Data Analysis Report

Venues of Freiburg im Breisgau

Inhalt

INTRODUCTION

An investment company would like to invest in the construction of a new hotel in Freiburg (Germany). The company has no idea in which neighborhood they would like to build the hotel. The most important thing is that the hotel generates the highest possible profit. This means that it should be built in a district where the maximum demand for a hotel exists. It is assumed that the maximum demand for a hotel will be in the neighborhood where there are currently the fewest hotels. So the Question is:

In which neighborhood in Freiburg should I build a Hotel to make the most profit?

To answer this question I web-scraped the names and coordinates of the different neighborhoods in Freiburg from Wikipedia. Then I made a query to "Four-square" in a radius of two kilometers to get all venues in the core of the different neighborhoods. I then grouped these venues into categories and sorted them by the percentage of all venues. At the end I could see in which districts the percentage of hotels is the highest and lowest and then make a recommendation.

DATASET

The **first** dataset will consist of the following variables:

- Neighborhood Name
- Neighborhood Latitude
- Neighborhood Longitude

I web-scraped this Data to an excel file from:

https://de.wikipedia.org/wiki/Stadtbezirke_von_Freiburg_im_Breisgau

The **second** dataset will consist of the following variables:

- Venue Name
- Venue Category
- Venue Latitude
- Venue Longitude

I did get this data by making an API-Request from “Foursquare” for a distance of 2000m far from the center of each neighborhood

The **third** dataset will consist of the following variables:

- Neighborhood Name
- Venue Name
- Frequence

I calculated the Frequence of the different categories from the dummy-dataset

The **fourth** dataset will consist of the following variables:

- Neighborhood Name
- 1st Most Common Venue
- 2nd Most Common Venue
- 3rd Most Common Venue
- 4th Most Common Venue
- 5th Most Common Venue
- 6th Most Common Venue
- 7th Most Common Venue
- 8th Most Common Venue
- 9th Most Common Venue
- 10th Most Common Venue

I merged the first and second dataset and made some modifications like grouping.