The battle of Neighborhoods

Capstone Project at Coursera Applied Data Science

(by Kostyantyn Baranov, 20th of September 2019)

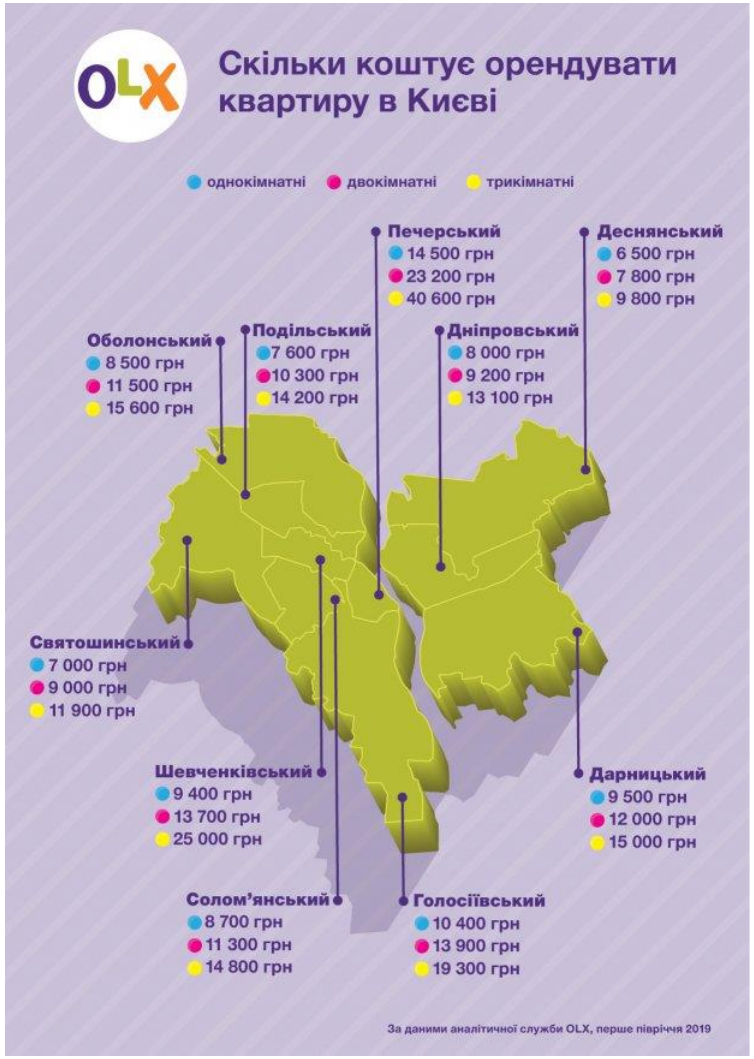
Table of contents and introduction

Table of contents

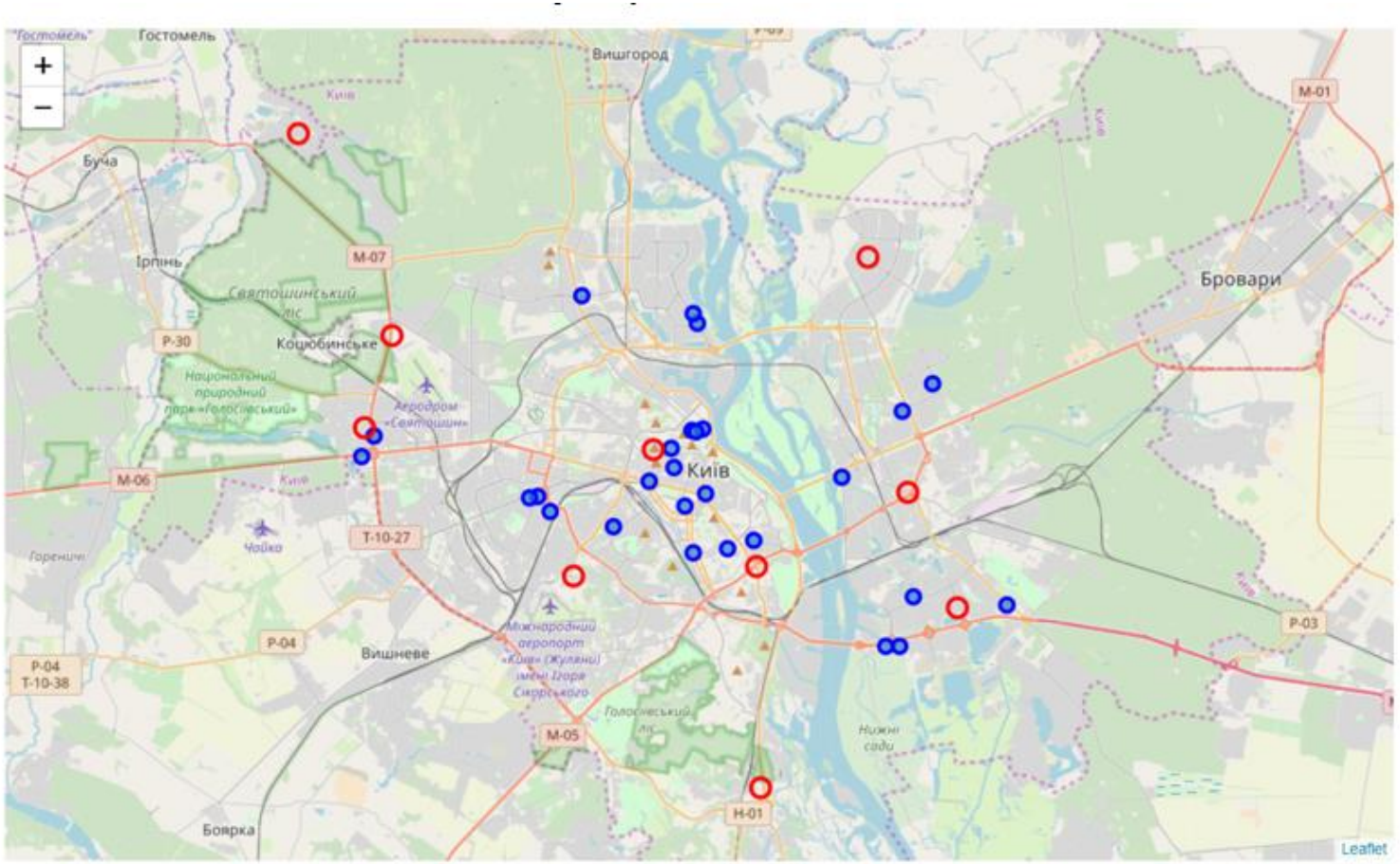
Starting such business as robotics club is not hard - couple of franchise offers is readily available. But the main challenge that would define success of this business is **choosing proper place to start** - that is where Foursquare and data analysis come into force.

Testing data with information on rent rate

	District	Lat	Lon	Rent
0	Darnicky	50.406387	30.648453	12000
1	Desnansky	50.517948	30.604093	7800
2	Dniprovsky	50.443046	30.623970	9200
3	Holoseevsky	50.349179	30.550377	13900
4	Obolonsky	50.557136	30.319664	11500



Train data (red circels) and test data (alternatives - blue circles)



Methodology of analysis

Step 1. Data preparation

Step 2. Checking correlation matrices

Step 3. Training Multiple Linear Regression Model

Step 4. Calculating predicted values of venue count

Step 5. Calculating min distance to existing rival

Results

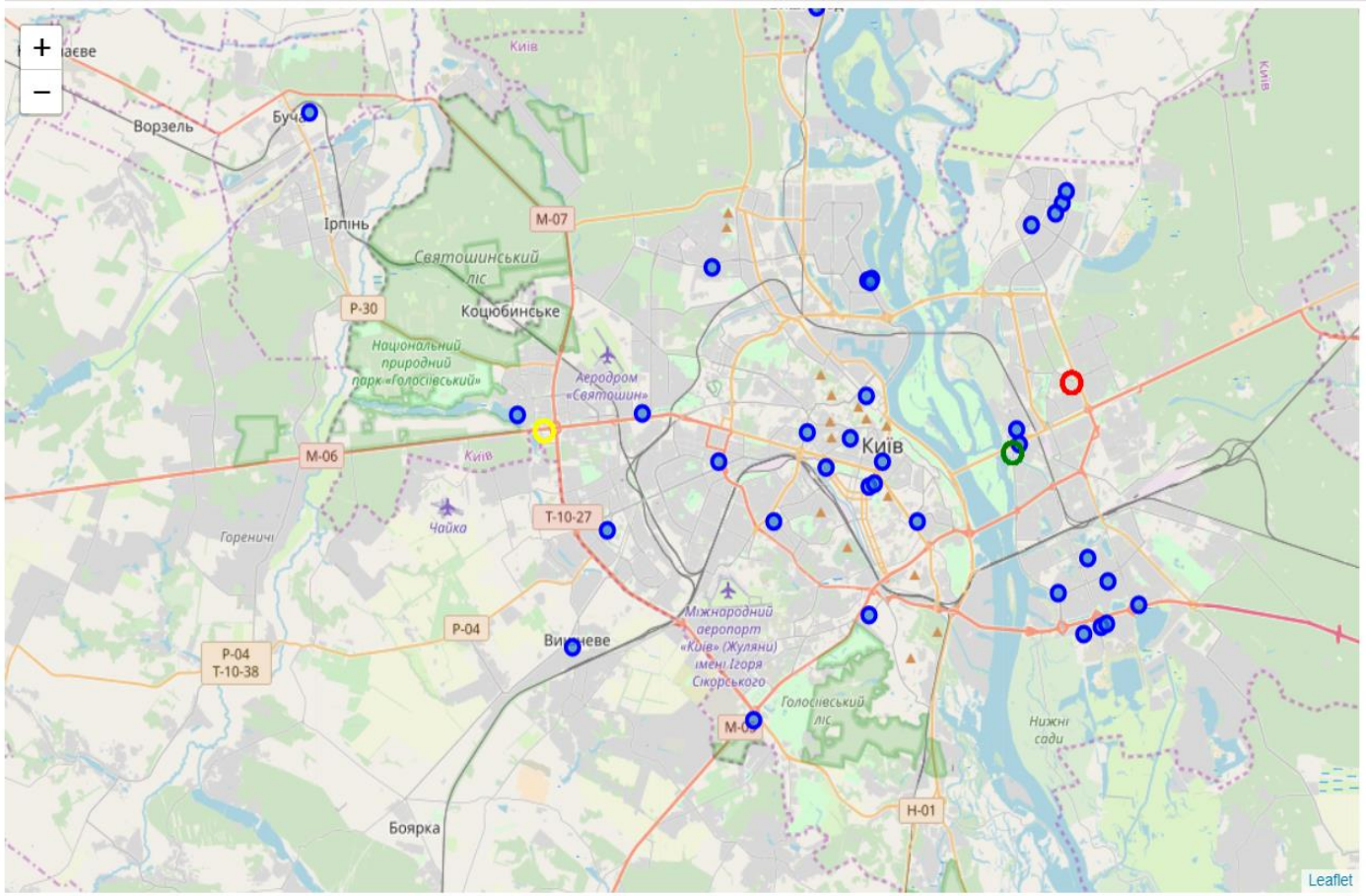
Best three alternatives (red, yellow and green) and existing rivals.

Red alternative is the best on multiple parameters:

-distance to rivals

-entertainment venues

-rent rate



What used for what

What used

- Folium
- Foursquare
- Linear regression model
- Math module and functions def

What received

- Presentation of locations on map
- Necessary data on venues around
- Let obtain expected number of venues and choose alternative with the
- Distance between alternatives

Final thoughts

Work on this project was motivated not only by studying but also had great practical use for own business expansion

Python with additional libraries like Pandas, Folium etc. proved to be a greatly universal and powerful tool even for beginners.

I am pretty sure (while passed this myself with own club) that such analysis is rarely performed prior or after starting of business, while this might save good bunch of money.

There are still some question left regarding choice of initial points (testing data should be larger) but that would be out of scope of this project. Best alternative found and it happed to be really best one.

Thank you for attention

And good luck with your projects