

DESCRIPTION OF DATA AND SOLUTION

A. Description of the data

- Scraping of Toronto neighbourhoods via Wikipedia
- Getting latitudes and longitudes data of these neighbourhoods via Geo-coder packages
- Using Foursquare API to get venue data related to these neighbourhoods

- NOTE: Here for finding the neighbourhoods that are most favourable for opening a gaming centre, we determine the venue data related to schools. the purpose behind the selection of schools venue is because maximum of the students and children are more feasible to be playful and more delighted to get into various kinds of games, moreover, it can be relieving tension time for students and also develop their curiosity in computer knowledge and applications.

B. Discussing the solution to the problem. *(In detail with examples)*

- Getting the list of neighbourhoods in Toronto extracting the list from wikipedia page.
- Web scraping by pandas html table scraping, pull tabular data directly from the web page into the data frame.
- Getting the coordinates to utilise foursquare to pull list of venues near neighbourhood with the CSV file provided by IBM.
- Visualise Toronto map using folium package to verify the coordinates.
- Use Foursquare to pull the venues, names, categories, latitudes, longitudes with the help of client ID and secret ID.
- Check the unique categories from the venues.
- Analyse each neighbourhood by grouping the rows by neighbourhood taking the mean of frequency of occurrence of each venue category.
- After clustering venues, perform clustering methods by k-means clustering, an unsupervised machine learning algorithm.
- Based on the result, recommend the ideal location to open a gaming centre

C. Target audience: schools and teenagers.