THE BIG APPLE VS HOLLYWOOD

IBM DATA SCIENCE CAPSTONE PROJECT

A. ABDELSALAM OCT 17, 2020

INTRODUCTION

- The fight over which of New York city or Los Angeles is the heart of the United States has been going on, and will most likely go on, for many years. Any of these cities alone has a tremendous influence on the rest of the world, maybe more so than any other city on this planet. From the huge impact that Wall Street has on world economics to the influence of Hollywood movies on the everyday lives of each of us, we could speak hours on end about them.

- New York city and Los Angeles are on the top of the lists of tourists wanting to visit for the first time the US, or for the regular tourists wanting to spend the holidays. Indeed, both are diverse and multicultural and offer a wide palette of experiences that is widely sought after by tourists. In the project we will try to group the neighborhoods of NYC and LA respectively and draw conclusions on what they both have to offer and if one city stands out more than the other.

BUSINESS PROBLEM

- From a business standpoint the aim is to help tourists in choosing their destinations depending on the type of experience proposed by one or the other city.

- Help the tourists identify what neighborhood to stay in.

- Help the locals looking to move from one city to the other, or immigrants, decide which city and/or neighborhood suits them best. Our analysis will hopefully provide them with a better breakdown of neighborhood compositions, in terms of culture, cuisine and facilitations.

- Finally, this could be the start of a more serious project for business owners looking to open up or branch their business.

DATA PRESENTATION

For this section we will need :

1. New York city data that contains the names of the neighborhoods, their latitudes and longitudes:
    - Data source: https://geo.nyu.edu/catalog/nyu_2451_34572
    - Description: This data set contains the required information. And we will use this data set to explore various neighbourhoods of New York City.

2. New York city data that contains the names of the neighborhoods, their latitudes and longitudes:
    - Data source: https://usc.data.socrata.com/dataset/Los-Angeles-Neighborhood-Map/r8qd-yxsr
    - Description: This data set contains the required information. And we will use this data set to explore various neighbourhoods of Los Angeles.

3. Venues related information such as names, neighborhood, latitudes and longitudes:
    - Data source : FOURSQUARE API
    - Description: This data API can be queried from within Python with specific search criteria. We will use this data set to explore various venues in neighbourhoods of both New York and Los Angeles.
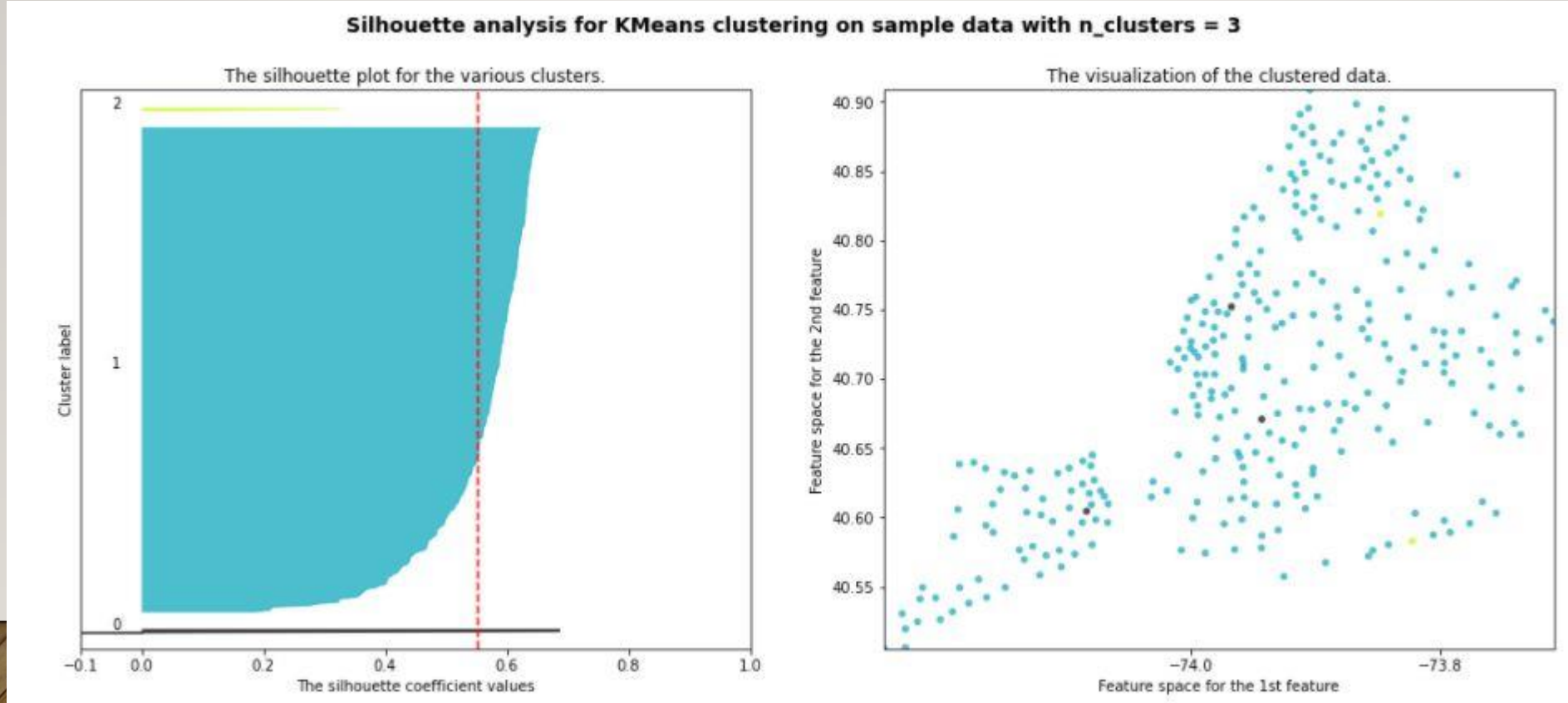
METHODOLOGY

- **Data fetching** : first step of the process is to gather the data from their source

- **Data pre-processing** : cleaning up the raw data and putting them in a dataframe so that it is easier to work with

- **Data visualization** : setting up of visualization tools and making sure every

- Collecting venue information : getting the venues information from the Foursquare API

- **One hot encoding** : converting the categorical data into numerical data for our machine learning algorithm to work with. Sorting the top 10 venues for each neighborhood

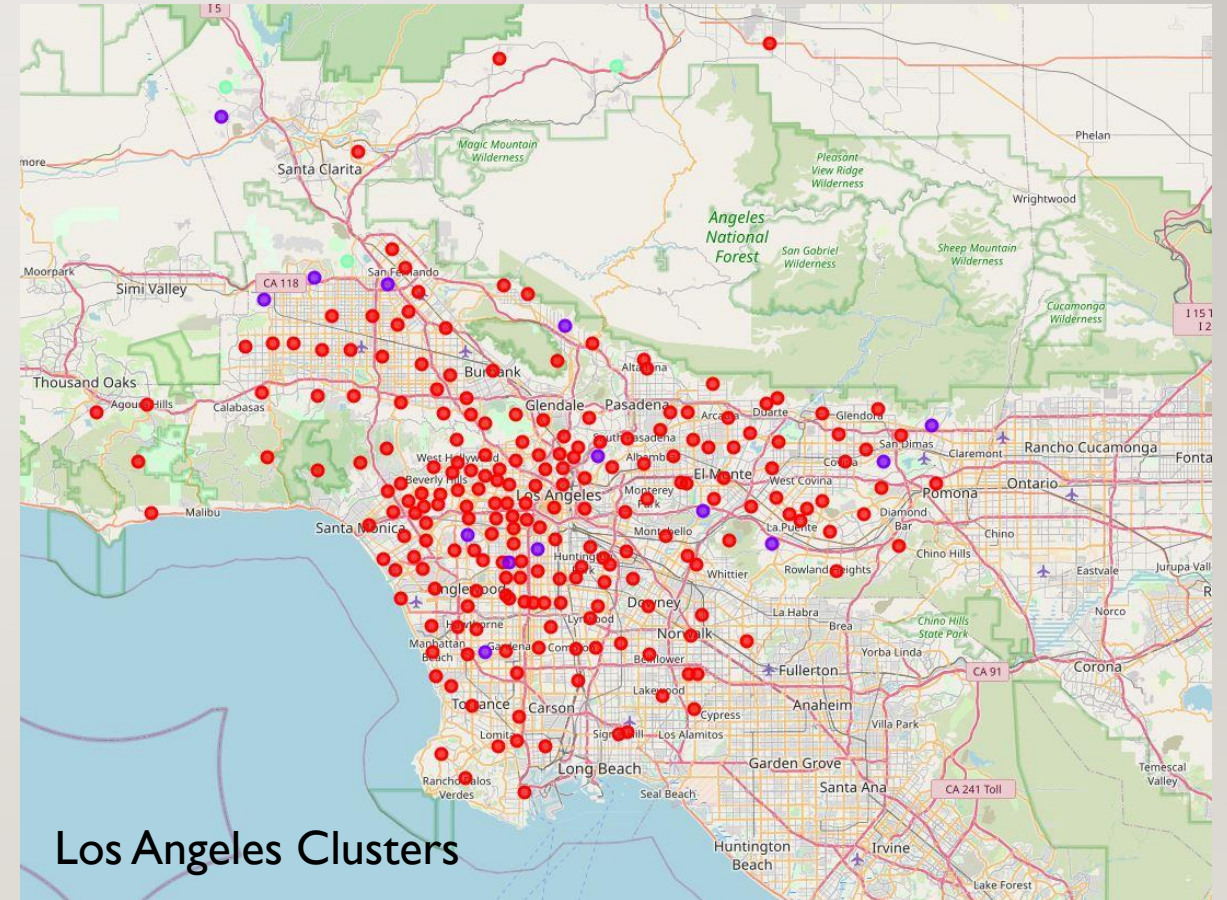- **Cluster model building:** building our cluster model to cluster similar neighborhoods together
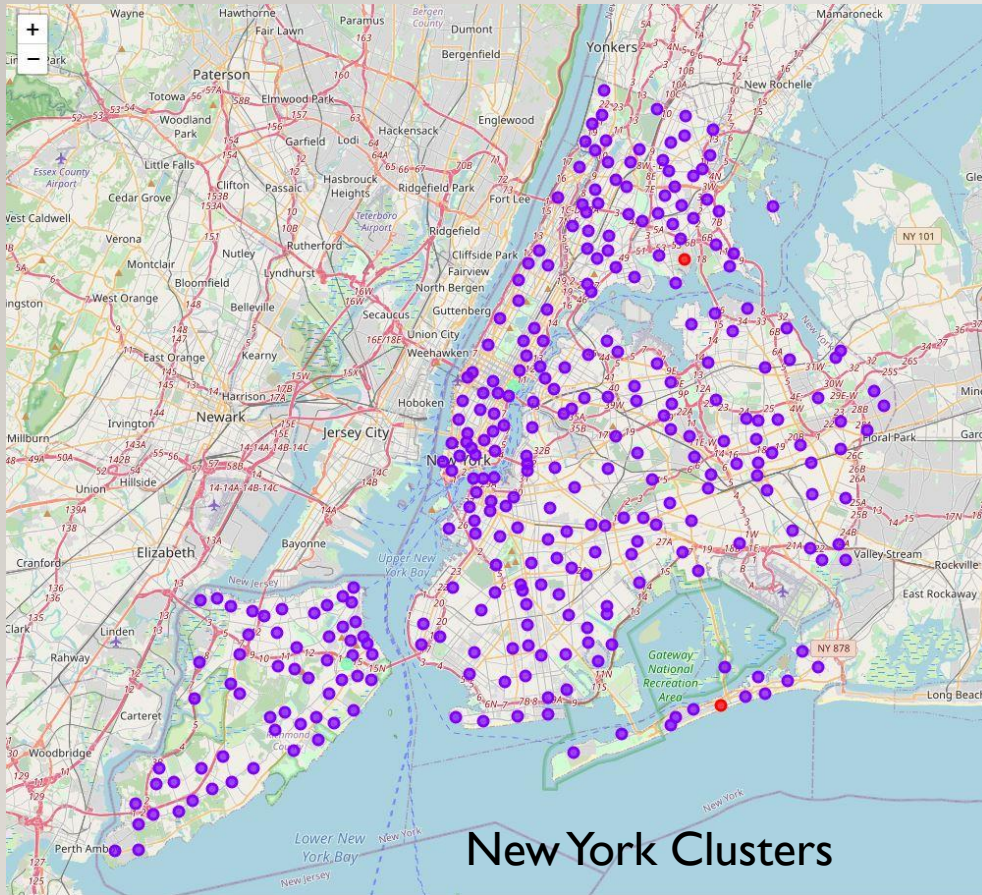
CLUSTERING MODEL

- Using the Silhouette coefficient score we will find the optimal number of clusters

- We then build our two models with them

Silhouette analysis for KMeans clustering on sample data with n_clusters = 3

RESULTS (PART 1)

New York Clusters

Los Angeles Clusters

RESULTS (PART2)

NY - Cluster 1

NY - Cluster 2

NY - Cluster 3

LA - Cluster 1

LA - Cluster 2

LA - Cluster 3

DISCUSSIONS

- Both cities' neighborhoods are clearly very multicultural. Common venues are deli's and supermarkets, Asian cuisine and international cuisine like Turkish or Peruvian.

- Both have parks to offer, as well as gyms and fitness centers, conveying a fit lifestyle in both cities.

- Los Angeles seems to offer specific food venues that dominate the rest : Italian for instance, clearly under-represented in New York

- However Pizza is one of the most recurring themes in New York, which comes to no surprise because NYC is renowned for their quality pizzas.

- Other cuisines such as, Chinese, Mexican are also more dominant in New York than Los Angeles.

- Furthermore, we can clearly see that NYC is more of a business city than LA, we see the words **bank, bus,** and **stations** recurring often in NYC and not a single time in LA. Again no surprise, NYC is the financial capital of the US and has one of the best transportation system in the world.

- What we can say about Los Angeles displays a more relaxed image. The word **coffee** is also more prominent in LA, commonly associated with gatherings and the outdoors.

- In fact, the words **yoga** and fitness related words occur more often in LA, and the absence of the word **fast-food** conveys a healthier lifetstyle in LA. This also seen by the lack of Pizza and Bar in the wordclouds, mainly associated by unhealthy habits.

CONCLUSIONS

- We have successfully gathered and explored the data, using the available datasets online as well as the data available from Foursquare .

- We were then able to cluster the neighborhoods with a kMeans clustering model in order to better compare the two cities.

- Those two cities are very much alike. Both are very multicultural, as shown by the various cuisine venues in both cities.

- Both are also considered to be fit cities with parks, gyms and fitness centers.

- There are however some differences. New york seems to have a stronger Italian and Chinese influence.
  - This could be a selling point for immigrants from these two countries for instance.

- New York also seems to be more of a business city, that could be a positive point for stakeholders looking to start or expand their business.

- More over, New York has better transportation system, more convenient for people with no cars, student or tourists alike. In fact this could mean selling point for tourists, as renting a car in Los Angeles could turn out to be expensive and hinder the vacation budget.

- On the contrary, Los Angeles seems to be a healthier city, reflected by the lack of fast-food, and a more relaxed city.

- Those results have been limited to 100 per neighborhoods and 500 m radius, therefore not completely representative, but give a general idea only.

All in all, the results of this project have shown that while there are similarities in both cities, there are some arguments that could favor one city over the other whether one is looking to spend the holidays, relax or start a business.