Capstone Project - The Battle of the Neighborhoods (Week 2)

Applied Data Science Capstone by IBM/Coursera

Data Analysis for Promotion of Vehicle Insurance in each borough of Munich city

1. Introduction

One of the vehicle insurance companies wants to increase the number of people having cars or motorbikes to use their company's vehicle insurances. They want to do so by providing customized offers and promotions to the people of Munich having cars. They have a tie-up with automotive manufacturers, petroleum stations, car/motor-bike repair shops, etc and also have contact with many common app owners.

The vehicle insurance company can know when people are near the venues like automotive repair shops, gas/petroleum stations, etc, and wanted to send customized offers/promotions when people are near the venues where they have a tie-up with.

Now they want to segregate the people having similar characteristics into clusters so that they can send customized offers/promotions to increase their insurances count thereby increasing their revenue.

we will help the insurance company by analyzing the number_of_venues they have a tie-up within each region, number of working professionals, number of new cars registered in each region of Munich in a year to segment the whole Munich region into different segments.

2. Data Source

The data being used are

1. Car-registration data in each region of Munich
2. Population data in each region of Munich
3. Working professional data in each region of Munich

All the data can be obtained from the below sources
https://www.opengov-muenchen.de/dataset/monatszahlen-kfz-neuzulassungen
https://www.opengov-muenchen.de/dataset/indikatorenatlas-bevoelkerung-einwohnerdichte-83r65mct

https://www.opengov-muenchen.de/dataset/bevoelkerung-stadtbezirken
https://www.opengov-muenchen.de/dataset/indikatorenatlas-bevoelkerung-hochbetagte-83r65mct

After cleaning some of the columns of the final dataframe looks similar to the below figure

	borough	residents	surface	population density	share_total_population	latitude	longitude
0	Allach - Untermenzing	30737	1545.17	20.0	2.10	48.195157	11.462973
1	Altstadt - Lehel	20422	314.57	65.0	1.39	48.137828	11.574582
2	Au - Haidhausen	59752	421.96	142.0	4.08	48.128753	11.590536
3	Aubing - Lochhausen - Langwied	42305	3406.02	12.0	2.89	48.165059	11.400221
4	Berg am Laim	43068	631.46	68.0	2.94	48.123483	11.633451

Fig:2.1 **Final dataframe**

Using the Foursquare API, we get the nearby venues in each region of Munich
The vehicle insurance company is assumed to have a tie-up with the following venues
Motorcycle Shop, Automotive Shop, Rental Car Location', Auto Dealership, Gas Station, Parking, Hotel, Motel
Number of these venues in each region is found using Foursquare API and grouped based on region as shown in below

	borough	Number of places
0	Allach - Untermenzing	3
1	Altstadt - Lehel	5
2	Au - Haidhausen	2
3	Aubing - Lochhausen - Langwied	1
4	Berg am Laim	3
5	Bogenhausen	2
6	Hadern	1
7	Laim	2
8	Ludwigsvorstadt - Isarvorstadt	1
9	Maxvorstadt	3

Fig 2.2 **Data from Foursquare API**

3. Methodology and Analysis

After cleaning and preparing the data, let us find the optimal clusters for promoting the vehicle insurance company in each borough of Munich city.

Firstly, we have to do exploratory analysis to the data. For that, we have to find the location of each borough on the map. We can visually plot the features and we have to find whether we have boroughs that possess similar characteristics that vehicle insurance can exploit for their company's promotion.

Secondly, We can reduce the number of features which we have by combining a few features to have more reasonable data.

Finally, We can perform clustering to find the best cluster among boroughs with similar characteristics.

3.1 Exploratory Data Analysis

We can visualize the location of each borough in the Munich to see how each borough is distributed in the Munich city.

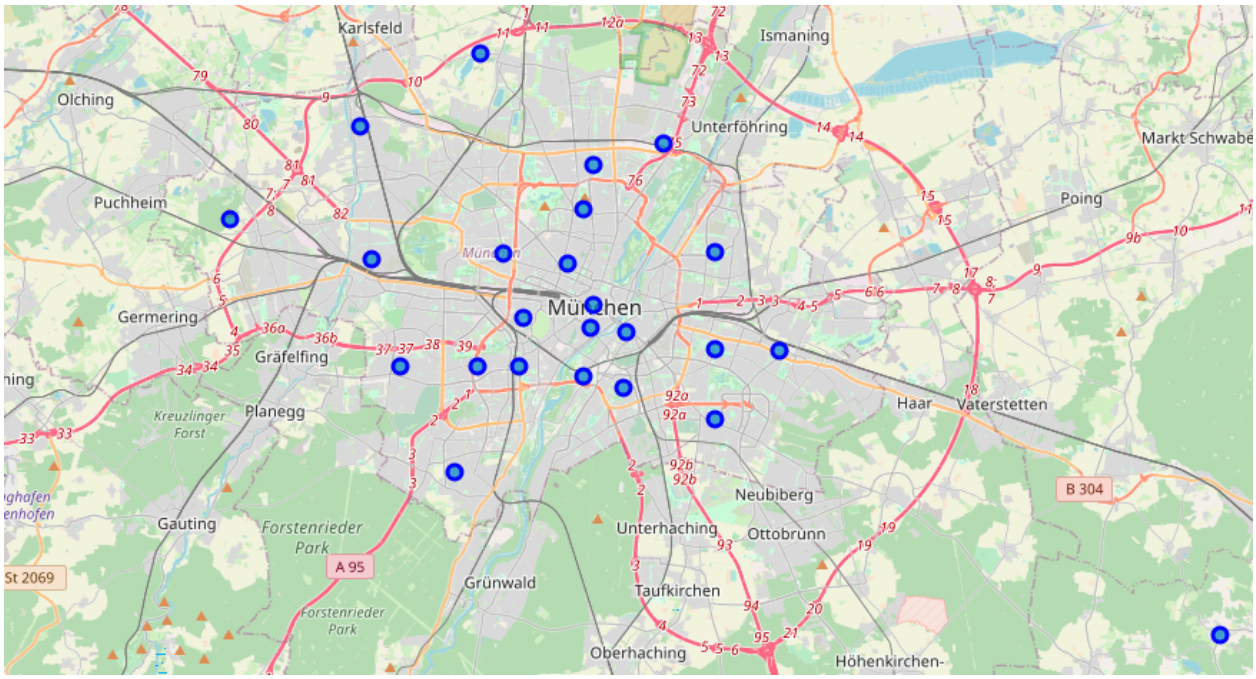


Fig. 3.1 Munich map with boroughs distribution

Additionally, some visualization of important features that helps to understand the characteristics of each borough in the Munich city.

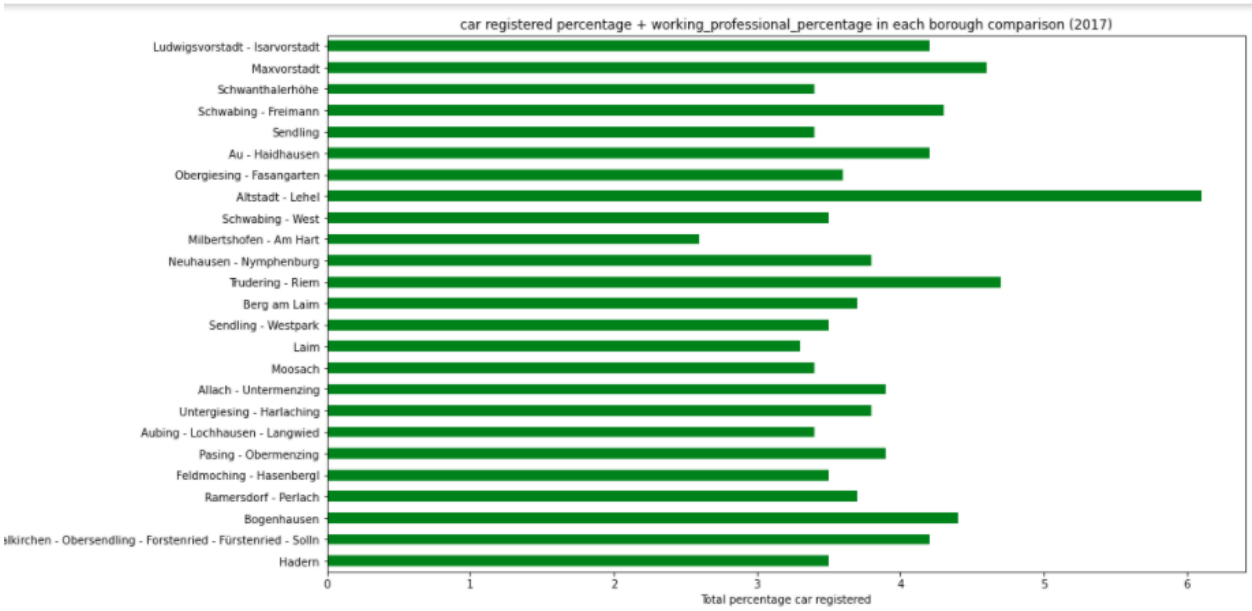


Fig 3.2 Car registered in each borough distribution

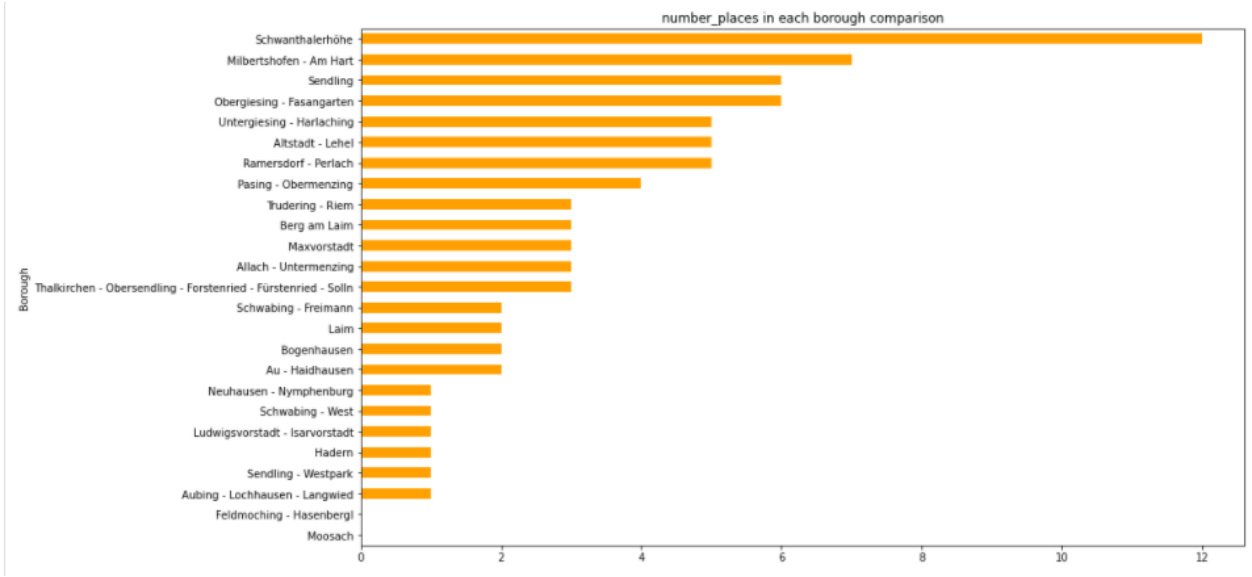


Fig 3.3 Number of venue_categories in each borough distribution

Fig 3.2 explains the distribution of the number of vehicles registered in each borough. This information is important to cluster the boroughs into groups later. Fig 3.3 explains the number of venue-categories that vehicle insurance companies have a tie-up with, more the number of venue_categories, easy for them to promote to more people in each borough.

3.2 Cluster Analysis

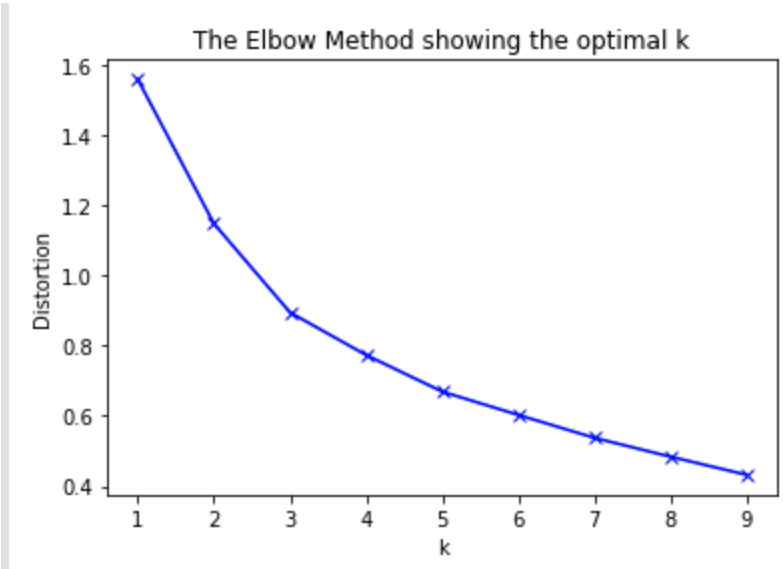
Before we go for cluster analysis we need to do the feature selection. In the dataset, we have the total population and the total number of non-working old age people (above 65 years) data, but we know that most of the working-professional use vehicles and their percentage in each borough influence the cluster analysis.

After doing the above step, the dataset for clustering looks like below

	no_cars_registered	total_work_professionals	number_places
0	565	15543	3.0
1	454	11887	5.0
2	710	36314	2.0
3	604	21244	1.0
4	544	23635	3.0

Fig 3.4 features used for clustering

To find the optimal clusters elbow method is used and k-means clustering is used for clustering the features



Based on the Elbow method 3 clusters were chosen as optimal clusters

Two maps were created one with number of vehicles registered and number of venue_category present in each borough

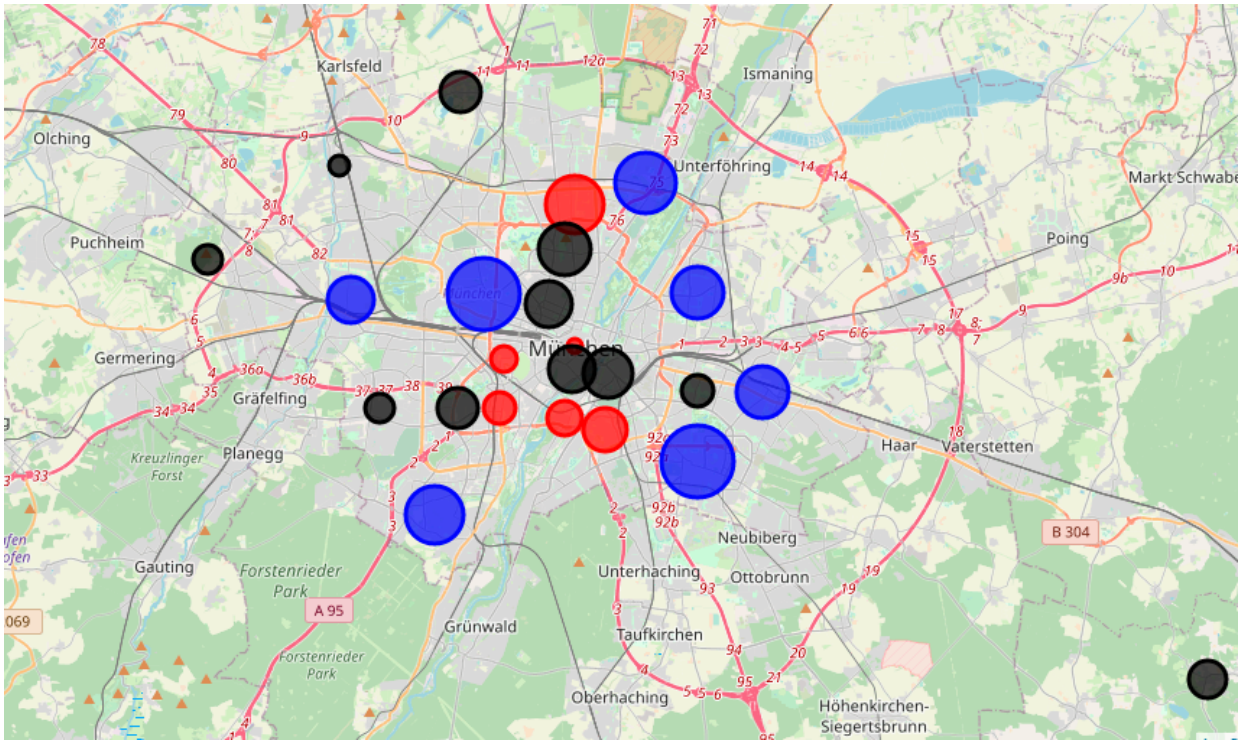


Fig 3.5 Clusters based on the number of vehicles registered in each borough

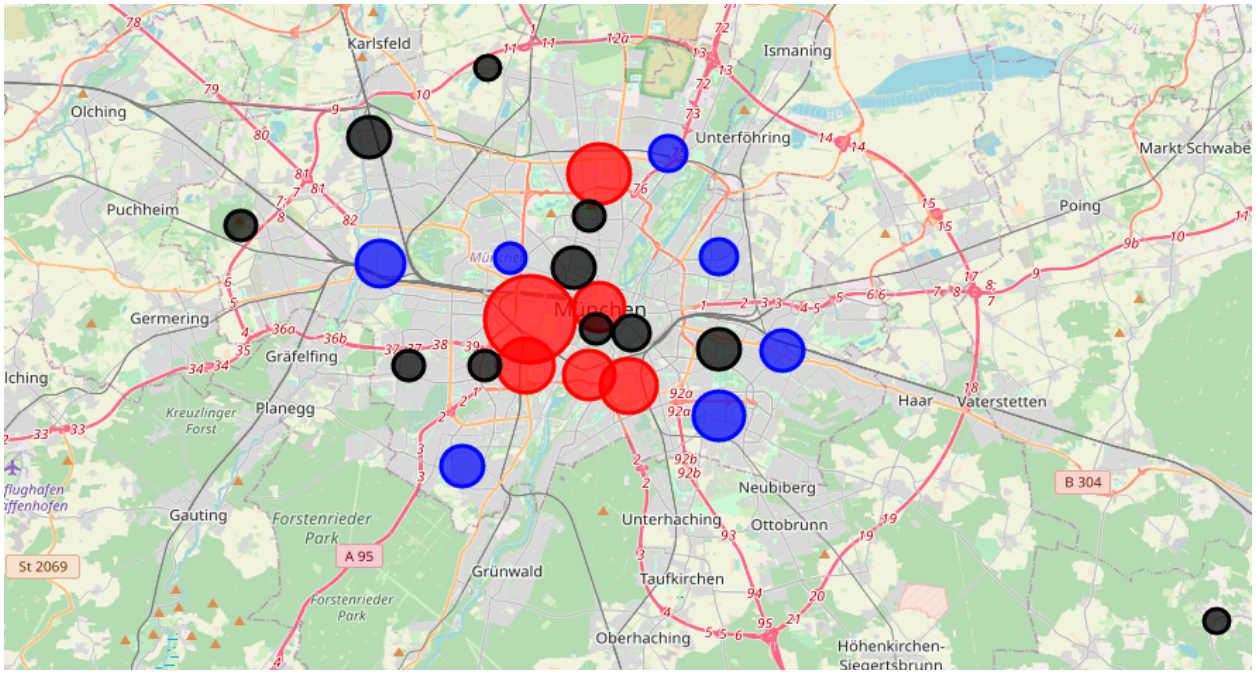


Fig 3.6 Clusters based on the number of venues present in each borough

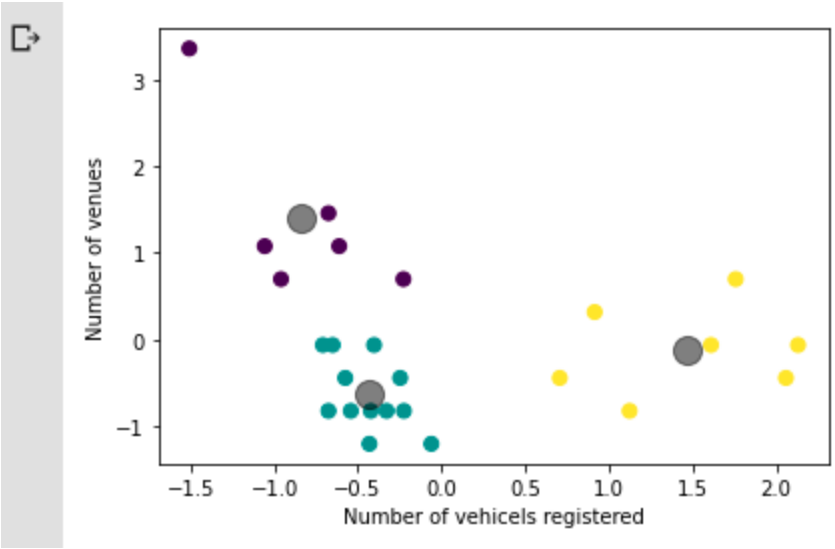


Fig 3.7 Scatter plot between number of vehicles registered and number of venues

Cluster maps in Fig 3.5 and Fig 3.6 show that the black cluster covers outer areas from Munich center whereas blue and red clusters cover the areas which are middle-distant and near to Munich center.

This can give an idea for the vehicle insurance company to prepare and promote suitable promotion offers to each cluster.

The scatter-plot between the Number of vehicles registered and the Number of venues registered in each borough shows clearly three clusters that can also be used for preparing/promoting suitable offers in vehicle insurance by the company.

4. Results and Discussion

From the clusters we can observe

cluster 0 - Have high number_venues where advertisement can be triggered

Cluster 2 - Have medium number venues where advertisement can be triggered

Cluster 1 - Have low number of venues where advertisement can be triggered

So we have to give a different set of offers/promotions for different clusters as the vehicles registered is high for cluster 0, medium for cluster 2 and low for cluster 1. Based on the above clusters, the Vehicle insurance company can try for different offers and they can use these initial clusters for promoting their company and see if the number of people opting for vehicle insurance or not.

However, one can perform further analysis of this particular cluster with additional features like net_income, how_freuqent people are visiting a venue category in the borough, number of accidents happening in the borough. With that, we will get more accurate clusters for advertising vehicle insurance.

5. Conclusion

To conclude, the basic data analysis was performed to identify different clusters for different kinds of offers and advertisements for vehicle insurance in the city of Munich. During the analysis, several important statistical features of the boroughs were explored and visualized. Furthermore, clustering helped us to highlight the group of optimal areas. Finally, we have three clusters based on number_venues present in the borough, number_vehicles registered, and number_working_people in each borough. I hope the clusters have approximately similar characteristics which can be used by vehicle insurance company for their promotion activities.