

Coursera Applied Data Science

Capstone

Ahmad Keewan
ahmedkeewan@gmail.com
<https://github.com/ahmedkeewan>

The battle of neighborhoods

Week 4

April 06, 2020

1. Introduction

Big data is a field that treats ways to analyze, systematically extract information from, or otherwise deal with data sets that are too large or complex to be dealt with by traditional data-processing application software.(wikipedia,Big_data).

Geometry and topology are very natural tools for analysing massive amounts of **data** since **geometry** can be regarded as the study of distance functions. Mathematical formalism, which has been developed for incorporating **geometric** and topological techniques, deals with point cloud **data** sets, i.e. finite sets of points.

Amman is one of the most crowded cities in the middle east, given its size and population, almost 4 million people live in the area of 1,680 km² with tons and tons of restaurants and venues, Amman is famous for its diversity, which makes opening a new small business kind of risky

1.1. Problem:

I need to compare different areas within the city of Amman, Jordan to recommend the hottest spots for an imaginary contractor who is planning to open a restaurant or any other business in a popular area.

1.2. Interest:

Anyone with small money capital planning to open a new business, would be very interested in checking the hottest or dense places before choosing where their business is going to take place, to avoid competition and possible loss.

2. Data

2.1. Required data:

The data that i need for the project is a geographical dataset that contains the different types of venues located in the city of Amman.

The dataset should consist of the Latitudes, Longitudes, Categories, Neighborhoods and the names of the venues that i will be exploring.

2.2. Data source:

To apply what I have learned during the last couple of weeks while preparing for the capstone, I will use Foursquare api to obtain my data, because it's free and kind of provides me with what I need.

2.3. How will it solve the problem:

Using the features extracted from the dataset which are the ones mentioned in point 1, we can compute the frequency of the different types of venues in each neighborhood, and accordingly use these frequencies to cluster and recommend the right spots, and by that we solve the problem.