

The Battle of Neighborhoods – Staff Deployment

1 Introduction

1.1 Background

In an increasingly globalized world, internationally operating companies are required to post their employees to different locations all over the world. There are multiple reasons for staff deployment across all types of industries, e.g. due to a lack of qualified personnel at foreign locations. Further reasons might be knowledge sharing across different locations, international project teams, improving the communication between the headquarter and abroad locations and many more.

1.2 Business Problem & Stakeholders

The goal of this project is to provide valuable insights and support for human resources departments that have to deal with staff deployment in terms of the following aspects and questions respectively:

- Based on the personal preferences of an employee, which neighborhood of the target city should be chosen to book a hotel or rent a project apartment for the duration of the deployment abroad?
- An employee might prefer a neighborhood with similar culture and lifestyle compared to where he/she currently lives.
On the other hand, employees might deliberately be looking for a specific neighborhood that differs from the one they currently live in.
- Enabling employees to live in a neighborhood they prefer, might result in increased willingness to be posted abroad as well as increased employee satisfaction.
- Are there neighborhoods that should be avoided due to very high crime rates?

The issues above will be addressed by segmenting and clustering neighborhoods based on venue data reflecting the culture and lifestyle of the respective area.

In general, the investigations performed in this project might be useful for all kinds of companies sending staff to locations abroad.

In this project, let us assume there is a Munich-based company regularly sending employees to locations in Detroit and Mexico City respectively for one of the reasons outlined in 1.1.

As Munich is known as a safe city with a very low crime rate, the company would like to make sure to post their employees only to areas with a low to moderate crime rate. Thus, prior to the clustering, neighborhoods with significant high crime (rates) are supposed to be excluded from the datasets.

2 Data

2.1 Data Sources

The following data sources will be used to solve the problem described in section 1.2.

City	Munich
Description	Dataset containing Munich districts
Source	Open Data Portal Munich
Link	https://www.opengov-muenchen.de/dataset/bevoelkerung-stadtbezirken
Format	CSV file

City	Munich
Description	Dataset containing Munich neighborhoods
Source	Open Data Portal Munich
Link	https://www.opengov-muenchen.de/dataset/bevoelkerung-stadtbezirksteile-muenchen
Format	CSV file

City	Detroit
Description	Dataset containing Detroit districts and neighborhoods (including shape data)
Source	City of Detroit Open Data Portal
Link	https://data.detroitmi.gov/datasets/neighborhoods
Format	KML file

City	Detroit
Description	Dataset containing crime incidents in Detroit (2019)
Source	City of Detroit Open Data Portal
Link	https://data.detroitmi.gov/datasets/rms-crime-incidents
Format	CSV file

City	Mexico City
Description	Dataset containing Mexico City districts and neighborhoods
Source	Data Portal of Mexico City
Link	https://datos.cdmx.gob.mx/explore/dataset/coloniascdmx/table/
Format	CSV file

City	Mexico City
Description	Dataset containing crime incidents in Mexico City (2019)
Source	Data Portal of Mexico City
Link	https://datos.cdmx.gob.mx/explore/dataset/carpetas-de-investigacion-pgj-cdmx/table/
Format	CSV file

In addition to the data sources displayed in the table above, the following sources are used to retrieve data:

- Foursquare API
- geopy library

2.2 Further Information

2.2.1 Neighborhood Data

The *explore venue* endpoint of the Foursquare API is used to retrieve venue data for the neighborhoods of Munich, Detroit and Mexico City based on geographic coordinates.

Munich

While geographic coordinates for Detroit and Mexico City are basically provided in the files that also contain the district and neighborhood data, the latitude and longitude data for the Munich neighborhoods is obtained using the *Nominatim* class of the *geopy* library.

Detroit

As can be seen from the table above, data for Detroit is contained in a KML file. This file will be converted to geojson format using *kml2geojson* library.

The required data will be extracted from the geojson file and saved in a data frame. Latitude and longitude values for Detroit will be derived from the polygon shape data utilizing the *shapely* library.

Mexico City

Both district and neighborhood data as well as the corresponding geographic latitude and longitude data is available in a single CSV file for Mexico City.

2.2.2 Crime Data

In addition to the neighborhood data, crime data is required for Detroit and Mexico City as neighborhoods with significantly high crime (rates) are supposed to be excluded from the datasets prior to the clustering.

The required data for crime incidents in Detroit and Mexico City is obtained from CSV files according to the table above.

2.2.3 Data Overview

Please see the following examples of the required data for each city.

Munich

district	neighborhood	latitude	longitude
Altstadt – Lehel	Graggenau	48.139616	11.579513

Detroit

district	neighborhood	acres	latitude	longitude	incidents	crime_rate
1	Castle Rouge	223.66	42.381679	-83.268125	207	0.93

Please note that *crime_rate* is calculated as incidents per acre.

Mexico City

district	neighborhood	latitude	longitude
Azcapotzalco	Petrolera	19.485156	-99.199994

district	incidents
Álvaro Obregón	8324

Due to data quality issues, crime data will only be considered on a district level for Mexico City.