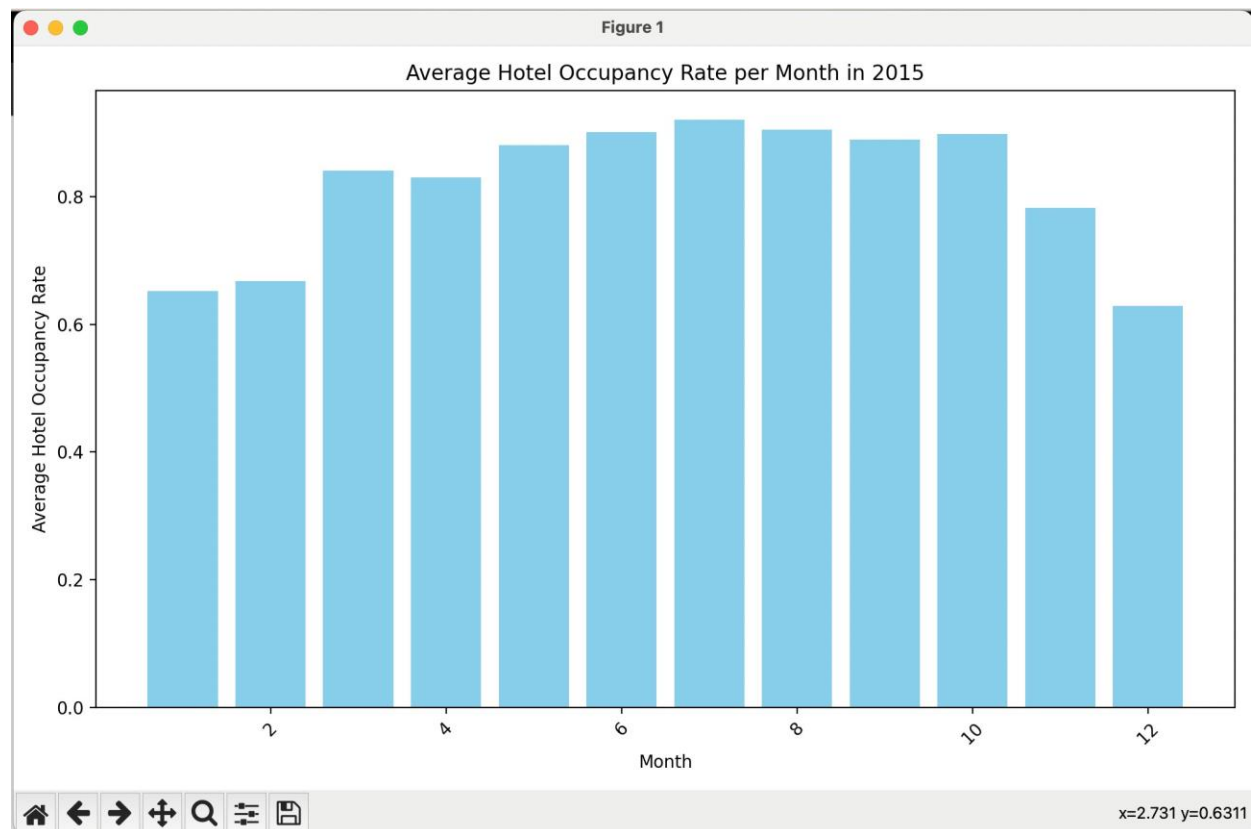The above image shows the correlation heat map to show the correlation between the variables.

MONTHLY ANALYSIS:
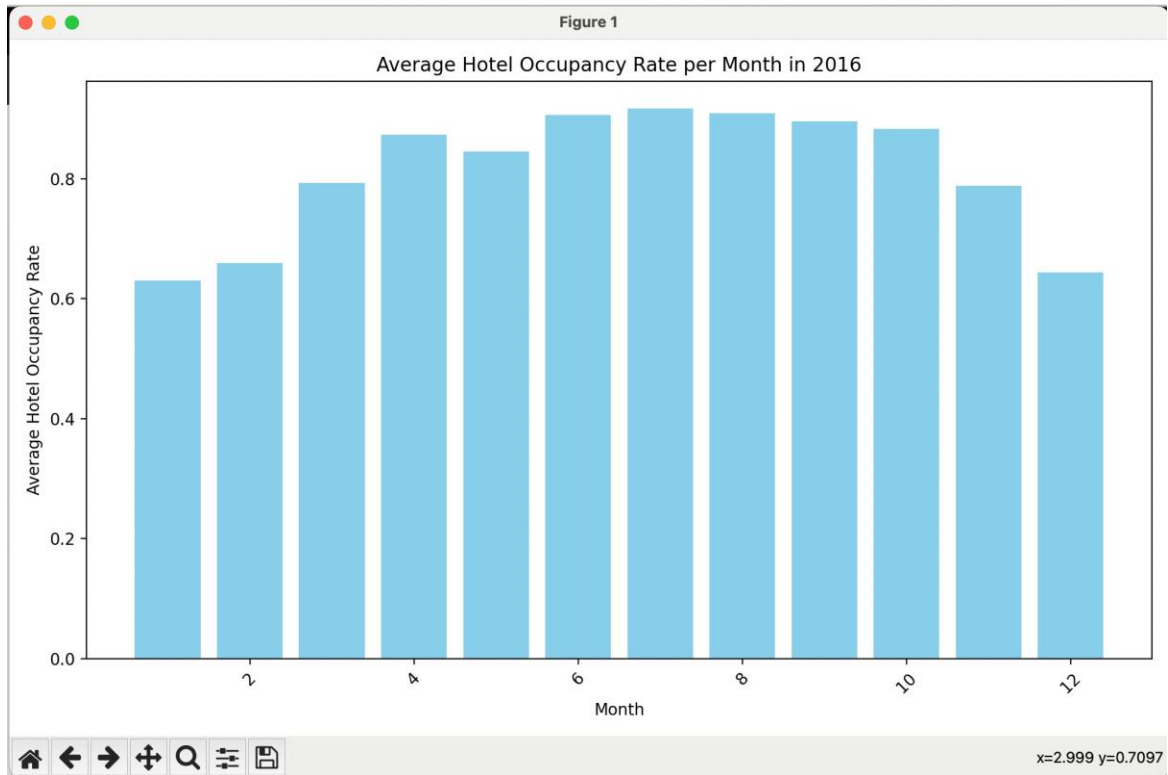

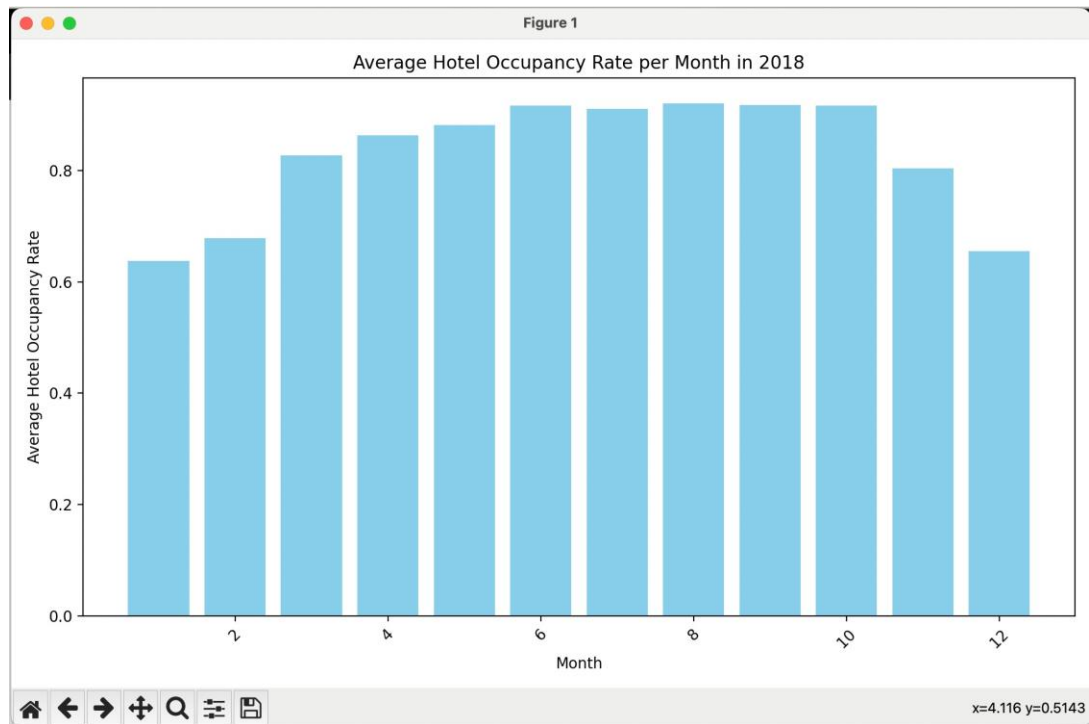The graph shows the average Hotel occupancy Rate per month in 2013.

Average Hotel Occupancy Rate per Month in 2014

The graph shows the average Hotel occupancy Rate per month in 2014.

Average Hotel Occupancy Rate per Month in 2015

The graph shows the average Hotel occupancy Rate per month in 2015.
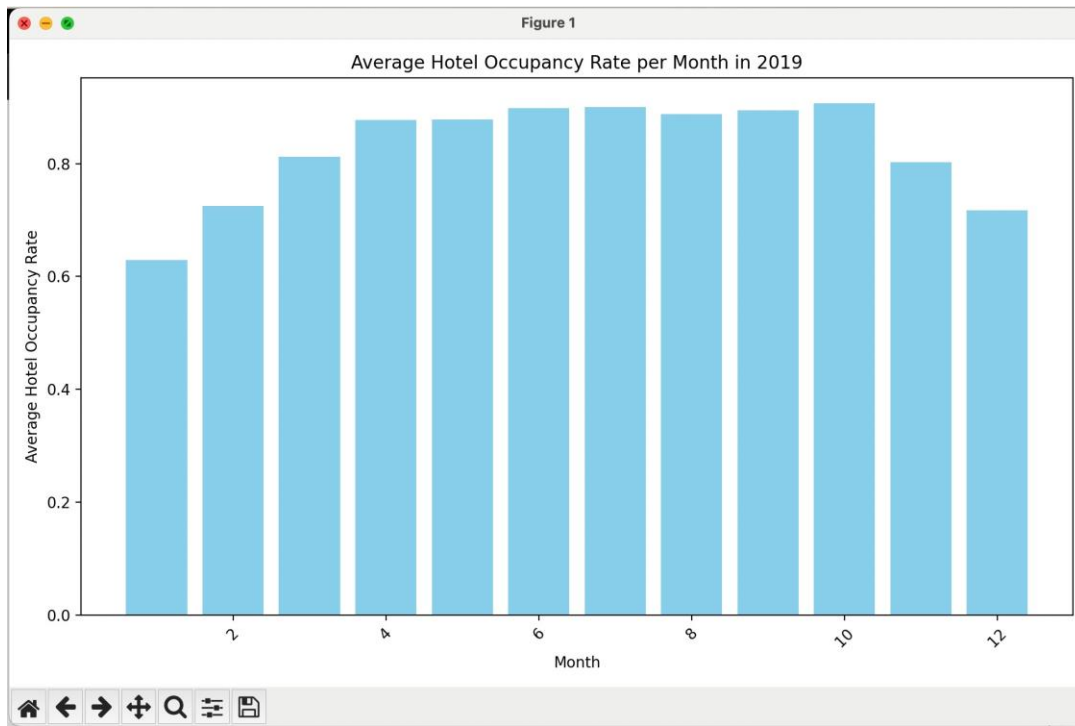

Average Hotel Occupancy Rate per Month in 2016

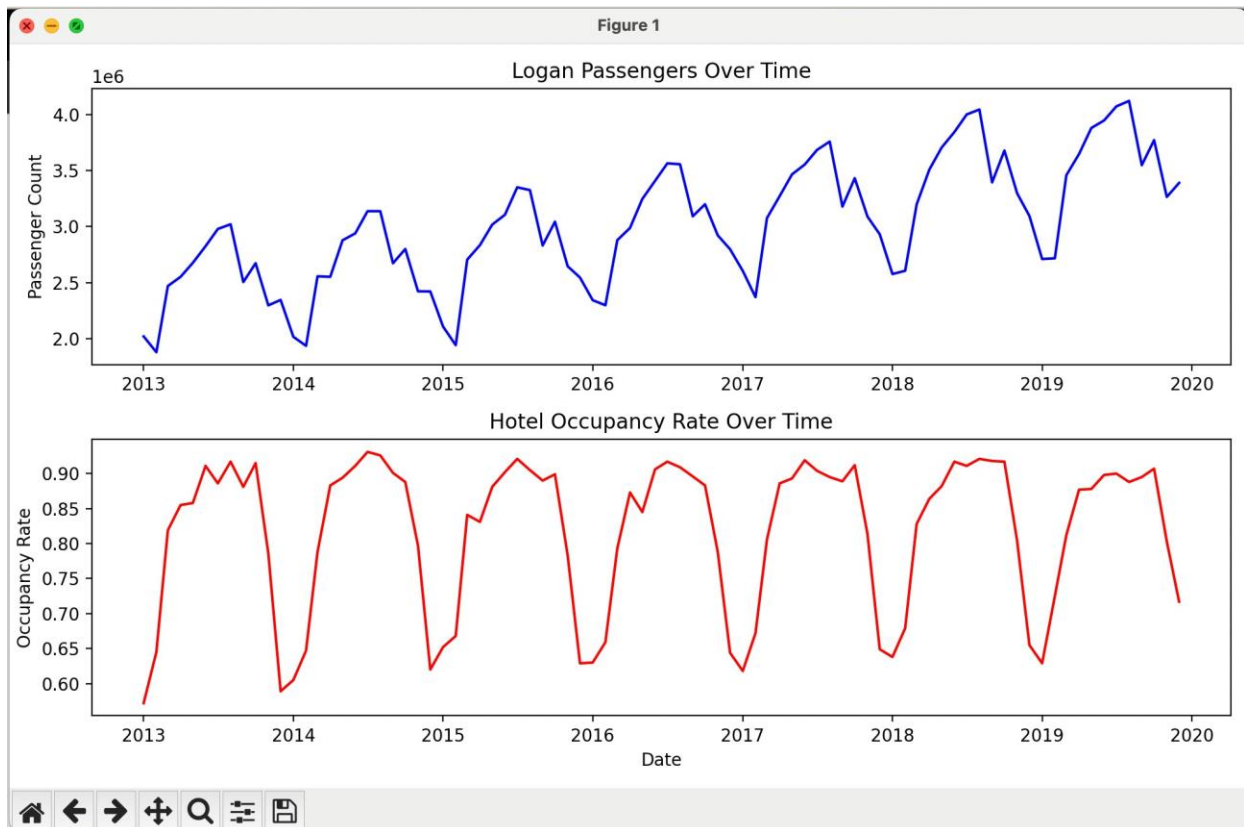Average Hotel Occupancy Rate per Month in 2017
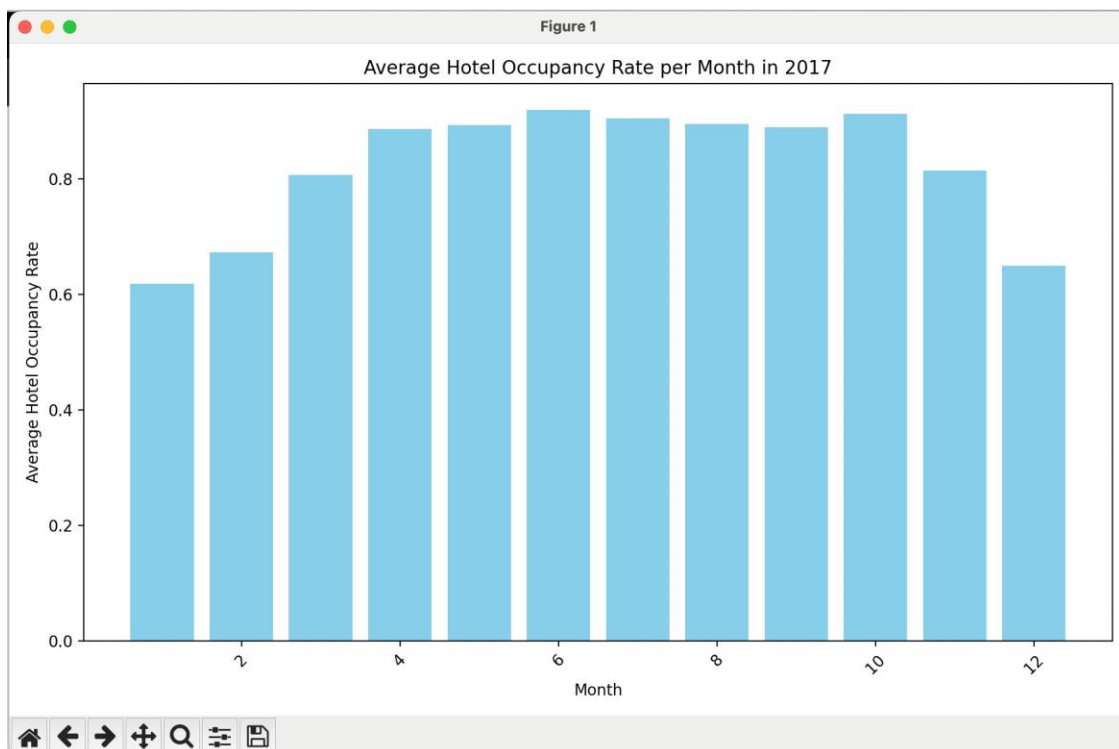
The graphs above show the average Hotel occupancy Rate per month in 2016 and 2017.

The graphs above show the average Hotel occupancy Rate per month in 2018.

Average Hotel Occupancy Rate per Month in 2019

The graphs above show the average Hotel occupancy Rate per month in 2019.

Time series analysis

This is the time series analysis of Logan passengers over time and time series analysis of Hotel occupancy rate over time.

ACF/PACF:

Peaks or significant spikes at certain lags indicate strong correlations between the time series values at those lags. It helps understand the seasonality or repeating patterns in the data.

PACF helps identify the direct relationships between observations at different lags, independent of other observations. Significant spikes

at certain lags indicate the direct influence of those lags on the current observation.
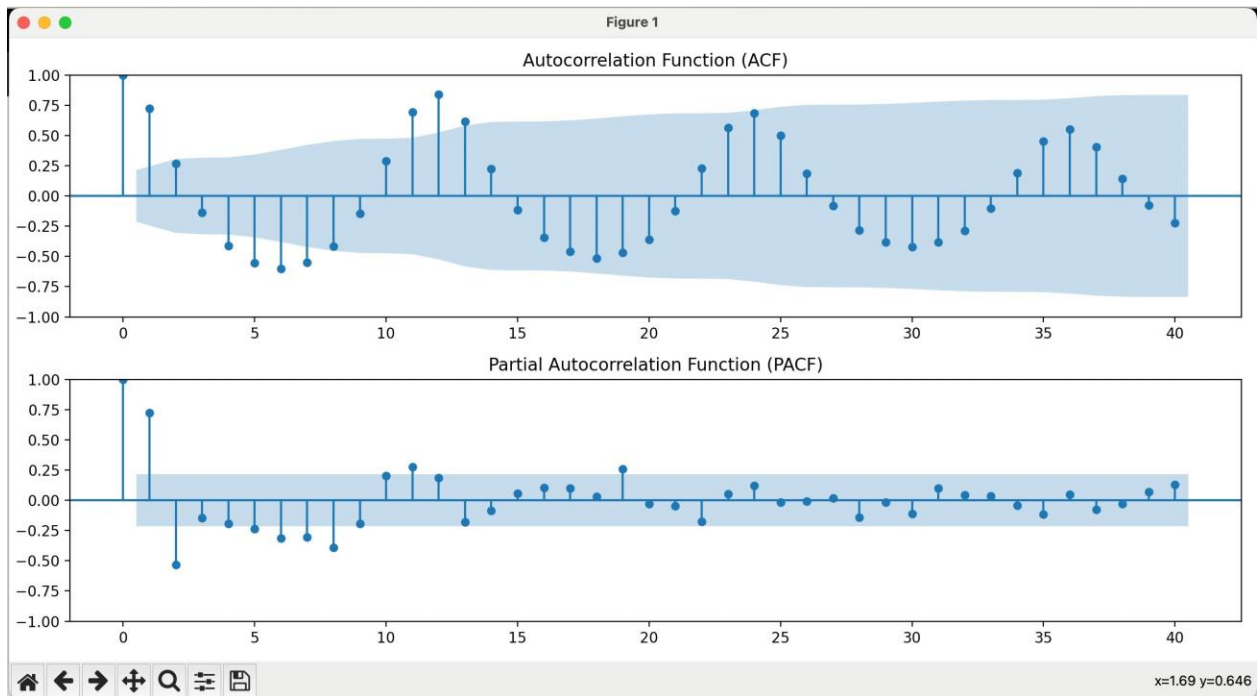
Original and Forecast of hotel occupancy rate.

Hotel Occupancy Rate - Original and Forecast
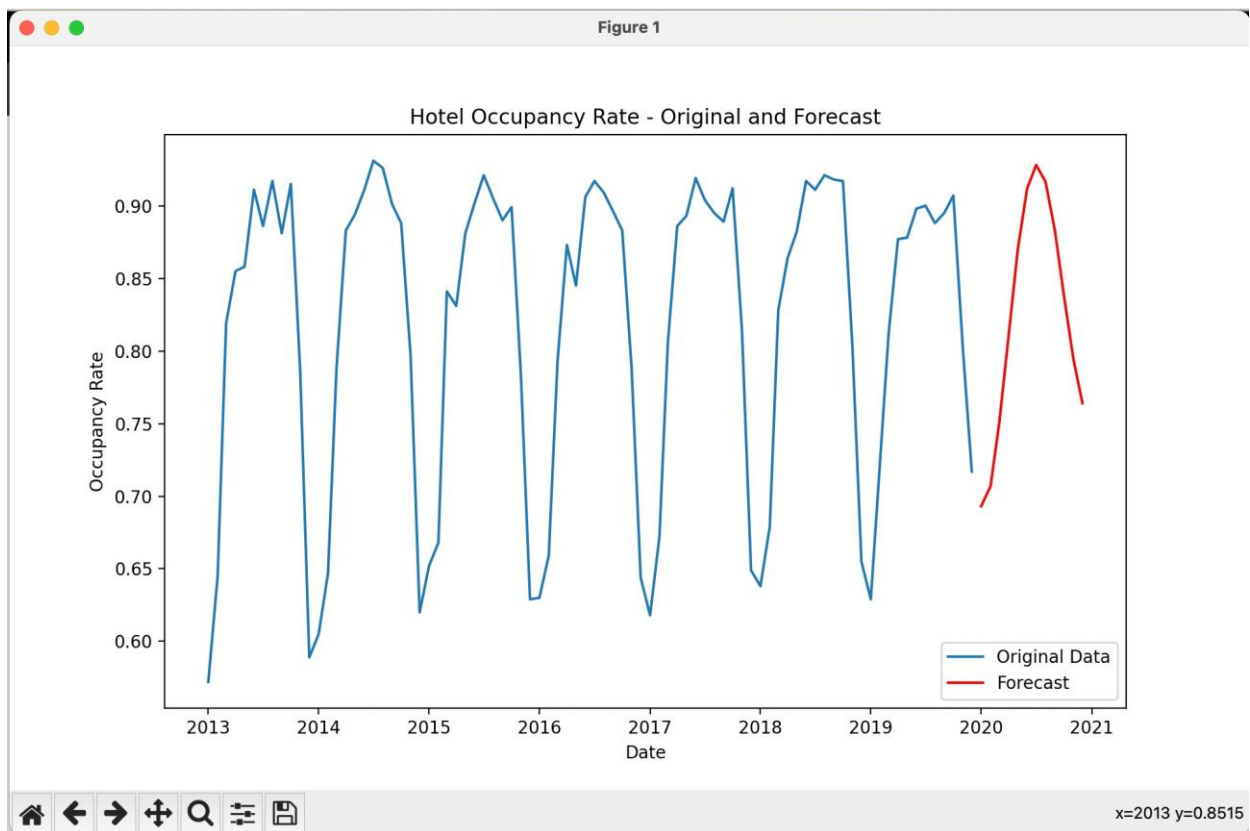
APPENDIX C code:

The code for this project is uploaded to the GitHub repository and the link to the repository is:

https://github.com/surajBR7/MTH_Project_3

CONTRIBUTIONS:

ALL THE GROUP MEMBERS EQUALLY CONTRIBUTED TO THE PROJECT