Coursera Capstone Project

Introduction:

For many investors, Toronto is a paradise. It's the biggest city in Canda. It's also one of largest economy, traffic center in the city. So, opening a shopping center may be a good way to make money. Shopping mall is very popular nowadays. It can attract many customers spending money in it, like taking food in restaurant, or buying clothes. Of course, many investors have come out of this idea and many shopping mall has been built. So the location of shopping center is much more important because a bad location may bring overlap between two shopping malls' influence radius.

Business problem:

This capstone project is aim to choose a proper location for shopping mall in Toronto to maximize stakeholders' benefits. Through using data science methodology and machine learning, we try to solve this problem that where is the best place to create a new shopping center.

Data:

We need following data to accomplish the task:

1. Toronto Neighborhoods data
2. Coordinates of Neighborhoods, including longitude and latitude of Neighborhoods
3. Venue data, especially related data to shopping mall.

Source of Data and methods to extract them:

The Wikipedia contains a Toronto neighborhoods data set: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M which contains a list of neighborhoods in Toronto. Through using Python and the beautifulsoup package, I will scrape the table list in the Wikipedia. Then, we can get coordinates of the neighborhoods through Python Geocoder package or directly, we have obtained the longitude and latitude of the neighborhoods in the following link: http://cocl.us/Geospatial_data

After that, we will get the venue data by using API of Foursquare. Foursquare City Guide, commonly known as Foursquare, is a local search-and-discovery mobile app developed by Foursquare labs inc. The app provides personalized recommendations of places to go near a user's current location based on users' previous browsing history and check-in history. This is a project containing different kinds of data science skills, from web scraping techniques, machine learning skills, working with APIs,

basic data cleaning, data wrangling to map visualization.