CAFÉ RATINGS

Most popular items influences on the overall café ratings

INTRODUCTION: BACKGROUND

The café and coffee shop industry is an important industry in terms of employment and contribution to the UK economy, according to "Investigating the Success of Independent Coffee Shops and Cafes in the UK: Findings from a Pilot Study" (by Jacqueline Douglas, Alexander Douglas, Michele Cano, David Moyes).

However, sustainability in terms of longevity is an issue. Despite the low barriers to entry into the industry, cafés are a very high risk business and most start-ups fail. According to the *Office for National Statistics*, only 34.6% of accommodation and food services survive longer than five years (*Office for National Statistics, 2016*).

It's clear that to survive coffee shops owners need to have some feedback from their clients. The feedback should be analyzed (in this research with the help of Machine Learning) and interpreted, as with this information, entrepreneurs can react and change their coffee shops according to their customers opinions and prosper in this business area.

INTRODUCTION: BUSINESS PROBLEM

In this research, the Foursquare Rating has been chosen as the measure of the customers' loyalty for the venues, which have placed their menus on the Foursquare Website.

The aim of this research project is to determine if the connection between Foursquare Rating (customers loyalty measure) and availability of the most popular items in a coffee shop menu exists (also the prices for these most popular items are being taken into consideration).

DATA SOURCES

1. Foursquare Rating is based on a number of signals that from social data mines (likes and dislikes, and positive versus negative tips). It's going to be used as the main measure of the project, as the goal is to find the connection between this metric and availability in the published menus the most popular of the coffee shop items.

2. Published on the Foursquare, coffee shop Menus with the items and their prices.

3. Foursquare Geospatial Data - latitude and longitude of the venues.

4. Some additional information: number of the coffee shop photos, number of guest tips, etc.

5. Wikipedia page with the list of the districts of London ordered by population density, based on population estimates for 2019.

https://en.wikipedia.org/wiki/List_of_English_districts_by_population_density

METHODOLOGY

Data Collection:

- Import of the Python libraries.
- Getting the list of the London districts from the Wikipedia.
- Foursquare API, getting the venues for the London districts from the previous step.
- Foursquare API, getting the data of every venue (venue's rating, URL, number of photos and tips, etc.).
- Web crawling, getting the menu (items and prices) for every venue.
- Data cleaning.
- Identification of the most popular items in the obtained menus.
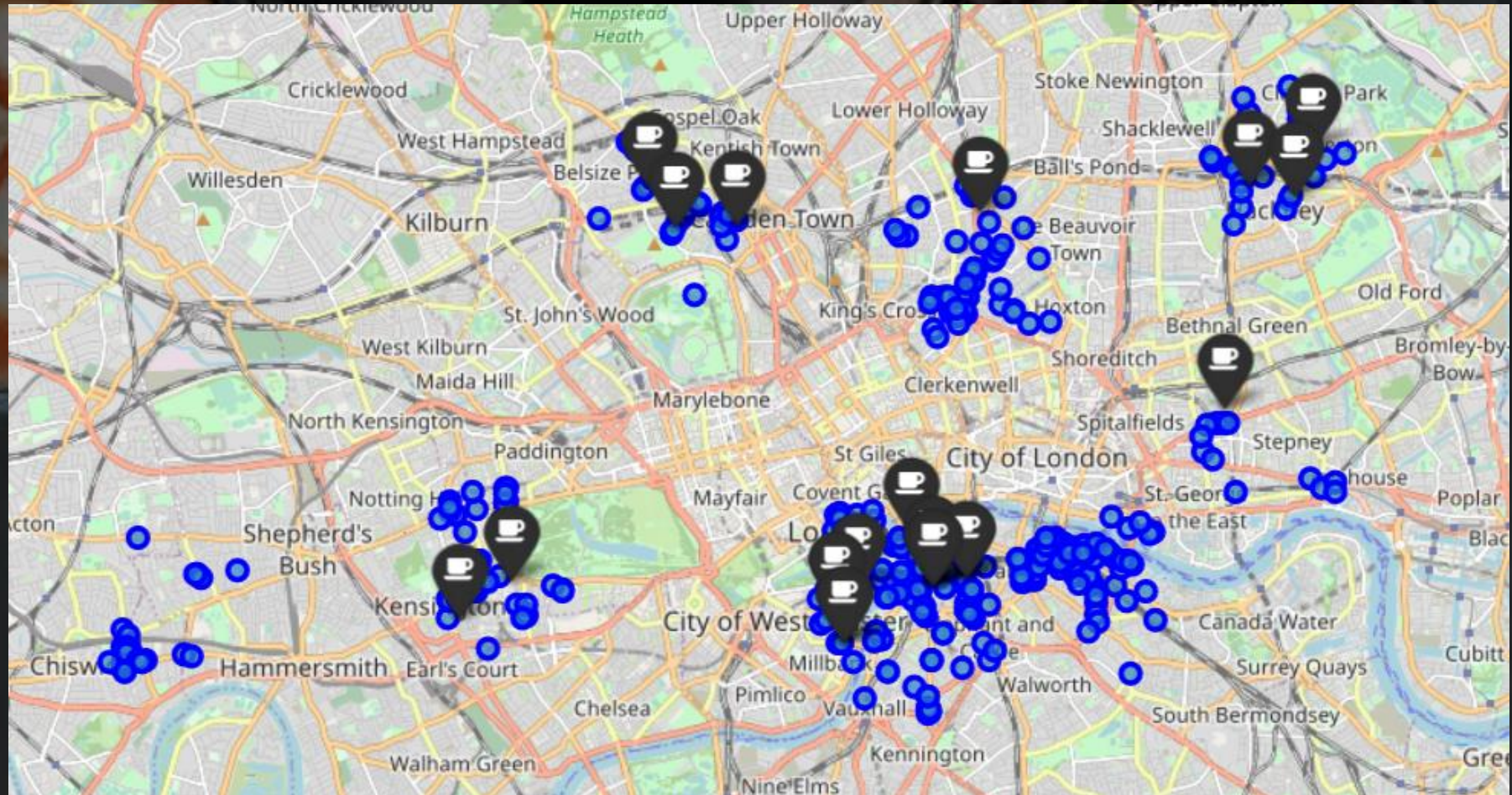- Representation on the map analyzed coffee shops.

Data Analysis:

- Data preparation.
- Linear Regression construction.
- Random Forest Regression construction.

THE MAP OF THE ANALYZED COFFEE SHOPS

Blue markers represent all the venues which have been found around neighborhood center.

Black icons represent venues with published menus, which will be analyzed.

RANDOM FOREST REGRESSION

Cappuccino, Macchiato, and Espresso have the highest influences on the coffee shop rating.

The higher, the more important the product. The importance of a product (feature) is computed as the Gini importance. Gini Importance or Mean Decrease in Impurity (MDI) calculates each feature importance as the sum over the number of splits (across all tress) that include the feature, proportionally to the number of samples it splits.
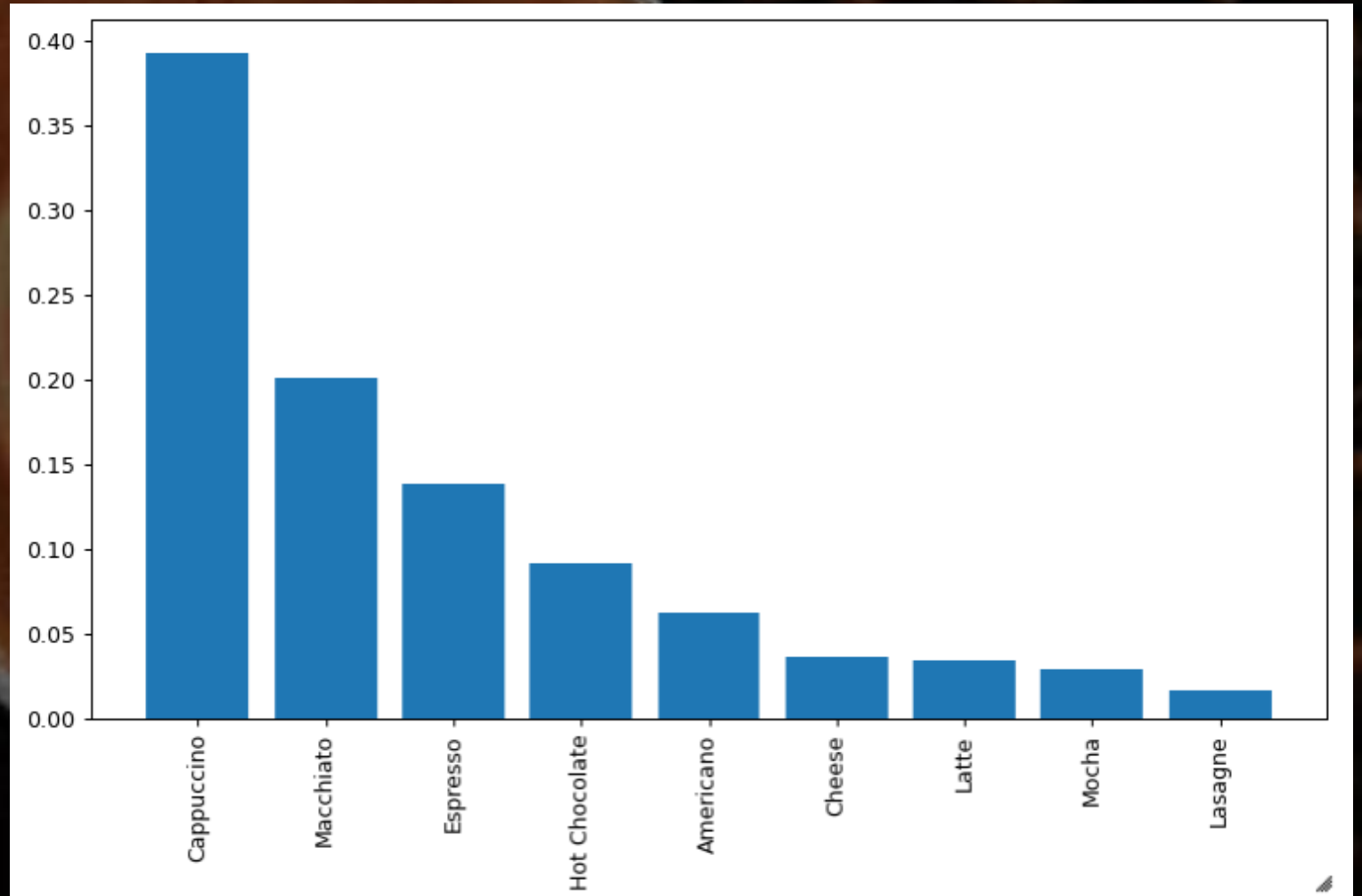
RESULTS, DISCUSSION AND CONCLUSION

Results:

• Cappuccino has the highest influence power on the venue's rating and guest's loyalty.

• Despite the fact, that Americano is the most popular item among the coffee shops, it is only on the fifth place of the products, ranked by importance.

Discussion:

• Clearly, before the research was obvious, that the coffee shops owners have to have available in café most popular coffee drinks, such as Cappuccino, Macchiato, Espresso, Hot Chocolate, and Americano. But which items are more important for their guests was undetermined.

• Now, entrepreneurs precisely know that the mentioned above 5 items are most important and their prices have an influence on the overall customer's rating. So, the recommendation for business owners is to keep the mentioned items on stock and carefully track their prices.'

Conclusion:

To sum up, in this research, we have determined the connection between venue's rating and availability of the most popular items (incl. their prices). We have identified the most important for coffee shop's rating items and numerically described their influence on the overall venue's ratings.