Coursera Capstone

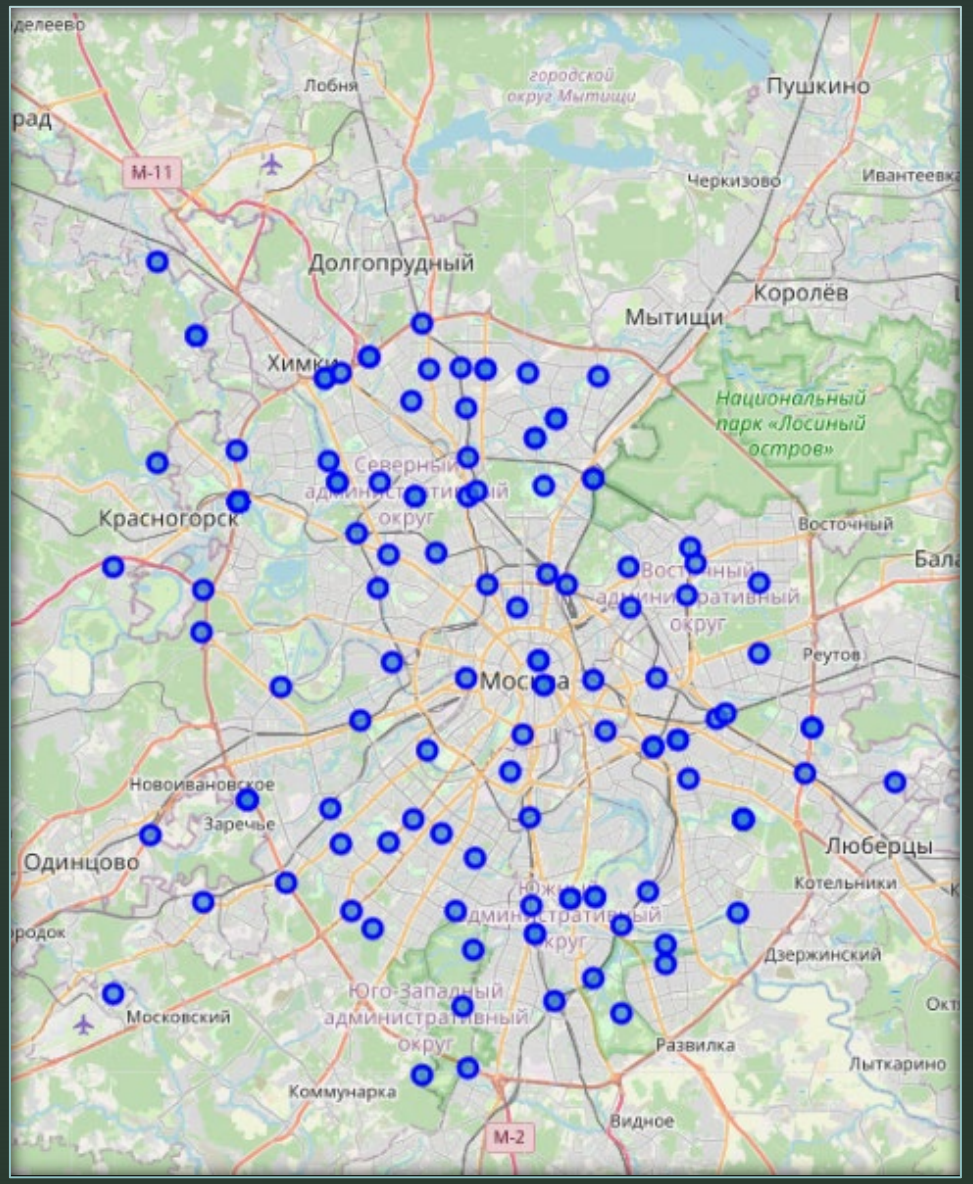
Finding the best district in Moscow for a new coffee shop

Introduction/Business Problem

- The enterpreneur is looking for a place to open a coffee shop in Moscow. The main purpose of the project is to determine which one of Moscow's 146 districs is better considering its population and existing venues (using Foursquare).

Data section

- The task requires several measurements: data on population in various districts of Moscow and data on venues located in these districts. For the latter part of the task we are gonna be using data from Foursquare, which requires us to set latitute and longitude for each district's center.
- Data on administrative regions in this project is used from GIS-lab source on Moscow districts and their coordinates https://gis-lab.info/qa/moscow-atd.html As some of data includes polygon coordinates, in order to find centre point we will need to do some modifications to the dataset.
- Population data will be used from statdata webpage http://www.statdata.ru/naselenie-moskvy-po-okrugam-i-rajonam As this data doesn't include OKATO codes to match datasets, this data was modified and uploaded separately.
- The main purpose of the data analysis and KPI for detecting the best district to open a new coffee shop will be assessed using population to density of coffee shops ratio, as it shows the district with the most people and least coffee shops. As there may be additional business costs to open a shop in a district, we will find TOP-5 best districts.

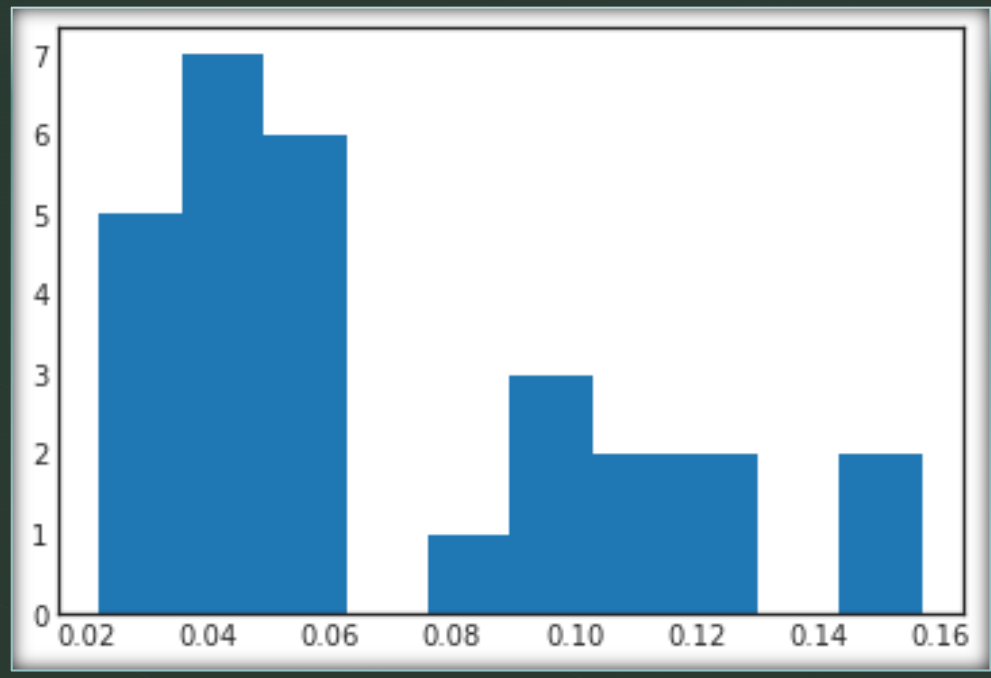


Map of districts

- In the project data was gathered using sources as described above. A centroid of each district was found using standard Python Libraries. The dataset for analysis
- In order to assess data a map of Moscow with all centroids of districts was depicted using pyplot.

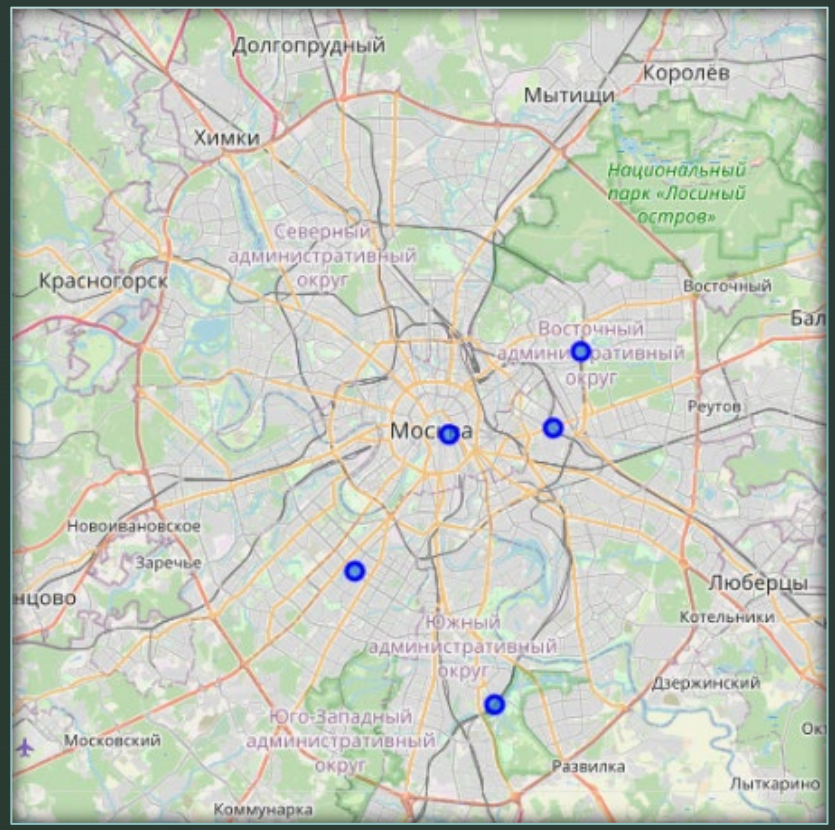
Data exploration and analysis

- There were 274 unique categories of venues after gathering data from Foursquare, but the problem required to use only coffee shops' density, so the next step was to find it for each district. Histogram plot of coffee shops' density:



Result

- Based on population and coffee shop density a Ratio was calculated to define top 5 neighborhoods for considering.



	Neighborhood	POPULATION	Coffee Shop	Ratio
0	Орехово-Борисово Северное	132307	0.031250	4233824.0
1	Перово	141269	0.035714	3955532.0
2	Академический	110038	0.033333	3301140.0
3	Измайлово	107883	0.033333	3236490.0
4	Таганский	123532	0.039216	3150066.0