WEEK5 Coursera Capstone Project - The Battle of Neighborhoods
Segmenting and Clustering Hotels of ALMATY city

Alfiya Kazanbassova

September, 2019

Table of Contents

1. INTRODUCTION

- **Almaty** is the largest city in Kazakhstan and it is the cultural and financial center of the country. The landscape of the city goes over the mountains where the nature views are awesome and of course there many venues for sightseeing. But despite of above features and cosmopolitanism the city is just on their first steps towards to the *world popular touristic destination*.
- This project will help potential tourists to find hotels for different types of trips based on *popular venues* analysis.

2. DATA

- For exploring the common venues of Almaty city **Foursquare location data** will be used as a source.
- As this report will be used by potential tourists the venues will be explored based on the Almaty city hotels data. The hotels data also can be explored from Foursquare location data.
- Let's assume the Hotels as the neighborhoods of Almaty.
- **Dostyk Street** will be the starting point of hotels and venues exploring. The above street is the center of the city.

3. METHODOLOGY

Most of the tourists before starting their trips to metropolises try to investigate the hotels in the center of location. And then they search the hotels which are more comfortable for their trip goals like short trip, business trip, entertain, etc.

So let's do the same in this research.

3.1. Used Python Libraries

- ○ **NUMPY** - library to handle data in a vectorized manner
- ○ **PANDAS** - library for data analsysis
- ○ **JSON** - library to handle JSON files
- ○ **GEOPY - NOMINANTUM** - convert an address into latitude and longitude values
- ○ **REQUESTS** - library to handle requests
- ○ **PANDAS.IO.JSON - JSON_NORMALIZE** - transform JSON into a pandas dataframe
- ○ **MATPLOTLIB.CM, MATPLOTLIB.COLORS** - plotting modules
- ○ **SKLEARN.CLUSTER - KMEANS** - kmeans from clustering
- ○ **FOLIUM** - map rendering library

3.2. Data Exploring

- ○ The starting point of the research will be the center location of the Almaty city. Let's define it as the Dostyk Street. As it is mentioned before the Foursquare API is used in all steps of this research.
- ○ After defining the Foursquare credentials the central location point should be converted to latitude and longitude coordinates.
- ○ The next step is to find hotels in radius of 2 km of the starting point.
- ○ Examine the row results and convert them to pandas dataframe with necessary formatting and visualizing (see *Picture 1*).
- ○ Next let's start to explore the hotels by venues. Firstly, only one hotel is explored to see preliminary view of the data. Then the function is used to explore all hotels in our selection.
- ○ All venues are distributed by hotels in another dataframe and grouped by mean of frequency of occurrence of each venue category.
- ○ Finally on the exploration stage new dataframe is created - 10 common venues defined for each hotel.
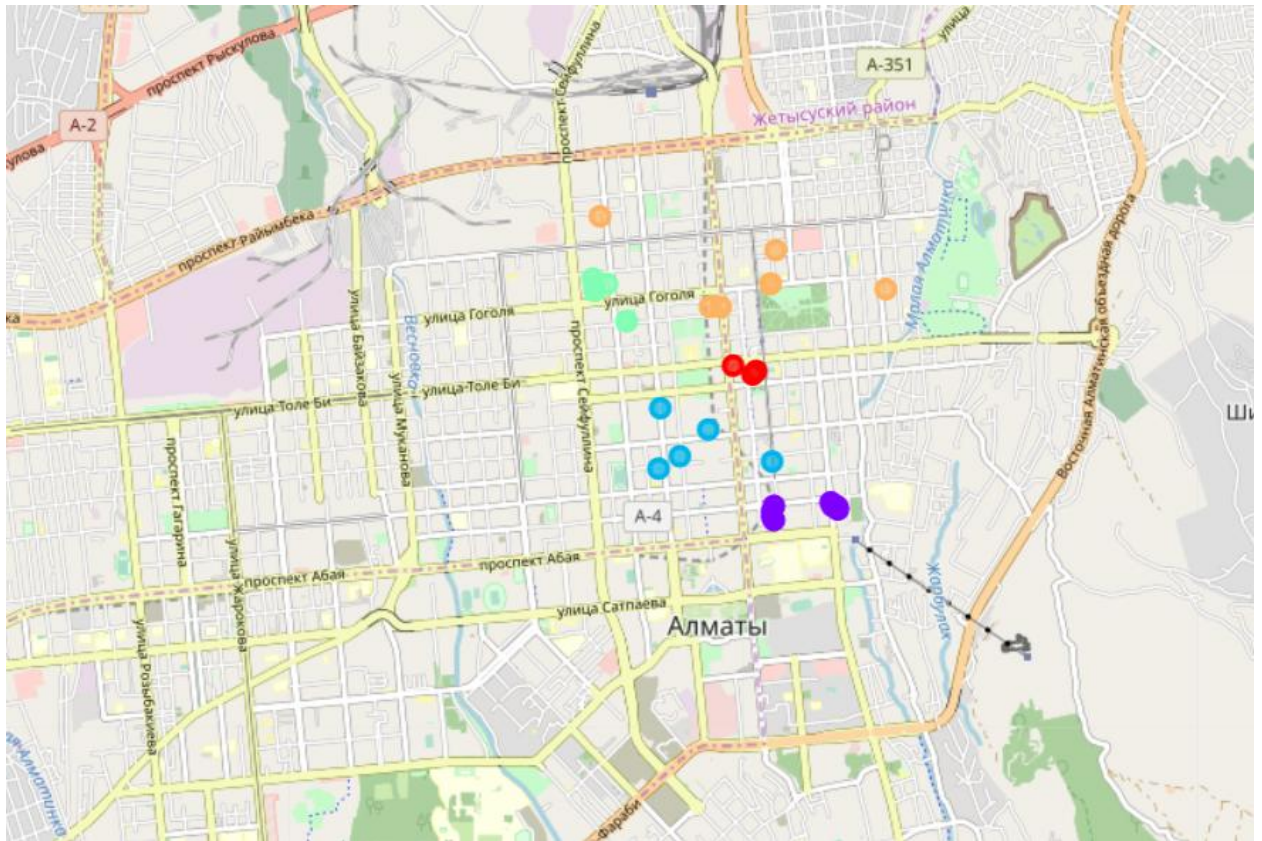
Picture 1. Hotels near starting point (radius 2 km)
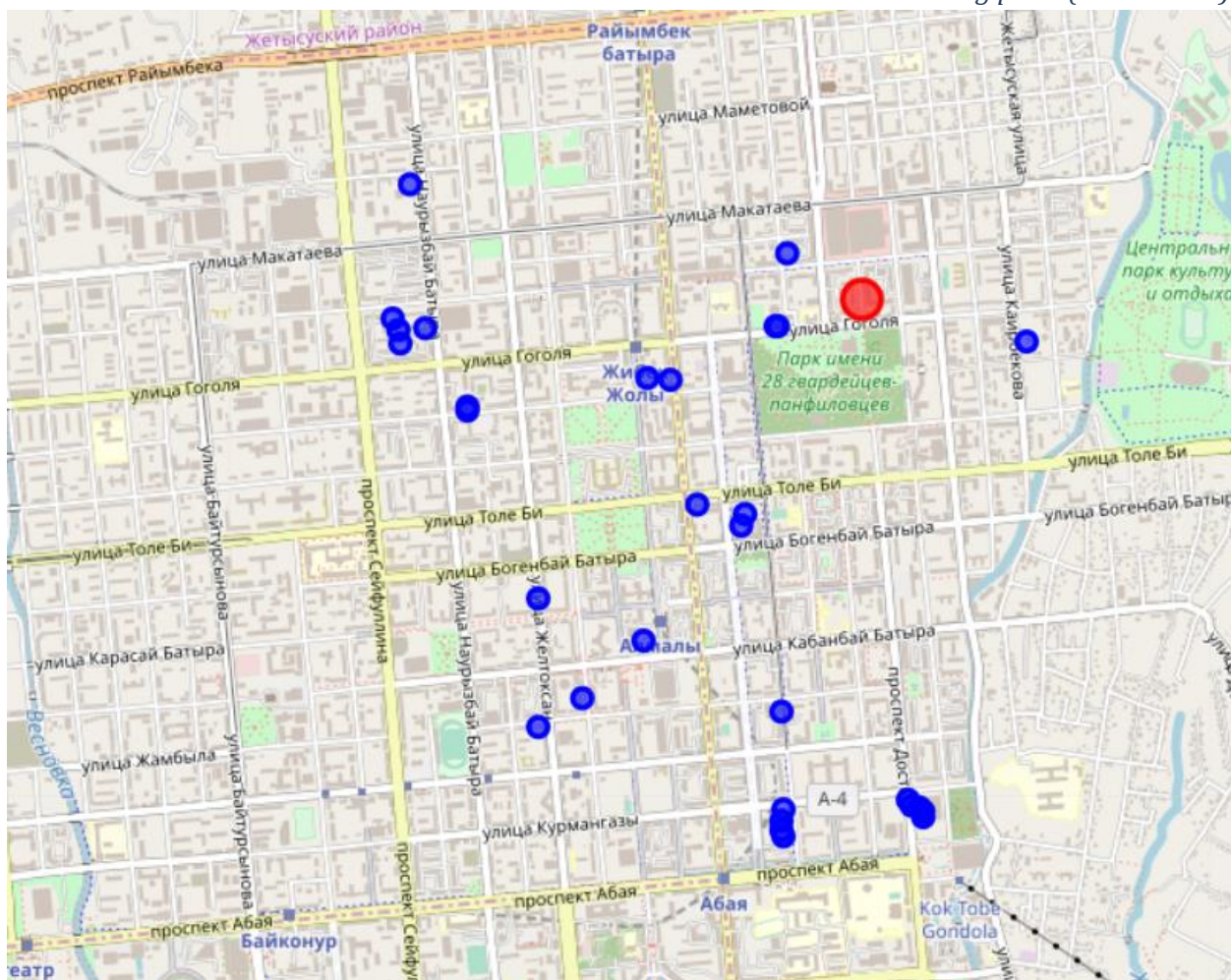
3.3. Data Analysis

Cluster analysis is used. Prepared data divided into 5 clusters with the help of Kmeans method (see *Picture 2*)

4. RESULTS

As a result of the research hotels are divided into **5 clusters** with the following feutures:

- ○ *CLUSTER 1* - **"Short Business Trip Hotels"**. The most common venues are small cafes, bars, fast foods - all necessary venues for short trip or transit tourists. There aren't any sightseing venues, so it can be assumed that only transit or business tourists visit these hotels.
- ○ *CLUSTER 2* - "**Long Business Trip Hotels".** As it is seen the most common venues are restraunts and steakhouses which require more free time.
- ○ *CLUSTER 3* - **"Culture Research Trip Hotels"**. There are theatres, walking areas, markets and stores in most common venues.
- ○ *CLUSTER 4* - **"Domestic Trip to Almaty"**. It seems that these hotels are visited by domestic tourists from other small cities of Kazakhstan for weekend shopping and so on.

- *CLUSTER 5* - **"Enjoy Fun Hotels"**. It seems that this hotels are visited by tourists who like entertainment.

5. DISCUSSION

5.1. Research findings

Almaty city is mostly defined as a business center, shopping center and fashion city of the region - as of most common venues are coffee shops, cafes, restaurants, shopping malls, etc.

5.2. Recommendations

Developing culture venues like Parks, Theatres, Concert Halls, Squares, Gardens in order to attract not only business tourists.

6. CONCLUSION

The research findings can be used in:

- Hotels booking applications.
- Used by city authorities to develop venues and open new venues.
- Used by potential investors and businessmen for their future projects in Almaty.