



IBM Developer  
SKILLS NETWORK

# Winning Space Race with Data Science

<Zandile Maleka>  
<09 August 2023>



# Outline

---

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

# Executive Summary

---

This report discusses the use of Foursquare Api to build upon the comparison between Paris and London and the correlation whether someone should start a business in either of the two cities. Various machine learning algorithms were used to compare and differentiate the two cities as well as other geolocation tools to check either of the city is good to start an Artificial Intelligence Business.

# Introduction

---

- The final course of the Data Science Professional Certificate consists of a capstone project where in all the skills and relevant knowledge that one has gathered from this 9 intense courses has to be applied on a final capstone project. The final problem as well as the analysis is left for the reader to explore and decide. The idea uses location data with the help of the foursquare api that can be leveraged into coming up with a problem that the foursquare location data to solve it or just in contrast to compare cities or neighbourhoods of one's own choice. London is the capital and largest city of England and the United Kingdom. Standing on the River Thames in the south-east of England, at the head of its 50-mile (80 km) estuary leading to the North Sea, London has been a major settlement for two millennia.

# Introduction

- Paris is the capital and most populous city of France, with an estimated population of 2,150,271 residents as of 2020, in an area of 105 square kilometres (41 square miles). Since the 17th century, Paris has been one of Europe's major centres of finance, diplomacy, commerce, fashion, science and arts. The main Goal of this project that I have chosen would be to evaluate the comparison between Paris and London as well as point out the differences. Another factor to be included is which city would be more ideal to start an Artificial Intelligence company and the various factors correlating to it as both cities are major cities and global hotspots in the world for tech companies

# Data Collection

---

- Various data sets were collected, reformatted and analysed in order to get the required results. Some of them include
- <http://www.cgedd.developpement-durable.gouv.fr/house-prices-in-france-property-price-index-french-a1117.html> - House Prices in France
- <https://www.kaggle.com/alphaepsilon/housing-prices-dataset> - Housing Dataset
- <https://data.world/datasets/real-estate> - Numerous Datasets for different categories
- <https://data.london.gov.uk/dataset?tag=start-ups> - Data sets for London
- <https://www.kaggle.com/tags/companies> - Various companies and their dataset

# Data Collection – SpaceX API

---

- <http://analytics.dkv.global/data/pdf/AI-in-UK/AI-in-UK-1000-UK-AI-Companies-Profiles.pdf>
- <https://craft.co/artificial-intelligence-companies-in-paris-fr?page=1>
- <https://clutch.co/fr/developers/artificial-intelligence> Finally to compare them various visualisation tools were used and articles referenced in order to reach the final conclusion

# Data Collection - Scraping

---

- The algorithm below gets the required latitude and longitude of London(similar code has been coded for Paris) using the GoogleMaps Geocoder API



# Data Wrangling

---

- An in-depth research of the dataset has been done and a thorough analysis of the various features and methods have been investigated to ensure the maximum accuracy of the model as possible. After reduction of the number of features in the data frame by replacing them with more useful data cluster analysis was done to find the best cluster of both Paris and London and then correlation and various other visual graphs were used to compare the two cities

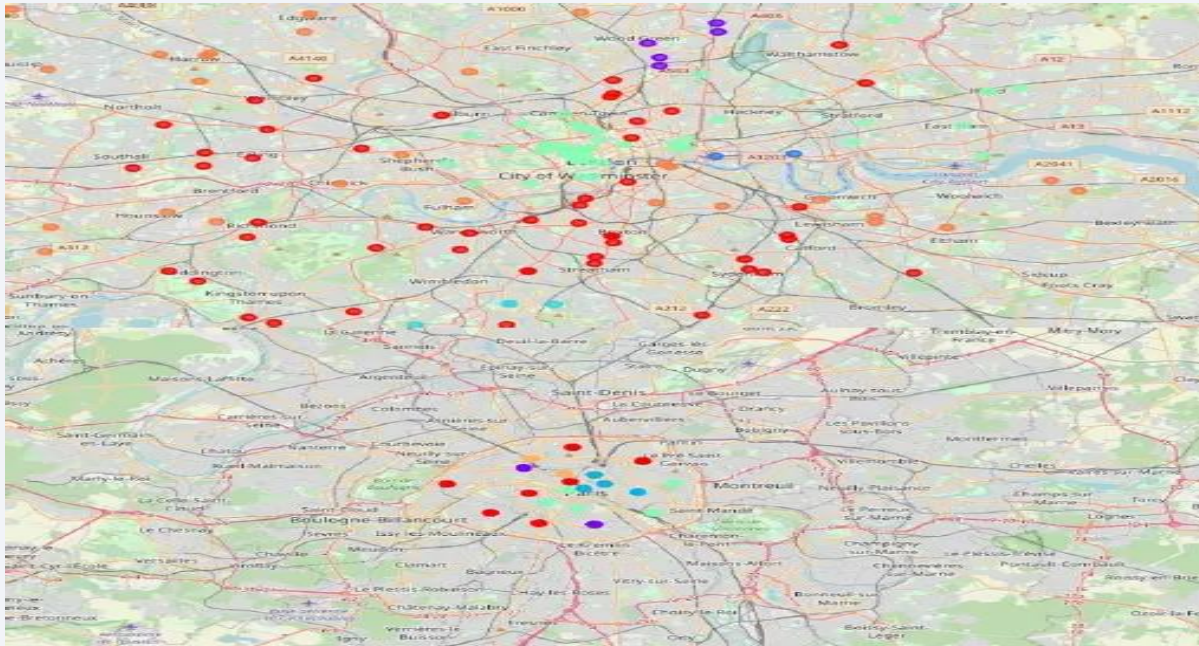
# EDA with Data Visualization

---

- Folium makes it easy to visualize data that's been manipulated in Python on an interactive leaflet map. It enables both the binding of data to a map for choropleth visualizations as well as passing rich vector/raster/HTML visualizations as markers on the map. It uses the Open StreetMap technology. The code below shows only of Paris but a similar code has been coded even for London

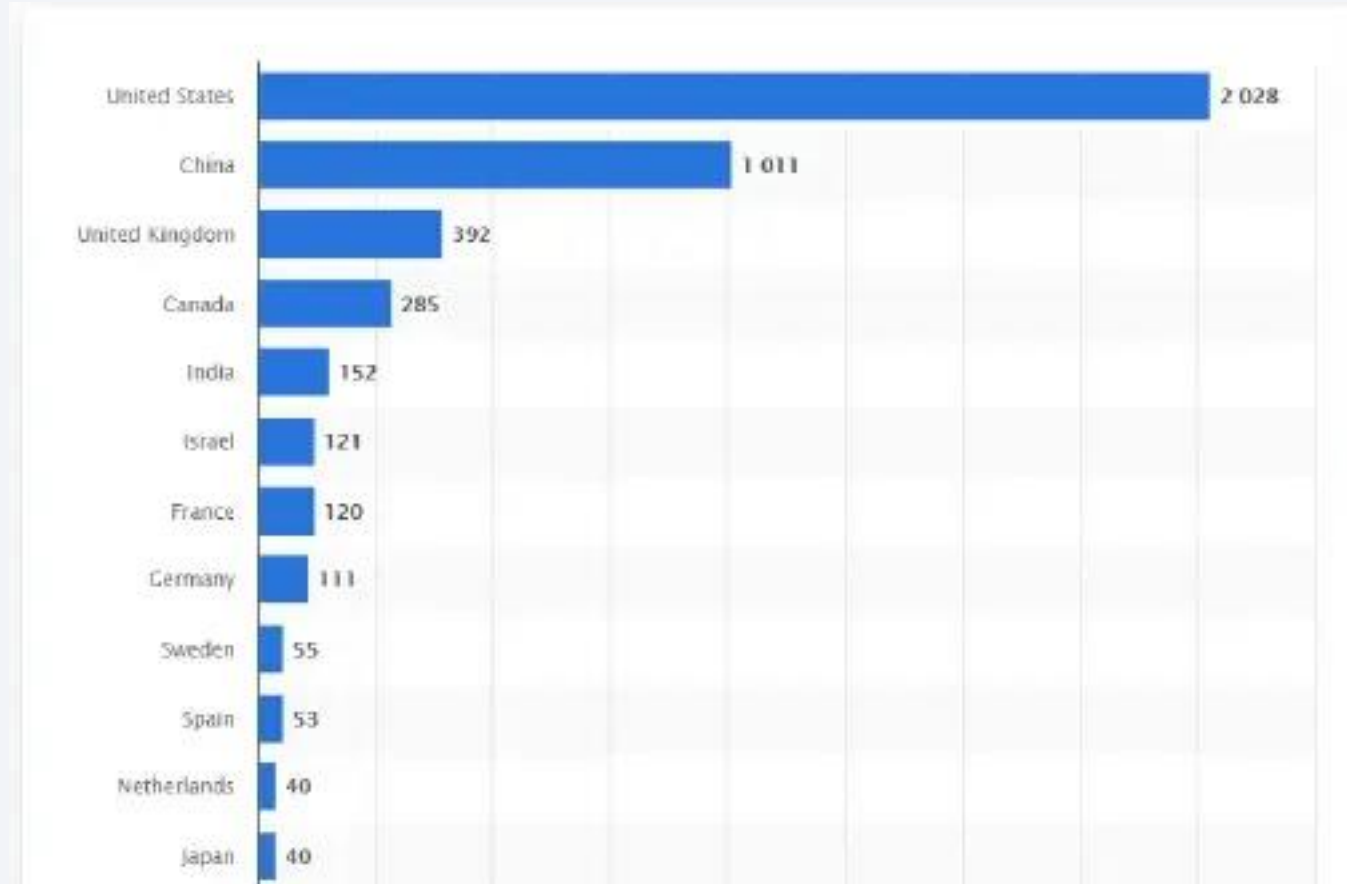
# Build an Interactive Map with Folium

---



# Build a Dashboard with Plotly Dash

---



# Predictive Analysis (Classification)

---

- For the artificial intelligence dataset after cleaning and webscrap-ping the number of startups were plotted and number of AI compa-nies in each country was counted. Our main focus however is on theUnited Kingdom specifically London and France specifically Paris.

# Results

- Similarities
- Both cities are multicultural and diverse in their own ways and share a rich history of their own.
- Most of the famous neighbourhoods have a restaurant as its top most Common Venue.
- Example : In Paris the Louvre is one of the most famous icons if not in the world also and its most common venue is the Plaza/French Restaurant.
- Similarly for the famous icon in London that is Westminster are Pubs and restaurants.
- The top 3 most common Venue points for London are
  - Coffee Shop
  - Hotel
  - Cafe
- The top 3 most common Venue points for Paris are
  - Coffee Shop
  - Pub
  - Cafe

## <Dashboard Screenshot 3>

---

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

# Discussion

---

- There are major challenges while constructing a dataset ie:
- The dataset for the Artificial Intelligence wasn't readily available and so had to be scrapped from multiple sources which often leads to inconsistency happening as well as errors.
- Only a random sample of 0.05 percent was taken into consideration. A good and optimal model would take a testing data and a training data and would train it on the complete dataset multiple times.
- The data obtained through the API calls would return different results each time its called. Multiple trials and error runs are required to get the desired result.
- The districts have too complex geometry which would bring an error in our analysis if the venues are too close to each other. This is one of the reason why Pipelines are required. However no doubt that if this process was to be repeated multiple times the desired outcome would have generated and a better comparison could have been made.



# Conclusions

---

- After an indepth review of the comparison between London and Paris and which city would be a better place to start an Artificial Intelligence Company or invest multiple conclusions can be drawn. One of them being that both cities are diverse in their own ways and boast a culture unlike no other. Artificial Intelligence is a booming topic and recently more people have started investing into it as well as companies automating their processes. Both cities offer a wide range of opportunities for anyone starting to invest in Artificial Intelligence or even start a company and various factors were shown. Finally a better model could be made by various other methods and much stronger Machine Learning Algorithms like KD Tree. Furthermore, clustering however did help us to highlight the most optimal venues and areas. Finally correlation does not imply causation and so any result here is subject to change on various other trends and opinions and datasets

Thank you!

