Choosing A Best Location in NYC For Operating a Restaurant

—

By Qi Leng

Background

I am helping a friend in New York on solving a business problem. However, New York is a big city and we do not want to blindly choose a location without research. We had three preferable locations in mind:

- the Vessel area
- Broadway area
- SOHO area

I helped him gather information on three aspects of each location: n**umber of WiFi hot spots** nearby, **crime rates** nearby, and the **price range** distribution of nearby restaurants.

Data Acquisition

- Kaggle's 'New York City Crime Dataset'. It has 20,000 rows and 24 columns, and in this project we are utilizing the column Longitude and Latitude as the geographical location for each crime happened. Each row represents a crime happened in a location in New York, and the columns are features of that incident, including locations, crime description and crime id, etc.
- Kaggle's 'WiFi Hot Spot In New York City'. It has 2,566 rows and 29 columns, with each column being the feature of a WiFi spot in New York, and each row being a WiFi hot spot in NYC.

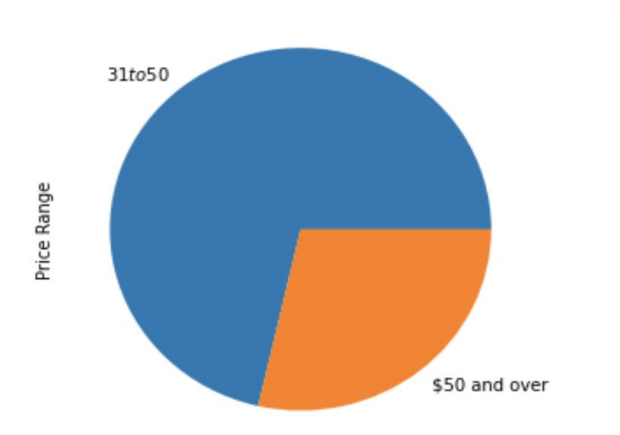
Data Analysis - Comparison Table

	Location	Number of Wifi	Number of Crimes
0	The Vessel	9	39
1	Broadway	68	153
2	SOHO	15	112

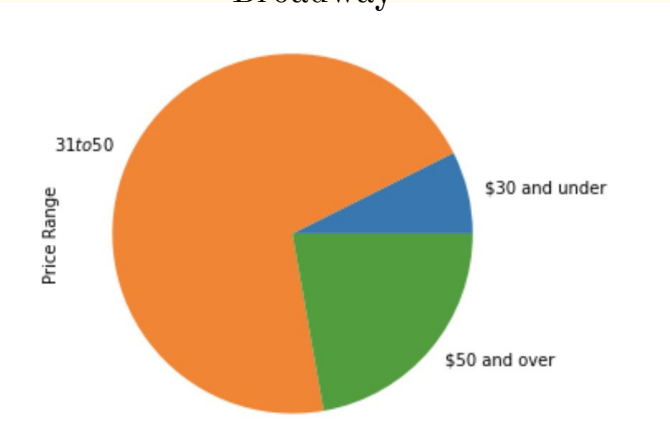
The Vessel is safe but it does not have sufficient WiFi hot spots. Broadway, on the other hand, is too dangerous. SOHO could be a great choice.

Data Analysis - Pie Chart

SOHO

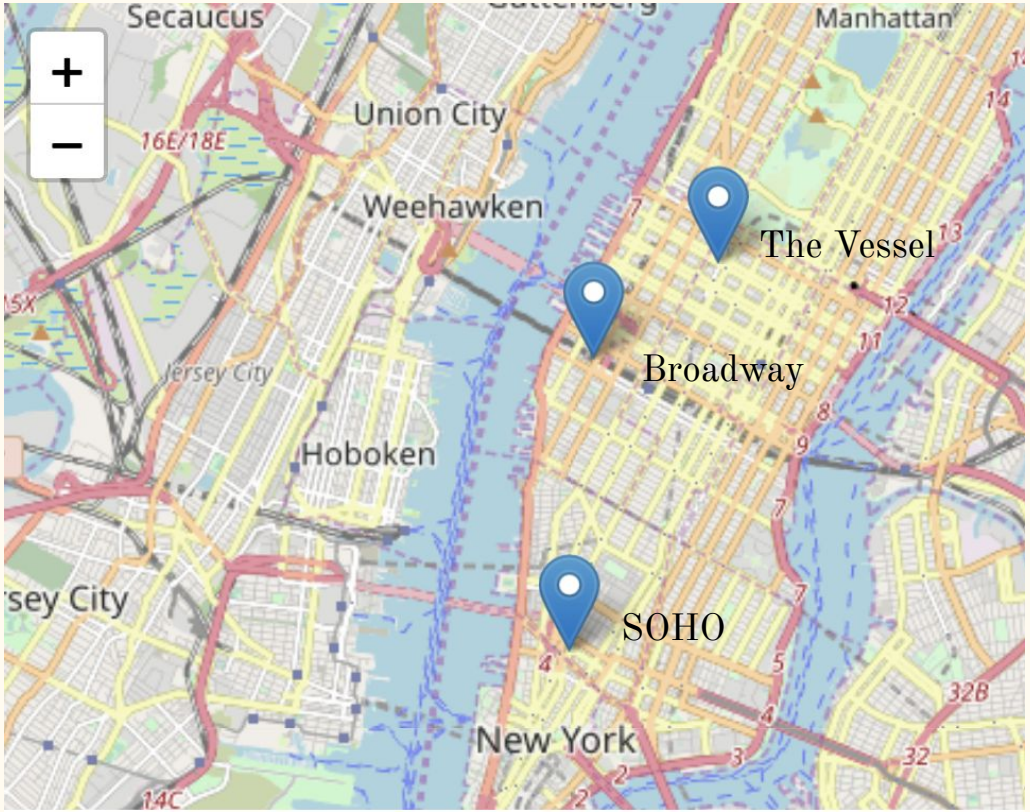


Broadway



Broadway's restaurants has less diversity of price range than SOHO's. The overall price average is higher in SOHO.

Data Analysis - Map



Three location choices are in Manhattan.

Conclusion and Future Directions

- Overall, SOHO is performing well on this competition. All three locations are in Manhattan, yet SOHO has features that are more preferable than the other two: SOHO has sufficient WiFi hot spots and less number of crimes, and it has higher price range.
- In the Future, I should explore more related dataset, and evaluate the features in multiple aspects: traffic, average rating, etc. I should also try a k-means clustering to predict the average rating and price on different areas, in order to choose the one location that is more preferable.