

# APPLIED DATA SCIENCE CAPSTONE

## BATTLE OF NEIGHBORHOODS

Muzothule Kubheka

# CONTENTS

## ▶ Introduction Section:

1. Discussion of the "background situation" leading to the problem at hand:
2. Problem to be resolved
3. Audience for this project.

## ▶ Data Section:

1. Data of Current Situation (current residence place)
2. Data required to resolve the problem
3. Data sources and data manipulation

## ▶ Methodology Section:

1. Process steps and strategy to resolve the problem
2. Data Science Methods, machine learning, mapping tools and exploratory data analysis.

## ▶ Results Section

Discussion of the results and how they help to take a decision.

## ▶ Discussion Section

Elaboration and discussion on any observations and/or recommendations for improvement.

## ▶ Conclusion Section

Decision taken and Report Conclusion.

# 1. INTRODUCTION SECTION :

## ▶ 1.1 Scenario and Background

- ▶ I am a data scientist residing in Singapore. I currently live within walking distance to many amenities and venues in the area,
- ▶ I have been offered a great opportunity to work in Manhattan, NY. Although, I am very excited about it, I am a bit anxious about the process to secure a comparable place to live in Manhattan. Therefore, I decided to apply the skills learned in the course to explore ways to make sure my decision is factual and rewarding. There are alternatives to achieve the answer using available Google and Social media tools, but it is rewarding doing it myself using Data Science tools.

## ▶ 1.2 Problem to be resolved:

- ▶ The challenge to resolve is being able to find an apartment unit in Manhattan NY that offers similar characteristics and benefits to my current situation. Therefore, in order to set a basis for comparison, I want to place subject to the following conditions:
- ▶ Top amenities in the selected neighborhood shall be similar to current residence (See item 2.1)
- ▶ Desirable to have venues such as Gyms, Restaurants, wine stores, and food shops
- ▶ As a reference, I have included a map of venues near current residence in Singapore.

## ▶ 1.3 Interested Audience

- ▶ I believe this is a relevant project for a person or entity considering moving to a major city in Europe, US or Asia, since the approach and methodologies used here are applicable in all cases.
- ▶ The use of FourSquare data and mapping techniques combined with data analysis will help answer key questions.
- ▶ Lastly, this project is a good practical case toward the development of Data Science skills.

## 2. DATA SECTION:

- ▶ **Description of the data and its sources that will be used to solve the problem**
- ▶ **2.1 Data of Current Situation**
- ▶ I currently reside in Singapore. I use Foursquare to identify the venues around the area of residence which are then shown in the Singapore map shown in methodology and execution in section 3.0 . It serves as a reference for comparison with the desired future location in Manhattan NY
- ▶ **2.2 Data Required to resolve the problem**
- ▶ In order to make a good choice of a similar apartment in Manhattan NY, the following data is required:
- ▶ List/Information on neighborhoods from Manhattan with their Geodata(latitude and longitude).
- ▶ Venues and amenities in the Manhattan neighborhoods (e.g. top 10)

## 2. DATA SECTION:


### ▶ 2.3 sources and manipulation

- ▶ The list of Manhattan neighborhoods is worked out during LAB exercise during the course. A txt file was created which will be read in order to create a data frame and its mapping.

### ▶ 2.4 How the data will be used to solve the problem

- ▶ The data will be used as follows:
- ▶ Foursquare and geopy data to map top 10 venues for all Manhattan neighborhoods and clustered in groups ( as per Course LAB)

### ▶ 2.5 Mapping of Data

- ▶ The following maps were created to facilitate the analysis and the choice of the palace to live.
  - ▶ Manhattan map of Neighborhoods
  - ▶ Manhattan map of clustered venues and neighborhoods
- 
- A series of white diagonal lines of varying lengths and thicknesses, located in the bottom right corner of the slide, creating a modern, abstract graphic element.

# 3. METHODOLOGY SECTION:

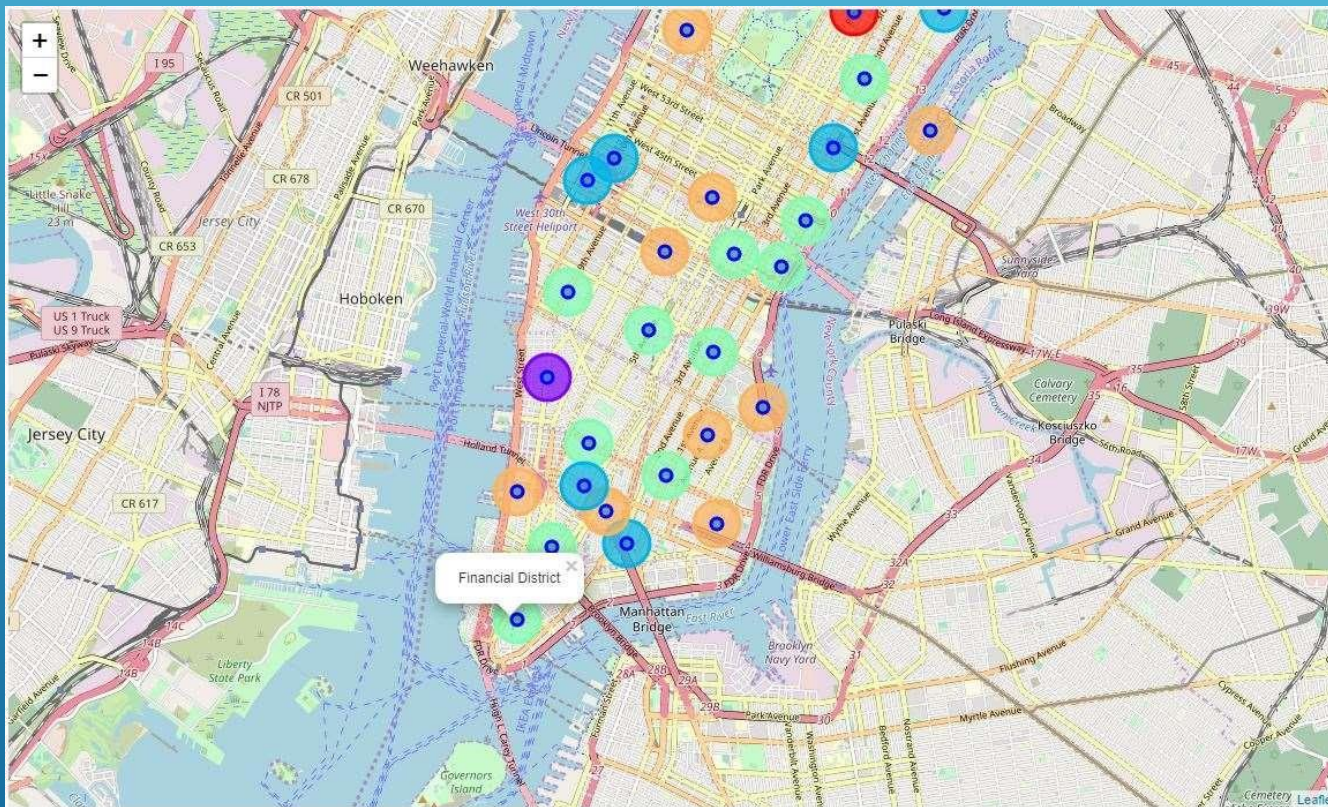
- ▶ This section represents the main component of the report where the data is gathered, prepared for analysis. The tools described are used here and the Notebook cells indicates the execution of steps.
- ▶ **The analysis and the strategy:**
- ▶ The strategy is based on mapping the above described data in section 2.0, in order to facilitate the choice of a candidate places for accommodation. The choice is made based on the demands imposed : similar venues to current residence. This visual approach and maps with pop-ups labels allow quick identification of location, thus making the selection very easy.
- ▶ The processing of these DATA and its mapping will allow to answer the key questions to make a decision:
- ▶ What are the venues of the best place to live?
- ▶ How venues distribute among Manhattan neighborhoods ?



# 4. RESULTS



# CLUSTERS OF NEIGHBOURHOODS IN MANHATTAN





# 3 resembles in amenities to our Current Residence

```
## kk is the cluster number to explore
kk = 3
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == kk, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[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	Mexican Restaurant	Lounge	Pizza Place	Café	Wine Bar	Bakery	American Restaurant	Park	Frozen Yogurt Shop	Spanish Restaurant
5	Manhattanville	Deli / Bodega	Italian Restaurant	Seafood Restaurant	Mexican Restaurant	Sushi Restaurant	Beer Garden	Coffee Shop	Falafel Restaurant	Bike Trail	Other Nightlife
10	Lenox Hill	Sushi Restaurant	Italian Restaurant	Coffee Shop	Gym / Fitness Center	Pizza Place	Burger Joint	Deli / Bodega	Gym	Sporting Goods Shop	Thai Restaurant
12	Upper West Side	Italian Restaurant	Bar	Bakery	Vegetarian / Vegan Restaurant	Indian Restaurant	Coffee Shop	Cosmetics Shop	Wine Bar	Mexican Restaurant	Sushi Restaurant
16	Murray Hill	Sandwich Place	Hotel	Japanese Restaurant	Gym / Fitness Center	Coffee Shop	Salon / Barbershop	Burger Joint	French Restaurant	Bar	Italian Restaurant
17	Chelsea	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bakery	Nightclub	Theater	Art Gallery	Seafood Restaurant	American Restaurant	Hotel
18	Greenwich Village	Italian Restaurant	Sushi Restaurant	French Restaurant	Clothing Store	Chinese Restaurant	Café	Indian Restaurant	Bakery	Seafood Restaurant	Electronics Store
27	Gramercy	Italian Restaurant	Restaurant	Thrift / Vintage Store	Cocktail Bar	Bagel Shop	Coffee Shop	Pizza Place	Mexican Restaurant	Grocery Store	Wine Shop
29	Financial District	Coffee Shop	Hotel	Gym	Wine Shop	Steakhouse	Bar	Italian Restaurant	Pizza Place	Park	Gym / Fitness Center
31	Noho	Italian Restaurant	French Restaurant	Cocktail Bar	Gift Shop	Bookstore	Grocery Store	Mexican Restaurant	Hotel	Sushi Restaurant	Coffee Shop
32	Civic Center	Gym / Fitness Center	Bakery	Italian Restaurant	Cocktail Bar	French Restaurant	Sandwich Place	Coffee Shop	Gym	Yoga Studio	Park


# VENUE SELECTION

- ▶ Using the "one map" above, I was able to explore all possibilities since the popups provide the information needed for a good decision.
- ▶ **Financial District** having Gyms, Hotels and Restaurants similar to New Delhi residence is **my preferable choice for a future residence**.
- ▶ Based on current New Delhi venue, I feel that Cluster 3 type of venues is a closer resemblance to my current place. **That means that Financial District is a better choice since it has Gyms, Hotels and Restaurant and is worth the conveniences it provides.**

# 5.0 DISCUSSION

- ▶ In general, I am positively impressed with the overall organization, content and lab works presented during the Coursera IBM Certification Course
- ▶ I feel this Capstone project presented me a great opportunity to practice and apply the Data Science tools and methodologies learned.
- ▶ I have created a good project that I can present as an example to show my potential.
- ▶ I feel I have acquired a good starting point to become a professional Data Scientist and I will continue exploring to creating examples of practical cases

# 6.0 CONCLUSION

- ▶ I feel rewarded with the efforts, time and money spent. I believe this course with all the topics covered is well worthy of appreciation.
  - ▶ This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools.
  - ▶ The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision thoroughly and with confidence. I would recommend for use in similar situations.
  - ▶ One must keep abreast of new tools for Data Science that continue to appear for application in several business fields.
- 
- A series of three parallel white diagonal lines in the bottom right corner of the slide, extending from the bottom edge towards the right edge.