

Moscow Pub chain business analytics



accomplished by Timur Taumurzaev

Business problem definition: pub chain foundation

What are the basic factors affecting business performance?

- Customers
- Competitors
- Expenses

Either of them can lead to success or failure.

All three components depend on Location.

Data acquisition: location is the key to success

How to assess the influence of factors?

Components:

Solutions:

- Customer location preferences
 - Public transport availability
 - Distance from city center
- Competitors → Number of competitors nearby
- Expenses → Commercial rent costs

Data acquisition: subway stations as objects of consideration

Dataset building and feature selection. Three sources of data. Four features.

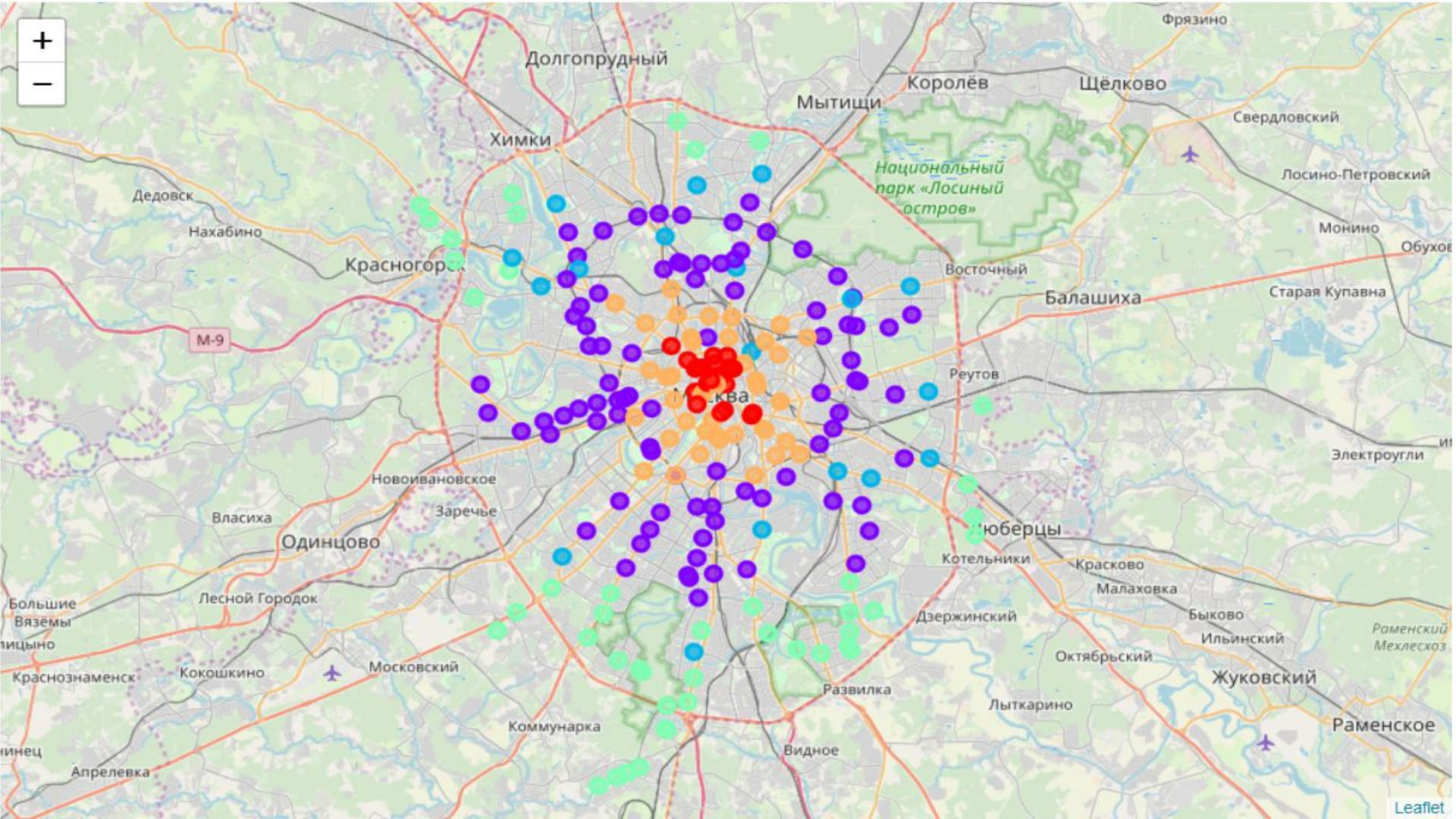
- Passenger flow - scrapped from mosmetro.ru
- Commercial rent cost nearby - scrapped from cian.ru
- Distance from the city center - calculated with [GeoPy](https://pypi.org/project/GeoPy/) library
- Number of competitors nearby - data obtained with [Foursquare API](https://foursquare.com/developers/)

The resulting dataset is build for 214 subway stations and 478 venues of competitors nearby within 700 meter radius.

Methodology

Machine Learning algorithm selection and analysis.

- Partitioning stations into five groups with k-means clustering algorithm.
 - Observing and wrangling data to search the most suitable cluster, according to stakeholders requirements through averaging values and visualizing it with area plot to visibly estimate clusters.
-

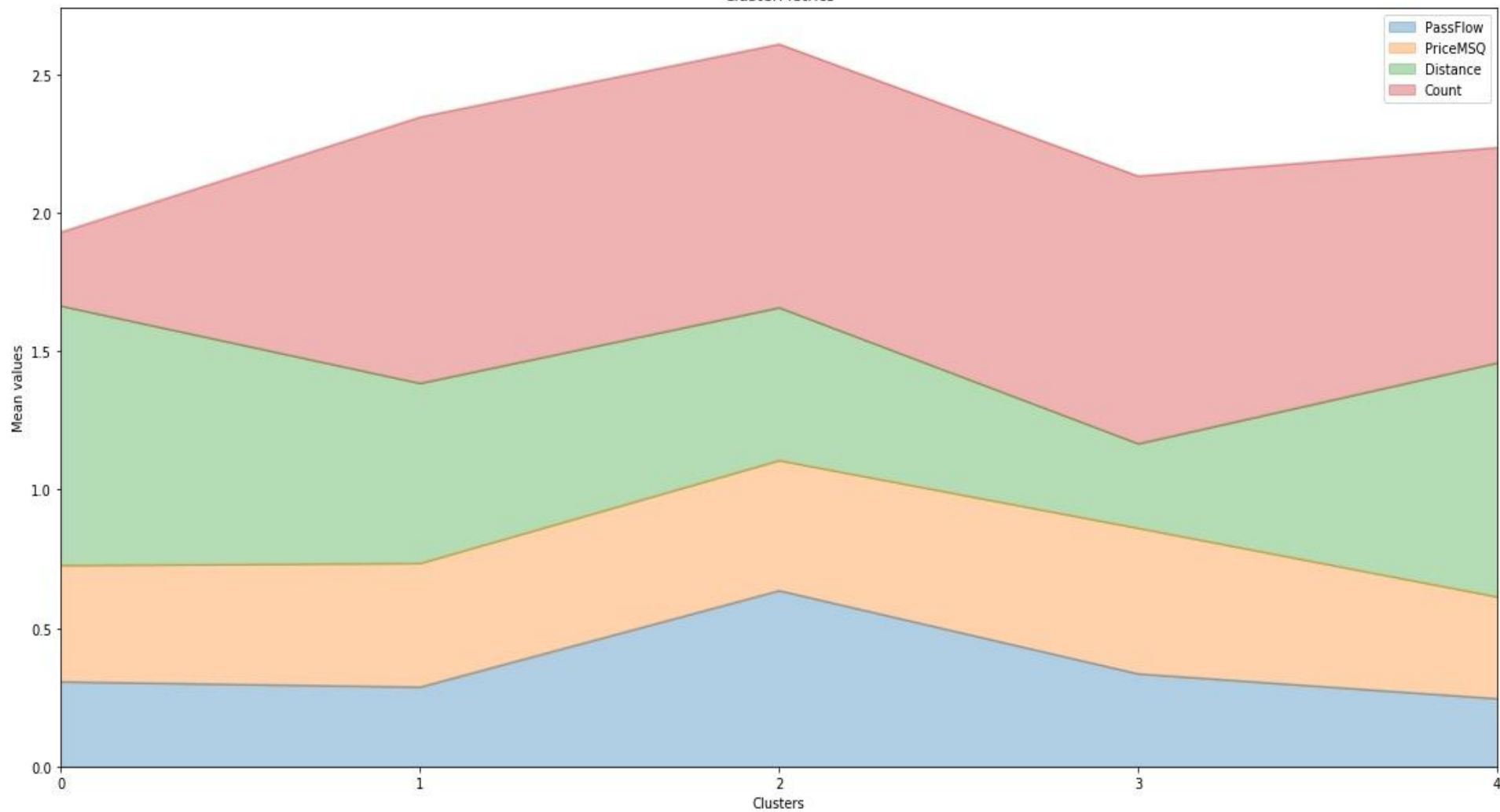


ML results with average feature values through each cluster

Data was normalized and standardized to $[0, 1]$ interval, where 0 corresponds to low quality feature and 1 to high quality

| Number of cluster | Passenger Flow | Rent cost | Distance | Competitors | Total result |
|-------------------|----------------|-----------|----------|-------------|--------------|
| 0 | 0.305 | 0.420 | 0.936 | 0.266 | 1.929 |
| 1 | 0.286 | 0.447 | 0.650 | 0.961 | 2.345 |
| 2 | 0.634 | 0.470 | 0.551 | 0.951 | 2.609 |
| 3 | 0.334 | 0.526 | 0.304 | 0.966 | 2.142 |
| 4 | 0.244 | 0.367 | 0.845 | 0.777 | 2.235 |

ClusterMetrics



Cluster 2

Cluster 2 has the highest average values summarized. It also has the highest values of 'Count', 'Passenger Flow' and worthy values of 'Distance' and 'Rent cost '. So we would strongly recommend stakeholder to pick locations from this cluster.

| Station | PassFlow | Rent cost | Distance | Count | Total |
|-----------------------|----------|-----------|----------|----------|----------|
| Выхино | 0.943880 | 0.564970 | 0.476871 | 1.000000 | 2.985721 |
| Комсомольская | 1.000000 | 0.238607 | 0.875619 | 0.866667 | 2.980892 |
| Текстильщики | 0.634955 | 0.490114 | 0.654738 | 1.000000 | 2.779807 |
| Петровско-Разумовская | 0.620524 | 0.501950 | 0.614195 | 1.000000 | 2.736669 |
| Новогиреево | 0.609300 | 0.544118 | 0.510839 | 1.000000 | 2.664257 |

Conclusion

Purpose of this project was to identify Moscow subway stations close to center with low number of pubs and bars nearby, high passenger flow and low commercial rent cost, in order to aid stakeholders in narrowing down the search for optimal location for a pub chain.

The final decision about the appropriate stations to open pubs nearby is made according to stakeholder requirements.

**Thank you for your
attention!**

the project by Timur Taumurzaev