


The logo features a large, dark blue, irregular shape on the left side of the slide, resembling a splash or a cloud. The text "IBM Data Science" is written in white, sans-serif font, positioned within the lower-left portion of this dark shape.

IBM Data Science



There are 9  
courses in  
this  
certification

- 1. What is Data Science ?
- 2. Open Source tools for Data Science
- 3. Data Science Methodology
- 4. Python for Data Science and AI
- 5. Databases and SQL for Data Science
- 6. Data Analysis with Python
- 7. Data visualization with Python
- 8. Machine Learning with Python
- 9. Applied Data Science Capstone


# 1. What is Data Science?

- Data science is the art of uncovering the insights and trends that are hiding behind data. It's when you translate data into a story. So use storytelling to generate insight. And with these insights, you can make strategic choices for a company or an institution.



## 2. Open Source tools for Data Science

- In this course, I have learned about various open source tools for Data Science.
- Skill Network Labs
- Jupyter Notebooks
- Apache Zeppelin Notebooks
- Rstudio IDE
- IBM Watson studio



### 3. Data Science Methodol ogy

- In this course I have learned about the major steps involved in tackling a data science problem. - The major steps involved in practicing data science, from forming a concrete business or research problem, to collecting and analyzing data, to building a model, and understanding the feedback after model deployment. - How data scientists think!.

# 4. Python for Data Science and AI

- In this course I have learned about Python Basics like types, expressions, variables, string operations, lists, tuples, sets, dictionaries, Loops, objects and classes, file handling, pandas and numpy.



## 5. Databases and SQL for Data Science

- In this course, I have learned about relational database concepts that helps to apply foundational knowledge of the SQL language, performing SQL access in a data science environment. The emphasis in this course is on hands-on and practical learning. I have also created some database instances in the cloud. I have done series of hands-on labs, practice building and running SQL queries in this lab. I have also learned how we can access databases from Jupyter notebooks using SQL and Python.

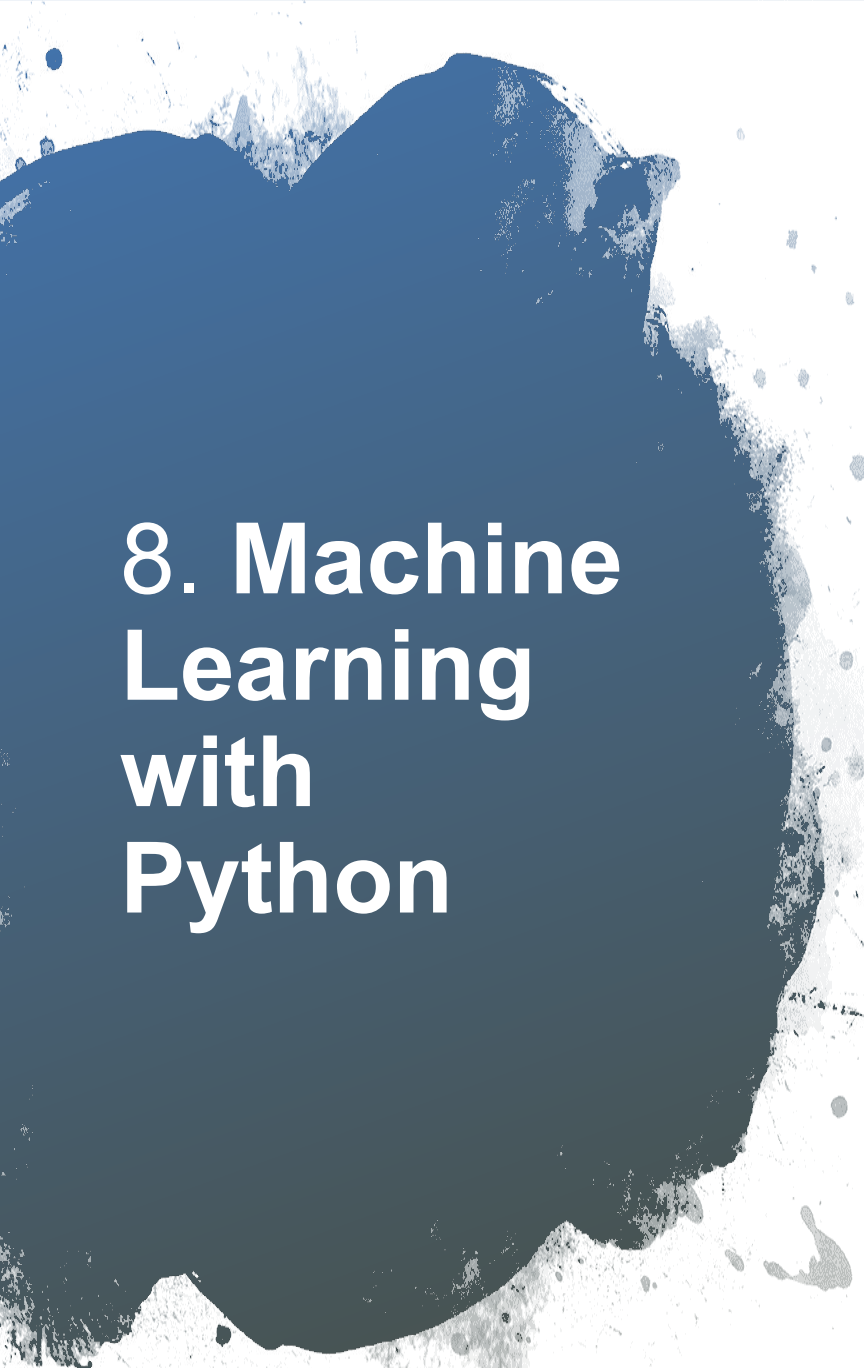
## 6. Data Analysis with Python

- In this course I have learned about Importing Datasets, Cleaning the Data , Data frame manipulation, Summarizing the Data. It includes following parts: Data Analysis libraries, use of Pandas, Numpy and Scipy libraries to work with a sample dataset. I have used this library to load, manipulate, analyze, and visualize cool datasets.



# 7. Data visualizati on with Python

- This course was all about several data visualization libraries in Python like Matplotlib, Seaborn, and Folium and how we can tell a compelling story by visualizing the data and findings from the data



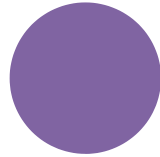
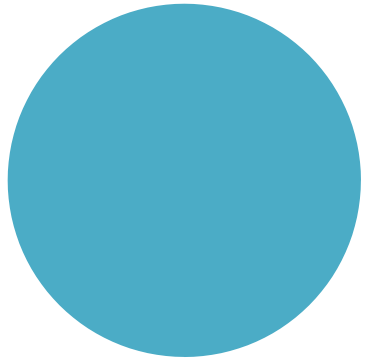
## 8. Machine Learning with Python

- In this course I have learned about some of machine learning topics like supervised and unsupervised learning, classification, clustering and some Python libraries like Sci-kit learn and Scipy.



## 9. Applied Data Science Capstone

- In this course I have learned about FourSquare API ( It is a restful API to retrieve the data about venues in different neighborhoods around the world and I have applied this learnings to complete my Capstone Project



# Capstone | Project |

# Background

New York City's demographics show that it is a large and ethnically diverse metropolis. It is the largest city in the United States with a long history of international immigration. New York City was home to nearly 8.5 million people in 2014, accounting for over 40% of the population of New York State and a slightly lower percentage of the New York metropolitan area, home to approximately 23.6 million. Over the last decade the city has been growing faster than the region. The New York region continues to be by far the leading metropolitan gateway for legal immigrants admitted into the United States.

Throughout its history, New York City has been a major point of entry for immigrants; the term "melting pot" was coined to describe densely populated immigrant neighborhoods on the Lower East Side. As many as 800 languages are spoken in New York, making it the most linguistically diverse city in the world. English remains the most widely spoken language, although there are areas in the outer boroughs in which up to 25% of people speak English as an alternate language, and/or have limited or no English language fluency. English is least spoken in neighborhoods such as Flushing, Sunset Park, and Corona.

# Introduction

With it's diverse culture , comes diverse food items. There are many restaurants in New York City, each belonging each belonging to different categories like Chinese , Indian , French etc. So as part of this project , we will list and visualize all major parts of New York City that has great Indian restaurants.

Queries that can be answered using this project?

- What is best location in New York City for Indian Cuisine ?
- Which areas have potential Indian Restaurant Market ?
- Which all areas lack Indian Restaurants ?
- Which is the best place to stay if I prefer Indian Cuisine ?

# Data to be used

1. Data source : [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)

Description - his data set contains the required information. And we will use this data set to explore various neighborhoods of new york city. Indian restaurants in each neighborhood of New York city.

2. Data source : Foursquare API

Description : By using this API we will get all the venues in each neighborhood. We can filter these venues to get only Indian Restaurants.

3. Data source :

<https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm>

Description : By using this geo space data we will get the New York Borough boundaries that will help us to visualize choropleth map.

# Approach

Collect the New York city data from [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)

Using FourSquare API we will find all venues for each neighborhood.

Filter out all venues that are Indian Restaurants.

Find rating , tips and like count for each Indian Restaurants using FourSquare API.

Using rating for each restaurant , we will sort that data.



# Libraries to be used

pandas and numpy for handling data.

request module for using FourSquare API.

geopy to get co-ordinates of City of New York.

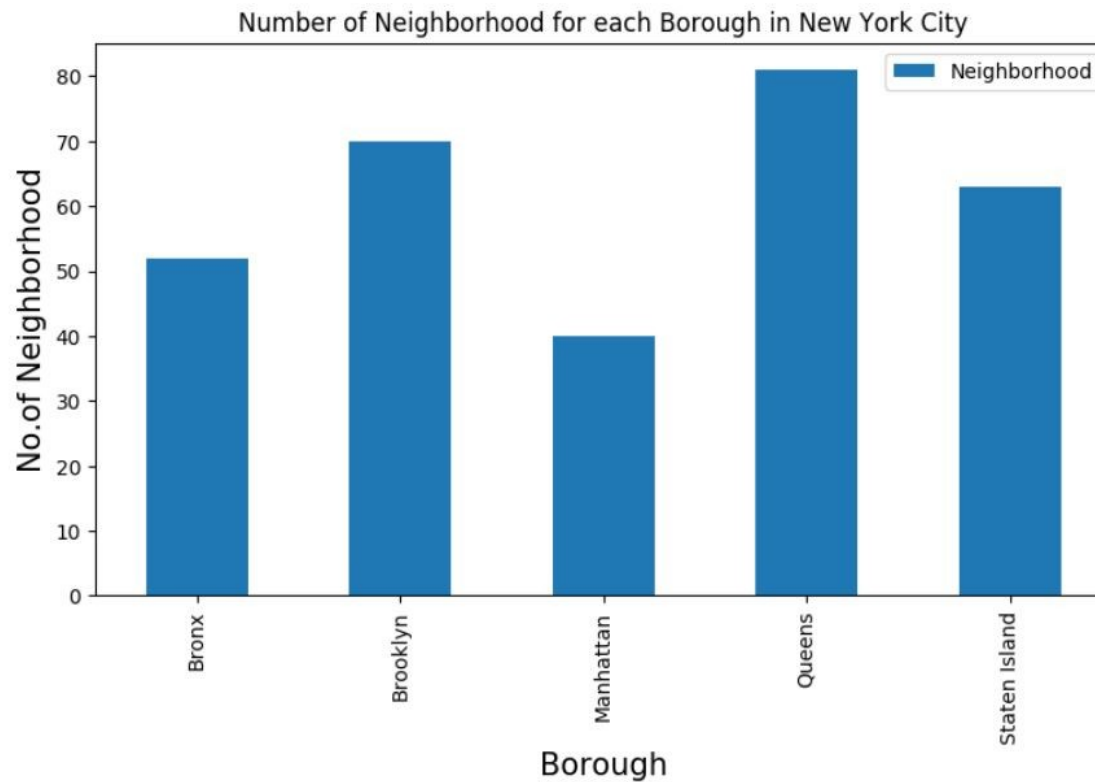
## Step 1

- Load data from [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset) in pandas Dataframe.
- Getting Latitude and Longitude for each address geopy library.

|   | Borough | Neighborhood | Latitude  | Longitude  |
|---|---------|--------------|-----------|------------|
| 0 | Bronx   | Wakefield    | 40.894705 | -73.847201 |
| 1 | Bronx   | Co-op City   | 40.874294 | -73.829939 |
| 2 | Bronx   | Eastchester  | 40.887556 | -73.827806 |
| 3 | Bronx   | Fieldston    | 40.895437 | -73.905643 |
| 4 | Bronx   | Riverdale    | 40.890834 | -73.912585 |

As result –

We have 306 rows like this.



Number of neighborhoods  
in each Borough

## Step 2

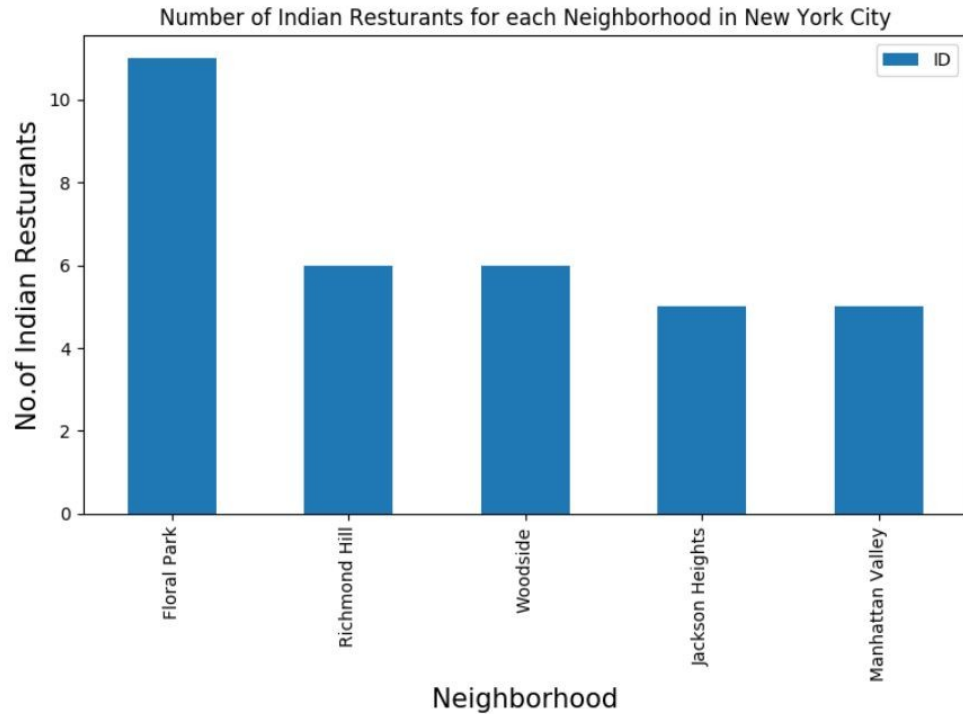
- Filter out which Borough and Neighborhood have maximum number of Indian Restaurants using FourSquare API.



**Result** – Queens has maximum number of Restaurants

## Step 2

- Filter out which Borough and Neighborhood have maximum number of Indian Restaurants using FourSquare API.



**Result** – Floral Park has maximum number of Restaurants

# List of Restaurants in Floral Park

---

|     | Borough | Neighborhood | ID                       | Name                            |
|-----|---------|--------------|--------------------------|---------------------------------|
| 103 | Queens  | Floral Park  | 4e4e3e22bd4101d0d7a5c2d1 | Kerala Kitchen                  |
| 104 | Queens  | Floral Park  | 4b647b56f964a520c4b62ae3 | Usha Foods & Usha Sweets        |
| 105 | Queens  | Floral Park  | 527ffc0811d2d329d5e49abd | Jackson Diner                   |
| 106 | Queens  | Floral Park  | 4b787c49f964a5209cd12ee3 | Santoor Indian Restaurant       |
| 107 | Queens  | Floral Park  | 4c0c01e0bbc676b00d6b4cd5 | Mumbai Xpress                   |
| 108 | Queens  | Floral Park  | 4c76ff35a5676dcb72671721 | Flavor Of India                 |
| 109 | Queens  | Floral Park  | 4df0f39dd4c04d0392c853ea | Sagar Chinese                   |
| 110 | Queens  | Floral Park  | 571af96a498e9e392d8d3786 | Namaste Authenic Indian Cuisine |
| 111 | Queens  | Floral Park  | 55d68c1b498ecf05fa196fe1 | Namaste Restaurant and Cafe     |
| 112 | Queens  | Floral Park  | 4c3e17f2ca012d7f82022fbe | Mushin's Halal Food [Gyro Cart] |
| 113 | Queens  | Floral Park  | 4e6bfe1c7d8b2c711b17bbe5 | Surya sweets and snacks         |

## Step 3

- Get likes, ratings, tips on each of Indian Restaurant using FourSquare API

|   | Borough | Neighborhood   | ID                       | Name                        | Likes | Rating | Tips |
|---|---------|----------------|--------------------------|-----------------------------|-------|--------|------|
| 0 | Bronx   | Woodlawn       | 4c0448d9310fc9b6bf1dc761 | Curry Spot                  | 4     | 8.1    | 11   |
| 1 | Bronx   | Parkchester    | 4c194631838020a13e78e561 | Melanies Roti Bar And Grill | 3     | 6.0    | 2    |
| 2 | Bronx   | Spuyten Duyvil | 4c04544df423a593ac83d116 | Cumin Indian Cuisine        | 13    | 6.1    | 9    |
| 3 | Bronx   | Concourse      | 551b7f75498e86c00a0ed2e1 | Hungry Bird                 | 8     | 6.9    | 3    |
| 4 | Bronx   | Unionport      | 4c194631838020a13e78e561 | Melanies Roti Bar And Grill | 3     | 6.0    | 2    |

# Result

- Restaurant with maximum like

```
: # Resturant with maximum Likes
indian_rest_stats_ny.iloc[indian_rest_stats_ny['Likes'].idxmax()]

: Borough                Manhattan
  Neighborhood            Midtown
  ID          49d91c12f964a520015e1fe3
  Name          The Kati Roll Company
  Likes                836
  Rating              8.8
  Tips                259
Name: 41, dtype: object
```



# Result

- Restaurant having maximum Rating

```
# Resturant with maximum Rating  
indian_rest_stats_ny.iloc[indian_rest_stats_ny['Rating'].idxmax()]
```

|                         |                          |
|-------------------------|--------------------------|
| Borough                 | Manhattan                |
| Neighborhood            | Tribeca                  |
| ID                      | 4bbb9dbded7776b0e1ad3e51 |
| Name                    | Tamarind TriBeCa         |
| Likes                   | 586                      |
| Rating                  | 9                        |
| Tips                    | 145                      |
| Name: 45, dtype: object |                          |

# Conclusion

Astoria(Queens),  
Blissville(Queens  
) , Civic  
Center(Manhatta  
n) are some of  
the best  
neighborhoods  
for indian cuisine.

Manhattan have  
potential Indian  
Resturant Market

Staten Island  
ranks last in  
average rating of  
Indian  
Restaurants.

Manhattan is the  
best place to stay  
if you prefer  
Indian Cuisine.

# Limitation

The accuracy of data depends purely depends on the data provided by FourSquare



# Thank you

Any queries ?