BUSINESS PROBLEM

In the given problem we’re trying to find the best location for a restaurant in Toronto.The solution derived from all the data present,will help everyone who could be looking for a place in the city for a new venture.In such big cities finding the optimum location is a daunting task,but with the help of data analysis of various locations,best possible locations could be pinned which could bring monetary benefits to the owners.Population of neighborhood will help in determining,the sites where highest no. of people would come to visit.Also,places with highly crowded restaurants have to skipped to decrease the entry barrier of a newplayer.Vicinity to the business centers will also play a key role.

Data science will be used to determine the neighbourhoods with the best quality of footfalls in the restaurant.All the stakeholders shall be benefitted from this location.

DATA

Our problem clearly requires a certain type of data set to come upon the best possible outcomes.

* Wikipedia will be used to get the postal codes of various locations
* Foursqaure API will be used to get exact locations,no. of restaurants and type of restaurants.

We are getting following data from these credible sources:

* Postal codes
* Longitudes and latitudes
* Distance from centre of the city
* Age group of people
* Income of people
* Existing restaurants in the neighbourhood

METHODOLOGY

We have taken following steps:

* We have collected data from aforementioned sources.
* Location (in terms of longitude and latitude) along with the category of restaurant.
* We have identified restaurants’ density in various areas.
* We’ll use K-mean to cluster areas with promising potential.
* Chosen clusters are made to be close to less than a km away from centre of the city.
* Map using Folium will be displayed to give a practical outlook of the chosen locations.

ANALYSIS

1. Identification and cleaning

We have to identify and capture the data from all mentioned sources.Some portion of the data is missing ,so we need to clean that portion out of our dataset.

1. Combining data sources

With postal address and longitude-latitude from different data source present with us,we need to combine all of them .

1. Sorting neighbourhood on the basis of latitude and longitude

From previous step,a resulting data set will contain data about neighbourhood ,its postal code and latitude-longitudes

1. Clustering

We have used K-cluster algorithm to cluster various neighbourhoods.Each cluster is analsyed on the basis of distinguishing features. No. of restaurants and types are our determining variables.

RESULT & CONCLUSION

In our result we have found 65 neighbourhoods inside the gerographical co-ordinates obtained using Foursqaure API.Out of the 5 clusters ,1 cluster shows the perfect density for opening of the restaurant.