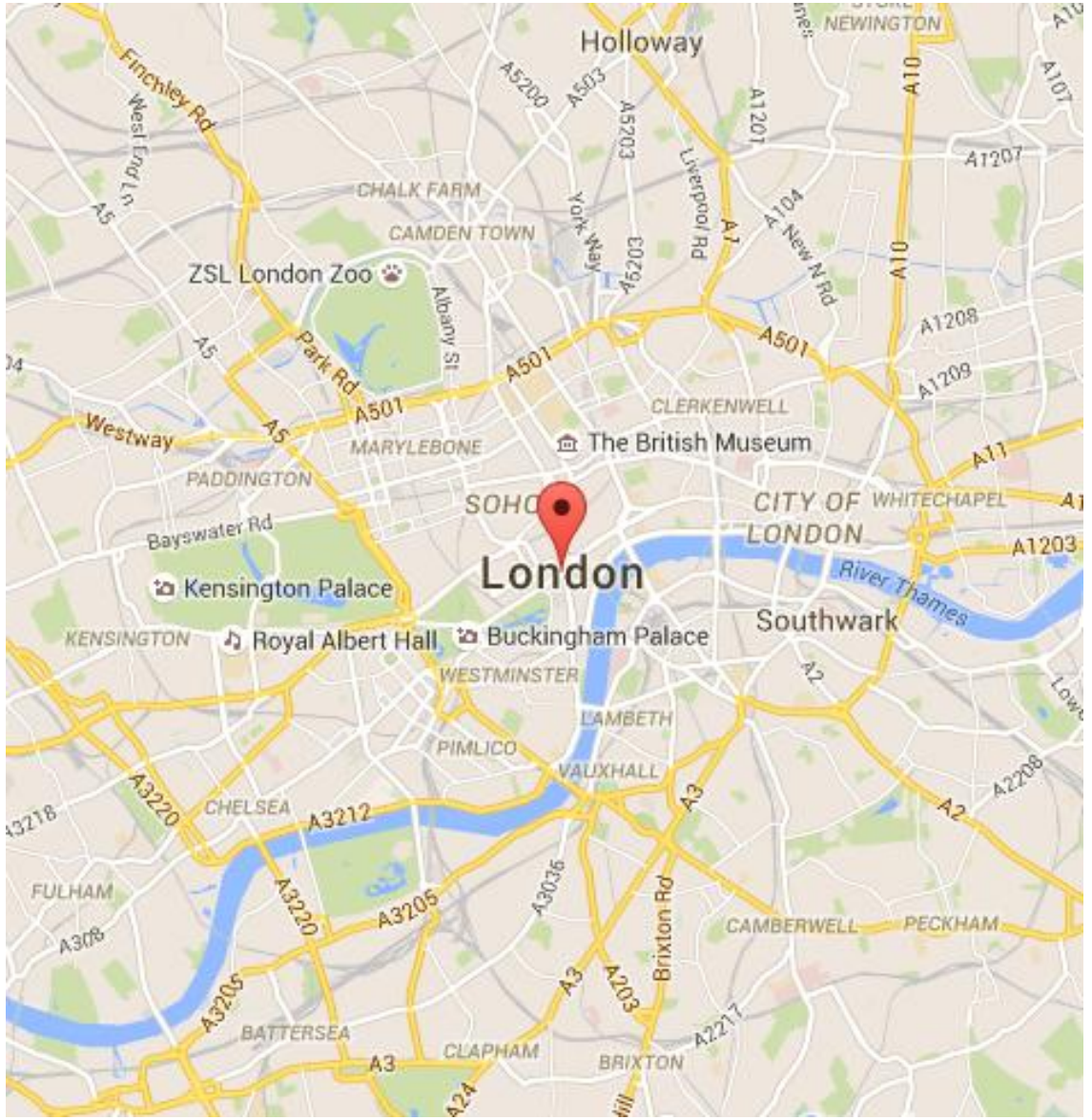
Analysis of Real Estate Market through the lenses of Geo-tagged Social Media Data

Goals:

- House price class predictions (by regions)
- Important features selections

Inputs:

- Foursquare check-ins
  https://figshare.com/articles/Foursquare_amp_Flickr_activities_in_20_cities/1584973
- House price information for London
  http://landregistry.data.gov.uk/



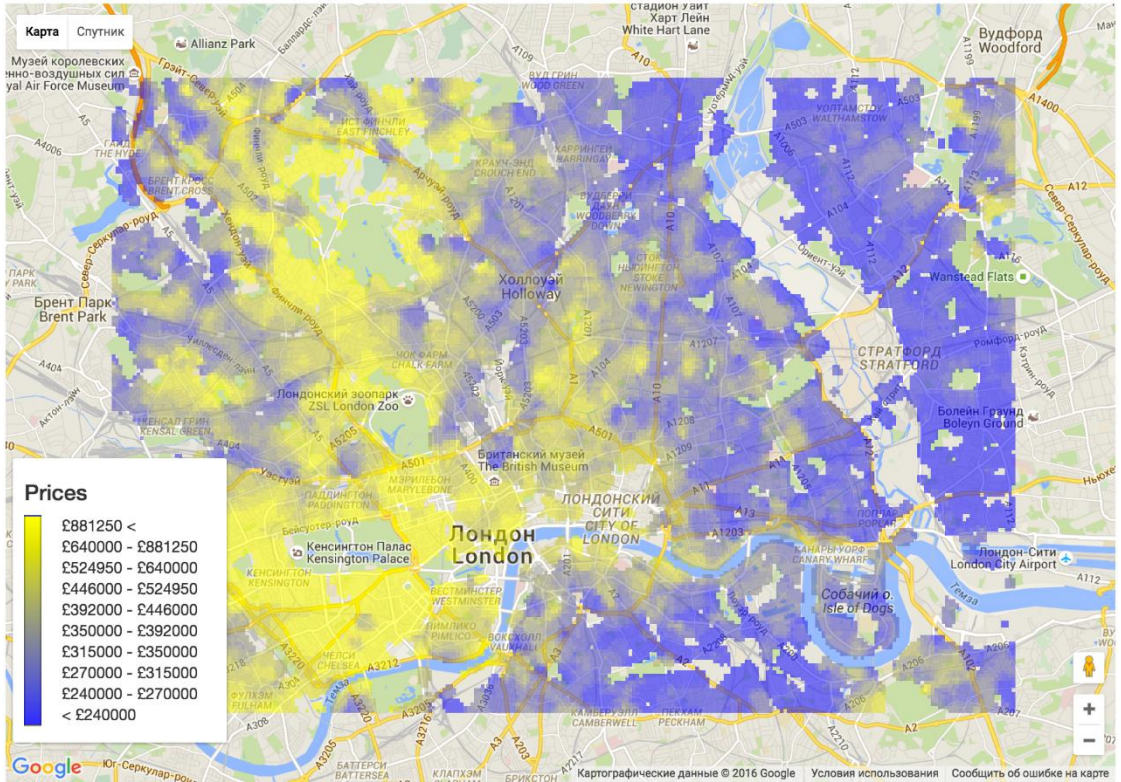
Used tools and approaches:

Analyzing data: Python
For preparing data : R/Python
Algorithm: xgboost, t-sne
Visualization: JavaScript

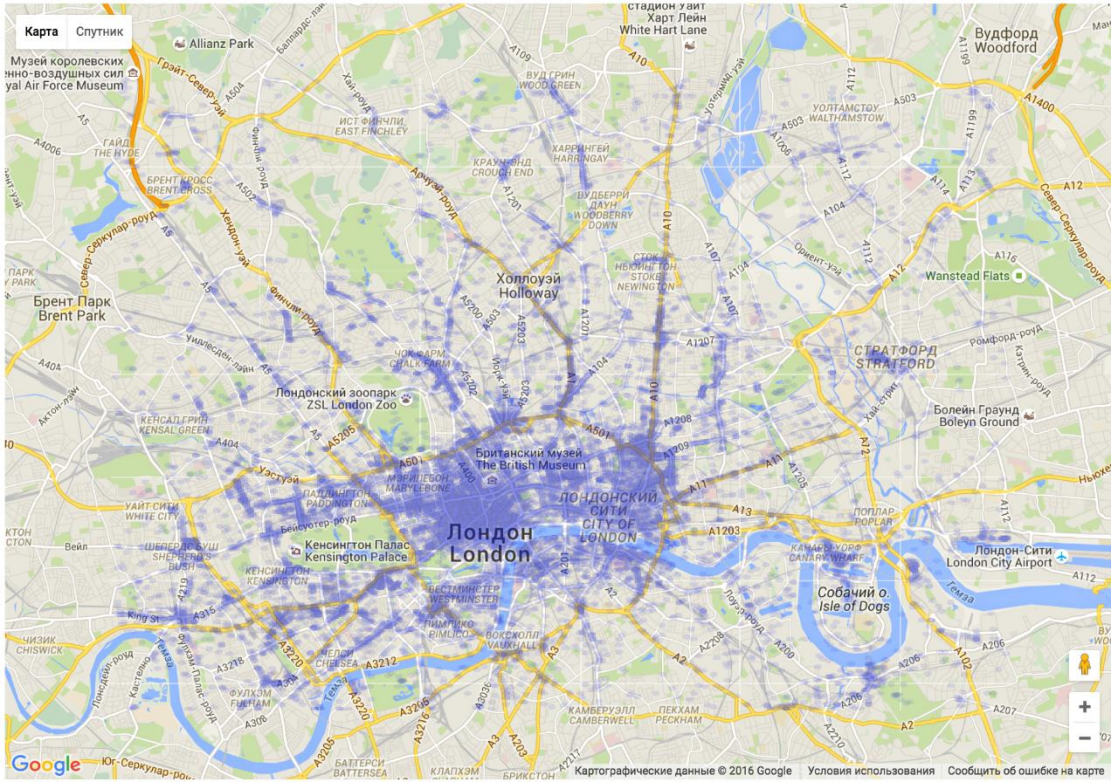
London map was divided by 75x75 meters grid.

Divided prices into 10 clusters (ignored errors when region appeared in nearby cluster).

Distribution of prices:

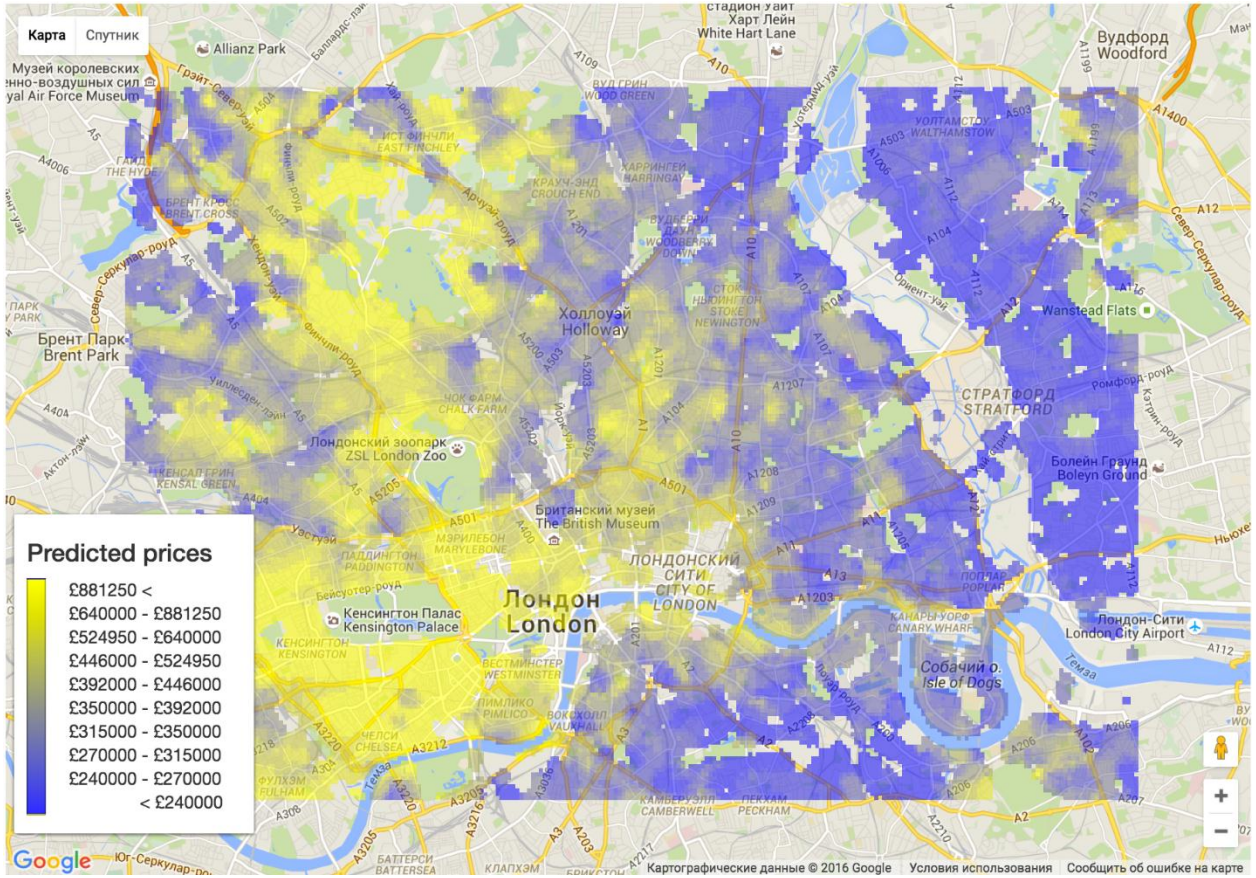


Distribution of check-ins:

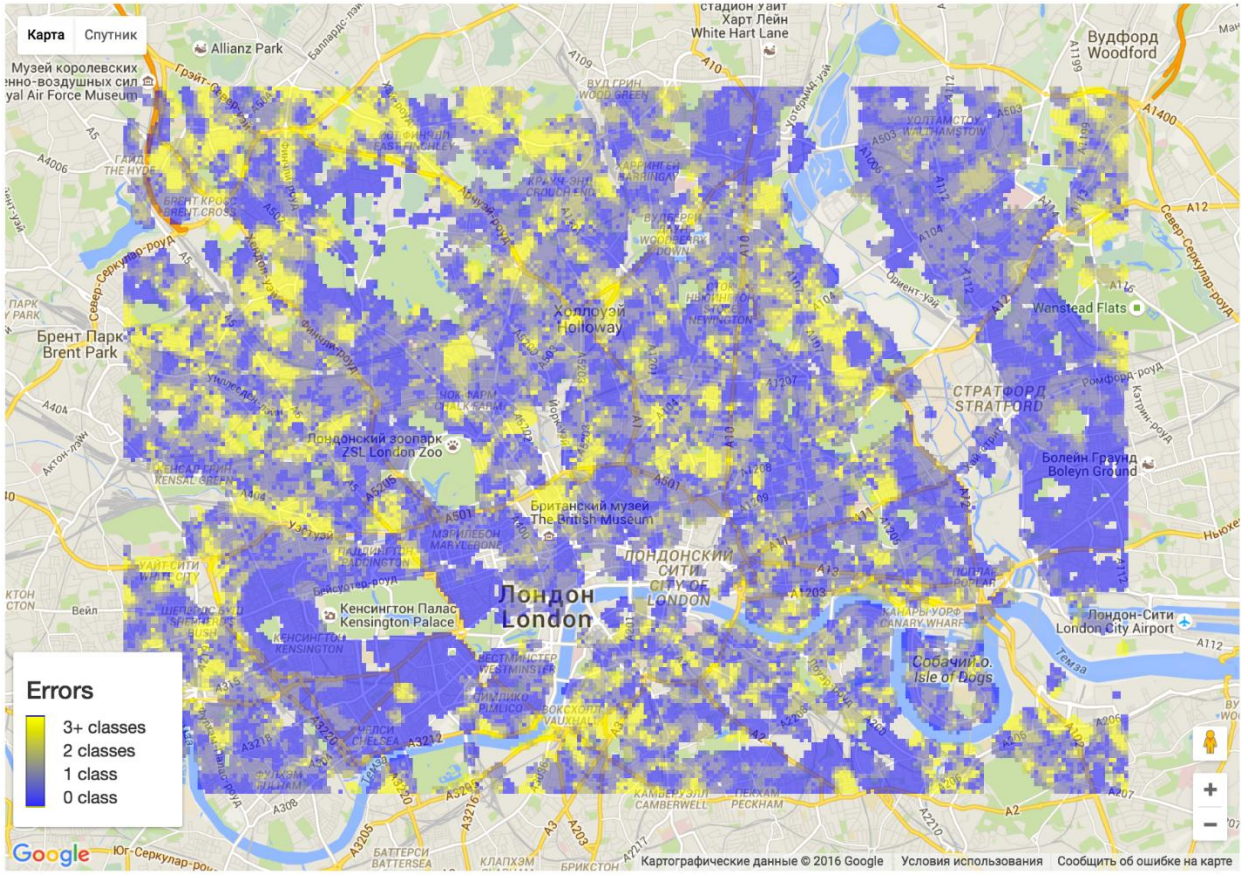


Final predicted map:

Prediction:



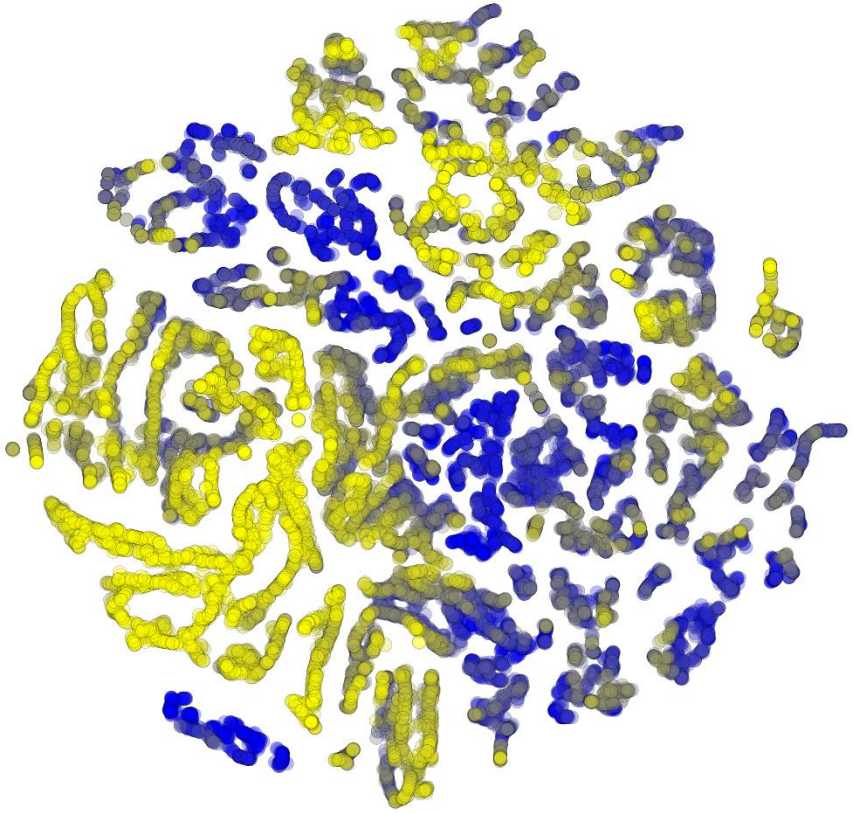
Errors in prediction:



http://recommender.mazurik.me/

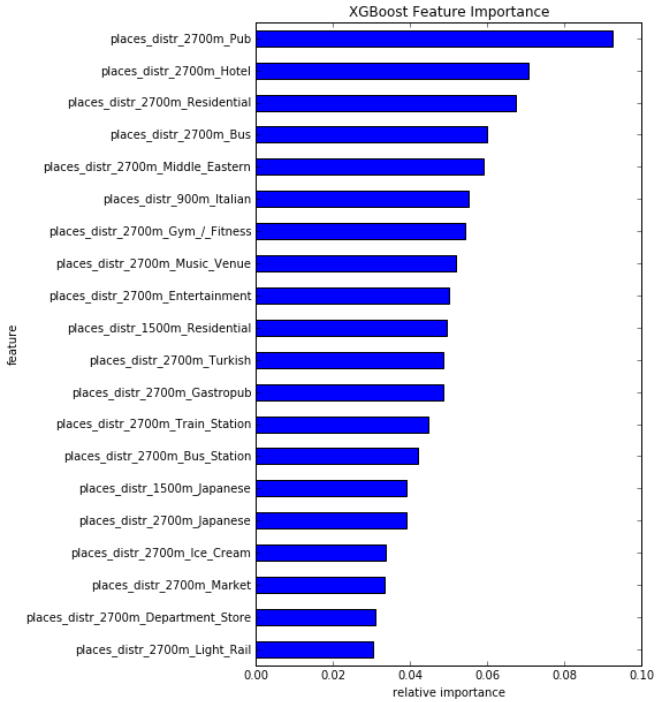
Feature selections

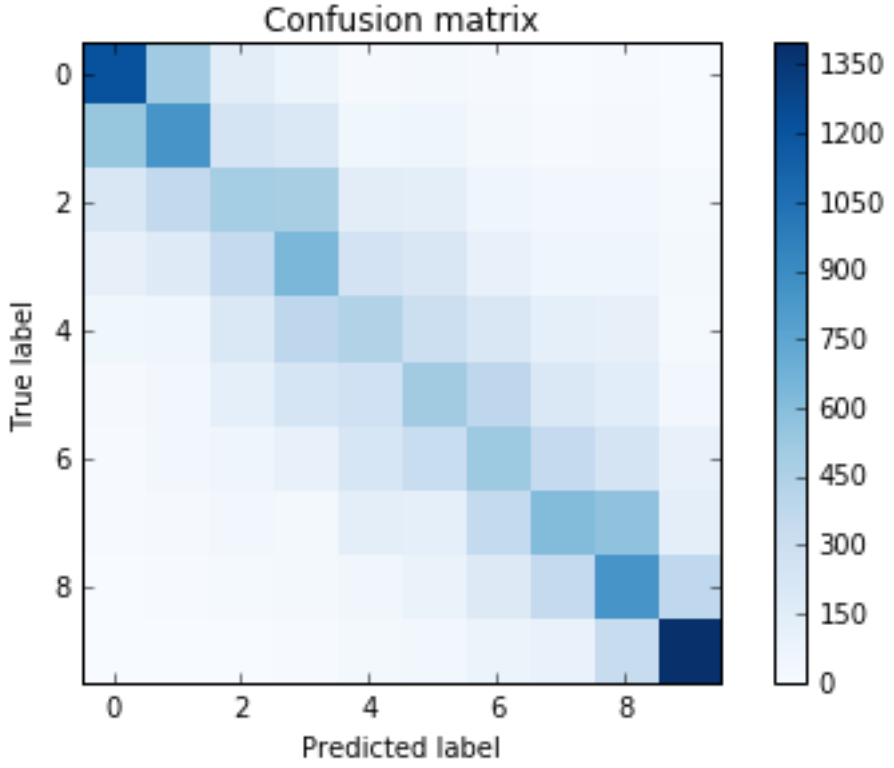
t-distributed stochastic neighbor embedding



Accuracy (top 2 classes): 74%

feature importance





Future work:

- Applying same model for other cities.
- Create recommendation where is better to build new pubs ☺
- Use people mobility as feature
- Predict price in future

Our Team

Vitaliy Radchenko

project mentor:

Dr. Dmytro Karamshuk

Vitalii Moshkivskyi

Aleksey Mazurik

Igor Tymchuk