Capstone Project- Establishing a Strategically Located Restaurant Around Manhattan Financial District in NYC

The Battle of the Neighborhoods: Last Week Assignment

Applied Data Science Capstone Course by IBM

Table of Contents

Introduction: Business Problem

With a view to running a successful business venture, the stakeholders we will be supporting through our analysis want to make sure they prepare thoroughly before starting a restaurant business. Specifically, in crafting their business plan, they require in-depth analyses and data-backed decisions before investing their money. The business problem requires us to target a geographical area in the Financial District, Manhattan, in which to establish a strategically located Japanese restaurant, without sacrificing the affordability of a potential commercial premise and choosing a well-served area located near a subway station. We will craft our strategy mindful that our stakeholders intend to establish the abovementioned restaurant activity to meet the needs of those wealth professionals that hang around the Financial District neighborhood, thus exploiting the availability factor leveraging on those areas and streets particularly lacking in the same type of business.

Our effort, however, isn't only limited to define the geographical boundaries around which to set the business but to provide our stakeholders with other data-backed information to help them make the most efficient choice.

Data Section

Based on the definition of our business problem, the factors that may potentially influence the positive outcome of the business venture, and our choices as well, are the following:

- Number of existing restaurants in the neighborhood (any type of restaurant);
- Number of and distance of Japanese restaurants from the predefined area;
- Property Price Data.

With the aim of achieving the abovementioned goals, the following data sources will be used. Following data sources will be needed to extract/generate the required information:

- Number of restaurants and their type and location in every neighborhood will be obtained using Foursquare API;
- NYC coordinates and distance parameters obtained through the use of several libraries and, again, Foursquare API;
- New York City Location Data in Json format;
- Manhattan Property Price Data obtained from New York City Department of Finance Website.

Methodology

Our analysis will be preliminarily focused on determining the best choice in terms of geographic availability and proximity to a predetermined location, with this aim in mind, our first steps in the data analysis process will be addressed by looking at the geographic distribution of similar business ventures over a given area. With this aim in mind, the initial steps of our analysis will be focused around some data wrangling operations and exploratory data analysis. Specifically, we will first obtain New York City data from a json format file which will be converted into a Pandas DataFrame. Following this preliminary step, the analysis will subsequently cover some descriptive statistics helping us to determine where our investors may achieve greater profits by analysing Manhattan property price data. According to these findings, we will conclude our analysis by providing our investors with a well-defined geographic area around which to establish the restaurant based on criteria of availability and affordability.
For the purpose of backing up our analysis and strengthen the derived findings, an additional attempt of performing some inferential statistics based on the Manhattan Property Price Data obtained from New York City Department of Finance website has been carried out, however, as shown below, due to the lack of some crucial data for two specific attributes, we abandoned the initial choice. Indeed, the correlation patterns with

other attributes of the data where too "weak" (even negative in several cases) to produce meaningful results.

As shown in the tables below, the single correlation coefficients are consistently low and do not allow us to perform linear regression operations. For this purpose, in fact, the attributes that were originally believed to be more correlated with property prices were "LAND SQUARE FEET" and "GROSS SQUARE FEET", unfortunately, however such data does not display after execution of the ".corr()" Pandas method due to the lack of such information in the initial dataset for the specific areas we filtered the dataset for throughout the exploratory process.

|	BLOCK	LOT	zip_code	COMMERCIAL UNITS	TOTAL UNITS	YEAR BUILT	TAX CLASS AT TIME OF SALE	SALE PRICE
BLOCK	1.000000	-0.390543	-0.204496	0.011631	-0.020452	0.142560	0.238092	-0.016331
LOT	-0.390543	1.000000	0.659132	-0.046297	-0.060462	-0.341313	-0.069797	-0.073864
zip_code	-0.204496	0.659132	1.000000	-0.024001	-0.035688	-0.436816	-0.048555	-0.049067
COMMERCIAL UNITS	0.011631	-0.046297	-0.024001	1.000000	0.505762	0.018751	0.270126	0.168845
TOTAL UNITS	-0.020452	-0.060462	-0.035688	0.505762	1.000000	0.013958	0.121580	0.277706
YEAR BUILT	0.142560	-0.341313	-0.436816	0.018751	0.013958	1.000000	-0.015855	0.026495
TAX CLASS AT TIME OF SALE	0.238092	-0.069797	-0.048555	0.270126	0.121580	-0.015855	1.000000	0.374793
SALE PRICE	-0.016331	-0.073864	-0.049067	0.168845	0.277706	0.026495	0.374793	1.000000

Results

After having defined a strategic point around which to establish our research and, potentially, our restaurant, we've designated a specific location to determine the exact coordinates which constituted the starting point of our "visual" analysis through the mapping process. To achieve more precision in the address variable we specify the address of the Fulton St. Subway Line as our main purpose is establishing the restaurant in a well-served and central area within the Financial District itself as show in the picture below. Subsequently, we've predetermined a range to define and highlight the physical boundaries of a potential new business. Thus, after having defined our initial boundaries, we have started mapping the restaurants of the same type within the area using Foursquare API as shown below.
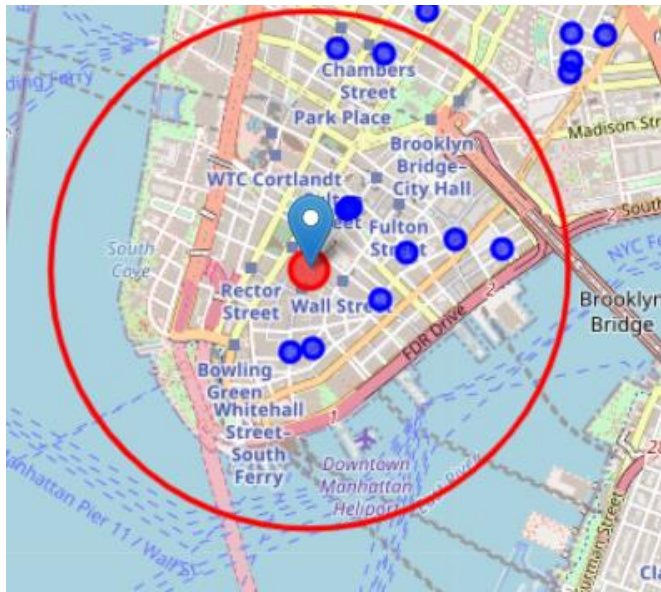
In the same picture we can appreciate a map showing all the restaurants of the same type (Japanese) around the area within a range of 1000 meters, however, for strategic reason is not only useful to consider restaurants of the same category but other restaurants as well so to offset the availability bias possibility.

Once again, using data obtained through Foursquare API, we have managed to obtain the locations of other restaurants of every cuisine and we have mapped them within the same area highlighting them in a different color for differentiation.
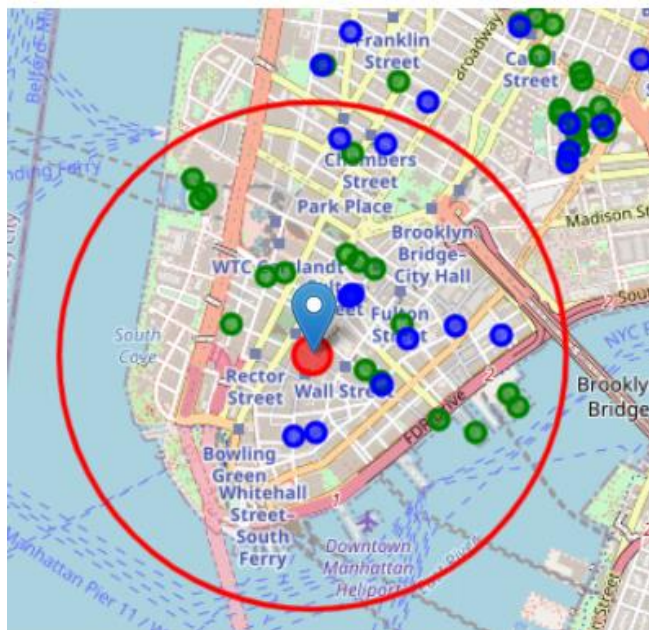
Such last step has really allowed us to graphically and geographically determine and highlight the gaps within our predefined area. Specifically, we could easily say that the area southwest Rector Street is particularly lacking in restaurants of every kind. We display below some descriptive statistics carried out on the same data.

Preliminary Analysis

```python
import numpy as np

print('Total number of restaurants in our predefined area:', len(all_restaurants
_filtered))
print('Total number of Japanese restaurants in our predefined area:', len(japane
se_df))
print('Percentage of Japanese restaurants in our predefined area: {:.2f}%'.forma
t(len(japanese_df) / len(all_restaurants_filtered) * 100))
```

```
Total number of restaurants in our predefined area: 39
Total number of Japanese restaurants in our predefined area: 24
Percentage of Japanese restaurants in our predefined area: 61.54%
```

Our preliminary analysis clearly shows a strong concentration of Japanese restaurant within our specific range. We therefore deepen our analysis so to take into account other fundamental factors concerning the affordability of a potential commercial venue around the area. With the aim of analyzing NYC property price data and, based on the geographical distribution of the venues in our predefined area, we want to analyze property prices as to determine the more convenient area to buy or rent a property to establish our restaurant. According to our visual analysis, the area most deficient in restaurants, especially Japanese, was that located further south-west of our area. Specifically, after some research online we've found that this neighborhood is "Battery Park". We therefore want to compare the property prices between Battery Park and Financial District as a commercial facility in the former neighborhood may be way cheaper to buy.

The results below show a data-frame and other descriptive statistics through a box-plot representation obtained after having filtered and grouped the NYC property price data by the zip-codes of the two areas:

Financial District zip-codes: 10004, 1000, 10006, 10038
Battery Park zip-codes: 10280, 10282

	BLOCK	LOT	COMMERCIAL UNITS	TOTAL UNITS	YEAR BUILT	TAX CLASS AT TIME OF SALE	SALE PRICE
zip_code							
10004	19.158537	1754.280488	0.024390	1.292683	1678.536585	2.048780	1.126281e+07
10005	32.034682	1231.023121	0.219653	5.757225	1848.728324	2.057803	1.053303e+07
10006	20.962963	1267.098765	1.962963	2.956790	1983.543210	2.049383	4.370808e+06
10038	94.762431	625.618785	0.342541	0.906077	1878.613260	2.099448	2.070025e+06
10280	16.000000	4300.681034	0.000000	1.000000	1061.827586	2.017241	1.029132e+06
10282	16.000000	4347.250000	0.000000	1.000000	2000.000000	2.000000	2.188501e+06

We can clearly spot a trend and see that the biggest outliers seem to occur within the zip codes related to the Financial district, while the last two seem to present more balanced price ranges and a potentially more convenient choice in terms of affordability of the commercial venue in the area.
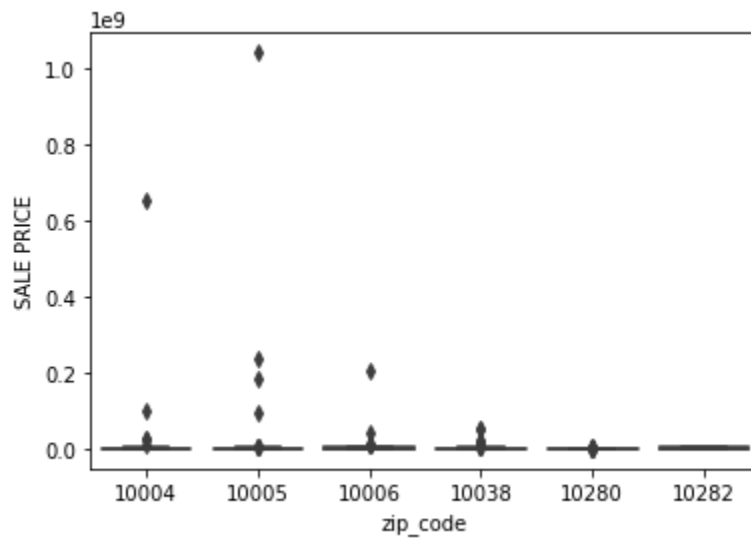
Discussion and Conclusion

Following a merely visual analysis of the venues belonging our category of interest within the predefined area, we have preliminarily determined the geographical boundaries that have allowed us to perform additional analysis concerning property prices in Manhattan and, more precisely, around the area that could allow our stakeholders to make a strategically better choice in terms of both affordability and proximity of the commercial venue. As can we clearly see from the data frame above, average prices differ widely across zip codes. Specifically, one of the zip codes (10280) concerning our area of interest, which has been shown before to lack in Japanese restaurants, shows lower sale prices as compared to most of the others). We could therefore conclude that renting or even buying a venue in the area bounded by the zip code "10280" may constitute a more efficient and strategic choice. We have thus highlighted the area most deficient in restaurants, especially Japanese one, which was that located further south-west of our area but still within the predefined range imposed by our stakeholders.

Additionally, a comparison of the property price data of the two districts has allowed us to help them determine a geographically more convenient choice to potentially buy or rent a commercial venue whereby to establish the restaurant. The strategic analysis that we have provided our stakeholders with has taken into account every predetermined factor: proximity to services and subway lines (specifically Fulton St. Subway Line stop); lack of potential competitors within the area and within a given distance, addressed the availability bias for our prospective customers.

- Introduction where you discuss the business problem and who would be interested in this project.

- Data where you describe the data that will be used to solve the problem and the source of the data.

- Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

- Results section where you discuss the results.

- Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.

- Conclusion section where you conclude the report.