IBM. Data Science

# Coursera Capstone Project

## THE BATTLE OF RESTAURANTS IN LONDON

By Valeriya Kovalenko

2020

# Overview

- **I**ntroduction and Business problem
- **D**ata Acquisition Approach
- **M**ethodology
- **R**esults
- **D**iscussion

# Introduction and Business problem

This project will compare London restaurant categories based on **Foursquare** data. There are a large number of **restaurants in London**, and therefore it is very difficult to understand where to go for a newly arrived person. Based on **Foursquare user reviews**, I group these restaurants so that Foursquare can determine where to go

# Data Acquisition Approach

In this project, I will use the Foursquare API to get the necessary data on the location of restaurants in London:

- Name and ID
- Location
- Category
- Count of Likes
- Coordinates of London

I also use the Foursquare API to get a list of all places in London

# Methodology

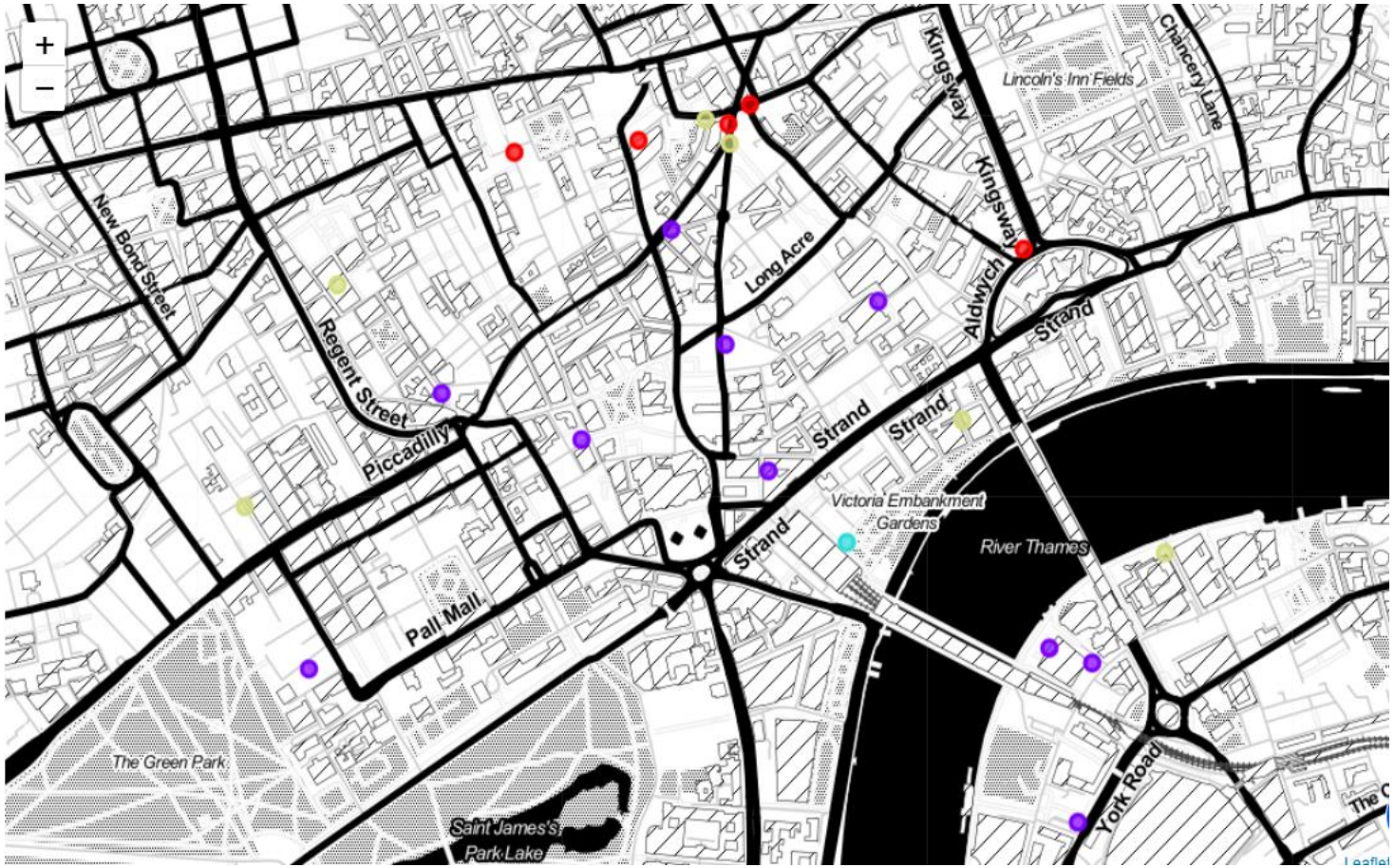
There are several steps in this research:

- The **coordinates** of London were obtained and **venues** from the site were requested.
- Created a **dataframe** with the name, id, category and coordinates. Excluded attractions that do not meet the conditions of restaurants.
- Then, the **number of likes** for each restaurant was requested and summarized from the site. Categorical variables for restaurants have been created to group them by the **number of likes**. User preferences are also taken into account by type of cuisine.
- In order to divide the restaurants into **groups**, the data collected from Foursquare was used.
- Next, **a k-means clustering algorithm** was created that groups restaurants into clusters so that people who want to eat in Moscow can easily see which restaurants are better, which cuisine is available and where in Moscow they can eat.

# Results

As a result of data processing and research in London, a **100** venues were found, of which only **22** are restaurants.

After collecting data on the number of likes and grouping restaurants by type of cuisine, a **cluster analysis** was performed.



# Results

Based on the number of **likes** and category, restaurants are divided into **4 clusters**.

**1.** Very high marks, only one bar

| | name | likes | likes_cat | categories_new | label |
|---|---|---|---|---|---|
| **4** | Gordon's Wine Bar | 1270 | high | cocktails wine | 2 |

**2.** High marks, mostly pubs and Italian food

| | name | likes | likes_cat | categories_new | label |
|---|---|---|---|---|---|
| **60** | Flat Iron | 655 | high | beer pub | 0 |
| **71** | Freud | 599 | high | cocktails wine | 0 |
| **72** | Pizza Pilgrims | 508 | high | italian food | 0 |
| **73** | The Delaunay | 789 | high | other | 0 |
| **79** | The Craft Beer Co. | 620 | high | beer pub | 0 |

**3.** Average ratings Asian and cocktail bars

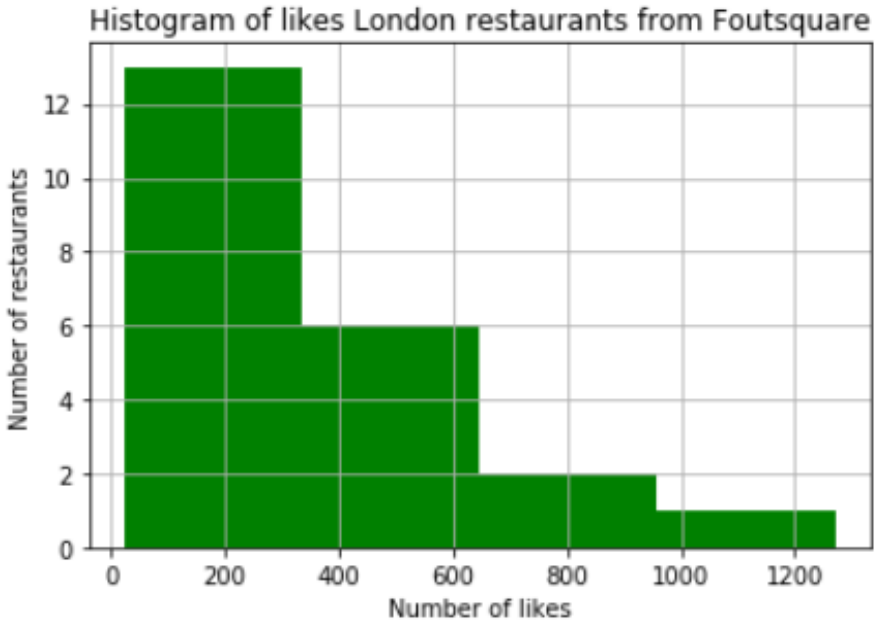
| | name | likes | likes_cat | categories_new | label |
|---|---|---|---|---|---|
| **23** | American Bar | 251 | low | cocktails wine | 3 |
| **70** | Cahoots | 366 | middle | cocktails wine | 3 |
| **77** | Punjab | 332 | middle | indian food | 3 |
| **81** | Kanada-Ya | 461 | middle | asia food | 3 |
| **93** | Gymkhana | 397 | middle | indian food | 3 |
| **94** | The Understudy | 322 | middle | beer pub | 3 |

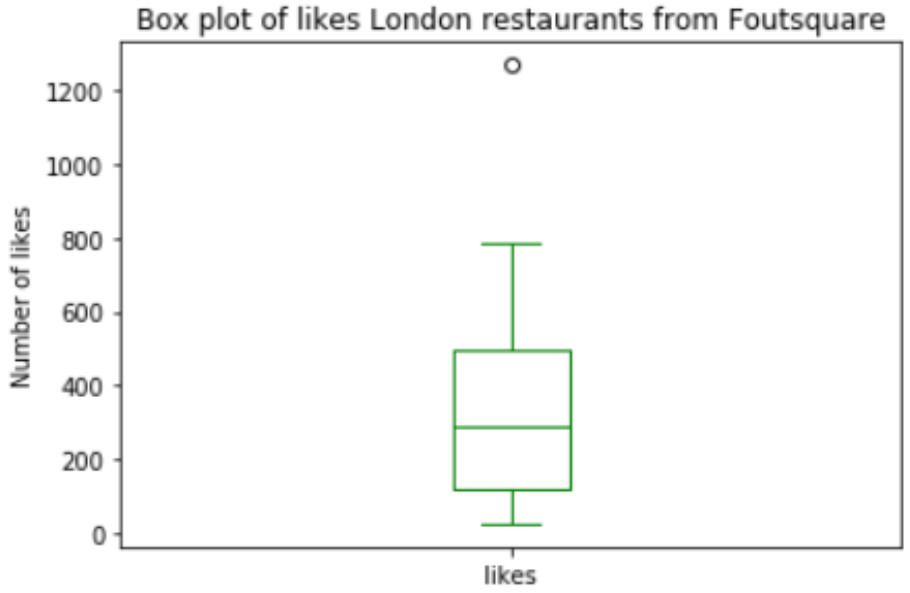
**4.** Low and very low, mainly Asian and cocktail bars

| | name | likes | likes_cat | categories_new | label |
|---|---|---|---|---|---|
| **5** | Tandoor Chop House | 74 | very low | indian food | 1 |
| **8** | Japan Centre | 175 | low | asia food | 1 |
| **12** | The Alchemist | 38 | very low | cocktails wine | 1 |
| **25** | Bar Americain | 126 | low | cocktails wine | 1 |
| **27** | SUSHISAMBA | 119 | very low | asia food | 1 |
| **29** | Fabrique | 134 | low | other | 1 |
| **52** | Dukes Bar | 156 | low | cocktails wine | 1 |

# Discussion

Quite interesting observations can be made from the results of data processing. The number of restaurants with low rating is greater than with high.




The maximum number of likes is **1270**, the minimum is **25**. At the same time, the median value is less than the average - **286** to **340**, which indicates that most restaurants are shifting towards a low number of likes.

The advantage of clustering from the usual classification is also quite obvious - classification by percentiles identified 4 groups, in which several restaurants received high marks. Thanks to clustering, we were able to see that one bar is especially loved by the residents of London, and it forms a separate cluster. It definitely increases the value of research in relation to classification.

# Conclusion

Gathering data on **London restaurants** from Foutsquare revealed the undisputed leader among restaurants - this is **Gordon's Wine Bar**. Obviously, people who come to London should visit this place most beloved by local establishments. Residents of London are obviously more attracted to Pubs and Italian restaurants than to Asian-Indian cuisine.

Thank you for the attention!