The Battle of Coffee Shops

Finding location for unique coffee shop in Warsaw

The Battle of Coffee Shops

Problem and background (1)

There are 2 ways to enter coffee shop market: you either buy franchise from a large chain or open an original coffeehouse.

- With franchise you'll get a fine tuned business model, brand and supply chain with prices negotiated for the large volumes
- With original coffeehouse you'll get an opportunity to make it from scratch

The Battle of Coffee Shops

Problem and background (2)

Interesting observation was made that a single location and small chain coffee shops co-exist with big chains.
They didn't disappear due to competition powered by the economy of scale.

There's the study showing that people are most likely looking for different experience when going to a big chain coffeehouse compared to a single location/small chain.

We'll be using data science tools to determine **how similar or dissimilar are spatial contexts for different types of coffee shops**. And **what type of coffee shop is the optimal choice for a given location represented by it's spatial context**.

The Battle of Coffee Shops

Data

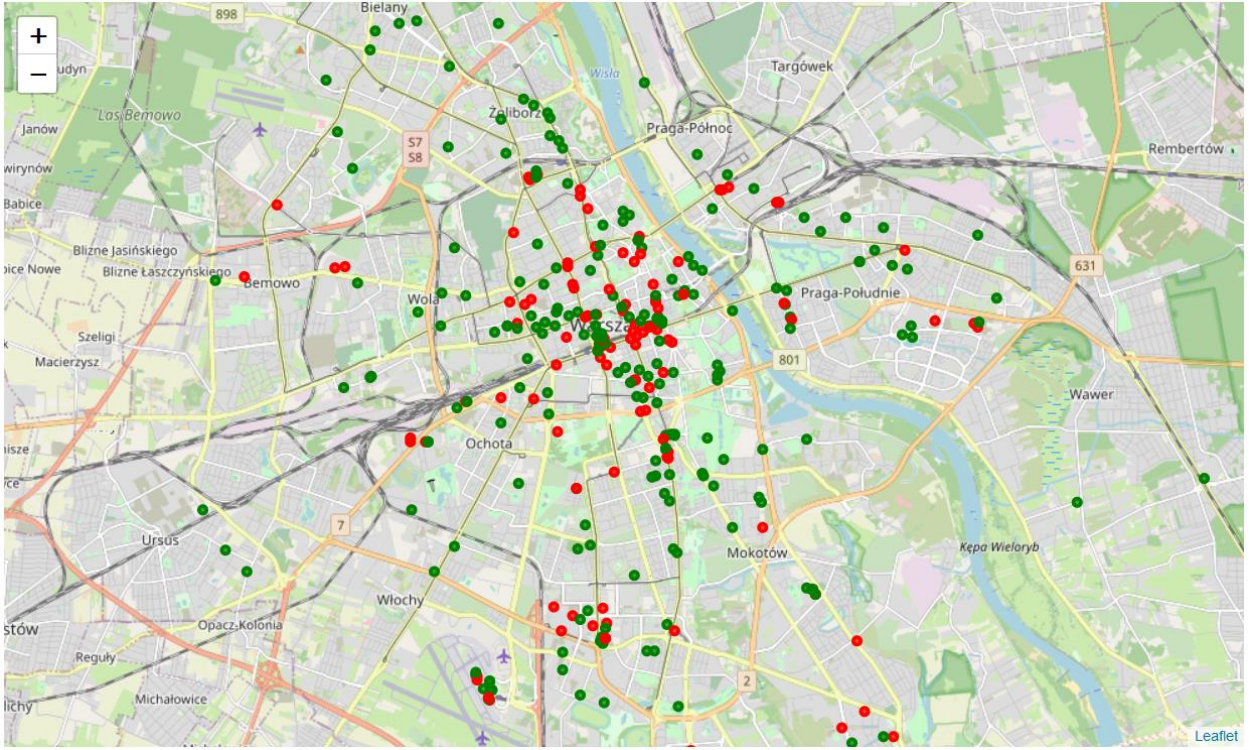
According to Tobler's first law of geography: "everything is related to everything else, but near things are more related than distant things". Following this idea will be doing study on spatial context of coffee shops in Warsaw.

Data sources:
Foursquare City Guide API will be used
to obtain coffee shops locations in Warsaw
to get spatial context for coffee shops - nearby venues
OpenStreetMap (OSM) data (obtained via Overpass API) will be used to enrich spatial context data by bringing more map features to our dataset
Geocoder Python library will help to get coordinates for Warsaw districts

The Battle of Coffee Shops

Main observation from data

There are areas where single location/small chain coffee shops (marked green on map) co-exist with big chains (marked red).



The Battle of Coffee Shops

Methodology

1)We'll manually classify coffee shops as below:

- Big chains (Costa, Starbucks, Nero) - class 0
- Single locations and small chains with big chains nearby - class 0
- All other single locations and small chains - class 1

2) We'll be using Logistic Regression classifier from scikit-learn - logistic regression will help to obtain actual probability values on top of predicted classes.

3) Finally we'll make use of our classifier to classify regular grid points in the part of Mokotów district of Warsaw. We'll visualize it using probability values on map to see final deliverable of this project.

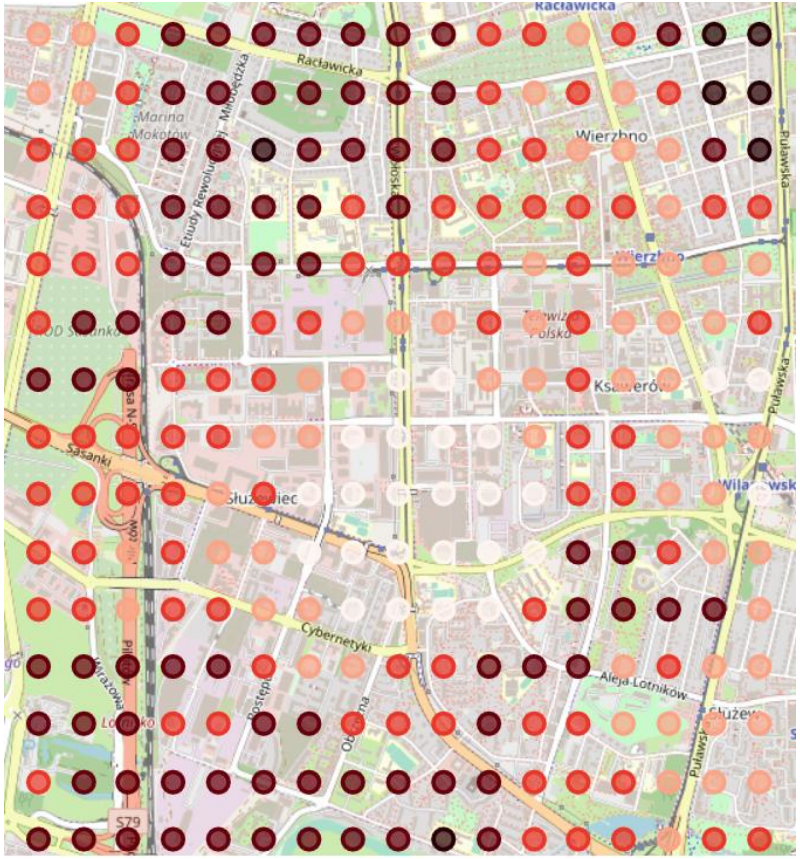
The Battle of Coffee Shops

Result

Darker point = more suitable point for unique coffee shop

Classified part of Mokotów district shows expected behavior:

- office and malls areas are not recommended as they are common location for big chain coffee shops
- living areas, park areas, leisure areas are recommended ones



The Battle of Coffee Shops

Conclusion

Areas of improvement for the same data are:

- optimization of the search radius for venues and OSM tags
- generalization of the model considering several cities of Poland

Obtained classifier most likely could be used as one contributing element of data science ensemble for selection proper location of unique coffee shop.

Model uses only documented OSM map features and Foursquare venues. And the main limitation is inability to separate single location coffee shops with big chains nearby.

It's expected that model could be improved with other datasets representing (for example) temporal features (people mobility, vehicles mobility).