

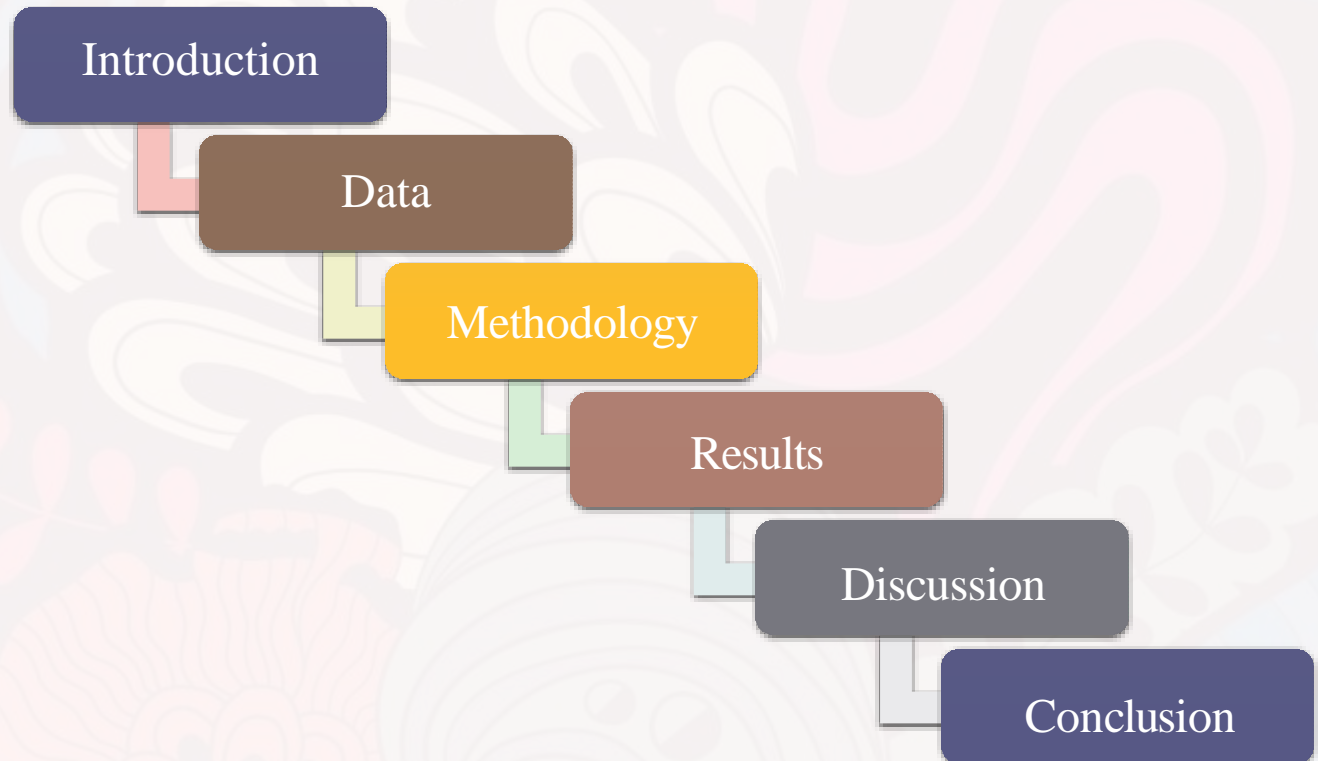
12/11/2019

COURSERA CAPSTONE PROJECT

By Shaurya Vardhan ,
December 2020



Agenda



12/11/2019

INTRODUCTION

Problem Description

- The aim is to help tourists choose their destinations depending on the experiences that the neighbourhoods have to offer and what they would want to have. This also helps people make decisions if they are thinking about migrating to London or Paris or even if they want to relocate neighbourhoods within the city. Our findings will help stakeholders make informed decisions and address any concerns they have including the different kinds of cuisines, provision stores and what the city has to offer.

Audience

- First Audience will be any person who plans to move to New York City or any other major cities.
- And the second one is Data Science training that needs a powerful case study to train his / her skill

12/11/2019

DATA

The heart of our project

Data Sources and Processing

We require geographical location data for both London and Paris. Postal codes in each city serve as a starting point. Using Postal codes we use can find out the neighborhoods, boroughs, venues and their most popular venue categories.

3.1 London

To derive our solution, We scrape our data from https://en.wikipedia.org/wiki/List_of_areas_of_London

This wikipedia page has information about all the neighbourhoods, we limit it London.

borough : Name of Neighbourhood

town : Name of borough

post_code : Postal codes for London.

This wikipedia page lacks information about the geographical locations. To solve this problem we use ArcGIS API

3.2 ArcGIS API

ArcGIS Online enables you to connect people, locations, and data using interactive maps. Work with smart, data-driven styles and intuitive analysis tools that deliver location intelligence. Share your insights with the world or specific groups.

More specifically, we use ArcGIS to get the geo locations of the neighbourhoods of London. The following columns are added to our initial dataset which prepares our data.

latitude : Latitude for Neighbourhood

longitude : Longitude for Neighbourhood

3.3 Paris

To derive our solution, We leverage JSON data available at <https://www.data.gouv.fr/fr/datasets/r/e88c6fda-1d09-42a0-a069-606d3259114e>

The JSON file has data about all the neighbourhoods in France, we limit it to Paris.

postal_code : Postal codes for France

nom_comm : Name of Neighbourhoods in France

nom_dept : Name of the boroughs, equivalent to towns in France

geo_point_2d : Tuple containing the latitude and longitude of the Neighbourhoods.

12/11/2019

METHODOLOGY

The heart of our project

Methodology

Package breakdown:

Pandas : To collect and manipulate data in JSON and HTML and then data analysis

requests : Handle http requests

matplotlib : Detailing the generated maps

folium : Generating maps of London and Paris

sklearn : To import Kmeans which is the machine learning model that we are using.

The approach taken here is to explore each of the cities individually, plot the map to show the neighbourhoods being considered and then build our model by clustering all of the similar neighbourhoods together and finally plot the new map with the clustered neighbourhoods. We draw insights and then compare and discuss our findings.

Data science Applied tools

- Web-scraping of sites is used to consolidate data-frame information which was saved as csv files for convenience and to simplify the report.
- Geodata was obtained by coding a program to use Nominatim to get latitude and longitude of subway stations and also for each apartments for rent listed.
- Geopy_distance and Nominatim were used to establish relative distances.
- Seaborn graphic was used for general statistics on rental data.
- Matplotlib used CM module that provides a large set of colormaps, functions for registering new colormaps and for getting a colormap by name, and a mixin class for adding color mapping functionality.
- Sklearn used KMeans for applying cluster on generated Data
- Maps with popups labels allow quick identification of location, price and feature, thus making the selection very easy

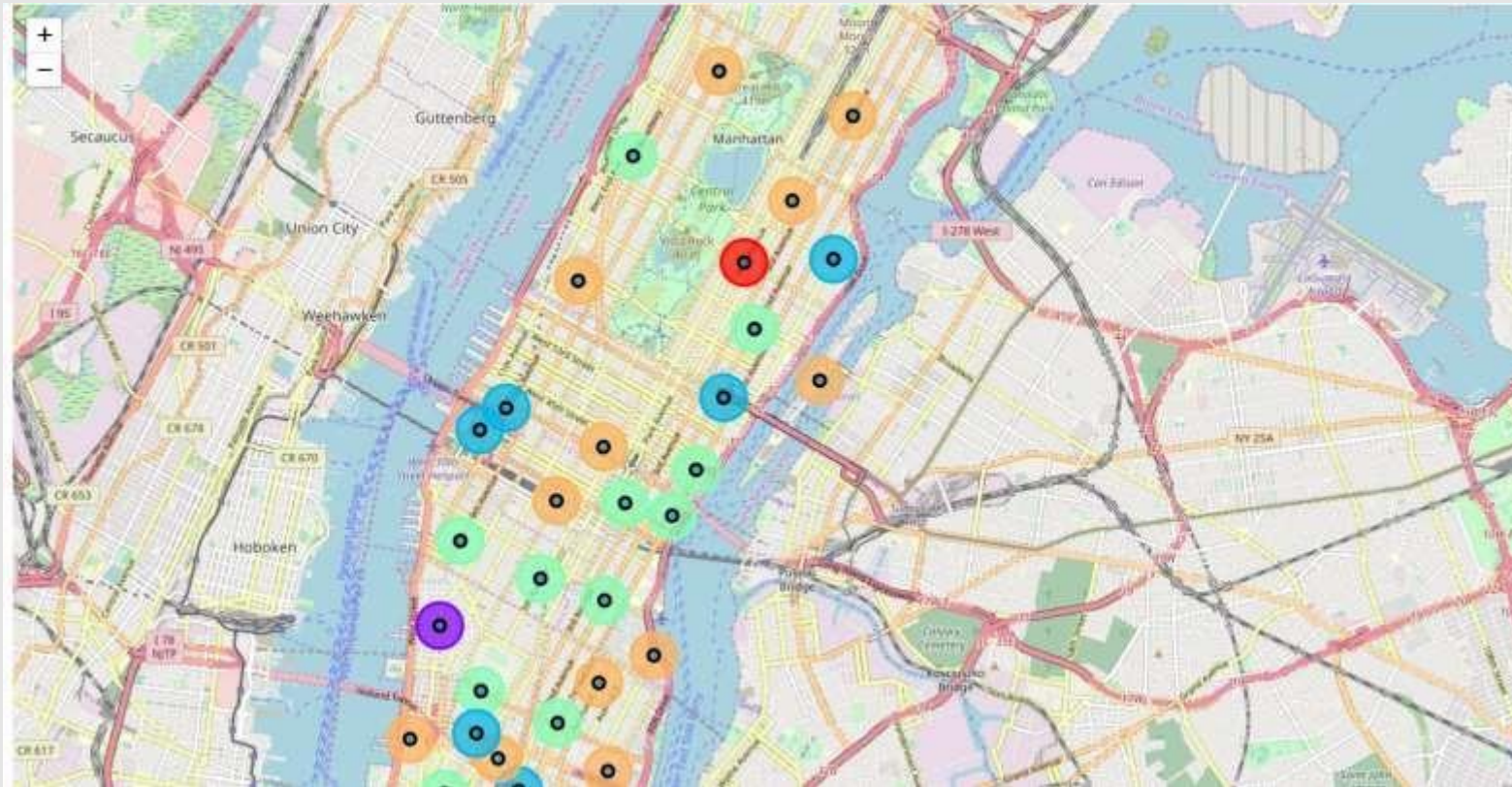
Execution

12/11/2019

EXECUTION

Run our codebase

Visualizing neighborhoods



Visualizing clustered neighborhoods



12/11/2019

CONCLUSION

Conclusion

The purpose of this project was to explore the cities of London and Paris and see how attractive it is to potential tourists and migrants. We explored both the cities based on their postal codes and then extrapolated the common venues present in each of the neighbourhoods finally concluding with clustering similar neighbourhoods together.

We could see that each of the neighbourhoods in both the cities have a wide variety of experiences to offer which is unique in its own way. The cultural diversity is quite evident which also gives the feeling of a sense of inclusion.

Both Paris and London seem to offer a vacation stay or a romantic getaway with a lot of places to explore, beautiful landscapes, amazing food and a wide variety of culture. Overall, it's up to the stakeholders to decide which experience they would prefer more and which would more to their liking.