Introduction to the problem

Marketers face a great challenge as the environment around us is oversaturated by all kinds of advertising. The only way how to succeed in such a world is to come up with innovative strategies that can approach the target group in more effective ways.

One such possibility is to rely on the ever-increasing importance of data in marketing decision making. Thanks to insights obtained from data sets, marketers can create offers with better customer response.

In the following project, we will have a look on how to specifically make use of data science in a fictive case of a real estate developer.

The ultimate purpose of the company's marketing is to prompt the highest rate of sold flats with the lowest media spend. One of the many possibilities is to tailor messages according to the reality of clients. Once the messages are created, we can easily deliver them to the proper audience through online marketing channels using location targeting.

But how to decide who should be exposed to which message?

Let's assume that a company wants to sell their flats in the new residence in the capital of Slovakia, Bratislava. The capital can be geographically divided between several boroughs which can be further split up into neighborhoods. But based on such divisions, we cannot accurately assess the quality of living in the neighborhoods.

One of the indicators of the quality of living in neighborhoods is civic amenities. There are more aspects than this but for the simplicity of the analysis, we will deal only with this parameter.

Based on this insight, we will group city neighborhoods based on the similarity of civic amenities. By doing this, we can create different advertising messages for each group according to its characteristic.

For example, the residence is situated in the city center with a high quality of social life with a lot of restaurants, shopping malls, schools, etc.

One part of our target group lives in the outer center with a lower amount of these amenities. Moving to a better socially equipped part of the city can be of a big benefit. Therefore, the ad version for this part of the city can rely more on presenting its good social life.

Another part of our target group lives in a similar area, so they would not respond to the same message es previous group. In this case, we could adjust the message and present the residence as maintaining the same level of living but in a modern building.

So, how many messages should the developer create? What should they focus on? And which are the neighborhoods with similar characteristics?

These are the question the project tries to answer by exploring the data obtained through the Foursquare API.

About the data

As mentioned in the previous stage, the analysis relies heavily on the data retrieved through the Foursquare API.

Foursquare allows developers to access their database and make a certain amount of calls per day. For this project, a Personal Tier account was created which allows for 99,500 regular calls (e.g. addresses, names of places ...) and 500 premium calls (e.g. ratings, photos...).

The data this project operates with were collected using the "explore" API endpoint. This function makes possible to gather descriptive data about a certain location within a defined radius. In this project, we were mostly interested in places, their names, types, and number of places in a category type.

To compile all the necessary information, the two following steps were done. First, we created a list of GPS locations for each of Bratislava's neighborhoods. Subsequently, we connected to Foursquare API in Python and ran a loop that retrieved data about each neighborhood.

The obtained dataset therefore contains:

- Boroughs in Bratislava

- Neighborhoods in boroughs

- Latitude and longitude coordinates of neighborhoods centers

- List of places within a neighborhood

    - Names of places

    - Address

    - Latitude

    - Longitude

    - Category

The data will be used for further descriptive analysis of each neighborhood and data modeling using K-means clustering.