IBM

Data Science Project Week 4th

INTRODUCTION

In this project, I will create a hypothetical scenario where an entrepreneur wants to build a Pub, so the young people in the region can have a pleasant place to go and have fun in the Toronto Area.
As the Pubs is a place to have fun with friends this entrepreneur might think of opening its business in areas where there is not places to have fun. With the purpose in mind, finding the location to is one of the most important decisions for this entrepreneur and I am designing this Project to help him find the most suitable location.

BUSINESS PROBLEM

The objective of this capstone project is to find the most suitable location for the entrepreneur to open a new Pub in Toronto, Canada. By using data science methods and tools along with machine learning algorithms such as clustering, this project aims to provide solutions to answer the business question : In Toronto, if na entrepreneur wants to open a Pub, where should they consider opening it?

TARGET AUDIENCE

An entrepreneur wants to find the best places to create a pub for people to have fun.

DATA

To solve this problem, we will need below data:

- List of neighborhoods in Toronto, Canada
- Latitude and Longitude of these neighborhoods
- Venue data related to Pubs. This will help us find the neighborhoods that are more suitable to open a Pub

EXTRACTING THE DATA

- Scrapping of Toronto neighborhoods via Wikipedia
- Getting Latitude and Longitude data of these neighborhoods via Geocoder package
- Using Foursquare API to get venue data related to these neighborhoods

PS*: I already have all the datas from the previuously lessons, i saved in a CSV format to use it in this final Project. If you wanna check, go to the "Segmenting and Clustering Neighborhoods in Toronto Part 1, 2 and 3"*