



# A Neighborhood Recommender for Upcoming Restaurants & Eateries

# Introduction/Background

- Every business, big or small is a vision backed by hard work, determination and aspirations.
- A drive to take an idea from its inception to fruition.
- Therefore, it becomes all the more necessary that one's blood and sweat don't go in vain.
- Presenting, RESTROLOGY: A Neighborhood Recommender for owners of upcoming Eateries & Restaurants.

# Introduction/Background contd.

- Restrology analyses local geographical data via Foursquare and recommends restaurant/eatery type(cuisines)
- Restrology takes one's neighborhood preference for a better shot at success
- Restrology analyses all the eateries and restaurants in every neighborhood of a city and then creates a list of top 10 spots (Restaurant/Eatery type) in every neighborhood

# Introduction/Background (contd.)

- Top 10 spots displayed against percentages of the total restaurants in that particular neighborhood
- A prospective business owner can use Restrology to see which cuisine based restaurants are lacking or what type of Restaurants are doing well(due to their high number) in which neighborhoods
- Restrology is like Astrology for Restaurants.

# Data

- Restrology, for the purpose of this project, will be restricted to New York City.
- Data sets used for the Project:
- *New York Data (Boroughs + Neighborhoods)*
- *Four Square City Guide Data (Venues)*

# NYC Data

- The first Data set we will be using would contain all the required geographical data about New York city.
- We would be using the same data set which was provided to us.
- [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)

# Foursquare API

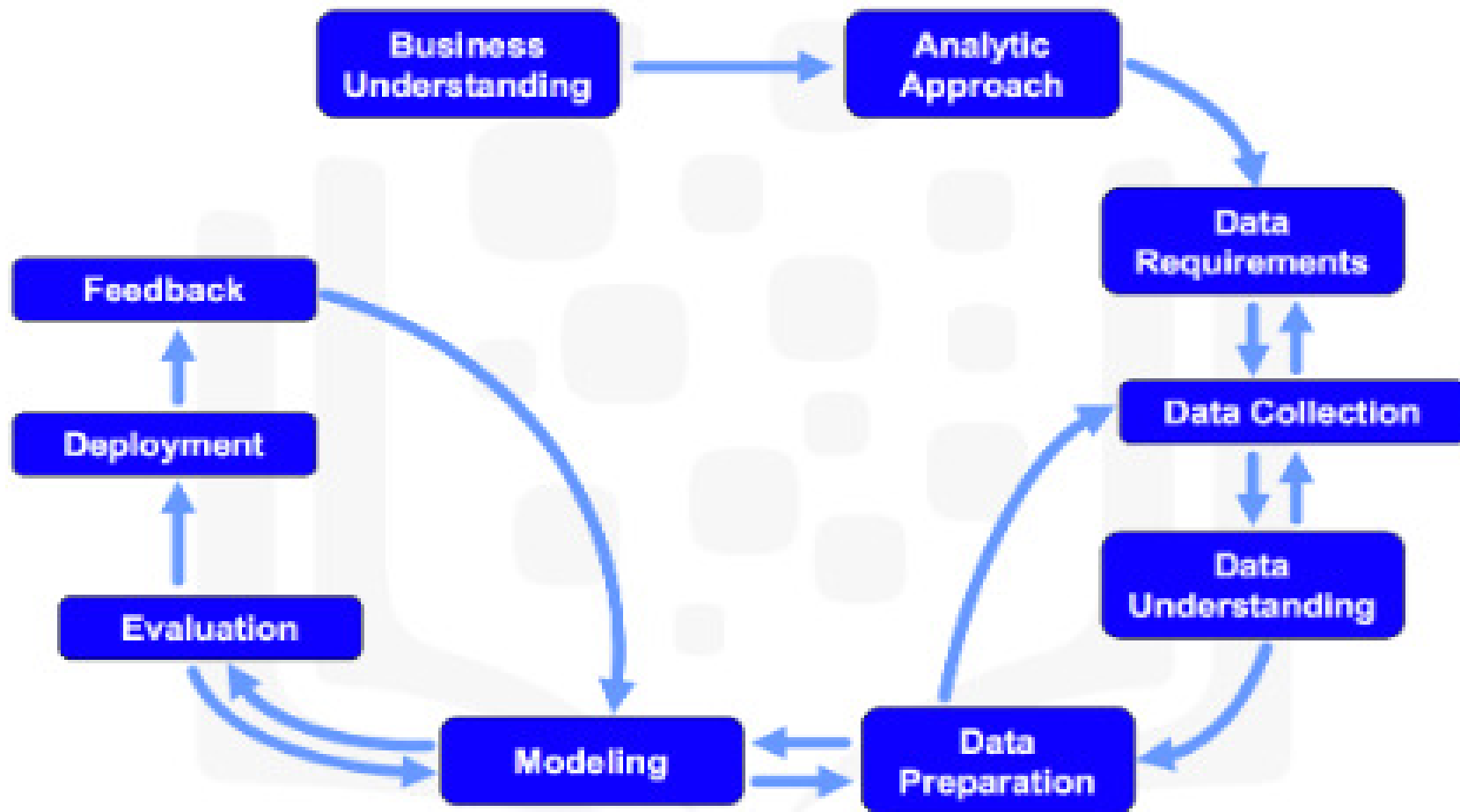
- Foursquare, is a local search-and-discovery mobile app which provides search results for its users.
- The app provides personalized recommendations of places to go near a user's current location.
- Also considers users' previous browsing history and check-in history.

# Foursquare API (contd.)

- The Foursquare API credentials you need to create are: your Client ID and your Client Secret.
- With a free personal developer account, you can make up to approximately 99 thousand regular calls per day.
- You still only get two photos and two tips per venue, which is just one more photo and tip compared to the sandbox account.



# Methodology



# Download & Explore Data

- Neighborhood has a total of 5 boroughs and 306 neighborhoods, all are extracted.
- Transform the Data into Pandas Data Frame.
- Segment the Neighborhoods of all 5 Boroughs.
- Define Foursquare credentials and version.
- Explore 'Venues' near every Neighborhood:

	name	categories	lat	lng
0	Lollipops Gelato	Dessert Shop	40.894123	-73.845892
1	Rite Aid	Pharmacy	40.896649	-73.844846
2	Carvel Ice Cream	Ice Cream Shop	40.890487	-73.848568
3	Cooler Runnings Jamaican Restaurant Inc	Caribbean Restaurant	40.898276	-73.850381
4	Shell	Gas Station	40.894187	-73.845862

# Download & Explore Data (contd.)

- Explore Neighborhoods in all 5 Boroughs.
- Output Data Frame:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bay Ridge	40.625801	-74.030621	Pilo Arts Day Spa and Salon	40.624748	-74.030591	Spa
1	Bay Ridge	40.625801	-74.030621	Bagel Boy	40.627896	-74.029335	Bagel Shop
2	Bay Ridge	40.625801	-74.030621	Cocoa Grinder	40.623967	-74.030863	Juice Bar
3	Bay Ridge	40.625801	-74.030621	Pegasus Cafe	40.623168	-74.031186	Breakfast Spot
4	Bay Ridge	40.625801	-74.030621	Ho' Brah Taco Joint	40.622960	-74.031371	Taco Place

# Analyze Each Borough + Neigh'd

- Group rows by neighborhood and by taking the mean of the frequency of occurrence of each category.
- Print the top 5 most common venues/Neigh'hd

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bath Beach	Pizza Place	Italian Restaurant	Fast Food Restaurant	Sushi Restaurant	Bubble Tea Shop
1	Bay Ridge	Italian Restaurant	Pizza Place	Greek Restaurant	Bagel Shop	American Restaurant
2	Bedford Stuyvesant	Deli / Bodega	Café	Pizza Place	Coffee Shop	Japanese Restaurant
3	Bensonhurst	Chinese Restaurant	Pizza Place	Ice Cream Shop	Donut Shop	Sushi Restaurant
4	Bergen Beach	Donut Shop	Wings Joint	Hawaiian Restaurant	Diner	Dumpling Restaurant

```
----Central Harlem----
                        venue  freq
0   African Restaurant  0.06
1   Seafood Restaurant  0.04
2   American Restaurant  0.04
3   French Restaurant   0.04
4   Chinese Restaurant   0.04
```

```
----Chelsea----
                        venue  freq
0           Coffee Shop  0.06
1       Ice Cream Shop  0.05
2   Italian Restaurant  0.05
3           Bakery      0.04
4   Seafood Restaurant  0.03
```

```
----Chinatown----
                        venue  freq
```

# Clustering

- We will be using k-means clustering.
- We would be using k-means to cluster all of our 5 boroughs:
  - Brooklyn
  - Manhattan
  - Queens
  - Bronx
  - Staten Island

# K-means Clustering

- K-Means can group data only unsupervised based on the similarity of customers to each other.
- Output Data Frames:

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	2	Coffee Shop	Sandwich Place	Bakery	Pizza Place	Diner
1	Manhattan	Chinatown	40.715618	-73.994279	4	Chinese Restaurant	American Restaurant	Vietnamese Restaurant	Ice Cream Shop	Dumpling Restaurant
2	Manhattan	Washington Heights	40.851903	-73.936900	3	Café	Bakery	Spanish Restaurant	Deli / Bodega	New American Restaurant
3	Manhattan	Inwood	40.867684	-73.921210	3	Mexican Restaurant	Café	Pizza Place	Bakery	Deli / Bodega
4	Manhattan	Hamilton Heights	40.823604	-73.949688	3	Deli / Bodega	Café	Pizza Place	Mexican Restaurant	Chinese Restaurant

# Restrology Output

- Finally, Restrology gives us the output like the following:

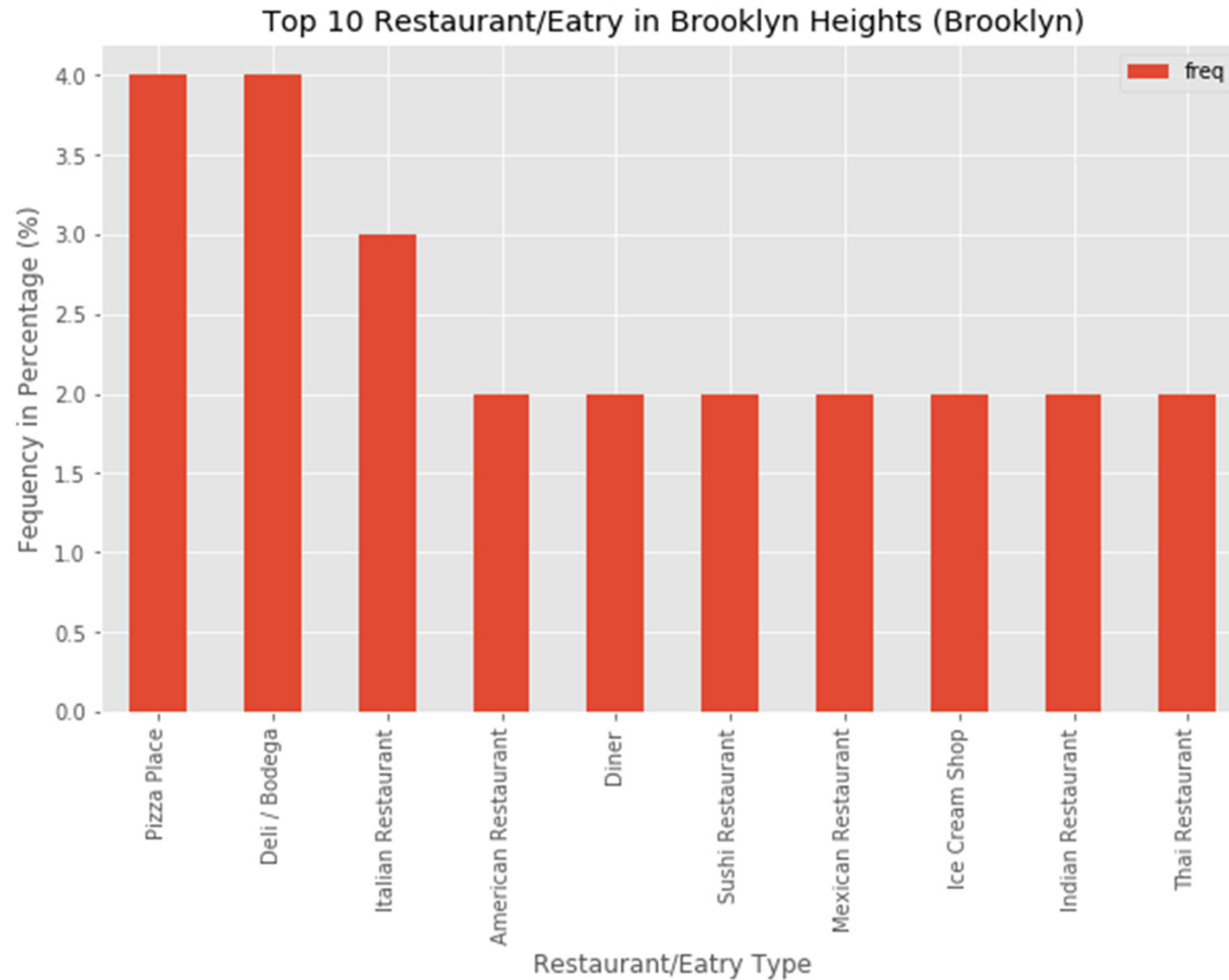
venue	freq
Pizza Place	11.0
Chinese Restaurant	7.0
Deli / Bodega	4.0
American Restaurant	2.0
Thai Restaurant	2.0
Sushi Restaurant	2.0
Sandwich Place	2.0
Japanese Restaurant	2.0
Italian Restaurant	2.0
Fast Food Restaurant	2.0

venue	freq
Thai Restaurant	18.0
Mexican Restaurant	15.0
Chinese Restaurant	8.0
Vietnamese Restaurant	5.0
Bubble Tea Shop	5.0
South American Restaurant	5.0
Colombian Restaurant	3.0
Snack Place	3.0
Malay Restaurant	3.0
Pizza Place	3.0

venue	freq
Italian Restaurant	19.0
Pizza Place	9.0
Deli / Bodega	9.0
Bakery	5.0
Dessert Shop	3.0
Spanish Restaurant	3.0
Sandwich Place	2.0
Donut Shop	2.0
Mexican Restaurant	2.0
Food & Drink Shop	2.0

# Results

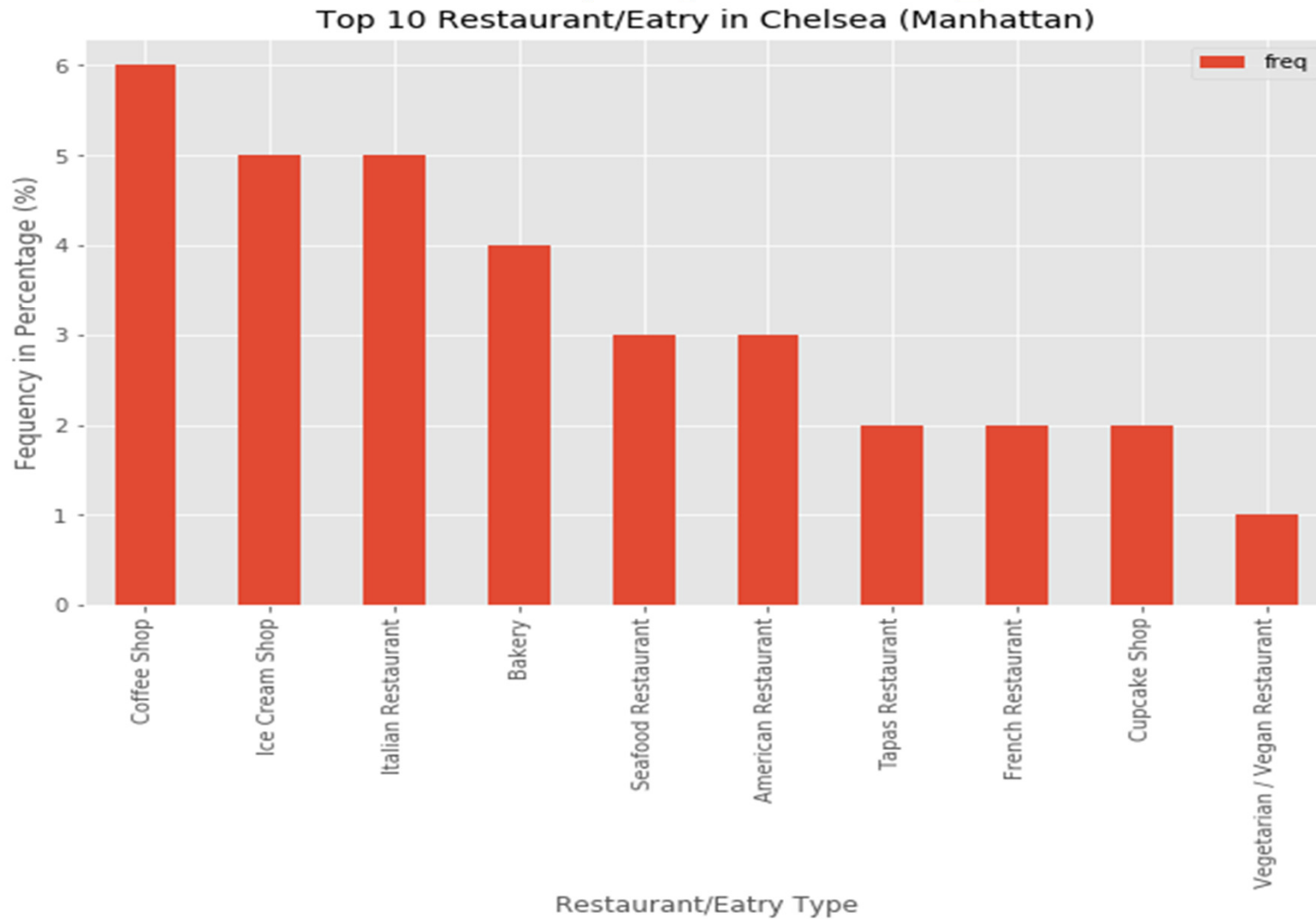
- Results derived, displayed using Bar Charts:





# Results (contd.)

- Results derived, displayed using Bar Charts:



# Recommendations

- First, we can integrate other data sets as well, which can add and quantify other factors for Restrology.
- Similarly, apart from the current data set, we can also integrate some Past data where we have access to closed business's data in the neighborhood.
- This would help us further strengthen Restrology to somehow quantify and predict the chances for 'Success' for any given Restaurant/Eatry.

# Recommendations (contd.)

- We can explore other features from services like Foursquare API to include and factor in parameters like 'Trending' to further enhance Restrology.
- Something which we were not able to do due to the fact that we have used a free developer account.
- Data from nearby cities can be integrated as well to uncover more trends and insights.

# Conclusion

- This concludes our journey with Restrology in this Capstone Project.
- During this journey we covered all the gospels in the field of Data Science and Machine Learning.
- A prospective Restaurant owner can make an informed decision and have a better chance of establishing a successful business.
- Given the extensive scope of the subject which seems limitless, we have merely made a dent and there are far bigger and tougher horizons for us to conquer.