

Battle of the Neighborhoods in Vilnius – the Case of New Gas Station

Aistė Kisielienė, January 10, 2020



Case of New Gas Station – Business Problem



How to find the **best neighborhood**
for a new gas station in Vilnius?



Best neighborhood with **highest traffic**
volume and **lowest number of gas stations**

Case of New Gas Station – Data Collection



Location data of 105 gas stations in Vilnius – <https://foursquare.com/>

- number of gas stations in each neighborhood



Data of 21 neighborhoods in Vilnius – <https://lt.wikipedia.org/wiki/Vilnius>

- number and names of neighborhoods
- coordinates of neighborhoods,
- population in each neighborhood



Data of traffic volume in each neighborhood – <https://github.com/vilnius/traffic>

- 50 traffic junctions in Vilnius with high traffic volume
- number of traffic junctions with high traffic volume in each neighborhood

Case of New Gas Station – Data Preparation

- 4 neighborhoods with no intensive traffic junctions and no data – data added
- Calculated values:
 - **number of gas stations per resident in each neighborhood** -
number of gas stations in each neighborhood divided by
population in each neighborhood.
 - **number of gas stations per one high traffic volume junction** –
number of gas stations in each neighborhood divided from
number of junctions with high traffic volume.



Case of New Gas Station – Data Preparation

- 3 neighborhoods with smallest number of gas stations per resident:
 - Naujoji Vilnia
 - Karoliniškės
 - Naujamiestis
- 3 neighborhoods with smallest number of gas stations per intensive traffic junction:
 - Naujamiestis
 - Šnipiškės
 - Vilkipėdė



Case of New Gas Station – Data Modeling

- **Data normalization** before modeling
- Clustering model - **kmeans**
- To find the best k in kmeans the **Silhouette coefficient** was calculated for **different models**.

Number of clusters	Value of Silhouette Coefficient
2	0.5313
3	0.4893
4	0.3659
5	0.4757
6	0.4242
7	0.4370
8	0.4637
9	0.4431
10	0.4343
11	0.3950



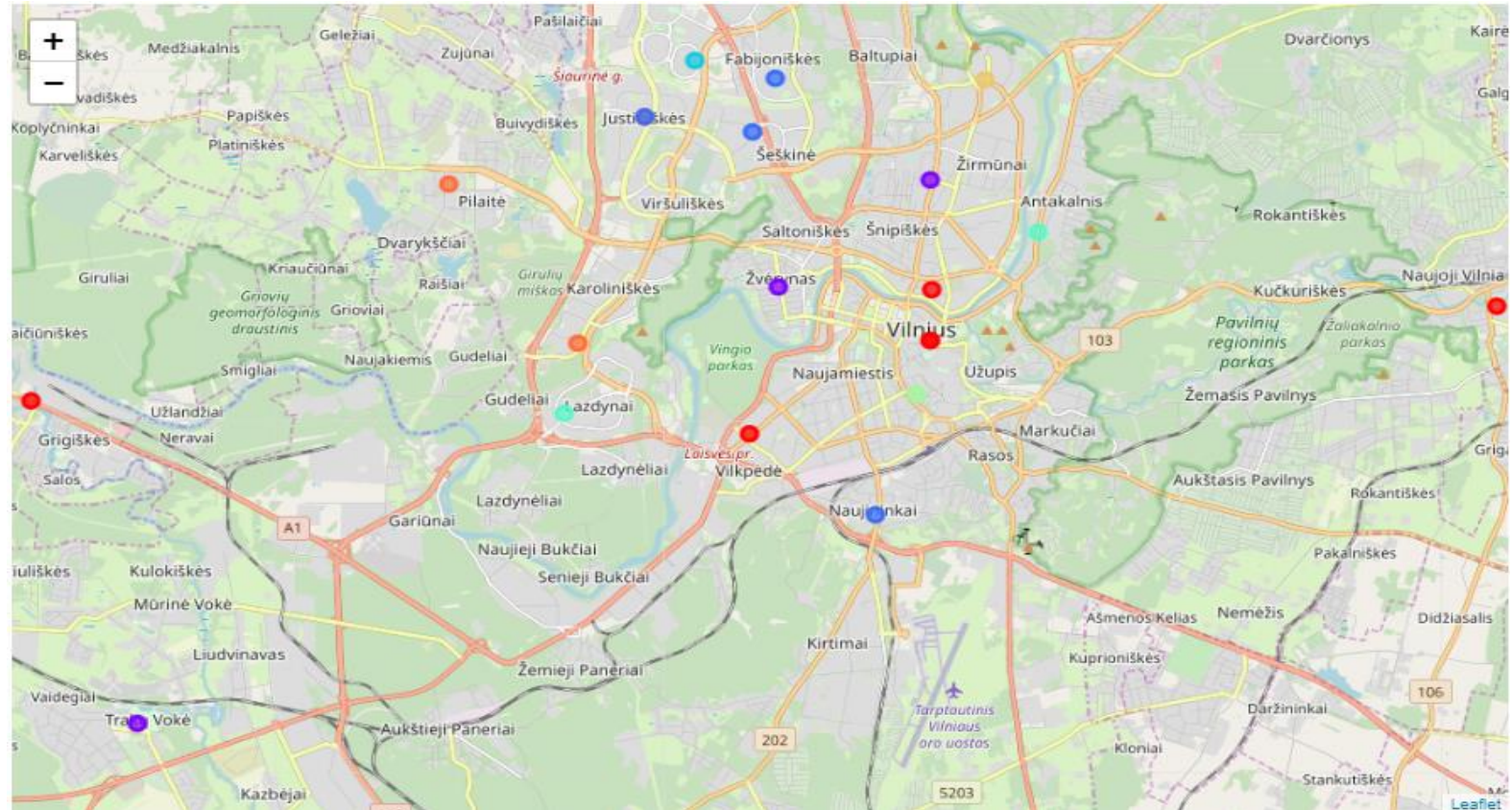
Case of New Gas Station – Clustering results

Clusters quantity **8** – average values

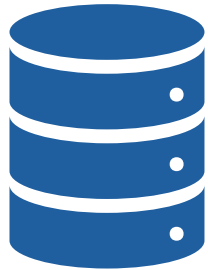


	Population	Latitude	Longitude	Count_gas_stations	Gas_station_per_person	Count_junctions	Gas_station_per_junction
Cluster Labels							
0	22896.000000	54.683866	25.265431	2.000000	0.000093	3.500000	0.733333
1	17317.666667	54.677369	25.213417	8.000000	0.000465	2.000000	3.888889
2	30738.750000	54.704344	25.246685	6.750000	0.000222	3.250000	2.062500
3	25674.000000	54.725942	25.231328	8.000000	0.000312	1.000000	8.000000
4	35930.500000	54.688059	25.255481	4.000000	0.000113	1.000000	4.000000
5	13054.000000	54.677718	25.281702	5.000000	0.000383	4.000000	1.250000
6	47410.000000	54.723249	25.297213	13.000000	0.000274	3.000000	4.333333
7	26043.000000	54.703721	25.200398	2.666667	0.000106	1.333333	2.000000

Case of New Gas Station – Clusters map



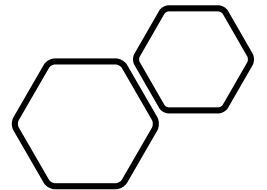
Case of New Gas Station – Clustering results



	Neighborhoods	Count_gas_stations	Gas_station_per_person	Count_junctions	Gas_station_per_junction
8	Šnipiškės	3	0.000155	6	0.5
12	Grigiškės	1	0.000086	1	1.0
14	Vilkipėdė	3	0.000121	6	0.5
15	Naujamiestis	2	0.000072	5	0.4
16	Senamiestis	2	0.000095	2	1.0
17	Naujoji Vilnia	1	0.000031	1	1.0

- **Cluster number 0**

- average value of **gas stations per person** is **0.000093** and average value of **gas stations per junction** is **0.733333**
- Silhouette coefficient - **0.4637**





Case of New Gas Station – Conclusions

- **6 neighborhoods in cluster 0:** Šnipiškės, Grigiškės, Vilkipedė, Naujamiestis, Senamiestis and Naujoji Vilnia.
- Naujoji Vilnia and Grigiškės should be crossed out because these neighborhoods were among 4 neighborhoods with no intensive traffic junctions' data and data was added.
- Senamiestis should be crossed out because of restrictions for building and a lot of protected heritage.
- **3 neighborhood to consider:** Šnipiškės, Vilkipedė and Naujamiestis.
- **First choice - Naujamiestis** with lowest number of gas stations per person and lowest number of gas stations per junction