Battle of the Neighborhoods – Battleground Berlin

Index

1. Introduction

Berlin is a city of many faces. It is a vibrant, modern metropolis. It is home to 3,5 Million people from all over the world. Its diversity is mirrored by its venues!

Berlin is also where many significant events in modern history have taken place. Until 1989, Berlin was divided and part of two countries: the Federal Republic of Germany (BRD) and the German Democratic Republic (DDR). More than 30 years ago, the city has had time to heal. But many differences are still visible today, for example in architecture. Also, although heterogenous compared to most other capitals in the world, areas in Berlin differ in income level of its inhabitants.

Foodwise, Berlin is very diverse. It is the birthplace of the turkisch style "Doener". From Japanese to Ethopian food, any cuisine from any country is available in Berlin.

2. The Business Case

An investor specialized in opening new venues (cafes, restaurants, bars) is planning to move onto the Berlin market. He would like to infuse his strategy with a data driven approach. The idea is to cluster districts based on venue type, price and ratings. Furthermore, the investor would like to see if there are any differences between parts of the city which were formerly east and west berlin.

3. The Data

Data Sources:

The data is assembled from several sources:

1) Firstly, the location data of Berlins destricts are downloaded in form of a GeoJSON file from the URL https://data.technologiestiftung-berlin.de/dataset/lor_bezirksregionen/en. The districts are superimposed on a map of Berlin using Folium.

2) Secondly, venue data is extracted using the Foursquare API. Relevant venue data includes:

- Location
- Type
- Price Level
- Rating

3) Thirdly, the data on districts formerly belonging to east or west berlin is scraped from the Wikipedia page regarding the districts of Berlin.

Data Processing:

The data will be processed in several ways:

1) Firstly, the city of berlin will be mapped using a heatmap which displays the mean price level and mean level of rating of each district (example: mean price level of each district).

2) Secondly, the city will be mapped using several heatmaps which display the number of a certain venue in each district (example: Heatmap depicting number of cafes in each district).

3) Thirdly the machine learning approach k clustering will be used for clustering and identifiying similarities between the suburbs.