Location Recommender System for Future Taxi Service

Introduction (Cover)

Part 1: Problem Description¶

A future taxi provider based on autonomous car fleet operating in Downtown of Toronto is looking for a best place for parking and charging his fleet . Customers are ordering the taxi service mainly from Restaurants, Cafes, Breweries and Groceries. The fleet operator wants to choose parking location to optimize time of the pick-up, especially for the best rated spots. This will optimize the perceived level of quality and should be competitive advantage for the company.

The operator should build the parking the closest to its customers to minimize the costs of opeartion. Finding the right location is the ultimate goal of proposed algorithm.

Part 2: Data Needed

1- We will utilize geolocation data on specific borough and the surrounding neighborhoods (latitude and longitude numbers). We will limit our search to the Downtown of Toronto. The Postal Codes that are into that borough would also be needed.

2- We will use the information about venues in different areas of Downtown and to gain that information we will use "Foursquare" location service (basic and advanced information about that venue such as category and popularity average price of the services).

Main Article

Part 1: Identifying Neighborhoods inside "Downtown, Toronto"

We will utilize postal codes of different areas inside Downtown to find the list of neighborhoods. We will get the required list of codes from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M..

Part 2: Foursquare Location Data for Venues in Neighborhoods

After finding the list of neighborhoods, we will connect to the Foursquare to gather information about venues with a chosen distance like 1 km from the center (measured by latitude and longitude of, not the walking distance for venues.)

Part 3: Processing the Data for venues inside the Downtown

After the data will be gathered, we will perform processing on that data to find needed features for each venue. The main feature is the ventue category which then will be one-hot-coded. After that dataset will be fully ready for machine learning purposes.

Part 4: K-Means Clustering

We will cluster neighborhoods using k-means clustering. We think that 4 clusters is enough for this project. After clustering we will update our dataset and create a column representing the group for each neighborhood.

Reporting Results

In this part we will focus on the centers of clusters and compare them for their "Total Restaurants" and their "Total Joints". The group which its center has the highest "Total Sum" will be our best recommendation to the contractor.