Research of schools abundance along new district of Almaty city

Problem description and background discussion

- **Problem description**

As every city nowadays, Almaty grows rapidly, new districts are to be built to accommodate new citizens. However, the main purpose of city's authority is not only to give people new houses but to maintain **facilities and infrastructure**.
For new families it is important to understand if there are enough schools for their children and is these schools are available to reach them in **acceptable time period** (30 minutes). Therefore, we can assume that schools need to be built in the same district area as houses.

- **Background discussion**

As an example of newly build district, Algabas district will be considered.
It was built 5 years ago and is still growing. Houses are not so expensive comparing to some other districts. Therefore, a lot of new families tend to buy appartments there. The most part of appartments were bought by young people who's children are going to school in the next couple of years.
It will be helpful for city's authorities to understand **what part of the district has not enough schools or which houses are too far away from schools**.

Data description and problem solving

- **Data description**

GoogleMaps API will be used to obtain and collect location data of schools in the district. After receiving the information in JSON it will be transformed into appropriate views.

- **Problem solving**

To help the authorities with decisions about new schools building, data will be clustered depending on their location. Map view will show parts of districts with lack of schools.

Data Acquisition and Understanding

Sample of data about districts received by Google API:


```
{
   "candidates" : [
      {
         "geometry" : {
            "location" : {
               "lat" : 43.2547398,
               "lng" : 76.7985669
            },
            "viewport" : {
               "northeast" : {
                  "lat" : 43.2727997,
                  "lng" : 76.80520539999999
               },
               "southwest" : {
                  "lat" : 43.2392942,
                  "lng" : 76.788168
               }
            }
         }
      }
   ],
   "status" : "OK“
}
```


These information is necessary for searching schools around the district

Sample of data describing schools


```
{
   "html_attributions" : [],
   "results" : [
      {
         "geometry" : {
            "location" : {
               "lat" : 43.2498531,
               "lng" : 76.8145555
            },
            "viewport" : {
               "northeast" : {
                  "lat" : 43.2511624302915,
                  "lng" : 76.81564868029152
               },
               "southwest" : {
                  "lat" : 43.2484644697085,
                  "lng" : 76.8129507197085
               }
            }
         },
         "icon" :
"https://maps.gstatic.com/mapfiles/place_api/icons/school-71.png",
…

}
```


These data helps to find what parts of district has schools

Clustering preparation

- To cluster the district depending on schools data should be transformed into dataframe as follows:

	geometry.location.lat	geometry.location.lng	geometry.viewport.northeast.lat	geometry.viewport.northeast.lng	geometry.viewport.southwest.lat	geometry.viewport.southwest.lng	icon
0	43.249853	76.814555	43.251162	76.815649	43.248464	76.812951	https://maps.gstatic.com/mapfiles/place_api/ic...
1	43.247521	76.800370	43.248861	76.801648	43.246163	76.798950	https://maps.gstatic.com/mapfiles/place_api/ic...
2	43.250591	76.784459	43.251891	76.785572	43.249193	76.782874	https://maps.gstatic.com/mapfiles/place_api/ic...
3	43.249814	76.812681	43.251139	76.813915	43.248441	76.811217	https://maps.gstatic.com/mapfiles/place_api/ic...
4	43.244706	76.789699	43.246292	76.791290	43.243594	76.788592	https://maps.gstatic.com/mapfiles/place_api/ic...

Clustering information

- After preparing data the K-means algorithm is used to divide all the districts' schools into 3 clusters:

	Cluster Labels	geometry.location.lat	geometry.location.lng	geometry.viewport.northeast.lat	geometry.viewport.northeast.lng	geometry.viewport.southwest.lat	geometry.viewport.southwest.lng	icon
0	1	43.249853	76.814555	43.251162	76.815649	43.248464	76.812951	https://maps.gstatic.com/mapfiles/place_api/ic...
1	0	43.247521	76.800370	43.248861	76.801648	43.246163	76.798950	https://maps.gstatic.com/mapfiles/place_api/ic...
2	0	43.250591	76.784459	43.251891	76.785572	43.249193	76.782874	https://maps.gstatic.com/mapfiles/place_api/ic...
3	1	43.249814	76.812681	43.251139	76.813915	43.248441	76.811217	https://maps.gstatic.com/mapfiles/place_api/ic...
4	0	43.244706	76.789699	43.246292	76.791290	43.243594	76.788592	https://maps.gstatic.com/mapfiles/place_api/ic...

Final results

- After drawing 3 clusters onto map we can resume that:

1) There is enough schools in northern and eastern parts of the districts
2) City authorities should think about building new school in the south-west part of the district