Capstone Project for IBM data science certification

BY FLORINA OPREA

Identifying best location to open a Pie Restaurant in USA

- **Introduction**
- **Data**
- **Methodology**
- **Result**
- Conclusion

Background and problem description

A Start-up business, Pie Friends Ltd., asked me to find a best place for them to open first restaurant. Nowadays the market is highly competitive and as the start-up, Pie Friends want to have a deep analysis which will provide a good understanding and help in reduction of risk.

A restaurant is a business which prepares and serves food and/or drinks to customers in return for money, either paid before the meal, after the meal, or with an open account. In this case we have a pie-oriented restaurant.

Possibly best location for this restaurant will be determined by studying and analyzing selected factors.

Data

Cities in United States with population density and coordinates
https://en.wikipedia.org/wiki/List_of_United_States_cities_by_populatio

Cities in United States with Per Capita Income

https://en.wikipedia.org/wiki/List_of_United_States_counties_by_per_capita_income

2014 New York City Neighborhood Names
https://geo.nyu.edu/catalog/nyu_2451_34572

FourSquare API to get list of venues.

Above sources are free for use or free with conditions (FourSqare).

Data

Fig. 1 Sample dataframe of cities with population density and location after cleaning

City	State	Population density in Km2	Radius	Latitude	Longitude
New York[d]	New York	10,933/km2	17363.755354	40.6635	-73.9387
Los Angeles	California	3,276/km2	21649.480363	34.0194	-118.4108
Chicago	Illinois	4,600/km2	15076.471736	41.8376	-87.6818
Houston[3]	Texas	1,395/km2	25248.762346	29.7866	-95.3909
Phoenix	Arizona	1,200/km2	22750.824161	33.5722	-112.0901

Methodology

Analytic approach

New York city has 5 boroughs, from which 2 were selected in this analysis, Queens and Staten Island and the exploratory analysis was done as described below.

Exploratory Data Analysis

United States with population density and coordinates data is gathered from Wikipedia page with BeautifulSoup and transformed into a dataframe with pandas library.

This dataset contains population density and coordinates which will further help in selecting the borough and will be used to get venues from FourSquare

Geopy and Folium libraries were used to create maps.

Cities in United States with Per Capita Income is processed in same way as previous dataset.

This dataset contains per capita income which is also a component the decision is based.

2014 New York City Neighborhood Names, a json file which contains neighborhood names and coordinates. Which was used to create a dataframe containing boroughs, neighborhoods and geographical coordinates.

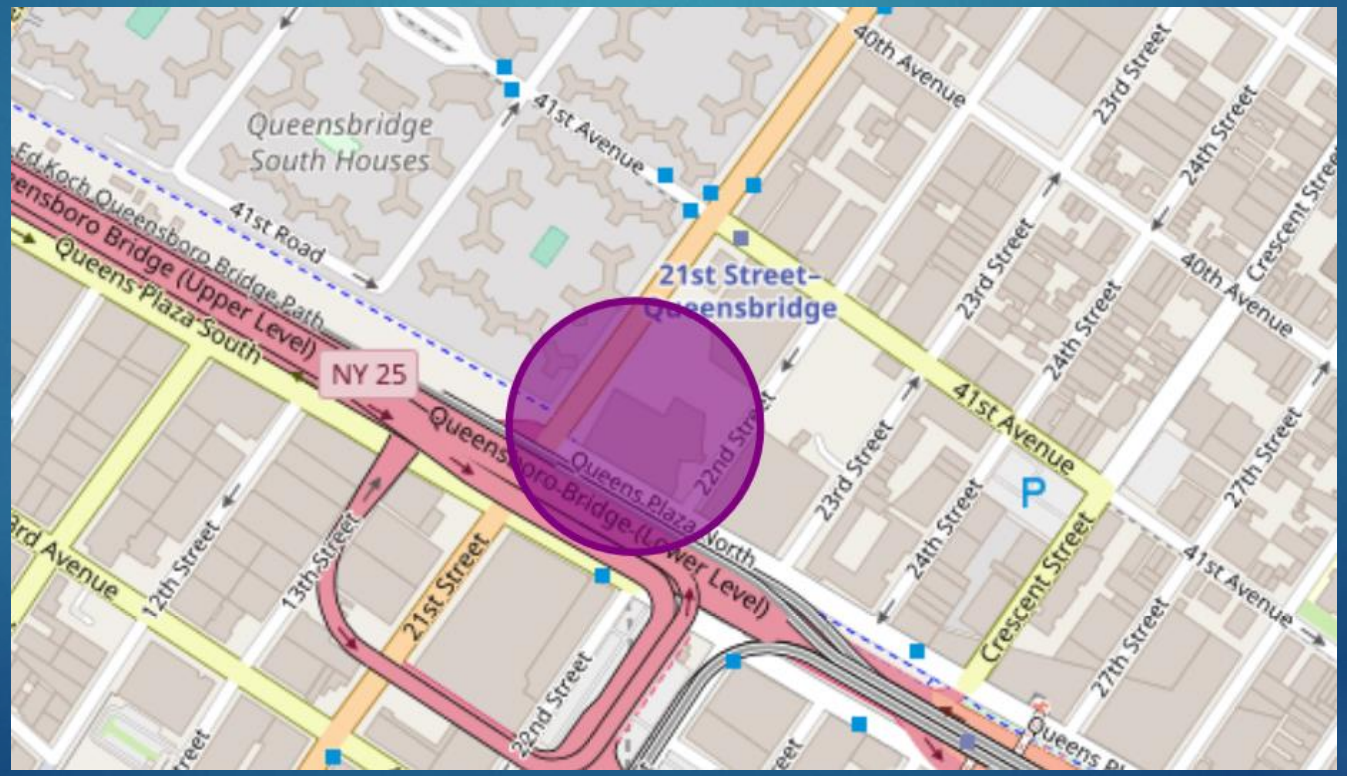
FourSquare API which was used to get venues in specific locations, which were further analyzed by adding weights.

Fig. 2 Map of New York with neighborhoods

Result

Based on the analysis done Queens- Queensbridge, Queens Plaza corner with 21/22nd was the selected location to open new restaurant

Fig. 2 Map of selected location



Conclusion

In this analysis I explored different boroughs and neighborhoods in New York to identify best location to open a new pie restaurant. Due to data limitations the result can be more or less correct. A more accurate and granular result can be obtained with a higher amount of data.

Thank You!