Project Summary and Description

Mathai Paul

gx3081

# Summary of Presentation

Then presentation starts by introducing the objective and the speaker. After which we immediately move to care. This section answers. The question “why care”. Highlighting that there is low revenue mainly due to low usage of Airbnb in the central regions of Singapore. The mean price of central region is higher and the availability too. Which is a situation that Airbnb does not want to be in. Also Highlighted is the fact that people stay longer in non-central regions since the availability is extremely low for these regions over the year. Primarily due to extensive demand and secondly due to lack of properties. People also tend to stay longer in home/apartments, which also lack in availability.

Moving to the “Do”. It was recommended to increase the number of properties in non-central regions. Specifically increase the number of houses/apartments and a private rooms. Even more specifically in the western region since the cost is even higher than central regions and availability is very less, while the demand seems to stay the same.

# Data Wrangling

Post importing the dataset into a DataFrame, first think I observed was that rank was not of much importance and with a lot of missing/default values of 0. Hence the column was dropped. Since I wanted to look at the numerical insights to find problems, dropping all the semantic columns like id, host details and reviews made the data less distracting. Next step was to identify the missing data in the available data and fill them with the median values. Except for the 2 entries of name there were no gaps in the data.

# Tools and Methods

Major Tools used were Jupyter and Python3, while the libraries to do the data analysis are numpy, pandas, matplotlib and seaborn. Beginning with Exploratory analysis by pair-plotting and multiple histograms to observe any interesting patterns. Where the unique relationship between room type-price, region-price and lack of properties were observed and plotted with the correct visualization. Based on the current data of minimum nights, plotting a LMPLOT shows that unless changes are made the non-central regions will continue to decline in usage. Further exploring the data continued to reassure the anomaly is real and that a change in these areas would increase revenue. The increase in revenue predicted, 20% increase in revenue by doubling the properties in non-central was concluded. The case for the figure was the higher cost of western region and the lack of full home/apartment options in non-central regions.