

Where should I live? An early analysis about life quality between two big EU cities: Rome and Berlin

Capstone Project - The Battle of Neighborhoods (Week 1)

Vincenzo Scoca

July 7, 2021

Contents

1	Dataset	2
1.1	Venue-Dataset	2
1.2	Neighbourhoods Dataset	2

1 Dataset

The main purpose of this work is to analyse the distribution of venues of different categories among the neighbourhoods of two main cities as Rome and Berlin.

Therefore, two datasets were needed:

- **Venue Dataset:** containing data related to the venue location and category;
- **Neighbourhoods Dataset:** containing all the relevant info about neighbourhoods zip-codes and coordinates;

1.1 Venue-Dataset

For each city we retrieved 50 venues for each of the following categories:

- Art;
- Food;
- Nightlife;
- University;
- Events;
- Outdoors Recreation;
- Professional Other Places;
- Residence;
- Shop Service;
- Travel Transport;

We used the Foursquare API ¹ to collect the raw set of data that has been refined removing the useless columns not for the analysis purpose. Indeed, only the following columns have been held in the clean version of the dataset:

- "id"
- "name"
- "categories"
- "location.postalCode"

The "id" and the "name" column have been held only for the element description, whereas the "categories" and the "location.postalCode" provided the data needed to carry out the study about the distribution of the category venues among the different neighbourhoods.

1.2 Neighbourhoods Dataset

To find the exact location of each neighbourhood, we retrieved two datasets providing the latitude and the longitude of each postal code for both Rome and Berlin:

- **Rome:** <http://download.geonames.org/export/zip/>
- **Berlin:** <https://gist.githubusercontent.com/iteufel/af379872bbc3bf5261e2fd09b681ff7e/raw/205e9a5f1a9f2fb70b58bf9022cab3cf6cc13a19/zipcodes.germany.sql>

In this way we could easily represent on our map each neighbourhood.

¹<https://developer.foursquare.com/docs/api-reference/venues/search/>