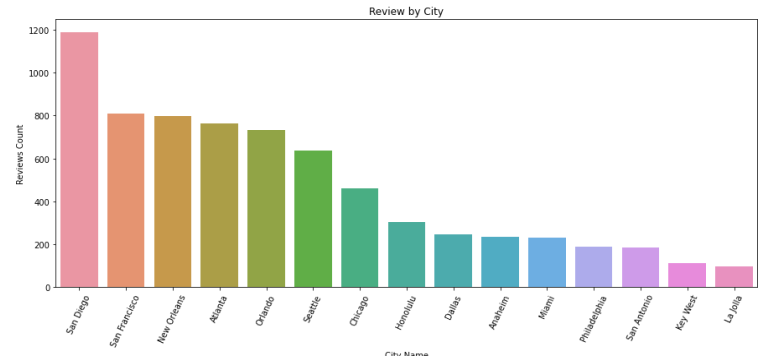


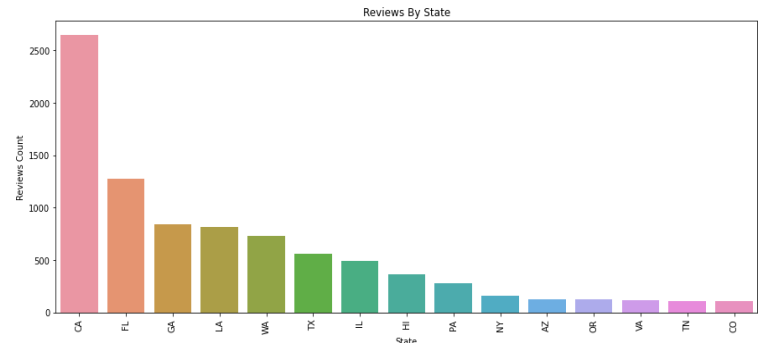
Exploratory Data Analysis:

Geography Dimension:

Based on the city dimension, from the bar chart below, it's clear to see TOP 5 cities with most reviews: San Diego, San Francisco, New Orleans, Atlanta, Orlando. And San Diego's review number is much more than the second place (San Francisco)

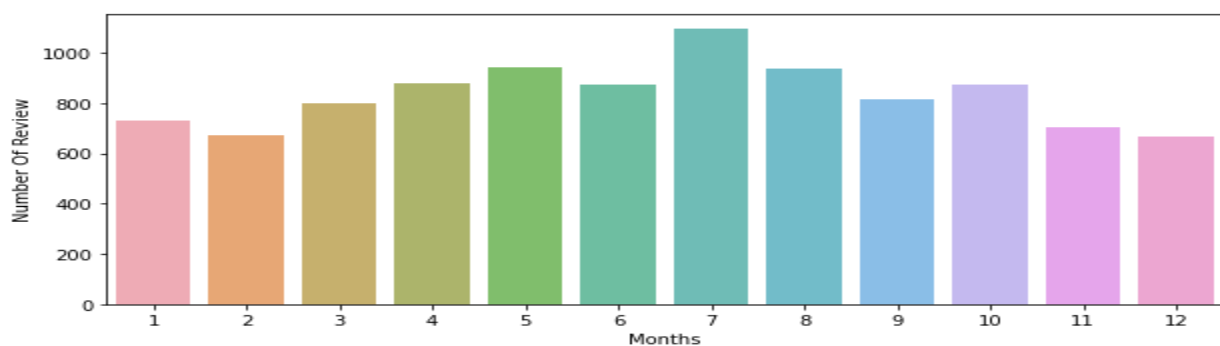
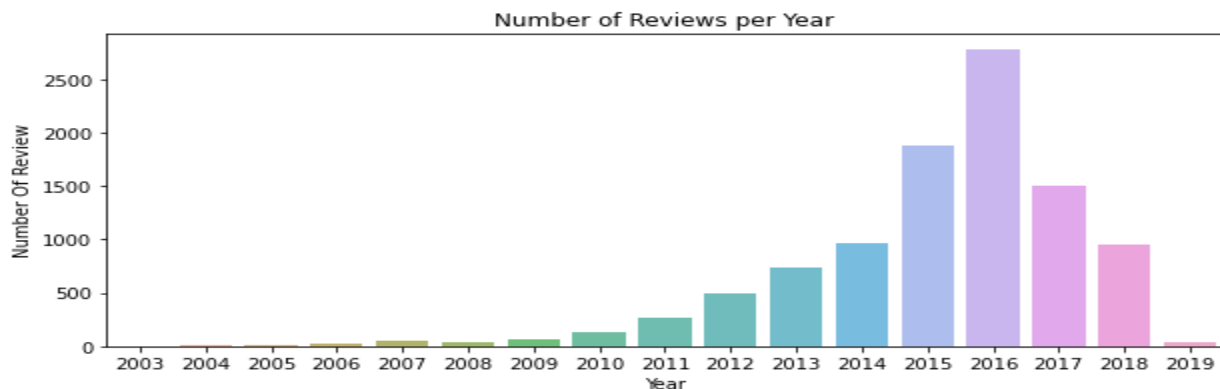


Based on the State dimension. California accounts for the vast majority of hotel reviews, and Florida the second most visited or reviewed state. It is reasonable to relate the hotel reviews with the degree of tourism. California and Florida are probably most people would like to visit during vacation due to their great sunshine and warm beach



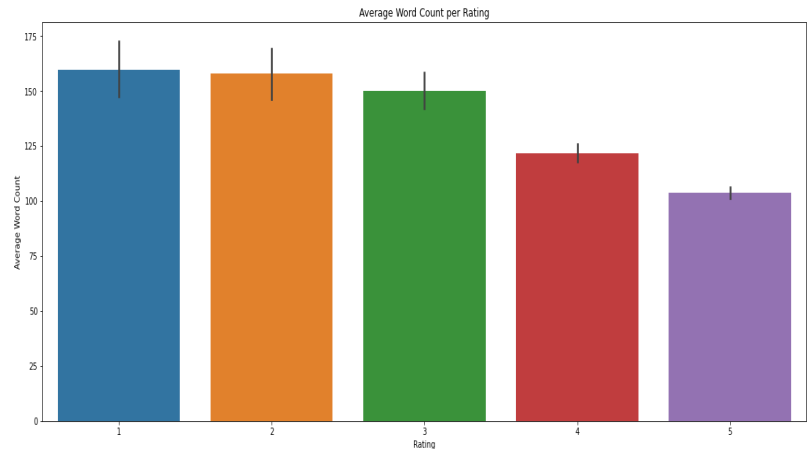
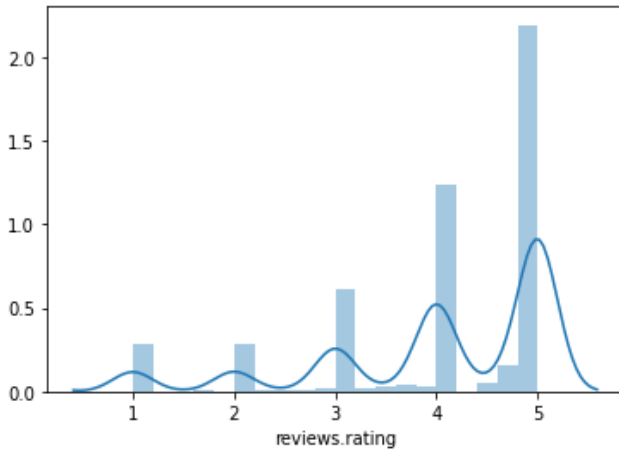
Date Dimension:

Based on the Year Dimension, most of the reviews from the time period between 2014 - 2018, with 2016 being the largest in volume. Based on the Month Dimension, most of the reviews occurred during the summer months.



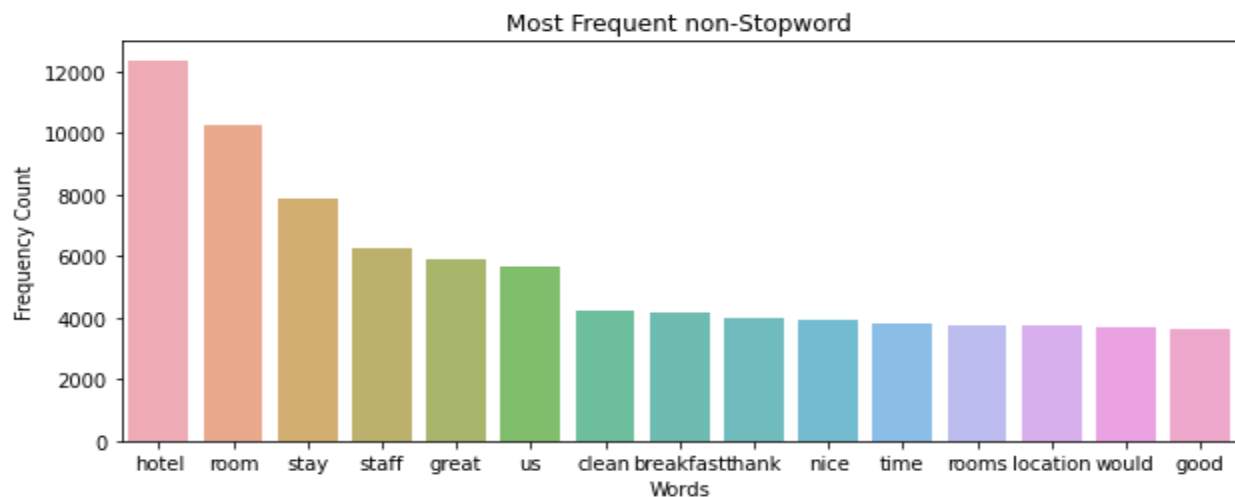
Rating Distribution:

According to the distribution plot below, we can see that most of the review ratings happen between 4 – 5 which means that most of the reviews got a high rating and were positive. Additionally we found that the lower the rating the higher the average word count.



Word Frequency:

Doing some simple tokenizing and eliminating stopwords, we found that the most frequent word was hotel. Additionally “great”, “nice”, and “good” being in the top 15 most frequent words makes sense given most of the reviews were positive (4s and 5s).



Next Steps:

We will do some more text analysis on the reviews. We hope to do some sentiment analysis to decipher the good and bad reviews. Additionally we will do a large ngram_range to see what are some of the words associated with the strong positive and negative words.