IBM Data Science Professional Certificate Capstone Project

Report

Finding the right location for an Indian restaurant in Manhattan with Data Science

Introduction

In the capstone project I want to find a good location for an Indian restaurant in Manhattan. Specifically, this report is targeted to stakeholders interested in opening an **Indian restaurant** in **Manhattan**, NY.

Since there are 2874 restaurants in Manhattan, I tried to find locations

- that are not already crowded with restaurants.
- where there are as few Indian restaurants as possible in the closer area around.
- Where the share of Indian restaurants in the neighbourhood is very little.
- which are as close to the centre of Manhattan as possible.

With Data Science I will try to find and present to the stakeholders the most promising neighbourhoods of Manhattan where to open up an Indian restaurant.

Data

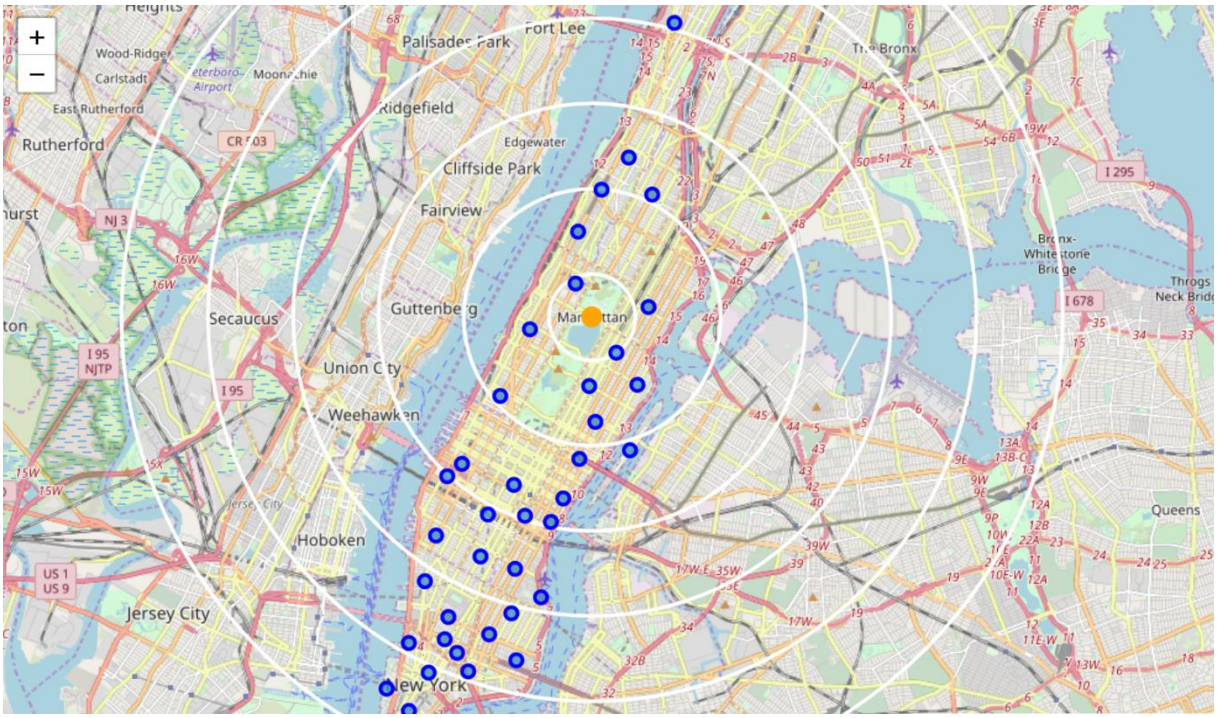
Based on the definition of the Business Problem, the decision was influenced by the following factors:

- total number of existing restaurants in the neighbourhood.
- total number of Indian restaurants in the neighbourhood.
- share of Indian restaurants in the neighbourhood.
- distance to the next Indian restaurant, if there are any.
- distance from city centre.

To find the most promising neighbourhoods to open an Indian restaurant in Manhattan I used the following data sources:

- The Information about all the neighbourhoods in Manhattan and their centres is available at the **New York Dataset**(https://cocl.us/new_york_dataset).
- To find all restaurants/Indian restaurants in each neighbourhood I used the **Foursquare API**.
- To define a choropleth map of the most promising neighbourhoods I need the **borders of each neighbourhood** which are available at https://raw.githubusercontent.com/ibuilder/NYCPolyline/master/manhattan.geojson

The first visual representation of the **New York Dataset** shows the different neighbourhoods of Manhattan.



Using the Foursquare API, I retrieved all the restaurants and Indian restaurants in Manhattan. In total there are 2891 restaurants in Manhattan and 274 Indian restaurants what makes a share of 9.48%. The visualization of all venues show how they are spread over Manhattan. The red dots represent restaurants the green one's Indian restaurants.

The first exploration and visualization of the data helped to get a feeling for the data.
We have gathered all the information we need to do our further analysis.

- We know all neighbourhoods and their centre location
- We know all restaurants of Manhattan and their location
- We know all Indian restaurants and their location
- We can visualize all locations and types of restaurants in Manhattan

This concludes the Data preparation phase and now we can continue with the analysis of the data to find the most promising neighbourhoods.

Methodology

Description of the methodology:

In the **first step** I want so see if I can **identify some areas in Manhattan with low density of restaurants/Indian restaurants that are as close as possible to the centre of Manhattan**.

Therefore, I **calculate additional figures for each neighbourhood to get a better understanding of the data**:

- Number of restaurants in every neighbourhood
- Number of Indian restaurants in every neighbourhood
- Percentage of Indian restaurants in every neighbourhood
- Distance from the centre of a neighbourhood to the next Indian restaurant

Then I will **use heatmaps to visualize**:

- the density of restaurants
- the density of Indian restaurants

and **choropleth maps to visualize**:

- the percentage of Indian restaurants in a neighbourhood
- the distance from the centre of a neighbourhood to the next Indian restaurant

In the **second step** I will use the identified areas and **generate a grid of cells** for those areas.
For every grid cell I will calculate some figures in order to define how good the location is and to be able to **filter them to get a map of all the areas that are promising to open up a Indian restaurant**.
For each grid cell the following figures will be calculated:

- Latitude
- Longitude
- Nearby restaurants
- Distance to next Indian restaurant
- Distance to centre of Manhattan

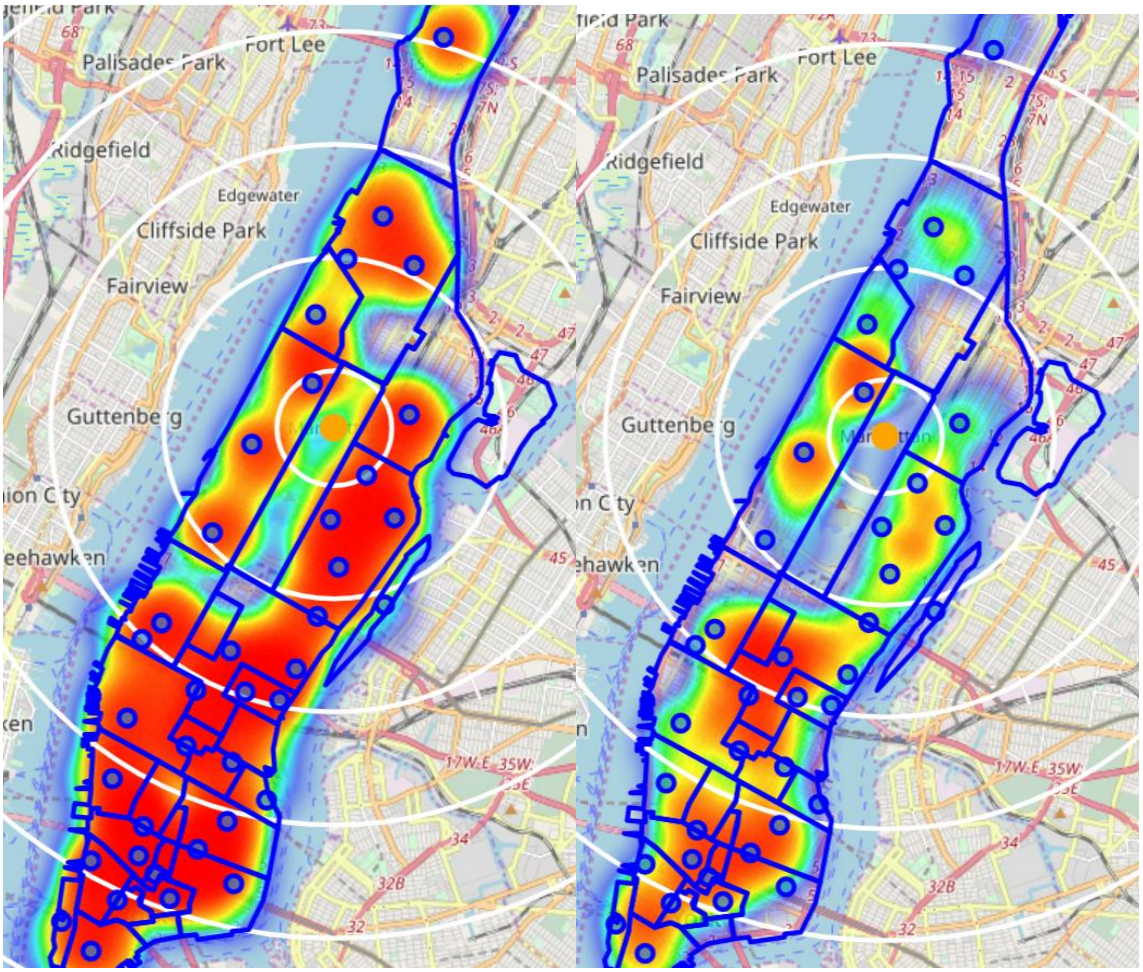
Then the generated data frame of all grid cells will be filtered for grid cell where:

- the next Indian restaurant is more than 500m away
- and there are no restaurants within a radius of 250m

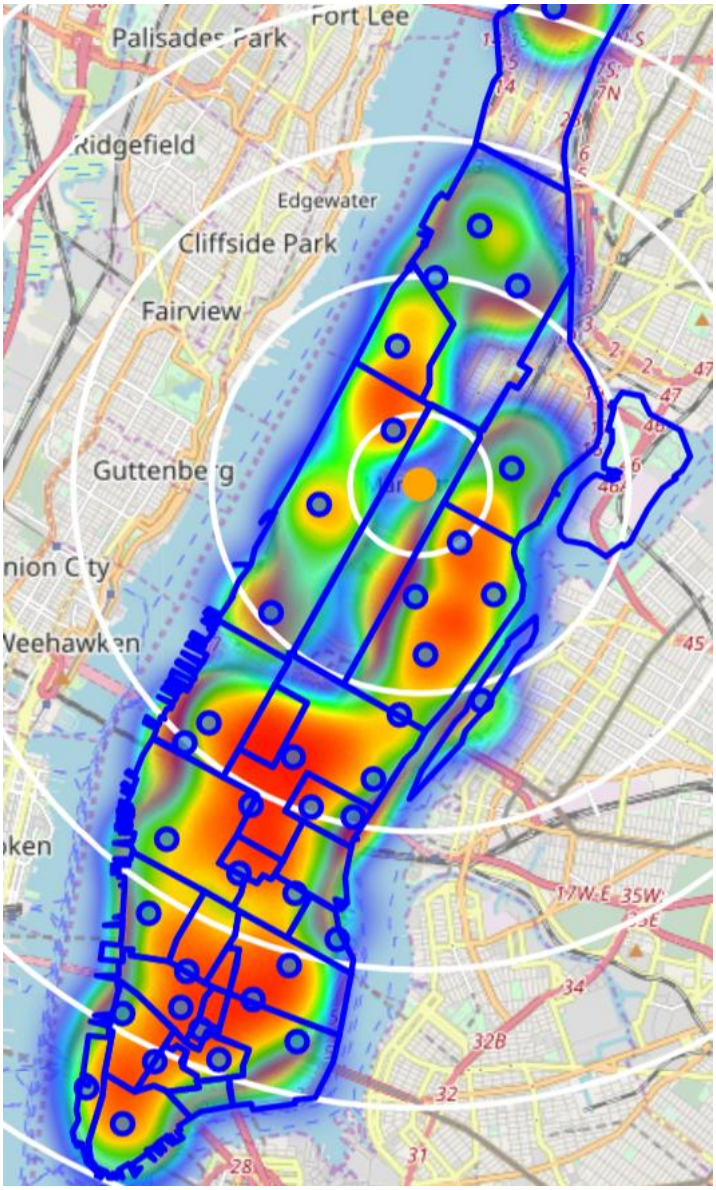
In the **final step** I will generate a **heatmap to visualize the filtered list of grid cells** which represent a map of all the **promising locations** to open an Indian restaurant in Manhattan.

Application of the methodology:

After calculating the additional figures for each neighbourhood, we can visualize the density of restaurants (left)/Indian restaurants (right) with a heatmap to see if we can spot some areas with low density close to the centre of Manhattan.



Overlaying both heatmaps makes it easier to see spaces in between the heatmaps.



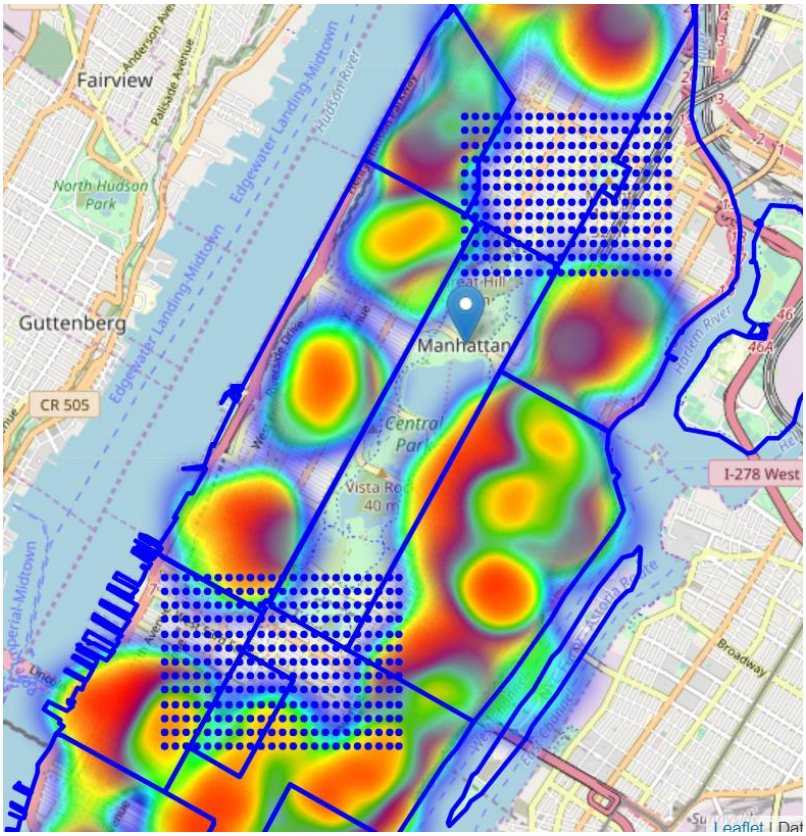
There are some Insights from the heatmaps maps. When having a look on the **density of restaurants/Indian restaurants in Manhattan** we can see that there are a few spaces with low density close to the centre of Manhattan.
In the **close area around the Center of Manhattan**:

- Bigger area north/north-east of central park
- Bigger area south/ south-west of central park

So, as we can see the areas which have an overall low density of restaurants are matching the areas with a low density of Indian restaurants quiet well in the closer area around the centre of Manhattan. We can see as well that the Heatmap of Indian restaurants is not that hot in general. With an overall share of round about 10 % the share of Indian restaurants is not that high in Manhattan.

Using choropleth maps I tried to visualize the percentage of Indian restaurants in a neighbourhood and the distance from the centre of a neighbourhood to the next Indian restaurant. Unfortunately, we can't take that much information from those plots cause the names of the neighbourhoods in the New York Dataset doesn´t match the ones of the geojson file. The areas with low density we identified in the heatmaps are exactly the ones that are named differently in the geojson file. So especially for those areas we can't see any information in the choropleth maps.

In the second step of the methodology I created a grid of cells covering the identified areas with a low density.

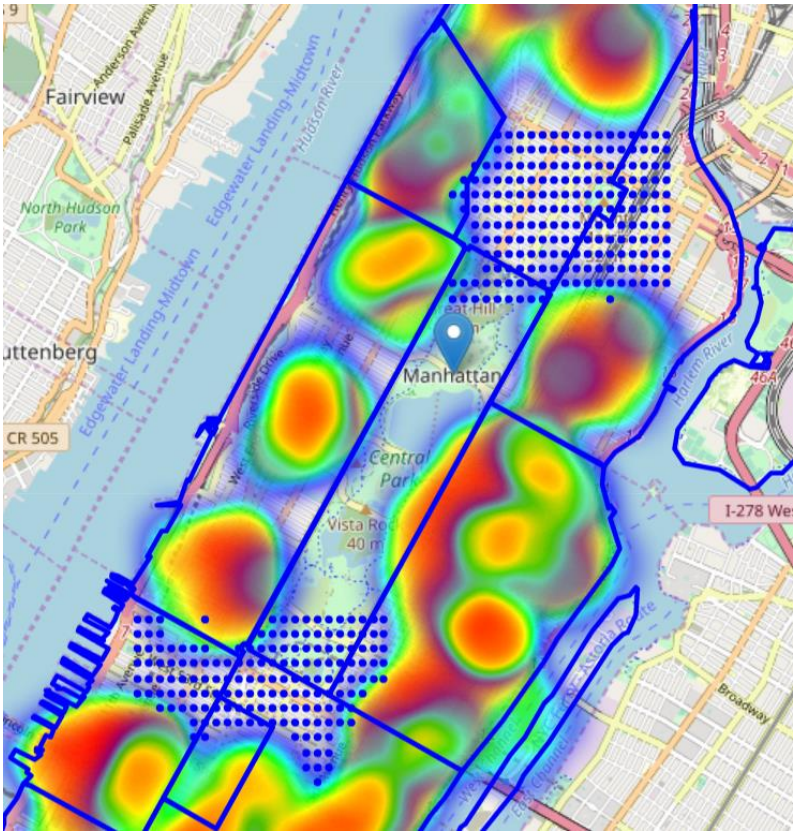


For each of the 539 generated grid cells I calculated the above defined figures. The following table shows an example of the data frame.

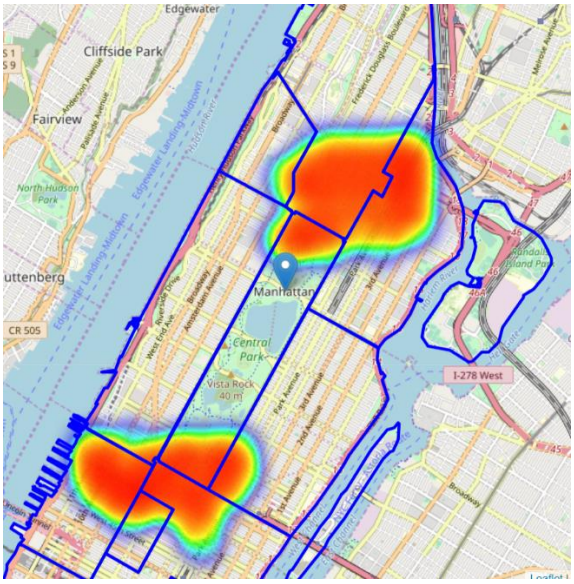
	Latitude	Longitude	Restaurants nearby	Distance next Indian Restaurant	Distance to Center
7	40.764449	-73.993485	0	767.0	6139.0
8	40.765649	-73.993485	0	959.0	5996.0
9	40.766849	-73.993485	0	1155.0	5858.0
10	40.768049	-73.993485	0	1353.0	5723.0
11	40.769249	-73.993485	0	1553.0	5592.0

This data frame can now be filtered according to the defined criteria. The number of restaurants within a radius of 250m should be 0 and the distance to the next Indian restaurant should be at least 500m. By

filtering the data frame, we got a total number of 372 locations. Visualizing this filtered data frame shows how it perfectly fits into the heatmaps of Manhattan.



We can also use the filtered data frame to plot a heatmap of the promising locations for an Indian restaurant. This map also represents the **final result**. It visualizes all the promising areas close to the centre of Manhattan to open up an Indian restaurant.
Be aware that the part of the heatmap overlapping with the central park needs to be ignored cause there it is obviously not possible to open up a restaurant.



Results and Discussion

The analysis shows some areas close to the centre of Manhattan where the density of restaurants/Indian restaurants is low even if you can find nearly 3000 restaurants in Manhattan.

The analysis presents two areas where you won't find any Indian restaurant within at least 500m radius and where there are no restaurants in at least 250m of radius.

From a perspective of competition the analysis is able to present two quiet large areas where it might be interesting to open up an Indian restaurant but it doesn't take into account if the rent is affordable or if there are spaces available to open up a restaurant or if it is an attractive neighbourhood.

Conclusion

The purpose of this analysis was to present attractive locations to the stakeholders to open up an Indian restaurant in Manhattan.

Therefore, the analysis used data science to calculate the density of restaurants/Indian restaurants. By visualizing those densities, we were able to identify two areas quiet close to the centre of Manhattan where the density of restaurants/Indian restaurants is very low.

This analysis will build the foundation for stakeholders for making a decision where to open up a Indian restaurant. For the decision additional factors needs to be taken into account like for example the rent, if there are available locations for a restaurant, the population density and the overall attractiveness of the neighbourhood.