Case Study

Capstone Project - The Battle of Neighborhoods

Finding the best spot for a new coffee shop

Manhattan

A home to mix cultures and habits

One in common for all- Coffee

Context

Investor looking to open a new coffee shop and wants to find the best location that will guarantee plenty of customers and little competition

Problem statement

Which of Manhattan's neighborhood is the most probable to be a successful choice?

Data selection and processing

Data is coming from an external source (https://cocl.us/new_york_dataset) and comes as a CSV file that needs to be loaded into Jupyter notebook

Raw data is transformed into Pandas dataframe to make further processing easy and clear

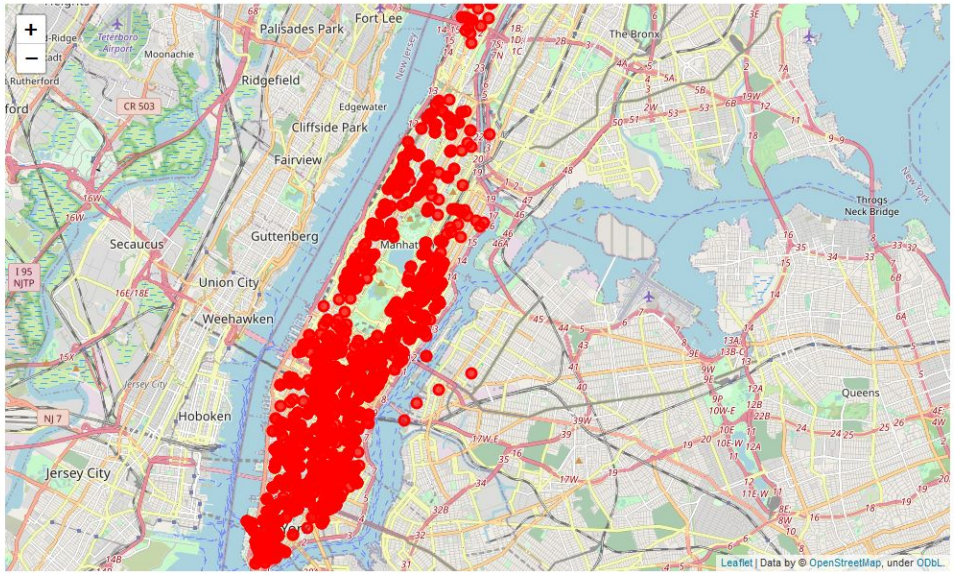
Dataframe is then processed and filtered leaving us with just enough information about five Manhattan's neighborhoods

Methodology

Map representing all coffe shops in Manhattan

```
map_newyork_coffeshop = folium.Map(location=[latitude, longitude], zoom_start=10)
addToMap(newyork_venues_coffeshop, 'red', map_newyork_coffeshop)

map_newyork_coffeshop
```



-Data was collected from https://cocl.us/new_york_dataset , acquired CSV file was cleaned and processed into Pandas dataframe

-FourSquare API was used to locate all coffeeshops in Manhattan

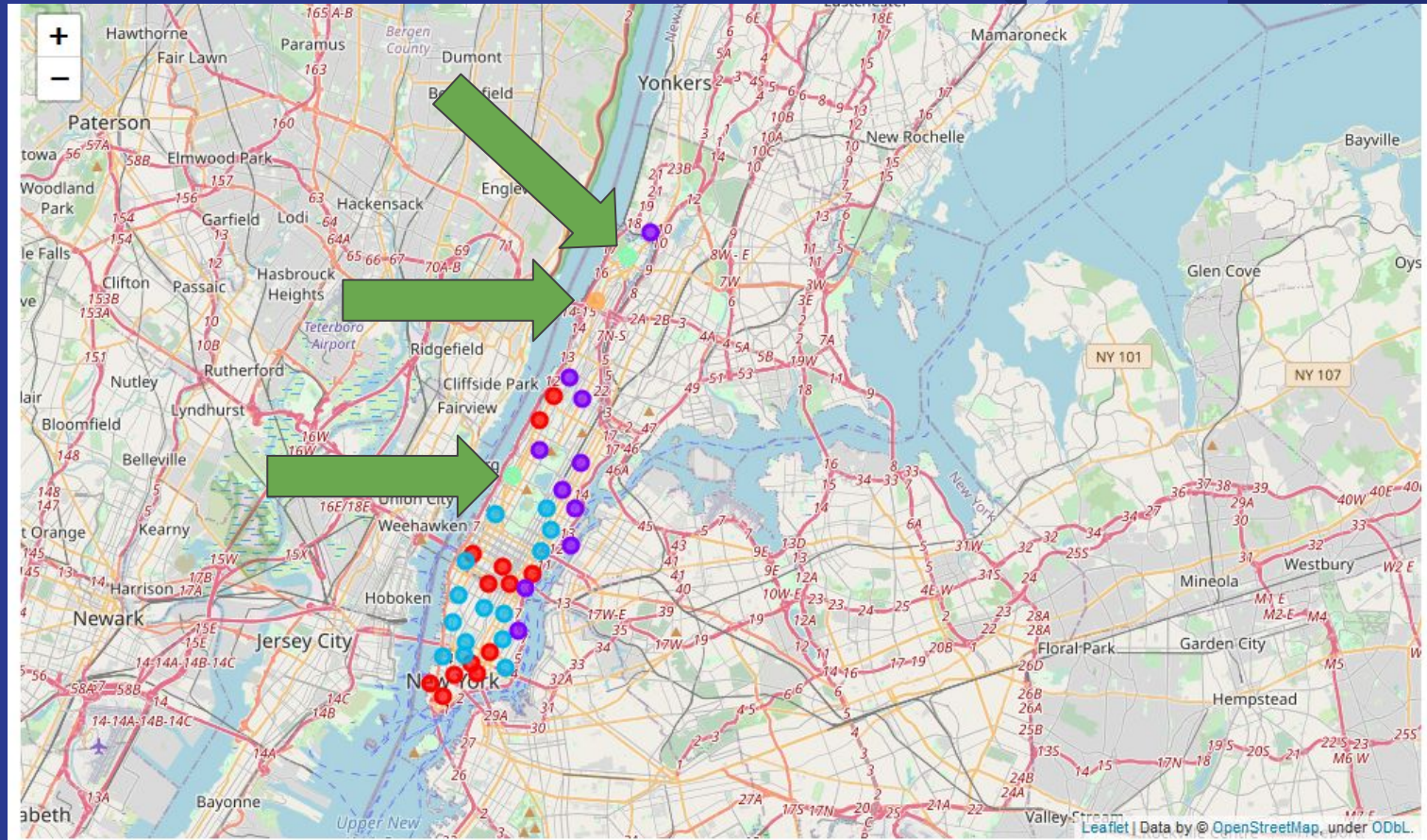
-Geo coordinates were assigned from Google Maps API to available burroughs

-Maps were created using Folium libraries

-Clusters were created and analyzed followed by final report and recommendation for a new coffee shop location

Results

3 out of 5 boroughs are not suitable for a new coffee shop due to already an abundance of those in the area. Chinatown, Financial District, East and West village, Upper East Side and Harlem are not suitable for a new coffee shop. However, the analysis showed that we may be successful with opening our shop in the vicinity of Upper West Side, Inwood and Washington Heights. The amount of coffee shops in these areas tells us to have a closer look at it and take them into consideration when making a final decision.



Discussion

Although the results were clear and could be used in real life, a fully comprehensive approach to such excersie would also require taking other factors into the account. Suggested areas are not very touristy and thus our targeted audience would be local habitants of these neighborhoods. That means we exclude tourists form the potential customer base and focus on the locals. We do not (yet) know the demographic characteristic of these neighborhoods and it would be great to include more data such as average income or age as minimum. If could then tell which of the three sugetsed neighborhoods have the demographic we are looking for would definitely get us closer to certaintity of our choice.

```
hape[1]))]]
```

]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	Inwood	Coffee Shop	Donut Shop	Bakery	Diner	Convenience Store	Dessert Shop	Deli / Bodega	Cupcake Shop	Creperie	Tea Room
12	Upper West Side	Coffee Shop	Donut Shop	Café	Deli / Bodega	Bakery	Food & Drink Shop	Food Truck	French Restaurant	Diner	Pizza Place

]:
```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 4, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.s
hape[1]))]]
```

]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Washington Heights	Donut Shop	Coffee Shop	Restaurant	New American Restaurant	Café	Bar	Bookstore	Bakery	Bubble Tea Shop	Flower Shop

Conclusion

To conclude, this excercise gave us a clear answer of where to look for potential new coffe shop localtion in Manhattan. The results could be use in real life but a fully comprehensive version of it would require a few more variables to consider- mainly demographic data to explore more. The results are a solid foundation for further analysis and considereation to a potential investor.