Capstone Project The Battle of Neighborhoods

REPORT

MLA Inc.

Table of Contents

Executive Summary

1. Introduction

The MLA Inc. founded in 2011, with headquarters in Chicago has experienced year on year a substantial growth on their business of supplying Gym and Gym appliances.

The innovative concept of Gym, associated with Digitization, AI, and his strong brand "GoGo Fitness", allow MLA Inc. to become the market leader in USA.

The MLA vision and strategy are to expand to markets outside USA. In 2019, the executive board from MLA Inc., chaired by Mr. J.L. Champs (CEO), has decided to expand their Gym concept to UK. The business plan is to start in 2020, with 5 new gym in the City of London, and grow from there to other main cities in the subsequent years.

The MLA business expansion to UK has enormous importance in shareholders strategy, therefore Mr. J. L. Champs has set this program at the highest priority for 2020.
Mr. Champs has assigned 3 of his champions for this program coded "The Battle of Neighborhoods".

The program main stakeholders and sponsors are Mrs. S. K. Voort (CMO), Mr. L. A. Raidillon (CDO), and Mrs. C. H. Nagasaki (COO).

Mr. Raidillon has agreed with his peers to bring on board his senior data scientist (Mr. A. R. Duarte) to help with business problem resolution.

1.1 Business Problem

Due to the high importance of "The Battle of Neighborhoods" program, it is from upmost importance to select the right locations for the first 5 Gyms in the City of London.

The stakeholders came up with following requirements (business problem):

Which neighborhood in the City of London has 'Gym' in top 10 venues?

Which neighborhood in the City of London has "Gym" in the top 5 most common venues?

How is the Gym market segmented in the City of London?

Which are the best recommended districts in the City of London to setup the 5 gyms?

1.2 Business Understanding

The Data Science team has been tasked to produce a report to highlight the Gym market segment distribution across the boroughs and neighborhoods of the City of London.

The report should feature the top 5 districts where Gym is within the top 5 most common venues.

The report should explore and cluster the neighborhoods in the City of London regarding Gym venue, and present findings and conclusions.

2. Methodology

This section describes the research methods and data sources used for analysis as well as the Data requirements, data collection, data preparation, and modeling.

2.1 Analytical Approach

The business problem requests customer segmentation per location.
The data science team decided to explore the data with a cluster algorithm to see whether a natural statistical separation exists.

The clustering technique separates the data set into significant groups or buckets. Cluster algorithms attempt to divide the data into distinct groups by minimizing the distance between data points within a cluster and maximizing the distance between clusters.

2.2 Data Sources

The team decided to leverage the Foursquare location data to solve the problem in combination with two other datasets.

The table below describes the data sources used in this analysis.

Data sources	Description
Foursquare	Using the Foursquare API (www.foursquare.com), one can search for specific type of venues or stores around a given location. The Places API offers real-time access to Foursquare' s global database of rich venue data and user content to power ones location-based experiences

UK Postcodes with Latitude and Longitude	This dataset contains all the UK postcodes with respective latitude and longitude, and can be found in the website https://www.freemaptools.com/download-uk-postcode-lat-lng.htm
The City of London postcode area	This dataset can be found in the Wikipedia https://en.wikipedia.org/wiki/EC_postcode_area

The diagram below depicts the high-level architecture used for this data science analysis.
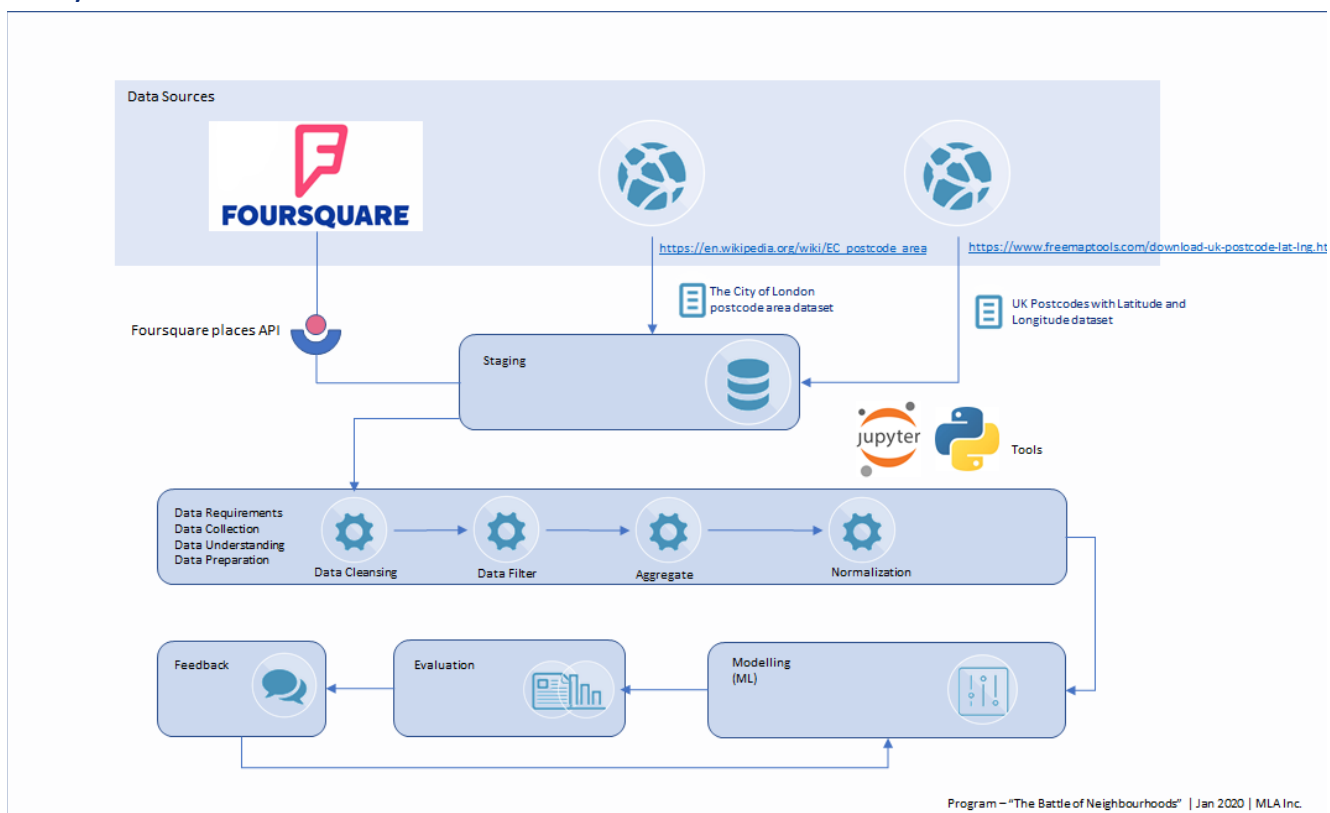
The data collected via Foursquare places API will be following format:

https://api.foursquare.com/v2/venues/_id/client_id=*****&client_secret=*******&v=version

The sample of the data collected in the UK postcode dataset and the City of London postcode dataset is shown below.

id	postcode	latitude	longitude
1	AB10 1XG	57.14417	-2.11485
2	AB10 6RN	57.13788	-2.12149
3	AB10 7JB	57.12427	-2.12719
4	AB11 5QN	57.1427	-2.0933
5	AB11 6UL	57.13755	-2.11223

Postcode district	Post town	Coverage	Local authority area
EC1A	LONDON	St Bartholomew's Hospital	City of London, Islington
EC1M	LONDON	Clerkenwell, Farringdon	Islington, Camden, City of London
EC1N	LONDON	Hatton Garden	Camden, City of London
EC1P	LONDON		*non-geographic*
EC1R	LONDON	Finsbury, Finsbury Estate (west)	Islington, Camden
EC1V	LONDON	Finsbury (east), Moorfields Eye Hospital	Islington, Hackney

2.3 Data Preparation

[TO BE DELIVERED IN WEEK2]

2.4 Modelling

[TO BE DELIVERED IN WEEK2]

2.5 Model Evaluation

[TO BE DELIVERED IN WEEK2]

3. Results

[TO BE DELIVERED IN WEEK2]

4. Discussion

[TO BE DELIVERED IN WEEK2]

5. Conclusions

[TO BE DELIVERED IN WEEK2]

6. References

[TO BE DELIVERED IN WEEK2]

7. Acknowledgments

[TO BE DELIVERED IN WEEK2]

8. Appendices

[TO BE DELIVERED IN WEEK2]