



# Preferred location map for the online pharmacy warehouses in London

IBM APPLIED DATA SCIENCE  
CAPSTONE PROJECT

BY ANTON GIL

MAY 2020

# Business problem - Intro

- ▶ The goal of this capstone project is not to develop a business plan of an online pharmacy, but rather to try to solve one of the core problems of this process which is the pharmacy warehousing optimal location detection. Therefore, the problem I am trying to solve through the usage of data science methodology, machine learning algorithm and location data is to define the optimal location of the pharmacy warehouses.
- ▶ The key assumption is that the optimal location for the warehouse should be in a place with a high density of population, hence a high demand for pharmacy products. With data available I will take the supply analysis approach and track the areas with the highest density of pharmacy stores, classifying those as the preferred locations for the warehouse location.
- ▶ This methodology should be applicable to number of countries, for the reason of data availability I will use the London city as a subject of this capstone. By London I surely mean the one standing on the river Thames.

# Data

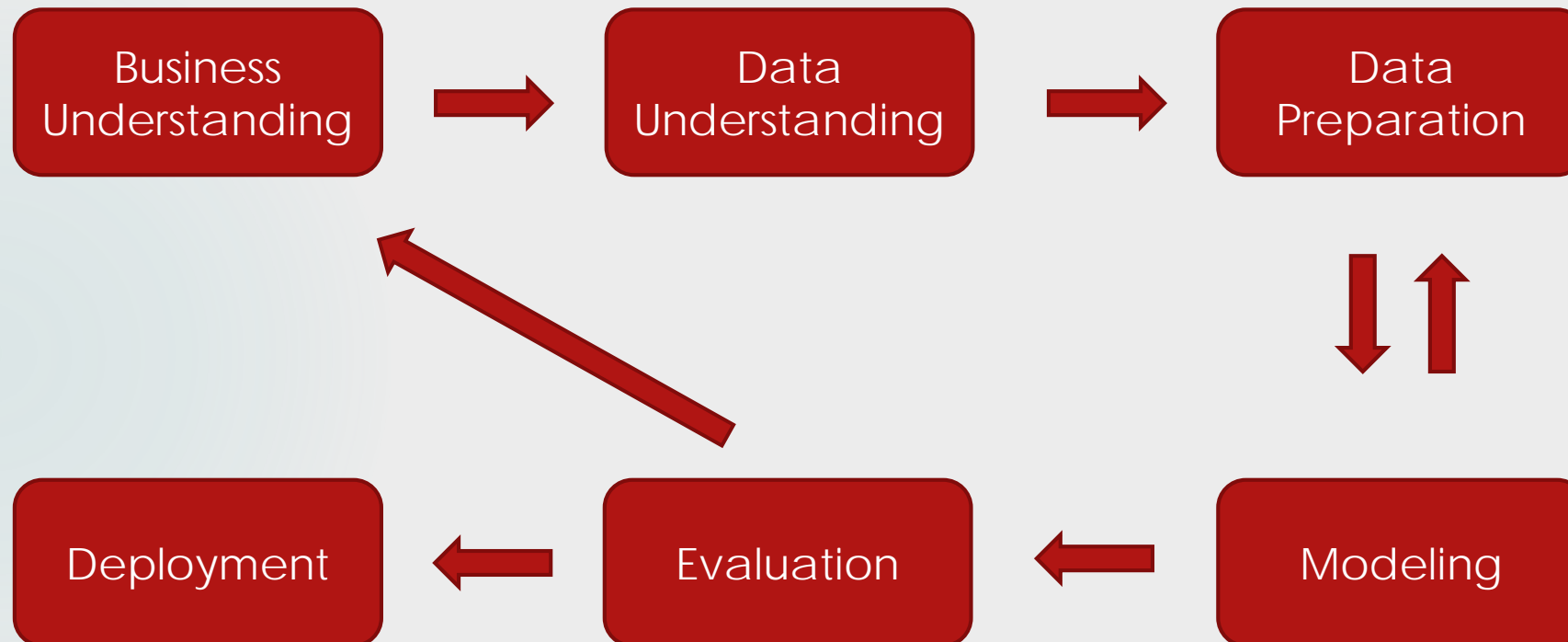
To solve the stated problem, I will require the data illustrated in the following sum-up table along with the purpose, extraction method and the source:

Data	Purpose	Extraction method	Source
List of neighborhoods in London	To define the area for the analysis	Web scraping (Beautifulsoup)	<a href="https://en.wikipedia.org">Wikipedia.org</a>
Latitude and longitude coordinates	City map plotting with pharmacy stores locations	Python Geocoder package	-
List of pharmacy stores with locations	To perform clustering on the neighborhoods	API Requests	<a href="https://www.foursquare.com">Foursquare</a>

# Methodology – CRISP DM

For this capstone I followed the data science methodology based on **CRISP-DM**.

It consists of 6 core steps:



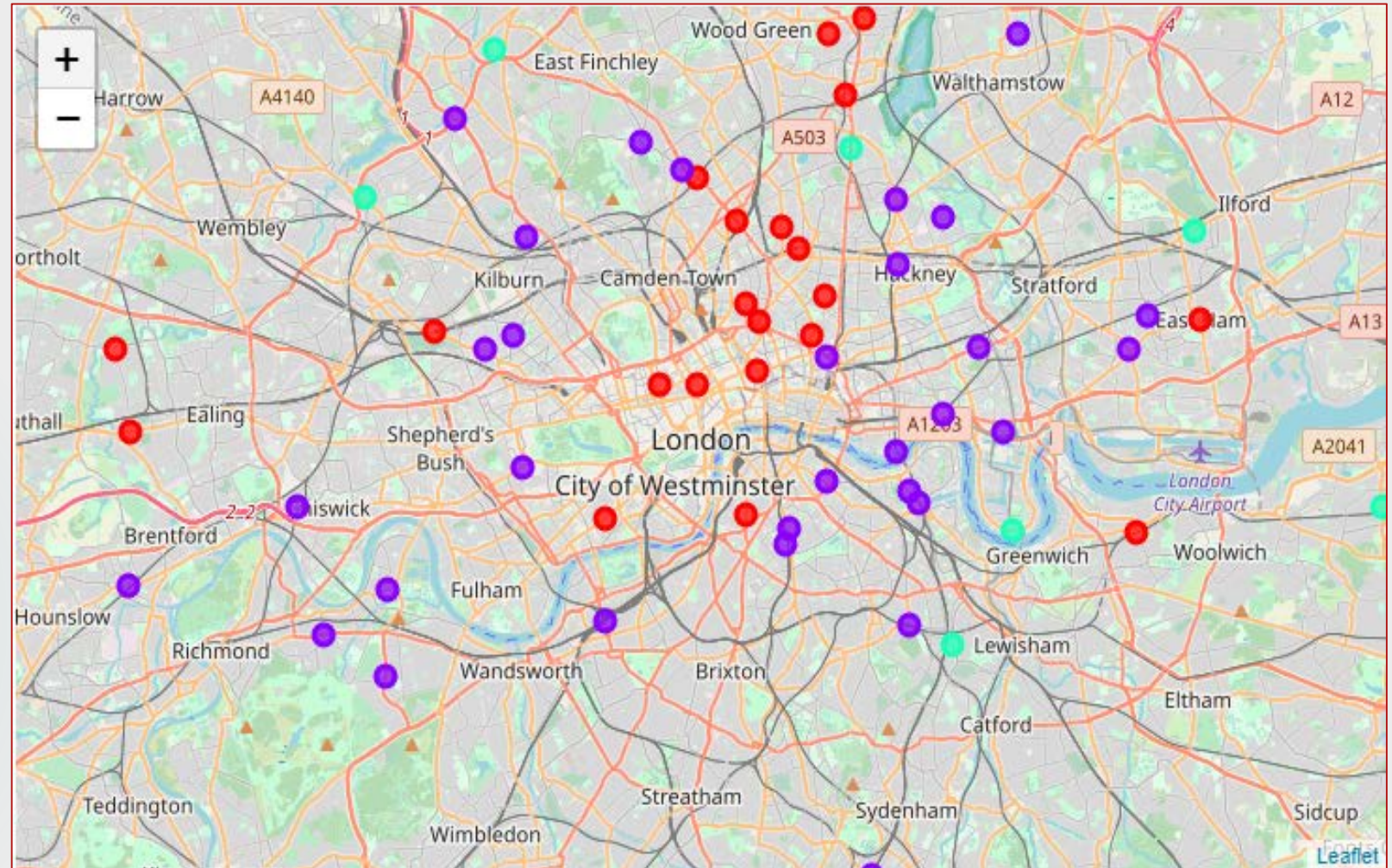


# Methodology – Approach

1. Conduct web scrapping with the use of BeautifulSoup from Wikipedia page to get the list of London neighborhoods;
2. Collect the geographical coordinates of each neighborhood location with the use of Geocoder from geopy library;
3. Build a map of London with the use of the latitude and longitude values of neighborhood and London itself along with the folium package;
4. Use Foursquare API to explore the specified neighborhoods for the Pharmacy shops;
5. Create a separate data feature with the use of one hot encoding to measure the frequency of occurrence.

# Results (1/2)

- ▶ Cluster 0 – the cluster of neighborhoods marked red is the one we are interested in, since it has the highest density of Pharmacy shops and these locations should be considered for the warehouse placement.
- ▶ Cluster 1 – the cluster of neighborhoods marked purple is the one that is a second tier with a moderate density of pharmacy shops.
- ▶ Cluster 2 – the cluster of neighborhoods marked light blue is the one with the lowest density of pharmacy shops.



# Results (2/2)

The preferred location for an online pharmacy warehouse in London, UK has been defined and characterized by the Cluster 0. The findings of this work should assist stakeholders in business plan implementation.

List of preferred neighborhoods

Elmers End	Harold Wood	Highams Park	Chipping Barnet
Islington	Highbury	Seven Sisters	Tottenham
Greenford	Worcester Park	Gidea Park	Kensal Green
Hackbridge	Holloway	Bedford Park	St Helier
Hanwell	Northolt	Upper Holloway	Canonbury
Hanworth	Ickenham	Barnsbury	Chelsea
Clerkenwell	Lambeth	Beckton	Charlton



# Summary

## Discussion

Based on the gathered results I do recommend to start the consideration of the online pharmacy warehouse deployment in the neighborhoods that got in cluster 0 as it is the one that has the highest density of current pharmacy shops. The high concentration of competitors surely might be a more complicated locating to start the online business, therefore if the business model analysis would suggest that the business should start from the places with the lowest competition possible the cluster 2 is the one to consider.

## Limitations and Suggestions for the Future Research

I would like to note that the approach used is market-supply-driven. Prior to making a decision it is important to consider the demand side of the market. The population density analysis and population age groups are factors to consider in further model development. From the business side in-depth market analysis is recommended to define customer preferences and readiness to use online pharmacy services on day-to-day basis.