

Vienna - where to settle for entertainment and recreational purposes as a student or young urban professional

Coursera: Applied Data Science Capstone project

Author: Jörn Grimmer

Date: December 2019/ January 2020

1. INTRODUCTION	1
1. DESCRIPTION & DISCUSSION OF THE BACKGROUND	1
2. PROBLEM	1
3. TARGET AUDIENCE	2
2. DATA ACQUISITION, PREPARATION & CLEANSING	2
1. DATA SOURCES	2
2. DATA PREPARATION	2

1. Introduction

1. Description & Discussion of the Background

According to a recent study, Vienna is the most attractive city worth living for the second year in a row. (Source: Economist Intelligence Unit, Global Livability Ranking 2019). This study ranks 140 cities for their urban quality of life based on assessments of stability, healthcare, culture and environment, education and infrastructure. No surprise, Vienna is attracting a lot of people starting their studies or working life.

2. Problem

In order to decide where to settle down, an individual is making use of a lot of criteria.

In our project we will try to find an optimal location so settle in Vienna, when your major interest is in recreational & sport facilities, restaurants and night life entertainment close to the university faculty of your choice.

The analysis is expected to come up with district clusters subject to recreational focus.

3. Target audience

This report is targeted at students & young urban professionals planning to settle in Vienna. We assume this target group is interested in general to settle close to spots of their daily live to minimize distances between every day locations.

2. Data acquisition, preparation & cleansing

1. Data sources

In order to perform this analysis we require the following data sources:

- 1) Geospatial data on Vienna, its districts and its universities. Austria publishes this type of data on data.gv.at.
Specifically I used datasets on Vienna, its districts, its university and colleges and their respective faculties:
 - https://www.data.gv.at/katalog/dataset/stadt-wien_bezirksgrenzenwien/ (BEZIRKSGRENZEOGD.csv)
 - https://www.data.gv.at/katalog/dataset/stadt-wien_universitaetenundhochschulen_standortewien/ (UNIVERSITAETOED.csv)
- 2) Venue data for specified districts of Vienna. Here, we used the **Foursquare data** as required by the Capstone course. On data.gv.at there are distinct sources on recreational spots like sports and swimming facilities. This data could be used to qualify the completeness of Foursquare data and or re-focus such an analysis. However, for scoping and timing reasons of this project, I made use of the Foursquare data only.

2. Data preparation

All relevant data on Vienna, its districts, its universities and colleges is provided in different formats, among them csv. I made use of the csv-file download, read the files with PANDAS and created according dataframes.

The original data set of Vienna's districts had 23 rows and 18 columns.

In order to be able to perform the analysis I prepared Vienna's district data by

- Dropping columns not required
- Sorting districts in ascending order according to the number of the district
- Extracting a latitude/ longitude datapoint from the polygon shape values to have a geospatial value to set a flag on a map for each district

After the preparation, the data set had 23 rows and 8 columns for further analysis.

	index	Borough	BEZNR	DISTRICT_CODE	STATAUSTRIA_BEZ_CODE	STATAUSTRIA_GEM_CODE	Longitude	Latitude
0	3	Innere Stadt	1	1010	901	90101	16.372641	48.216617
1	6	Leopoldstadt	2	1020	902	90201	16.403453	48.231919
2	1	Landstraße	3	1030	903	90301	16.396617	48.207387
3	19	Wieden	4	1040	904	90401	16.369165	48.200713
4	18	Margareten	5	1050	905	90501	16.359449	48.196617
5	20	Mariahilf	6	1060	906	90601	16.363064	48.201827
6	0	Neubau	7	1070	907	90701	16.338725	48.208537
7	2	Josefstadt	8	1080	908	90801	16.349147	48.215158
8	5	Alsergrund	9	1090	909	90901	16.361652	48.231918
9	16	Favoriten	10	1100	910	91001	16.383819	48.185157
10	15	Simmering	11	1110	911	91101	16.425986	48.185575
11	17	Meidling	12	1120	912	91201	16.341743	48.188466
12	21	Hietzing	13	1130	913	91301	16.214234	48.206523
13	10	Penzing	14	1140	914	91401	16.209138	48.264112
14	22	Rudolfsheim-Fünfhaus	15	1150	915	91501	16.327324	48.205005
15	4	Ottakring	16	1160	916	91601	16.276206	48.227037
16	8	Hernals	17	1170	917	91701	16.285159	48.256800
17	7	Währing	18	1180	918	91801	16.295017	48.249609
18	11	Döbling	19	1190	919	91901	16.356813	48.282287
19	9	Brigittenau	20	1200	920	92001	16.373612	48.261269
20	13	Floridsdorf	21	1210	921	92101	16.437762	48.316811
21	12	Donaustadt	22	1220	922	92201	16.507839	48.273446
22	14	Liesing	23	1230	923	92301	16.280553	48.159055

Table 1: Districts of Vienna

A similar preparation was required for the data on Vienna's universities and colleges. Here, the original data set had 158 rows and 6 columns.

In this case we prepared the dataset by

- Dropping columns not required and
- Extracting the latitude/ longitude datapoints

After the preparation, the data set had 158 rows and 3 columns.

	NAME	Longitude	Latitude
0	FH Technikum Wien	16.377856	48.239443
1	FH Technikum Wien	16.426908	48.269503
2	FH Campus Wien	16.382288	48.157733
3	Fachhochschule des bfi Wien	16.403446	48.219132
4	Fachhochschule des bfi Wien	16.426908	48.269503
5	FHWien-Studiengänge der Wirtschaftskammer Wien	16.349201	48.226579
6	Lauder Business School	16.352469	48.242701
7	FH Technikum Wien	16.355891	48.200143
8	Technische Universität Wien	16.363088	48.200171
9	Akademie der bildenden Künste Wien	16.361984	48.199806
10	Universität Wien	16.348993	48.233515

Table 2: Universities and Colleges of Vienna (First 10 entries only)

The venue data was received by using the Foursquare API.

The requested data is provided in a json format. Thus it needed to be cleansed and structured into a PANDAS dataframe.

Below is an example of a venue data pandas dataframe used for further analysis.

Out [40]:

	name	categories	lat	lng
0	ZWE	Jazz Club	48.216341	16.374444
1	Palais Hansen Kempinski Vienna	Hotel	48.216335	16.368463
2	Adria	Beach Bar	48.214945	16.375037
3	Tel Aviv Beach	Beach Bar	48.217081	16.373421
4	Feuerdorf	BBQ Joint	48.215972	16.373495

Table 3: Example of Foursquare Venue Data