Optimal Location for opening Supermarket

Vaibhav Agrawal

IBM Capstone Data Science Project- 10th July' 21

Introduction:

An opening of a Supermarket is an interesting and high profit-making business avenue. However, it depends on how successfully you can run business of the Supermarket Store.

On a very broad level success of any supermarket store depends on:

a) Extremely important is - Right location of Store to make sure that customers can easily come and buy products.

b) Customer base - based on Supermarket products range, one should focus on right customer group. For example, the household's income.

c) Demand - There must be high demand which again depends on the population in the area where store is located, the customer base which one is trying to target and lastly how many such stores are already available in the vicinity.

In this project, we will attempt to solve the problem of a supermarket chain owner/ franchise owner and help them to identify which area / neighborhood in Toronto, Canada, they can open their new store. This will cater to supermarket chain owner, franchise owner for supermarket.

Thus, using the data science & machine learning techniques, this project tries to give a recommendation for an optimal location for opening of a supermarket.

Data:

Through this project, focus will be on below factors to decide optimal neighborhood for opening the store:

a) Type of Neighborhood, for example, business & offices, airports, re-creational, residential etc. - Most preferred option to target residential area as it will have maximum customer base.

b) Population & their income – For larger customer base, the neighborhood must have moderate to high population density and decent household income.

c) Current market penetration i.e., how many stores are already in the area

To work on above factor and solving the business problem, below data sets will be used:

First, we must identify the neighborhood for Toronto city. The Wikipedia page has list of neighborhoods.

We will use three-digit postal code to identify neighborhood.

Next, to use foursquare location API, we will also need latitude and longitude for each neighborhood. Using this geo-codes and Foursquare location API, we will explore each neighborhood. We will try to cluster the neighborhood based on different category of venues. This will help us further to find the residential areas.

Further, we will use census data to find out population per neighborhood and household income. This is needed to understand potential market for opening the Supermarket.

Sample screen shots:

1) **Neighborhood with latitude and longitude**

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

2) **Census data**

```
[ ]  df_census.head()
```

	PostalCode	Borough	Neighbourhood Number	Population	Population density per square kilometre	Land area in square kilometres	Total - Household total income groups	Under $5,000	$5,000 to $9,999	$10,000 to $14,999	$15,000 to $19,999
0	M1B	Scarborough	263	90290	6208	45.74	26825	290	240	420	720
1	M1C	Scarborough	134	12494	2403	5.20	3700	60	25	45	60
2	M1E	Scarborough	411	54764	8570	19.04	19855	315	540	815	970
3	M1G	Scarborough	137	53485	4345	12.31	18445	435	455	685	1170
4	M1H	Scarborough	127	29960	4011	7.47	10765	615	220	255	450

References-

https://www.toronto.ca/city-government/data-research-maps/open-data/

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M