Battle of neighborhoods week_1_assignment

Development of a reference method for market exploration targeting to sell food and drink ingredients

1. Description of the problem

Our customer is a multinational company traditionally involved in the distribution of food and drink ingredients on local and international markets. This time they are interested in finding new opportunities in central Europe. After a discussion with the responsible representatives of the company we received an assignment to develop a simpler reference method to explore small territories of towns. The target of the exploration is to find potential customers, who can be interested in buying the food and drink ingredients as new products of our partner.

2. Description of the data

This part of the capstone project the data collection phase of our work is presented. Hungary was chosen as the target country of our investigation. After the selection of 8 Hungarian towns Foursquare API was applied to explore a sample from the area of each town. The sample territory was a circle with the radius of 500 m. Inside of the circle all venues were explored that the API can identify. Among these venues those businesses were manually selected and counted, which were considered to act as potential customers to buy food and drink ingredients (except for alcoholic ingredients, therefore bars and pubs were not taken into consideration). For instance restaurants, bakeries, coffee shops were marked as target customers. The selected and counted venues were collected into a data frame, which will be the base of the further analysis.

If our multinational customer is satisfied with the reference method we start a next discussion about the development of a complete data science solution planned to explore and analyze whole cities (not small samples of them) and give much more detailed market analysis.

Further explanation can be found in the ipynb file of this section.

Battle of neighborhoods week_2_assignment

Development of a reference method for market exploration targeting to sell food and drink ingredients

1. Methodology

After the collection of the necessary data about the 8 Hungarian towns, K-means clustering was applied to select the towns into appropriate groups due to the found venues in their test circle and the best value of K was identified. The clusters were visualized on a map with the application of Folium. Moreover classical exploratory data analysis was carried out, where several general features of the dataset were investigated. Further explanation can be found in the ipynb file of this section.

2. Results

K-means clustering and Folium visualization showed that applying K=4 is the most ideal clustering strategy to select the tows. Visualizing the correlation matrix of the data it can be concluded that high correlations among the appropriate venues can be good starting points for planning very detailed and targetted sales and marketing strategies. Visualizing the correlation between the population of towns and their venues and restaurants in the test cirlcle we guess a linear correlation between these parameters, which can be a good base to plan linear regression models in the further phase of the work.

3. Discussion

In general our work was designed to present a small part of a huge methodology system to our customer. Our results show a starting direction towards a much more detailed market analysis to explore sales and marketing possibilities in Hungarian towns. If our multinational customer is satisfied with this direction and our test system he can give a further assignment where we will carry out much more detailed data analysis and we will collect more data about whole towns.

4. Conclusion

Fuorsquare API, general data preparation methods, exploratory data analysis and K-means clustering were successfully applied to build a reference project targeting market exploration.