

Opening a delivery service in New York focused on Indian Restaurants

Introduction/Business problem

Our client has decided to open a new delivery service, to compete with Uber Eats by focusing on Indian Restaurants only and having expertise in their service for people who are true enthusiasts of Indian food. This is a delivery service made by professionals and they know that they can deliver, however they are also huge fans of optimization and therefore they have hired me, Saeed Molavi, a data scientist to help them with the task of setting up a number of delivery centers from where the drivers will be dispatched. Their business will start in New York and depending on how that works they will choose to expand. They have explicitly stated that their objective is to be as close as possible to the Indian Restaurants of the city as possible, as they want to optimize for delivery time, and this requires them to be in close proximity to the restaurants.

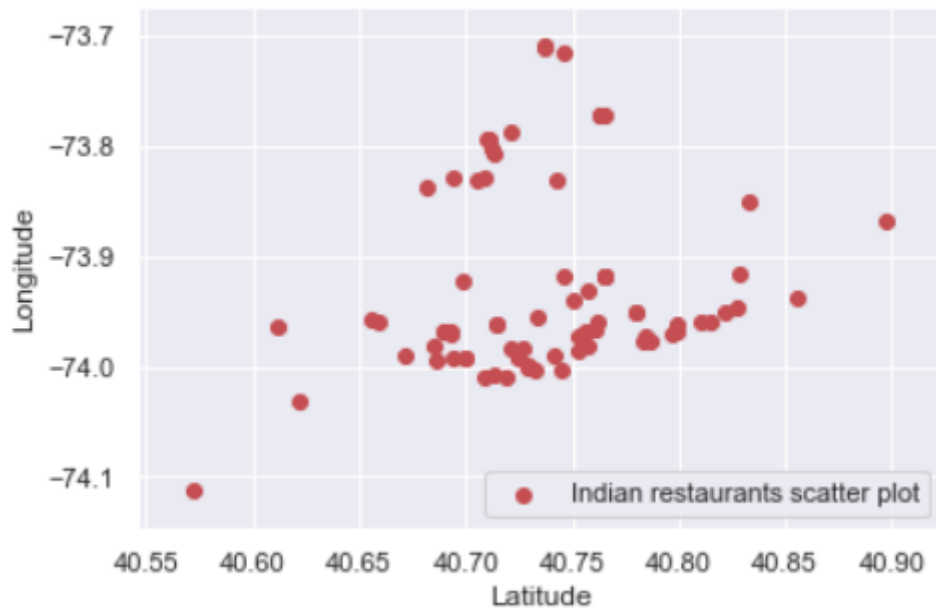
Data

For this task, I will use location data from NYU Spatial Data Repository, as well as the Foursquare API to explore neighborhoods in New York City and try to find the best spots. I will use the explore function of the API and use a k-means clustering algorithm to determine where the delivery centers should be placed in order to optimize the distance to the Indian restaurants in town. I will also use the Folium library to visualize the neighborhoods of New York and their emerging clusters. I will use Pandas in Python to perform these tasks and try to walk you through each and every step I take to solve the problem for the client.

Methodology

I used the New York City dataset to get a clear overview of the city and get an overview of the latitude/longitude data of the neighborhoods of the city. NYC is a large city and has a lot of different neighborhoods. After this I called the Foursquare API to explore the neighborhoods, segment them and then get an overview of the different businesses in the city.

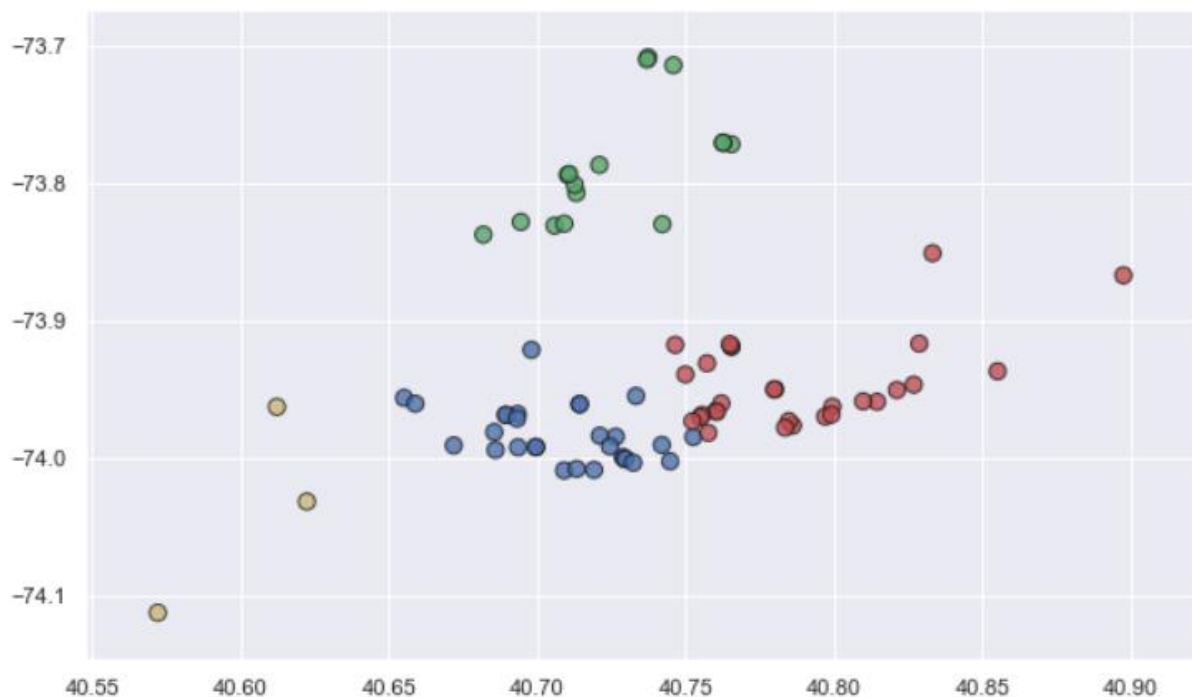
Having this in place, I filtered away everything except the Indian restaurants and their respective neighborhoods. After this, the real work started: I started by plotting all the Indian restaurants on a Seaborn plot with latitude on the x-axis and longitude on the y-axis:



Then I used K-means cluster, a machine learning method, to cluster the restaurants by their proximity to 4 different cluster centers.

Results

With the k-means algorithm, I presented the 4 cluster to the client. I told them that one of the clusters (the yellow cluster in the plot) may be unnecessary, but they insisted that one of the restaurants in that cluster is essential for people in that neighborhood due to its popularity so they might go for all 4 delivery centers regardless. See the plot below:



Another approach would be to go for just 3 cluster, but the client has decided not to.

Discussion

I believe another approach, that would require more time for the data scientist, is to classify the restaurants based on their importance for the delivery service. We could for example filter away all restaurants that are not 'good enough' to consider when setting up the delivery service. For example if the delivery service clients wants to focus on a specific segment of the market, say high-end or fast-food, we could remove the irrelevant restaurants in order to further optimize the result.

Conclusion

This has been a fun project for me, the data scientist. I have made used of maps, machine learning, as well as some nice plotting to deliver results to the client (who by the way is very happy with the results). The optimization of the location of each delivery hub will improve the speed of deliveries and thus lead to more revenue for the client. Everyone is happy!