

Recommendation of a Manhattan Neighborhood for a Restaurant Location

Eric Luis Barroso Cavalcante

Introduction

- The business problem that is addressed in this project is the choosing of a Manhattan neighborhood to install a restaurant of a chosen cuisine or type. Some variables influence the location of a restaurant in an area, such as: population density, median household income, median rent and number of restaurants already established.
- The specific cuisine or type of restaurant is chosen by the stakeholder. Based on the neighborhoods' population density, household income, median rent and rate of neighborhood area by the number of restaurants of the type specified, the analysis provides a list of the neighborhoods from the most to the least promising one to locate the type of restaurant picked by the stakeholder

Sources of data

- For the necessary latitude and longitude of the Manhattan neighborhoods it has been used the file `newyork_data.json` provided in the Lab "Segmenting and Clustering Neighborhoods in New York City" (week 3). This file contains the geographical coordinates for each neighborhood of the five boroughs of New York City.
- The data of population, area, median household income and median rent for each Manhattan neighborhoodIt has been obtained from the address <http://www.city-data.com/indexes/neighborhoods/NY/1/> by means of web scraping.
- The names and categories of restaurants have been acquired using the Foursquare API.

Methodology

- The data is gathered in two DataFrame objects:
- "manhattan_data": contains as columns the neighborhoods, their latitudes, their longitudes, their approximate radii, their population density, their median household income and their median rent.
- "manhattan_venues": contains as columns the neighborhoods, the name of the venues, their latitudes, their longitudes and their category.

Methodology

- The DataFrame objects “manhattan_data” and “manhattan_venues” are manipulated to produce ultimately the DataFrame object “df”, which contains as columns the population density, median household income, median rent, rate of neighborhood area by number of the restaurant type chosen by the stakeholder, number of the restaurant type chosen by the stakeholder and score of each neighborhood, and it is ordered from the largest to the smallest score.

Results and Discussion

- For the kind of restaurant chosen by the stakeholder ("Italian Restaurant"), our analysis shows that the best suitable neighborhood to install a restaurant of the cuisine is Wal Street and the worst is East Harlem. Of the six most adequate neighborhoods to place a Italian Restaurant four of them do not have a unique restaurant of this kind.
- The fact that Wall Street is the first and Chinatown is the second best recommended neighborhood, despite the zero number of Italian restaurants in the latter, is related to the other three variables (population density, median household income and median rent), which are all higher in Wall Street. The median household income of Wal Street is 83% larger than Chinatown's and the median rent in Wall Street is 174% higher than Chinatown's.

Results and Discussion

- The neighborhoods of Turtle Bay, West Village, Noho, Upper East Side and Upper West Side already have a considerable number of Italian restaurants, however it is still much more recommended to locate a restaurant of the kind in these areas than in the East Village, Washington Heights and Inwood due to economic factors.
- Another discussion that can be made with the result shown in the DataFrame object "df" is related to the Greenwich Village and Little Italy neighborhoods. Greenwich Village has only one less Italian restaurant than Little Italy, nevertheless it is much better positioned than Little Italy. This can be attributed to the median household income and to the median rent of the Greenwich Village, which are far superior than those of Little Italy. On the other hand, in this case we detect a limitation of the present project since it does not consider as variable, in its current version, the number of tourists that visits a neighborhood. This limitation it is certainly a point to be addressed in a next version of the project.

Conclusion

- The main purpose of this project has been to rank the Manhattan neighborhoods best suited to have a certain type of restaurant in its area. The variables taken into account to perform this ranking are the density of population, median household income, median rent and the rate of neighborhood area by the number of restaurants of the type specified. The larger are these variables for a neighborhood the greater is its rank.
- One interesting feature of the project is that allows the stakeholder to choose from a variety of 84 kinds of restaurants, i.e., the same number of possibilities contained in the Foursquare API.
- One limitation detected by the current version of the project is that it does not take into account as a ranking variable the number of tourists that visit a neighborhood. We will address this limitation in the following version of the project.