Advertising campaign

How can we choose where do it using data analysis?

Introduction

- A home delivery company would like to start a marketing project to promote its services near to Toronto. First of all, firm decides to identify every districts of the city and to check for each of them the main restaurant in the area.

- The goal of this analysis is understanding where the advertising campaign have to take place. Than, firm will try to group restaurants by proximity because different marketing strategies will be tested to identify the best way according to behavior of the citizens of the neighborhood.

Dataset

- In this analysis are used data relating to the different districts near to Toronto, after an activity of web scraping from a public wikipedia page. Starting from this database, we can identify restaurants where the advertising campaign will be taken and the subsequent clustering.

Postcode	Borough	Neighborhood	Latitude	Longitude
M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
M1E	Scarborough	Morningside, Guildwood, West Hill	43.763573	-79.188711
M1G	Scarborough	Woburn	43.770992	-79.216917
M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Methodology

- First of all, we import library useful to our analysis, such as pandas, numpy, random, request, matplotlib and more over(detailed in our report step by step). Pandas is early used to build a dataframe of neighborhoods and their latitude and longitude. In particular we used classic method of reading csv and transforming it into a dataframe (103 columns, 5 rows). Next step is to discover what restaurants are near to our geolocations. In order to do that, we used foursquare site. In particular through an api we can extract venues and their characteristics.

- Data are extracted in a json file that has to be converted in a dataframe using jnormalize. Now we can filtered dataframe because a clear structure allow us to do a correct analysis. Then, after loading folium library, we can show in a very simple, but detailed, map every neighborhood and their restaurants. This phase is important to do a first evaluation about our marketing strategy. In fact, we can observe where are concentrated more restaurants and in general their disposition.

- Now, one of the most important phases of our analysis is k-means clustering. This method allow us to segment restaurants in group useful in the next step of the marketing strategy. In particular we defined 5 clusters and we showed them into a map with different colours. Clustering is necessary beacuse our scope is to starting an advertising campaign and we want to understand how we can test different strategy in every group with similar characteristics. This simplify also next analysis of customers behaviour in every single neighborhood.

Results

Results of our analysis are based on the final process of clustering. In fact we can support firm that has commissioned this project through the definition of 5 peer groups where it can develop its marketing strategy. The importance of this analysis is connected to the ability of simplify decisions that firm have to take. In fact is too expensive starting a campaign in every restaurants and is less efficient applying the same strategy to every neighborhoods. Moreover this analysis is a base for a next evaluation about customers' behaviours.
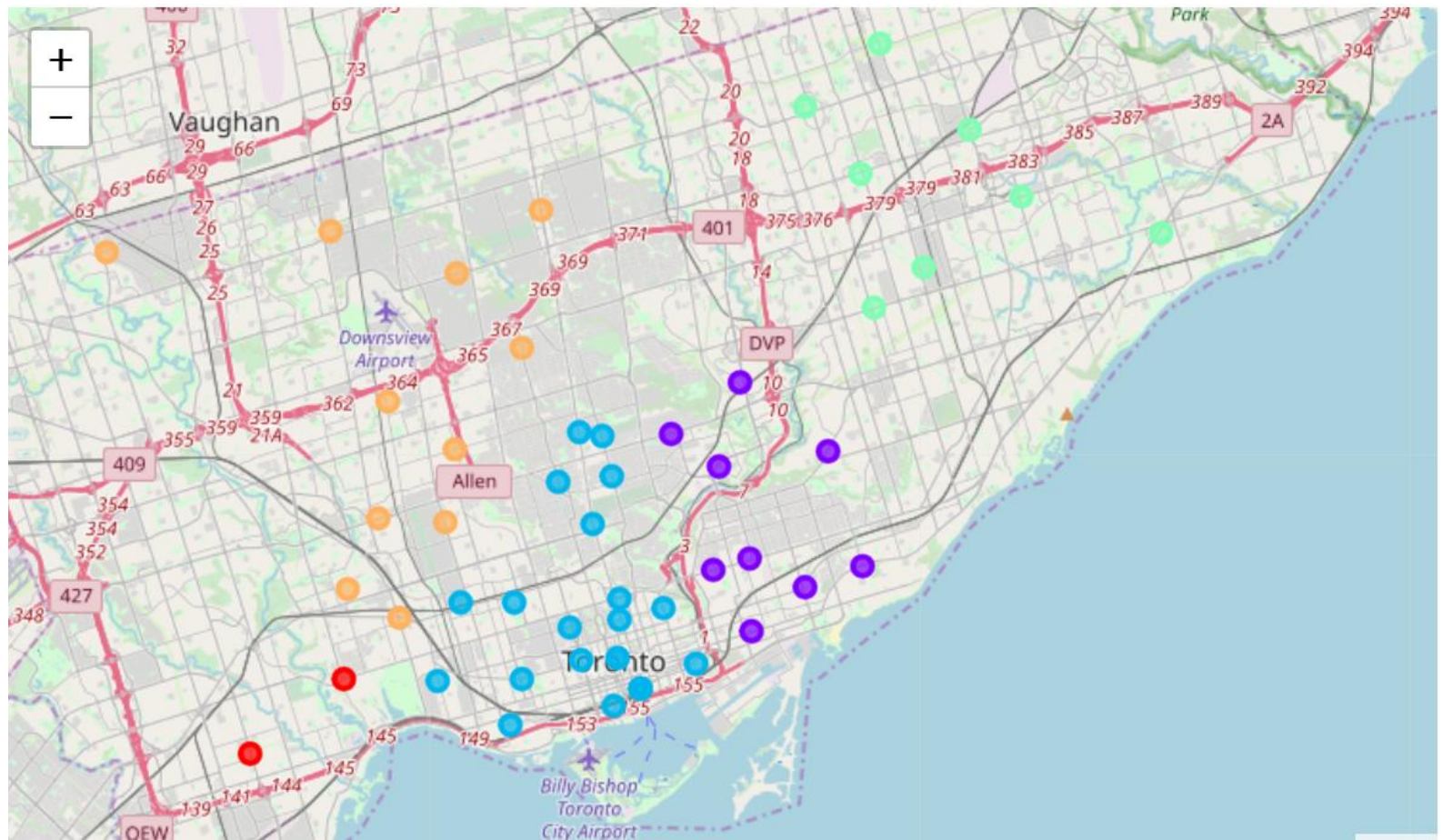
Discussion and Conclusion

- In order to being useful, this analysis have to be supported by a successive evaluation. In particular, we can apply 2 different strategies: first of all, firm can test for each group a lot of advertising campain and than select for every single group the best one; second, the firm can develop a trend analysis in every group, also using foursquare site in order to understand if there are trends that can justify a particular marketing strategy.

- In order to being useful, this analysis have to be supported by a successive evaluation. In particular, we can apply to different strategies: first of all, firm can test for each group a lot of advertising campain and than select for every single group the best one; second, the firm can develop a trend analysis in every group, also using foursquare site in order to understand if there are trends that can justify a particular marketing strategy. Moreover, we can evaluate if there are already delivery services in theese areas. In fact if we want to start with a new business is not enough to develop our strategy but is important to analyze how our competitors work and how they can react to our presence.

- As every analysis, results are just the beginning for other evaluations and analysis. Power of results are simply an opportunity to discover new aspects and develop project in different areas of study. Ability of a data scientist is to observe reality and to image results of his analysis as a starting point for a more complex structure. This were the initial goals of this analysis and i hope to have reached them.