Data Science Specialization (IBM)

Capstone Project :- Final Report

Recommendation Of Best And Similar Areas In Ahmedabad City Based On Different Group of Categories

Mayur Dangar - 2019

Content

1) Introduction and business problem

Objective:

The idea of this Data-Science project is to build a recommendation system for different types of categories like Entertainment , Fast Food , Indian Food , Sports etc.

Target Audience:

The Business Problem is related to many tourists who travels unknown cities/countries. But while travelling, most of the times people have to search for stay, food and other activities in famous area of that City. They always wish that they get all the same required things like services, enjoyments, clubs, restaurants, hangout places etc near the area City/Country. So Is there a way we can recommend them best area in new Cities/Countries.

Introduction / Business Problem :

So Here in this project, we are going to recommend the best areas in terms of their service,

search for the potential explanation of why an area is popular, the reviews of the places in the areas, or anything else related to areas.

Success criteria of the project are :

- define common cluster/class values for similar categories.
- deliver optimized model for these classes.
- provide a list of similar Area within the chosen city.
- show the recommended areas on a map.

2) Data Gathering, Cleansing and Exploratory Data Analysis

In order to be able to segment and compare different areas in any city of India. we need the whole country data along with latitude and longitude (coordinates) of each city and its areas.

1) Data Gathering

I have used **pgeocode** python library to get all the geolocation data including cities and areas details.

In Data Gathering part :-

1. First I set the country using **Nominatim** API of pgeocode "IN".
2. To get all the data of India I used **DOWNLOAD_URL** API of pgeocode which will be zip file , so extracted it using **ZipFile** library.
3. After Extracting I converted it into pandas Dataframe for further cleansing process.

2) Data Cleansing

Before we do any kind of experiments with the Data it needs to be cleaned.The Data cleansing can be done using Python Pandas.

In Data cleansing part ,

- Remove unwanted / missing rows.
- Fill missing values.
- Format data in a particular way that computers can understand.

3) Geo Visualization of all the Ahmedabad Areas

I used **Folium** library to make it easy to visualize data that's been manipulated in pandas on an interactive leaflet map.

- It enables both the binding of data to a map for choropleth visualizations as well as passing rich vector/raster/HTML visualizations as markers on the map.
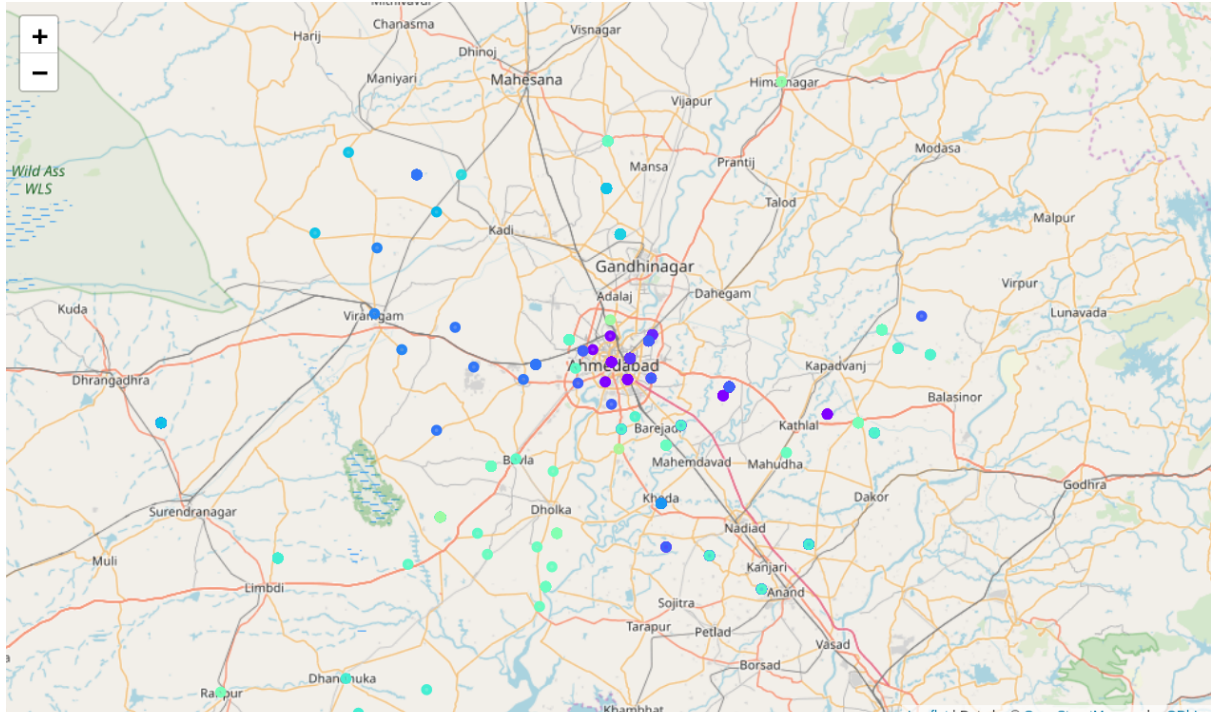
4) Get venues list of all Areas in Ahmedabad City based on given choices of categories

Foursquare API :-

1. To get all venues list I have use Foursquare API. Foursquare API is used to get venue data for each area in Ahmedabad .
2. Foursquare API can give the full details about a venue including location, tips, and categories. Important for this project are mainly the categories of venues (e.g. Hotels, Bars, Coffee Shops).
3. For this the explore function will be used to finally get the most common venue categories in each neighborhood.

Clustering and Recommendation :-

After getting all the venues for all areas within Ahmedabad City I used **K Means** clustering algorithm to get clusters of Areas based on the frequencies of the venues.

Assigned the related cluster numbers to the areas.

Recommended 5 most popular places for particular category groups like Asian Food , Cafe , Fast Food , Veg Only Food, Entertainment, Stores, Other Facilities by sorting the data based on the sum of frequencies of all categories.