

An early analysis about life quality between two big EU cities: Rome and Berlin

Capstone Project - The Battle of Neighborhoods (Week 2)

Vincenzo Scoca

July 9, 2021

Table of Contents

- 1** Introduction
- 2** Dataset
- 3** Methodology
- 4** Results
 - Venue Category Distribution
- Venue Category District Distribution
- Venue Category Correlation
- District Clustering
- 5** Conclusion

City Quality Life Report

- Many reports and rankings about city life quality are currently available on books, internet and other sources.
- But do not always provide a clear evidence of how these scores are affected by the actual city structure.
- The aim of this work is to provide an early analysis about the distribution of ten main service categories, among the neighbourhoods of two main EU cities as Rome and Berlin, to extend the quality analysis provided by the Teleport Company^a

^a<https://teleport.org/>

Dataset

Data

- **Venue-Dataset:** data about 10 venue categories for each city retrieved on Foursquare;
- **Neighbourhoods Dataset:** two datasets providing the latitude and the longitude of each postal code for both Rome ^a and Berlin ^b;

^a<http://download.geonames.org/export/zip/>

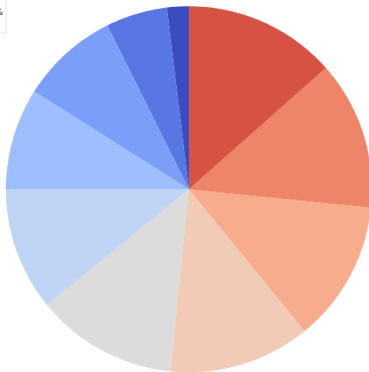
^b<https://gist.githubusercontent.com/iteufel/af379872bbc3bf5261e2fd09b681ff7e/raw/205e9a5f1a9f2fb70b58bf9022cab3cf6cc13a19/zipcodes.germany.sql>

Data Analysis

Main points

- Analysis the venue category distribution within the overall city area and within the city districts;
- Clustering of similar districts;

Results - Venue Category Distribution - Rome



Results - Venue Category Distribution - Berlin



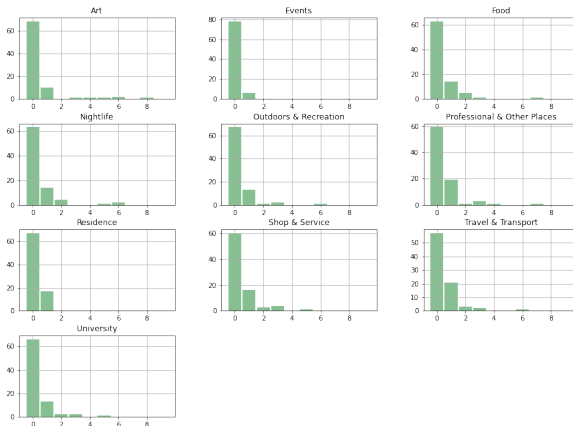
Results - Venue Category Distribution - Final Thoughts

Venue Category Distribution

Both Rome and Berlin are characterized by same venue categories distribution over the city area

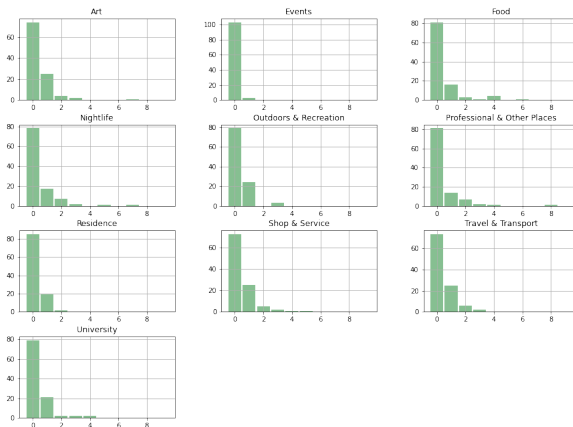
Results - Venue Category District Distribution - Rome

Rome venue-district distribution



Results - Venue Category District Distribution - Berlin

Berlin venue-district distribution

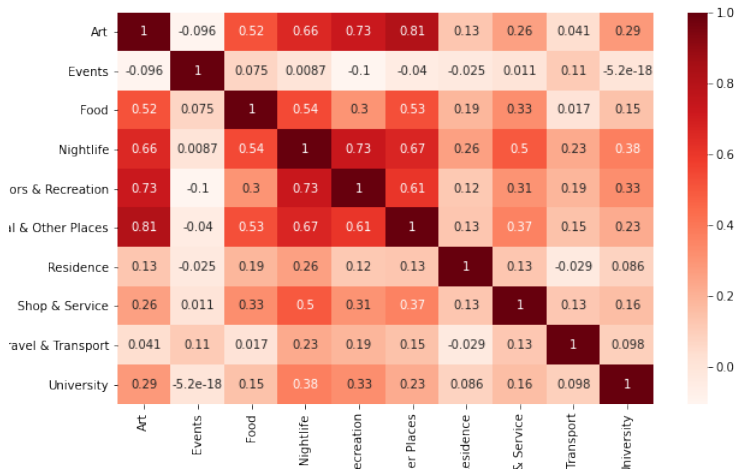


Results - Venue Category District Distribution - Final Thoughts

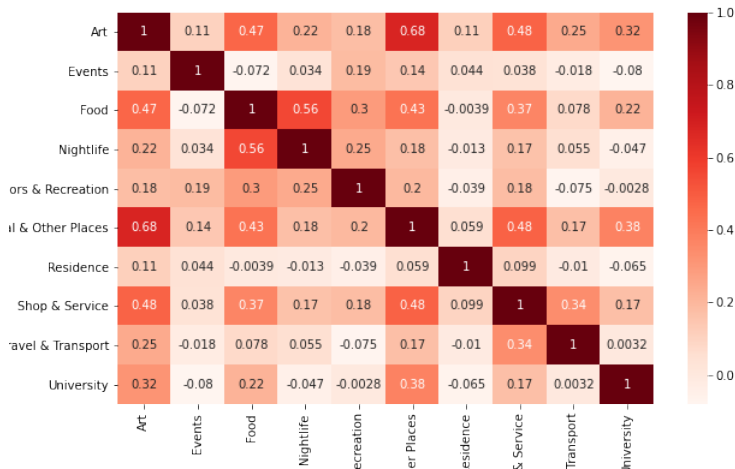
Venue Category District Distribution

- Both Rome and Berlin are characterized by same venue categories distribution over the city districts as well;
- Each district has in average less than one venue for every category;

Results - Venue Category District Distribution - Rome



Results - Venue Category District Distribution - Berlin



Results - Venue Category District Distribution - Final Thoughts

Venue Category District Distribution

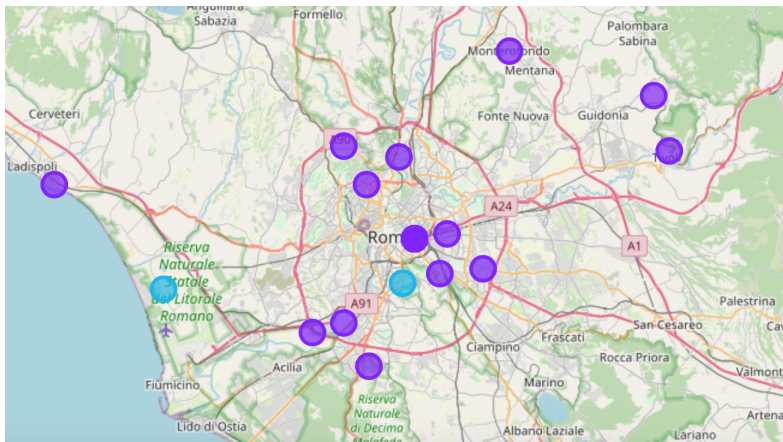
- Only few categories show a stronger correlation, especially for Rome;
- Berlin districts do not provide any clear evidence about categories correlation;

District Clustering

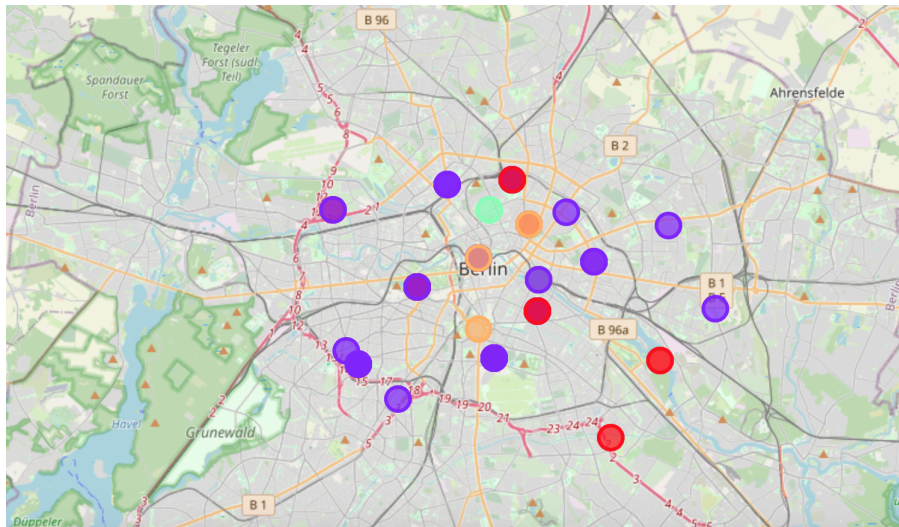
District clustering

- We grouped similar districts in terms of the venue categories;
- We used the **K-Means** algorithm and we estimated the optimal k by means of the **Elbow Method**
- For both cities the number of optimal cluster is $k = 5$

Results - District clustering - Rome



Results - District clustering - Berlin



Conclusion

Take home messages

- We provide some insights about how to start a territorial analysis about the city services and how to use the outcome of such study.
- This work represent an early analysis about the city quality score, is not a comprehensive study;
- All the results must be extensively mastered with further elaborations to be considered as a real support for the city evaluation;

Thank You !