Hotel Data, Early Reservations

AaronKohn

Bellevue University

DSC530: Data Exploration and Analysis

June 2021

Hotel Data, Early Reservations

Hotel reservations are at times made the same day as arrival while other are made earlier. Some reservations are made more than a year and half prior to arrival date. Using the Hotel booking demand dataset (Antonio, Almeida, and Nunes, 2018) this analysis is designed to explore the relationship between reservations, demand and price.

**Outcome of Exploratory Analysis**

Using the data cited above there appears to be a difference in the reservation times for busy seasons. Looking at the CDF for LeadTime there is an obvious increase in early reservations for the summer half of the year. For this study summer is defined as week 13 through 39. This difference is statistically significant with a p-value less than 0.001. Price difference is evident as well between summer and winter. With a p-value less than 0.001 as well.

These findings are duplicated when testing correlations between the number of reservations for each week and the above variables. However, it is interesting to note that LeadTime and ADR (Average Daily Rate) are negatively correlated. With Pearson’s correlation equal to -0.0656 and a p-value less than 0.001. This seems to suggest that booking early may lower hotel prices.

**Analysis Limitations**

The most obvious step that would need to be taken in order to have a more complete analysis is to conduct a time series analysis. This would provide insight into seasonal fluctuations as well as year to year change.

Another area that requires further attention is dealing with cancelled and incomplete data. There appears to be reservations that were not completed. The variables were collected from various sources and then calculated to determine the values. This seems to have led to some interesting values as evident in the outlier that was removed. The average daily rate was more than 10 times higher than the next closest rate. This instance is coded as cancelled with a one day stay and nonrefundable. There is no indication as to how long the reservation was originally for. Cancelled data may be missing price as well since no money was paid. The average rate includes all services charged at the stay and may not necessarily reflect the price of the room alone.

**Other Variables**

The analysis focused on the above-mentioned variables. There are most probably many factors involved in pricing and reason for early reservations. Type of room is provided but is given in category code. There is no explanation as to what the code indicates. It can be assumed that room type has an effect on pricing.  
 Something that may influence reservation time is the reason why someone is coming. An event that cannot be missed may lead one to book in advance. This information may not be possible to collect in a reliable way.

**Incorrect Assumptions**

When starting the analysis, the assumption was that the three variables would be positively correlated. It was surprising to find that although demand was correlated with both early reservations and price those two were negatively correlated. Furthermore, although the correlation between price and reservations is small further analysis would need to be done to control for demand. Reserving early may have an even greater effect on price than already indicated.

**Challenges and Further Study**

As discussed above a time series analysis would have been appropriate. Merging variables to combine dates in order to create a datetime object would require further research on how to correctly merge the dates.

Although regression analysis was preformed further learning is needed in order to understand the results. The purpose of the study was not to try to predict ADR or LeadTime rather to understand their relationship with demand and each other. Using the model for prediction dose reduce the RMSE a little indicating a small relationship but not much predictive power.

**Conclusion**

Following the analysis of the data it can be said that demand is positively correlated with both price and early reservations. This does match common belief that at times of higher demand one is likely to reserve early to ensure availability. Secondly, it seems that price goes up closer to the time of stay. The exact effect of reservation time on price would require further analysis.

References

Nuno Antonio and Ana {de Almeida} and Luis Nunes, (2019). Hotel booking demand datasets. *Data in Brief 22.* Retrieved from [Hotel booking demand datasets - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S2352340918315191#bib5)