IBM Data Science Capstone Project

Location analysis for a bar in Vienna

IBM Applied Data Science

Philipp Schardax

Course provided via coursera.org

The business problem

THE CASE

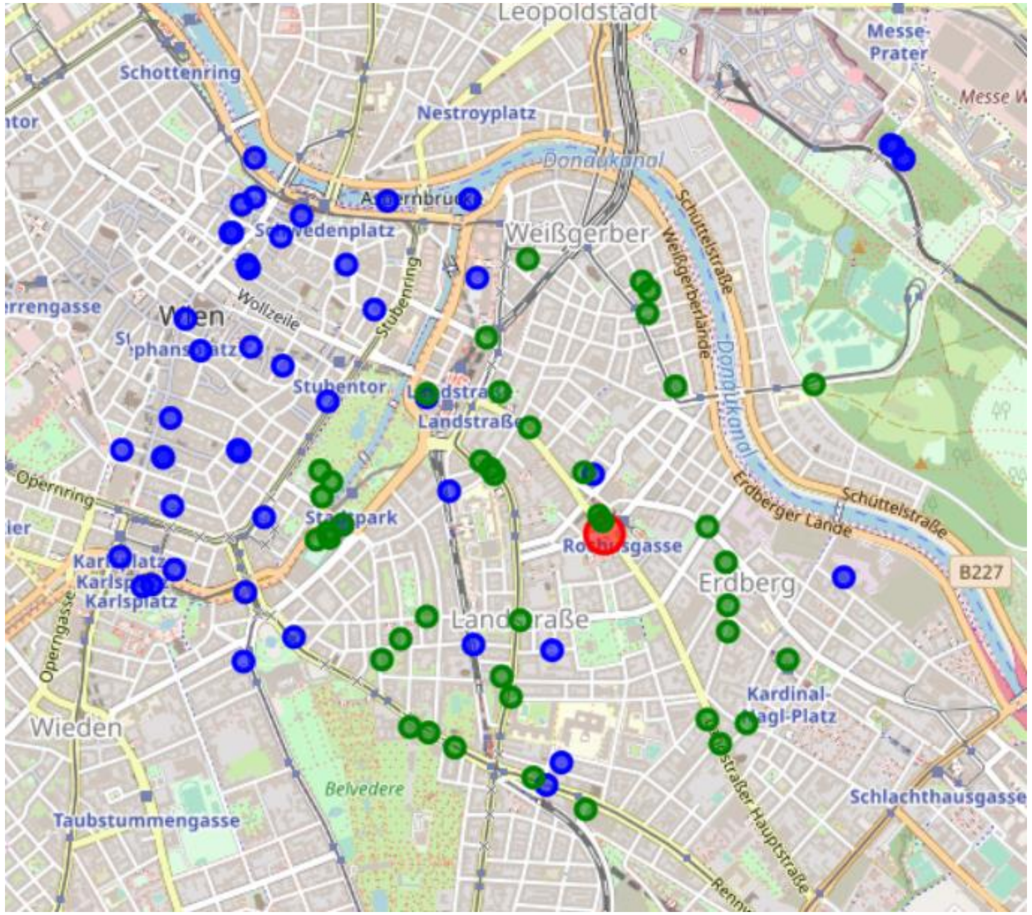
- Client wants to open a **bar in Vienna**
- Locate it **southeast** of Vienna‘s center
- **Avoid competition** with other close bars
- Find an **affordable venue**

APPROACH

- Retrieve location data from **Foursquare**
- Employ **publicly available data** of real estate transactions for analysis
- Create a map using **GeoJSON**
- Identify **promising areas**

The result:
A Folium map displaying all bars and nightclubs in the area, and real estate prices for each district in question.

Dataset 1: retrieved from Foursquare API



DATASET 1

- Data from Foursquare API of **50 bars and 50 nightclubs**
- Retrieved for **categories** „Bar“ and „Nightclub“
- Within 1500 meters of the area in question
- Includes names, adresses, **latitude, longitude…**
- Were filtered, cleaned and displayed using **folium**

Dataset 2: public data about real estate transactions

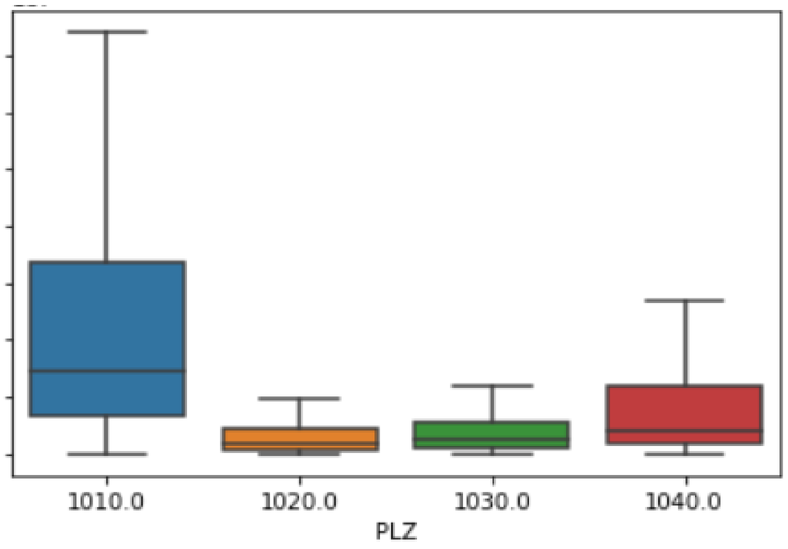
	A	B	C	D	E	F	G	H	I	J	K
1	KG.Code	Katastralge	EZ	PLZ	Straße	ON	Gst.	Gst.Fl.	ErwArt	Erwerbsda	Widmung
2	1616	Stammers	3638	1210	Peter-Berr	09-Nov	.1346	2371	Kaufvertra	27.10.201	W
3	1803	Inzersdorf	4076	1230	Pfarrgasse	87	1551/400	5365	Kaufvertra	16.10.201	GB
4	1609	Jedlesee	920	1210	Bellgasse	25	587/178	224	Kaufvertra	03.12.201	Eklw
5	1616	Stammers	91	1210	Stammers	4	.94/1	684	Kaufvertra	26.06.201	GB
6	1603	Donaufeld	791	1210	Leopoldau	87	1405/26	8	MA 64-Bes	27.10.201	GS
7	1616	Stammers	1521	1210	Alte Bahntrasse		2956	10683	Kaufvertra	06.11.201	SwwL
8	1613	Leopoldau	5470	1210	Grellgasse	3	1914/17	672	MA 64-Bes	21.10.201	GB
9	1201	Auhof	172	1140	Rosenweg	4	833	395	Kaufvertra	11.12.201	W
10	1658	Hirschstett	990	1220	BREITENLE	162	416/119	636	Kaufvertra	28.08.198	WI/GBI
11	1658	Hirschstett	378	1220	ZIEGELHOI	21	382/3	3000	Kaufvertra	09.10.198	WI/GBI
12	1806	Mauer	3153	1230	Bertegasse	26	1153/32	496	Kaufvertra	26.01.198	WI/GBI
13	1806	Mauer	1783	1230	GEBIRGSG	84	779	2227	Kaufvertra	11.08.198	WI/GBI
14	1806	Mauer	3239	1230	ROSENHÜ	193	1160/34	624	Tauschver	22.12.198	WI/GBI
15	1009	Mariahilf	1089	1060	WEBGASSE	42	632/1	418	Kaufvertra	26.06.198	WIV/GBIV
16	1009	Mariahilf	402	1060	LINKE WIE	106	807/1	684	Kaufvertra	24.07.198	WIV/GBIV
17	1803	Inzersdorf	1657	1230	KINSKYGAS	14	513/22	640	Kaufvertra	22.10.198	WI/GBI
18	1009	Mariahilf	794	1060	LINKE WIE	120	818/1	1789	Kaufvertra	18.01.198	WIV/GBIV
19	1009	Mariahilf	124	1060	LEHARGAS	13	220	656	Kaufvertra	16.02.198	WIV/GBIV
20	1009	Mariahilf	879	1060	RAHLGASS	6	1600/6	701	Kaufvertra	12.01.198	WIV/GBIV
21	1009	Mariahilf	694	1060	MARIAHILI	93	603/1	2373	Kaufvertra	05.11.198	WV/GBV
22	1009	Mariahilf	955	1060	STUMPERC	7	932	1488	Kaufvertra	21.08.198	WIV/GBIV
23	1803	Inzersdorf	4010	1230	RICHARD-STRAUSS-ST		1568/41	63959	Kaufvertra	06.11.198	BG/IG
24	1658	Hirschstett	1230	1220	BREITENLE	170	416/122	758	Kaufvertra	05.10.198	WI/GBI
25	1514	Währing	347	1180	WÄHRING	158	162/2	227	Kaufvertra	23.02.198	WIII/GBIII
26	1514	Währing	750	1180	SCHOPENH	59	295/1	359	Kaufvertra	10.12.198	WIII/GBIII
27	1658	Hirschstett	641	1220	ARNIKAWE	13	421/10	667	Kaufvertra	29.08.198	WI/GBI
28	1514	Währing	1085	1180	JOHANN N	3	293/119	372	Kaufvertra	26.01.198	WIII/GBIII

DATASET 2

- Retrieved from https://www.data.gv.at/auftritte/?organisation=stadt-wien
- contains data of **52.000 real estate transactions** in 2019
- **Messy** data
- Does not contain m² of property sold
- Looped Latitude and longitude conversion timed out
- Was cleaned, filtered and used for **statisitical analysis**
- GeoJSON retrieved from https://github.com/codeforamerica/click_that_hood/tree/master/public/data

Analysis: Boxplot and grouped means

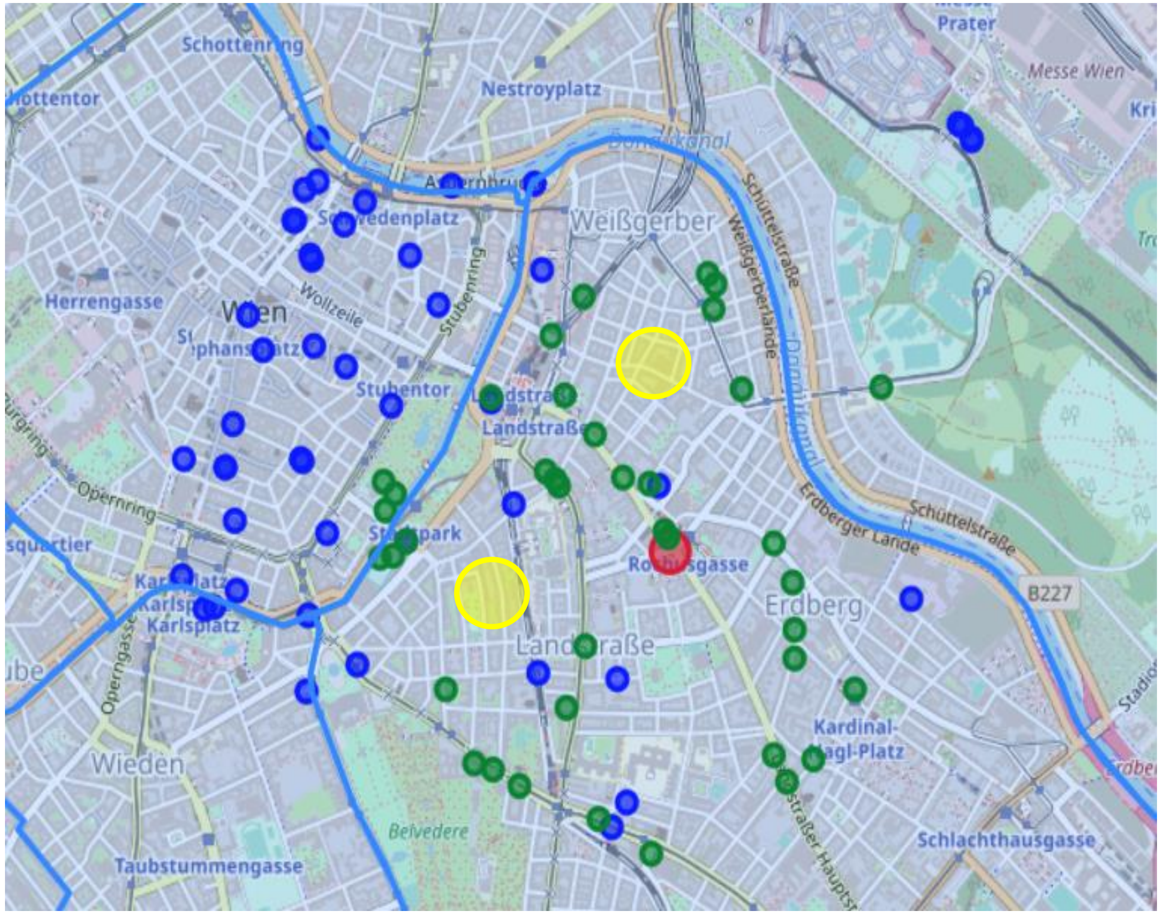
```
PLZ
1010.0     6474874
1020.0     1051841
1030.0     1371054
1040.0     2190004
Name: Kaufpreis €, dtype: int32
```



ANALYSIS OF DATASET 2

- **Pre-processing** the Pandas dataframe
- Filtering only for relevant **property categories**
- **Dropping** rows without values
- **Grouping** by districts
- Calculating the **means for each district**
- Using a boxplot to visualize **statistical distribution** and outliers

Result: a map of promising areas



INTERPRETING THE MAP

- **Competition:**
  - Dark blue: nightclubs
  - Green: bars
- Blue lines: **different districts**
- Yellow: **most promising areas** for the new bar
- → a bar opened around one of the areas hgihlighted in yellow avoids direct competition by close bars and nightclubs, and benefits from lower real estate prices in Vienna‘s second district