Battle of The Neighborhoods Capstone

Coffeeshop Project

Part One

Introduction

As a part of the final IBM Capstone Project, we have to determine what data scientists go through in real life. Objectives of the final assignments were to define a business problem, look for data in the web and use Foursquare, and Google location data to compare different districts within Jeddah city to find out out which neighborhood is suitable to start a **coffeeshop** business. As preparation for this assignment, I will go through the problem discussion, data preparation, and final analysis section, moreover; a detailed codes and images are given in GitHub and link can be found at the end of the post.

Business Problem

Jeddah is the second largest city in Saudi Arabia, it is located at the red-sea coast in the western province, it is far from the holy city of Makkah by 70km. However; in this project I will try to find a suitable location for to start a coffeeshop business point at specific districts that are located between street of Tahlia south to north, Madinah rod east to west. Since there are lots of coffeeshops in the areas mentioned I will try to detect locations that are not already crowded with coffeeshops. I will also particularly be interested in areas with no or few coffeeshops in the neighborhood. I would also prefer locations as close to Salamah, Zahra districts circle as possible, assuming that first two conditions are met. I will use the data science powers to generate a few most promising neighborhoods based on this criterion. Advantages of each area will then be clearly expressed so that best possible final location can be selected.

Data

Based on the criteria of the business problem, the factors that will affect our decision are:

1. Number of existing **coffeeshops** in the neighborhood.
2. Number of and distance to **coffeeshops** in the neighborhood, and
3. Distance of neighborhood from Salamah, Zahra center

We decided to use regularly spaced circular grids of locations, centered around Salamah, Zahra Circle center, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

1. Centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Google Maps API reverse geocoding.
2. Number of **coffeeshops** and their type and location in every neighborhood will be obtained using Foursquare API.
3. Coordinate of center will be obtained using Google Maps API geocoding of well-known Salamah, Zahra districts in Jeddah city.

Neighborhood Candidates

We'll create latitude and longitude coordinates for centroids of our candidate neighborhoods. We will create a grid of cells covering our area of interest which is around 12x12 kilometers centered around Salamah and Zahra center.

First, we find the latitude and longitude of Salamah and Zahra Districts center, using specific, well known address in Google Maps geocoding API.