Applied Data Science Specialisation - IBM Course

Capstone Project Documentation

CLUSTERING SMART CITIES AROUND THE WORLD BASED ON THEIR VENUES

Bhaskar Rao

1. Introduction

A smart city integrates technology, knowledge, infrastructure and governance to become more responsive to its residence and thereby improving their quality of life. They do so by improving city services and develop strategies for safety, transport, communication and health. As more than 50% of all the people in the world live in an urban environment, it becomes more and more important for city planners, government official and stakeholders to apply modern solutions and development strategies, often utilizing technology, to become a 'smart city'. To facilitate and tract their smart city initiatives, city officials generally refers to the benchmarks provided by independent bodies. One such organisation is Eden Strategy Institute who rate top 50 smart cities based on vision, leadership, policies, initiatives etc. City leadership should also have information about how similar is their city to one of the smart city. This can be done by either looking at transport spend, population, health facilities or education facilities. Apart from these macro measures, a city's characteristic can also define by it's social life. Social life can be gauged by looking at the popular social hotspots around the city, which can be determined by the popular places/venues around it. This can be a new approach of finding similarities between cities and grouping them together based on their social dynamics, which will give the leadership a new perspective to compare their city.

2. Problem Statement

Top 50 smart cities around the world are selected and analysed. A novel way to cluster them together, based on their social signature (social venues), is explored to generate insights. City officials can use the clusters generated to compare how similar or different is their city to a smart city in terms of social life. This will help them evaluate and plan strategies more suitable for their residence, keeping in mind their social signature.

3. Data

A list of 50 top global smart cities is extracted by web-scrapping https://www.smartcitygovt.com/. This data provides the ranking of these cities and their individual score for different dimensions (Vision, Financial Initiatives, Innovation Ecosystem etc.). The latitude and longitude of these city are calculated using the *geopy* python library. These city latitude and longitudes are queried against the foursquare data to get the most popular venues around their city centre. These venues are them used in the clustering algorithm.