Find The Best location for a Tool Shop in Toronto

Introduction:

From big construction sites to small garage adjacent to a house, the common thing that can be found is a toolbox.But the tools in the box vary based on the application from heavy duty to lightweight.With the right tools in the right hands any problem can be solved in a split of seconds.
Tools are a necessity in every industry especially in the Core fields like construction,manufacturing,electrical to name a few.Tools distribution has a huge potential as there are no specific audience,it is used everywhere.

Problem:
One of my family members is thinking about starting a business of his own,well he has been in this tools field for the past 40 years and he decided to start a tool shop in toronto.So Finding a best location for the tool store is the problem at hand.
From his experience I found out that garages,service centres,gyms have slightly higher consumption of tools than the rest of the segments.

Data Acquisition and Cleaning:

1. **Toronto neighbourhoods broken down by postal code**

   source:https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
   Scraped the data from the wikipedia using Beautiful soup and made necessary dropped columns having borough as ‘Not Assigned’.

2. **Load Geospatial Coordinates of toronto and merged with Postal codes**
   Source:CSV
   Merged the two dataframes on ‘Postal Code’
3. **Found out the venues in toronto based on the latitude and longitude.**
   Source:Foursquare
   Using foursquare venues surrounding a radius of 500m has been found out with respect to latitude and longitude of toronto postal codes.

4. **Get all venues in Toronto**
5. **Create a list of potential customers for tool shop**
   using a list containing potential customers categories in our dataset
6. **One hot encoding**
   Using one hot encoding data is prepared for clustering

After clustering and plotting the geo spatial map of the results shows the best neighborhood to open a tool shop.

METHODOLOGY

MACHINE LEARNING MODEL USED:CLUSTERING

“Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. In simple words, the aim is to segregate groups with similar traits and assign them into cluster”

Here the objective was to make separation based on the venue categories from the list of potential customers for the tool shop.In order to perform clustering K-Means algorithm was used.

K-Means Algorithm

“Kmeans algorithm is an iterative algorithm that tries to partition the dataset into *K*pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group”

Inorder to find most suitable value of K **Silhouette Analysis** was used.

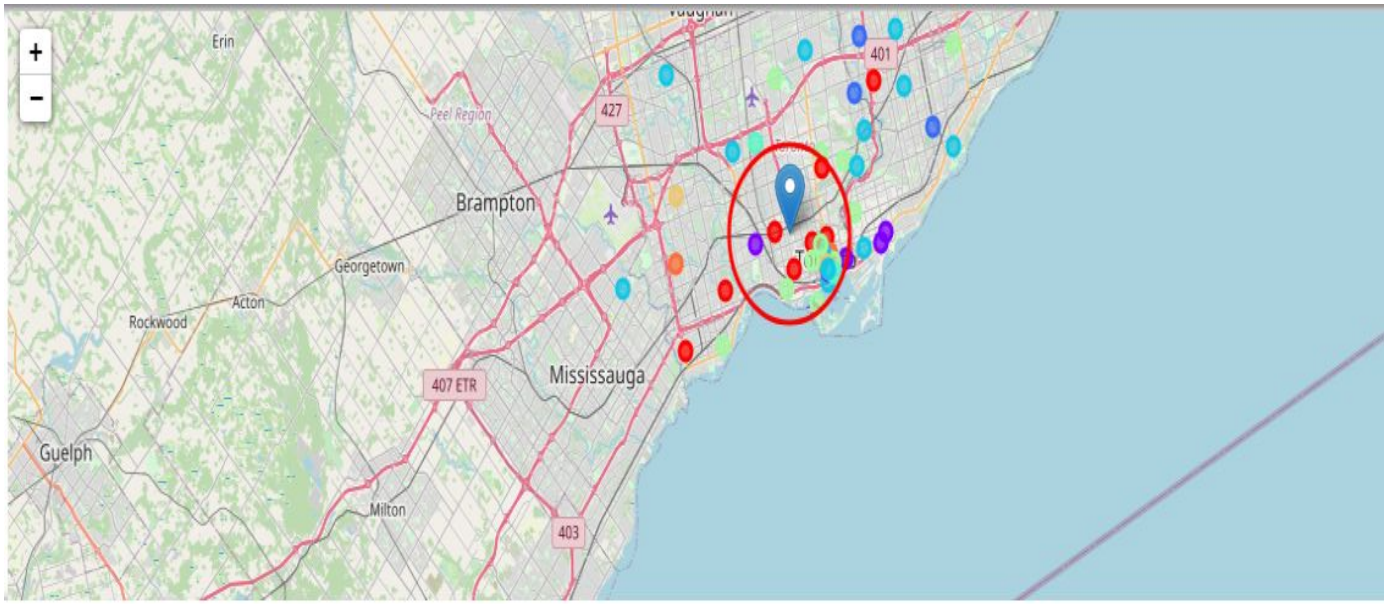
“The silhouette value is a measure of how similar an object is to its own cluster (cohesion) compared to other clusters (separation). The silhouette ranges from −1 to +1, where a high value indicates that the object is well matched to its own cluster and poorly matched to neighboring clusters. If most objects have a high value, then the clustering configuration is appropriate. If many points have a low or negative value, then the clustering configuration may have too many or too few clusters.”

Using silhouette analysis value of K was found out to be 9 with a silhouette value=0.42.

RESULT

After determining the density of each cluster I found out that Cluster 1 has highest density. Inorder to find the geographic centroid of this cluster average of the latitude and longitude was taken.

Using foursquare reverse lookup coordinates ,the optimum location for the tool shop was found to be 'Christie' in 'Downtown Toronto'.



CONCLUSION

After fair evaluation of the data at hand and running silhouette analysis to determine the number of clusters,we were able to plot the clusters in the map using folium library and find the optimum location for opening a tool shop in toronto.
The desired location will be Christie in Downtown Toronto.
Expected outcome from this analysis to the stakeholder will be:

- Lower delivery Cost
- Customer accessibility
- Overall lower run costs
- Increase in overall business
- Overall greater customer satisfaction