# **Housing Sales Prices & Venues Data Analysis of Almaty**

**Alexey Startsev**

**September 12, 2020**

## 1. A description of the problem and a discussion of the background.

Almaty is the largest city in Kazakhstan, with a population of about 2.000.000 people (about 11% of the country's total population) and it has a population density of 2.636 people per square kilometer. As a resident of this city, I decided to use Almaty in my project. The city is divided into 8 districts in total. Almaty generates approximately 20 per cent of Kazakhstan's GDP. The nation is the most powerful economically in Central Asia and Almaty is a key financial center. The economy of Almaty city and Almaty Region continues to grow, and is expected to increase by nearly 6.5 percent per year. As you can see from the figures, Almaty is a city with a high population density and it has a big investment potential.

From the investors point of view, we expect from them to prefer the districts where there is a lower real estate cost, high population density and the type of business they want to install is less intense. And from the city resident point of view, it's mostly valuable to chose districts where real estate values are lower and the social places density is high.

However, it is difficult to obtain information that will guide investors in this direction, nowadays. When we consider all these problems, we can create a map and information chart where the real estate index is placed on Almaty and each district is clustered according to the venue density.

## 2. A description of the data and how it will be used to solve the problem.

To solve the problem I will use below data:

* I found the population of Almaty districts data from <https://journals.openedition.org/belgeo/28865?lang=en>, cleaned it and uploaded to Github repository
* I used Forsquare API to get the most common venues of given Borough of Almaty.
* There are not too many public datas related to demographic and social parameters for the city of Almaty. Therefor you must set-up your own data tables in most cases. In this case, I collected latest per square meter Housing Sales Price (HSP) Averages for each Borough of Almaty from housing retail web page (<https://krisha.kz/content/analytics>).
* I used Google Map, ‘Search Nearby’ option to get the center coordinates of the each Borough.

## 3. Methodology section

### As a database, I used GitHub repository in my study. My master data which has the main components Borough, Average House square meter Price, Latitude and Longitude informations of the city:

|  | **Borough** | **Avg-House-SQR\_M-Price-USD** | **Latitude** | **Longtitude** | **Population** | **Density-of-population-per-square\_km** |
| --- | --- | --- | --- | --- | --- | --- |
| **0** | Alatau district | 874 | 43.278479 | 76.849038 | 195348 | 11152 |
| **1** | Almaly district | 1048 | 43.252580 | 76.912461 | 204879 | 1877 |
| **2** | Auezov district | 929 | 43.229568 | 76.837504 | 277338 | 11754 |
| **3** | Bostandyk district | 1172 | 43.164182 | 76.929771 | 316020 | 3179 |
| **4** | Medeu district | 1337 | 43.147694 | 77.033535 | 188899 | 746 |
| **5** | Nauryzbay district | 841 | 43.172516 | 76.837540 | 91172 | 1308 |
| **6** | Turksib district | 778 | 43.339788 | 76.985542 | 209789 | 2783 |
| **7** | Jetysu district | 836 | 43.281011 | 76.952791 | 157999 | 4001 |

### I used python folium library to visualize geographic details of Istanbul and its boroughs and I created a map of Almaty with boroughs superimposed on top. I used latitude and longitude values to get the visual as below:

### Изображение выглядит как текст, карта Автоматически созданное описание

### I utilized the Foursquare API to explore the boroughs and segment them. I designed the limit as 100 venue and the radius 3000 meter for each borough from their given latitude and longitude informations. Here is a head of the list Venues name, category, latitude and longitude informations from Forsquare API.

Изображение выглядит как снимок экрана

Автоматически созданное описание

### In summary of this data 311 venues were returned by Foursquare. Here is a head of merged table of boroughs and venues.

Изображение выглядит как снимок экрана

Автоматически созданное описание

### 122 unique categories were returned by Foursquare, then I created a table which shows list of top 10 venue category for each borough in below table.

Изображение выглядит как снимок экрана

Автоматически созданное описание

### We have some common venue categories in boroughs. Due to this reason I used unsupervised learning K-means algorithm to cluster the boroughs. K-Means algorithm is one of the most common cluster method of unsupervised learning. First, I will run K-Means to cluster the boroughs into 5 clusters. Here is my merged table with cluster labels for each borough.

### Изображение выглядит как снимок экрана Автоматически созданное описание

## 4. Results section

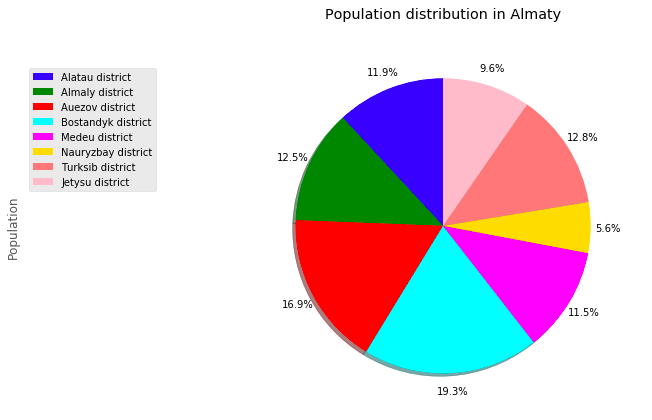
### You can see a clustered map boroughs of Almaty in the below.

Изображение выглядит как текст, карта

Автоматически созданное описание

### Also here is a several dependency charts:

### 1. Population distribution in Almaty



### 2. Avg House Price and Population Density Dependancy

Изображение выглядит как снимок экрана

Автоматически созданное описание

## 5. Discussion section

As I mentioned before, Almaty is a big city with a high population density in a narrow area. The total number of measurements and population densities of the 8 districts in total can vary. As there is such a complexity, very different approaches can be tried in clustering and classification studies.

I used the Kmeans algorithm as part of this clustering study. For more detailed and accurate guidance, the data set can be expanded and the details of the neighborhood or street can also be drilled.

I also performed data analysis through this information by adding the coordinates of districts and home sales price averages as static data on GitHub.

I ended the study by visualizing the data and clustering information on the Almaty map. It could be used by city government and investors.

## 6. Conclusion section

As a result, people are turning to big cities to start a business or work. For this reason, people can achieve better outcomes through their access to the platforms where such information is provided.

Not only for investors but also city managers can manage the city more regularly by using similar data analysis types or platforms.

Thanks, Alexey Startsev

## 7. References section

1.<https://en.wikipedia.org/wiki/Almaty>

2.<https://journals.openedition.org/belgeo/28865?lang=en>

3.<https://developer.foursquare.com/>

4.<https://krisha.kz/content/analytics>

5.<https://www.google.com/maps/place/Almaty/@43.2178605,76.6639808,10z/data=!3m1!4b1!4m5!3m4!1s0x38836e7d16c5cbab:0x3d44668fad986d76!8m2!3d43.2220146!4d76.8512485>

6.<https://www.linkedin.com/pulse/housing-sales-prices-venues-data-analysis-ofistanbul-sercan-y%C4%B1ld%C4%B1z/>