

Prospects of a opening an Vegan Restaurant in Manhattan

As part of the IBM Data Science Capstone Project, I have decided to work on the problem on finding the best location to start a new Vegan restaurant in New York. Main objective of this section was to define the problem and discuss the ways in which the data can be found. The data will be got from different sources including the web, open source .CSV files and Four Square data. Analysis will be done on this data to find out which is the best neighbourhood suitable for starting a new vegan restaurant. In this report, the step-by-step process from explaining the problem, preparing the data, analysing, and thus deriving the conclusion will be discussed in detail.

Introduction: Description of the problem.

New York City has been described as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports. New York City is composed of five boroughs, each of which is a county of the State of New York. The five boroughs - Brooklyn, Queens, Manhattan, the Bronx, and Staten Island.

New York City's food culture includes an array of international cuisines influenced by the city's immigrant history. The city is home to "nearly one thousand of the finest and most diverse haute cuisine restaurants in the world", according to Michelin. Hence it will be an easy choice to start a new Vegan Restaurant in New York, but with limited knowledge of the demographics starting a restaurant in an already overcrowded market will be futile.

Hence the criteria set by the clients to look for the best neighbourhood to start the restaurant will be as follows:

- *Most concentrated areas of Vegetarian/Vegan restaurants in Manhattan.*

Target audience:

1. Business personnel who wants to invest or open a Vegan restaurant or any other type of new restaurant in an already crowded New York market. This analysis will be a comprehensive guide to start or expand restaurants targeting the young to middle aged crowd.
2. Freelancer Chef who loves to have their own restaurant as a side-line.

Data:

Mainly two types of data will be used for this project.

1. The first one is the file containing the neighbourhoods of New York. Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and longitude coordinates of each neighborhood.

Luckily, this dataset exists for free on the web. Here is the link to the dataset: https://geo.nyu.edu/catalog/nyu_2451_34572

a) And this will be converted to a DataFrame.

Assumptions made to a DataFrame of neighbourhoods in New York:

- Dataframe will consist of three columns: PostalCode, Borough, and Neighborhood

- Only the cells that have an assigned borough will be processed. Borough that is not assigned are ignored.
- More than one neighborhood can exist in one postal code area. These two rows will be combined into one row with the neighborhoods separated with a comma as shown in row 11 in the above table.
- If a cell has a borough but a Not assigned neighborhood, then the neighborhood will be the same as the borough.

b) Adding geographical coordinates to the neighborhoods

Next important step is adding the geographical coordinates to these neighborhoods. To do so we will be extracting the data present in the Geospatial Data csv file and combining it with the existing neighborhood dataframe by merging them both based on the postal code.

Neighbourhood level data from a variety of other sources are also available through the City's mapping application and here on the Open Data portal.

Each data point in this file is presented for the City's neighbourhoods, as well as for the City of Toronto as a whole. The data is sourced from several Census tables released by Statistics. The general Census Profile is the main source table for this data, but other Census tables have also been used to provide additional information.

c) Scrap the distribution of population from Wikipedia

Another factor that can help us in deciding which neighborhood would be best option to open a restaurant is, the distribution of population based on the ethnic diversity for each neighborhood. As this helps us in identifying the neighborhoods which are densely populated with Indian crowd since that neighborhood would be an ideal place to open an Indian restaurant.

- 2) The other type of data is from FourSquare to get the location data of venues in the neighbourhoods. Foursquare API is very useful online application used by many developers & other applications like Uber etc. In this project we will use it to retrieve information about the places present in the neighborhoods of New York. The API returns a JSON file and we need to turn that into a data-frame. Here we will choose 100 popular spots for each neighbourhood within a radius of 1km.

Results and Discussion:

Results

We have reached the end of the analysis, in the result section we can document all the findings from above clustering & visualization of the data. In this project, as the business problem started with identifying a good neighborhood to open a new Vegan restaurant, we looked into all the neighborhoods in Manhattan, analysed spread of Vegan restaurants in those neighborhoods to come to conclusion about which neighborhood would be a better spot for opening a new Vegan restaurant. I have used data from web resources like Wikipedia, geospatial coordinates of Manhattan neighborhoods, and Foursquare API, to set up a very realistic data-analysis scenario. We have found out that there are 14 neighbourhoods have already got a least density of Vegan restaurants. they include:

1. Central Harlem
2. Lincoln Square
3. Clinton
4. Greenwich Village

5. Tribeca
6. West Village
7. Morning Heights
8. Gramercy
9. Battery Park City
10. Financial District
11. Noho
12. Civic Centre
13. Turtle bay
14. Hudson yards

So it would be a good idea to start in the above areas as they do not have enough Vegan restaurants and it will be a good business investment.

With a cluster value of 3, Marble Hill is the worst neighbourhood to open a new Vegan restaurant as this is already an over saturated area of vegan restaurants. We can then use the same project to analyse the best options for other restaurants and even other type of business establishments like gym, cinema, etc

Discussion

I really enjoyed doing this data analysis as it gave me the required ideas to work out the most suitable place to open a restaurant. and this can be used to analyse opening any type of restaurant. The biggest drawback is the clustering is based only on the information provided by Foursquare and the data is not up to date. Nevertheless, it certainly provides us with some good insights, preliminary information on possibilities and a head start into this business problem by setting the step stones properly. Furthermore, this may also potentially vary

depending on the type of clustering techniques that we use to examine the data.

Conclusion:

As part of the IBM Data Science Capstone Project, I decided to work on the problem on finding the best location to start a new Vegan restaurant in New York. Analysis was done on this data to find out which is the best neighbourhood suitable for starting a new vegan restaurant. The datasets contents were manipulated and visualised using the different methods in Python. I also applied machine learning technique to predict the output given the data and used Folium to visualize it on a map. We can then use the same project to analyse the best options for other restaurants and even other type of business establishments like gym, cinema, etc. Hopefully, this project helps as an initial guidance to take head on more complex real-life challenges using data-science