

# **Exploration of Zurich's hotels and tourist attractions**

Data Analytics

**Team members:** Adalia Fernanda Aneiros Gutiérrez, Diego Arturo Gonzalez Juarez,  
Stanislaw Zapala

# **Content**

- 1. Introduction**
- 2. Materials and methods**
- 3. Results and discussion**
- 4. Conclusions**
- 5. Bibliography**
- 6. Appendix**

# Introduction

The tourism industry plays a vital role in the global economy. However, finding optimal accommodations and tourist attractions based on individual preferences and constraints remains a complex task for travelers. This report seeks to address this problem by analyzing hotel and tourist attraction data from Zurich, employing advanced data processing, statistical analysis, and machine learning techniques.

The **objective** of this project is to create a data-driven methodology to enhance the experience of tourists by providing optimized recommendations for hotels and attractions. The **research question** focuses on understanding the relationships between key factors and leveraging these insights to build an interactive recommendation system.

# **Materials and methods**

# Data Scraping → Hotels data

Extracting information from websites by accessing their HTML structure and can be exported in CSV, JSON and XML format

The screenshot shows a web-based hotel search interface. At the top, there's a search bar with 'Zürich', date range '15 Jan – 12 Feb', guests '3 guests', and a 'Filter' button. Below the search bar are three listing cards:

- 105 m<sup>2</sup> Apartment · 4 guests · 2 bedrooms**  
Mainaustrasse, Zürich  
1,9 km to centre · Mühlebach, District 8  
 **CHF 220** per night **View deal**  
Promoted by HousingAnywhere
- 90 m<sup>2</sup> Apartment · 4 guests · 2 bedrooms**  
Geranienstrasse, Zürich  
2,7 km to centre · Mühlebach, District 8  
 **CHF 271** per night **View deal**  
Promoted by HousingAnywhere
- Apartment · 3 guests · 1 bedroom**  
Attic loft with rooftop terrace!  
1,7 km to centre · Sihlfeld, Wiedikon  
 **CHF 80** per night **View deal**  
★ 4.4 (9 ratings)

To the right of the listings is a map of Zürich showing various neighborhoods and price points for accommodations. Price labels are placed over specific locations, such as CHF 126, CHF 122, CHF 176, CHF 105, CHF 322, CHF 1457, CHF 80, CHF 541, CHF 108, CHF 271, CHF 70, and CHF 1'880.

Data scraping 1/2 (0.5 pt.)

# Data preparation

## Data types handling:

1. Regular expressions:

```
df_hotels['no_reviews'] = df_hotels['no_reviews'].str.replace(r'\(|\)|reviews|review', '', regex=True)
df_hotels['no_reviews'] = pd.to_numeric(df_hotels['no_reviews'])
df_hotels.head()
```

2. Encoding of categorical values:

```
dicc_review = {'Acceptable' : 1, 'Average' : 2, 'Good' : 3, 'Excellent' : 4}
df_hotels['review_cat'] = df_hotels['review_cat'].map(dicc_review)
df_hotels.head()
```

3. Elimination of missing values and duplicates

# Data obtension

## Attractions data

[opendata.swiss](#) 

Daten Organisationen Showcases Handbuch Über uns 

Startseite > Organisationen > Stadt Zürich > Sehenswürdigkeiten in der...

### Sehenswürdigkeiten in der Region Zürich (Zürich Tourismus)

 Abonnieren

Zuletzt aktualisiert  
13. Dezember 2022

Nutzungsbedingungen



Organisation  
Stadt Zürich

Kategorien  
Umwelt, Wirtschaft und Finanzen

### Beschreibung

Sehenswürdigkeiten in Zürich und der Region unterteilt in Architektur, Kirchen, Denkmäler, Plätze und Strassen, Aussichtspunkte und Kunstobjekte im öffentlichen Raum (KiöR).

kior zuerichtourismus denkmal kirche kunstobjekt aussichtspunkt platz architektur

### Ressourcen



[Link zu API-Endpoint: Sehenswürdigkeiten](#)

Preview

Herunterladen

Data enrichment 2/2 (0.5 pt.)

# Data preparation

## Data types handling:

1. Elimination of missing values (variables with a high proportion of missing values)

```
for i, val in enumerate(na_count):
    if val > 0.3:
        drop.append(columns[i])
        print(columns[i])

df_att.drop(columns= drop, inplace= True)
df_att.head(5)
```

2. Convert the json format of the variables into python dictionaries

```
address_df = pd.json_normalize(df_att['address'])
address_df.head()
```

3. Extract the English version from a column

```
def get_english_version(sub_df):
    result = re.search("en': '(.*)', 'it'", sub_df)
    if result:
        return result.group(1)
    return None
```

# Non-graphical analysis (hotels data)

## .info() method

```
#   Column           Non-Null Count   Dtype  
--- 
0   web-scraper-order    152 non-null    object  
1   web-scraper-start-url 152 non-null    object  
2   name                  152 non-null    object  
3   address                152 non-null    object  
4   description              152 non-null    object  
5   review_cat                152 non-null    category
6   price                   152 non-null    int64  
7   no_reviews                 152 non-null    int64  
8   Latitude                  152 non-null    float64
9   Longitude                  152 non-null    float64
dtypes: category(1), float64(2), int64(2), object(5)
memory usage: 12.2+ KB
```

Insights:

- The dataset consists of 152 non-null entries
- There are textual, numerical and categorical types of data

## Summary statistics:

	price	no_reviews
<b>count</b>	152.000000	152.000000
<b>mean</b>	231.355263	1827.802632
<b>std</b>	145.510766	1829.342664
<b>min</b>	35.000000	1.000000
<b>25%</b>	139.750000	367.250000
<b>50%</b>	203.500000	1334.500000
<b>75%</b>	268.000000	2626.500000
<b>max</b>	1022.000000	9769.000000

Insights:

- The average price of the hotels is 231.36.
- High level of variability in the prices.
- The distribution of prices is right skewed.
- 1827.80 reviews for each hotel in average
- The median number of reviews is 1334.5, which suggests a skewed distribution.

## .value\_counts() method

	count
review_cat	
4	85
3	57
2	9
1	1

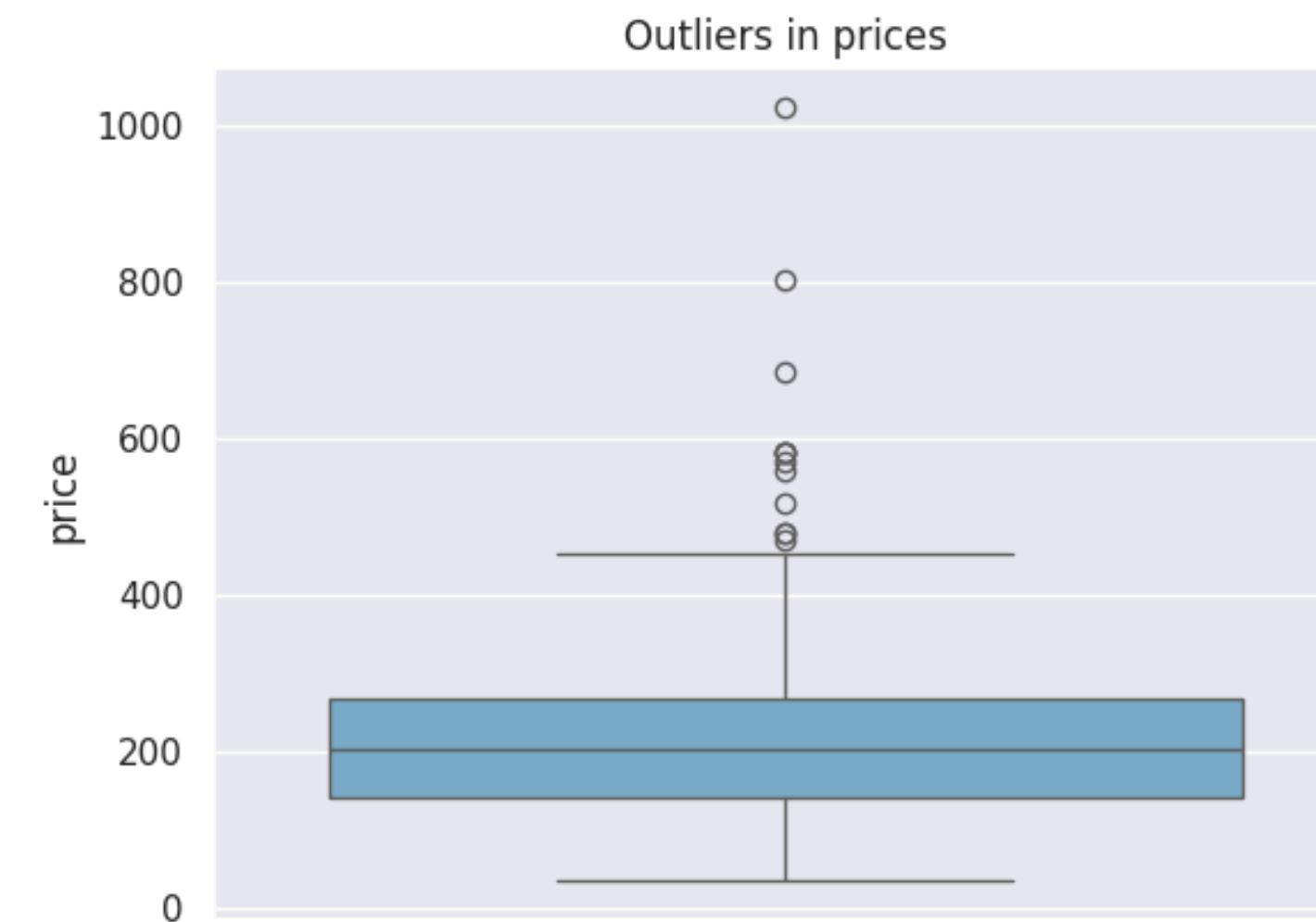
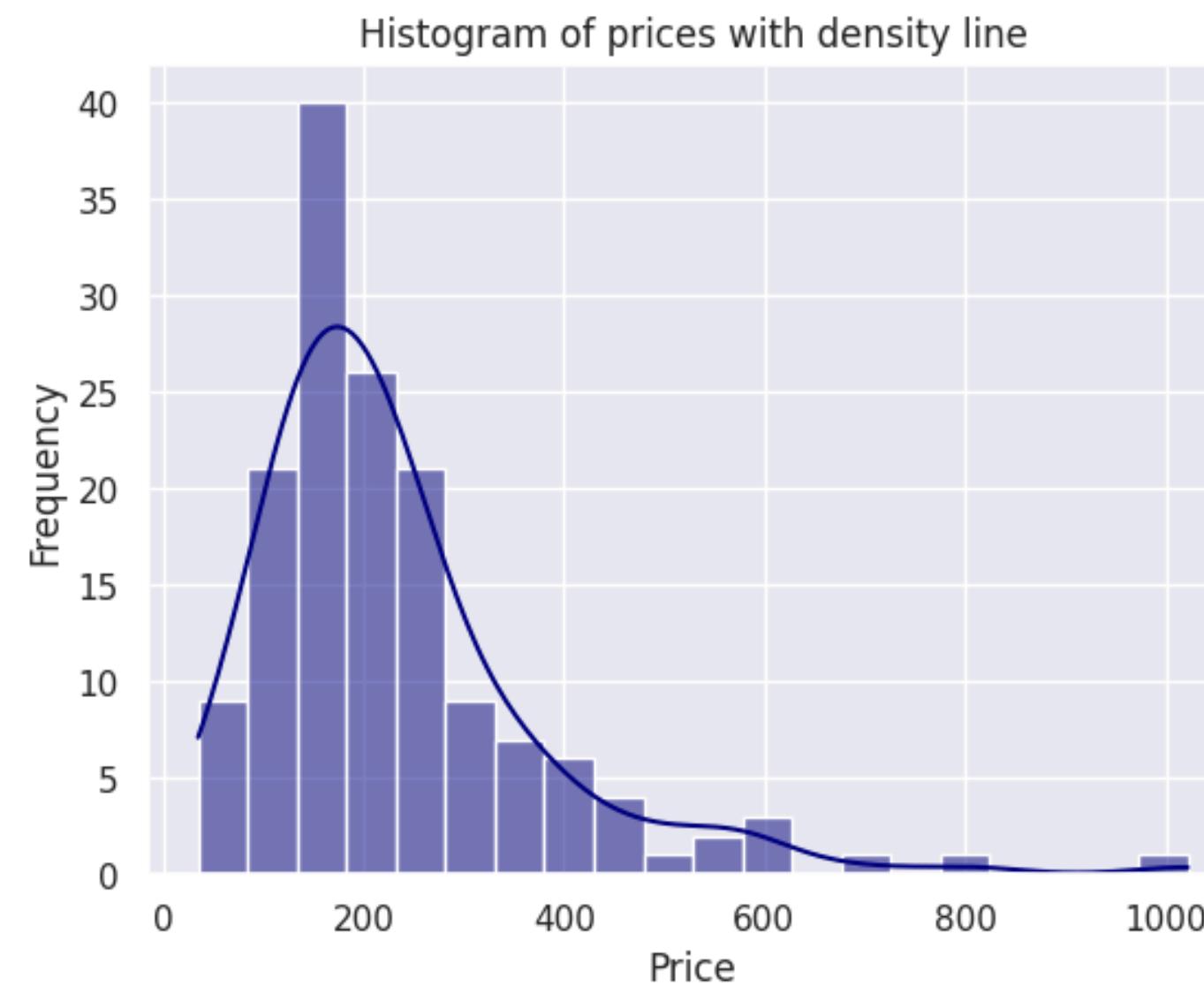
## Correlation matrix:

	Price	Number of reviews	Review category
Price	1.000000	-0.372821	0.139150
Number of reviews	-0.372821	1.000000	0.017792
Review category	0.139150	0.017792	1.000000

Insights:

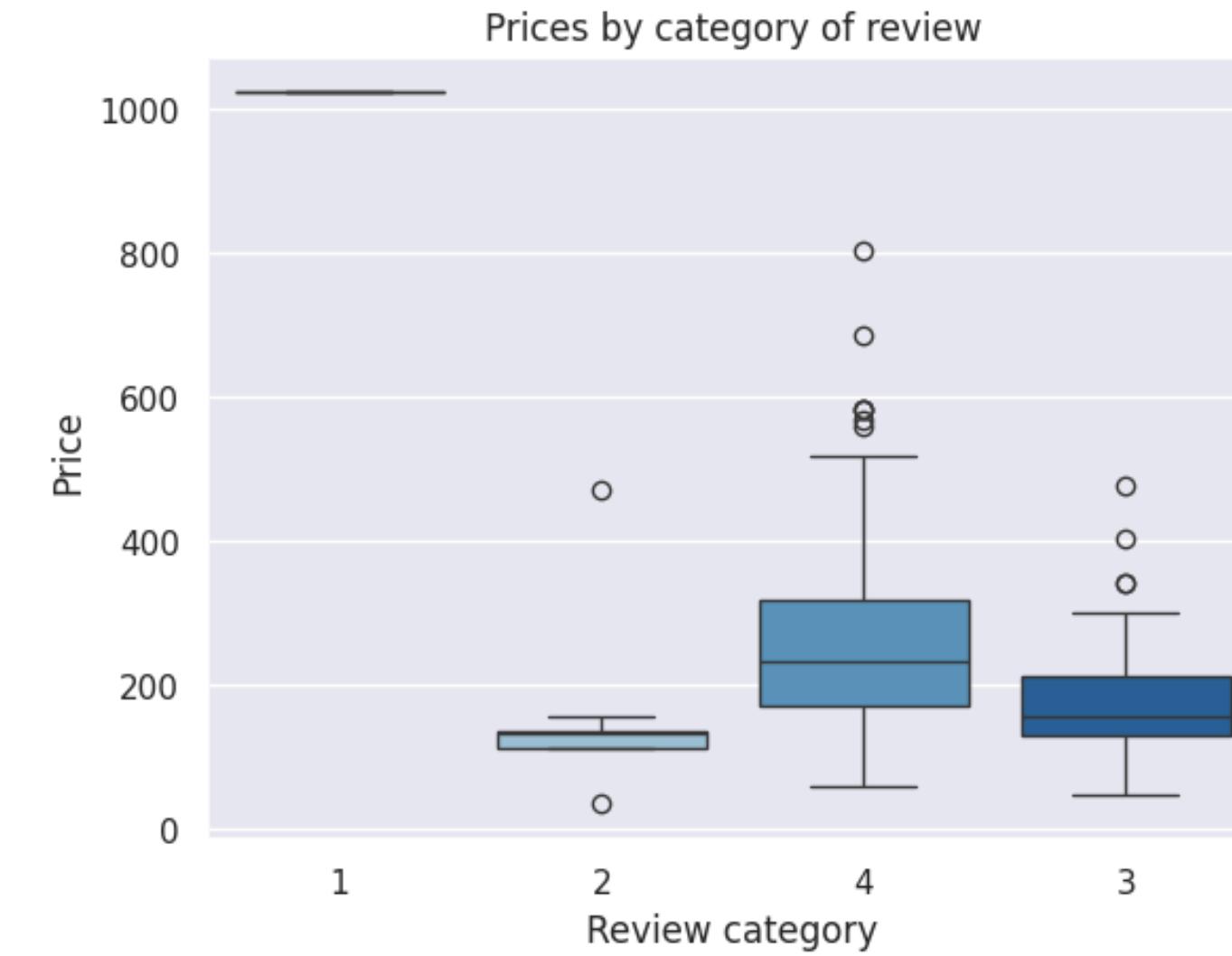
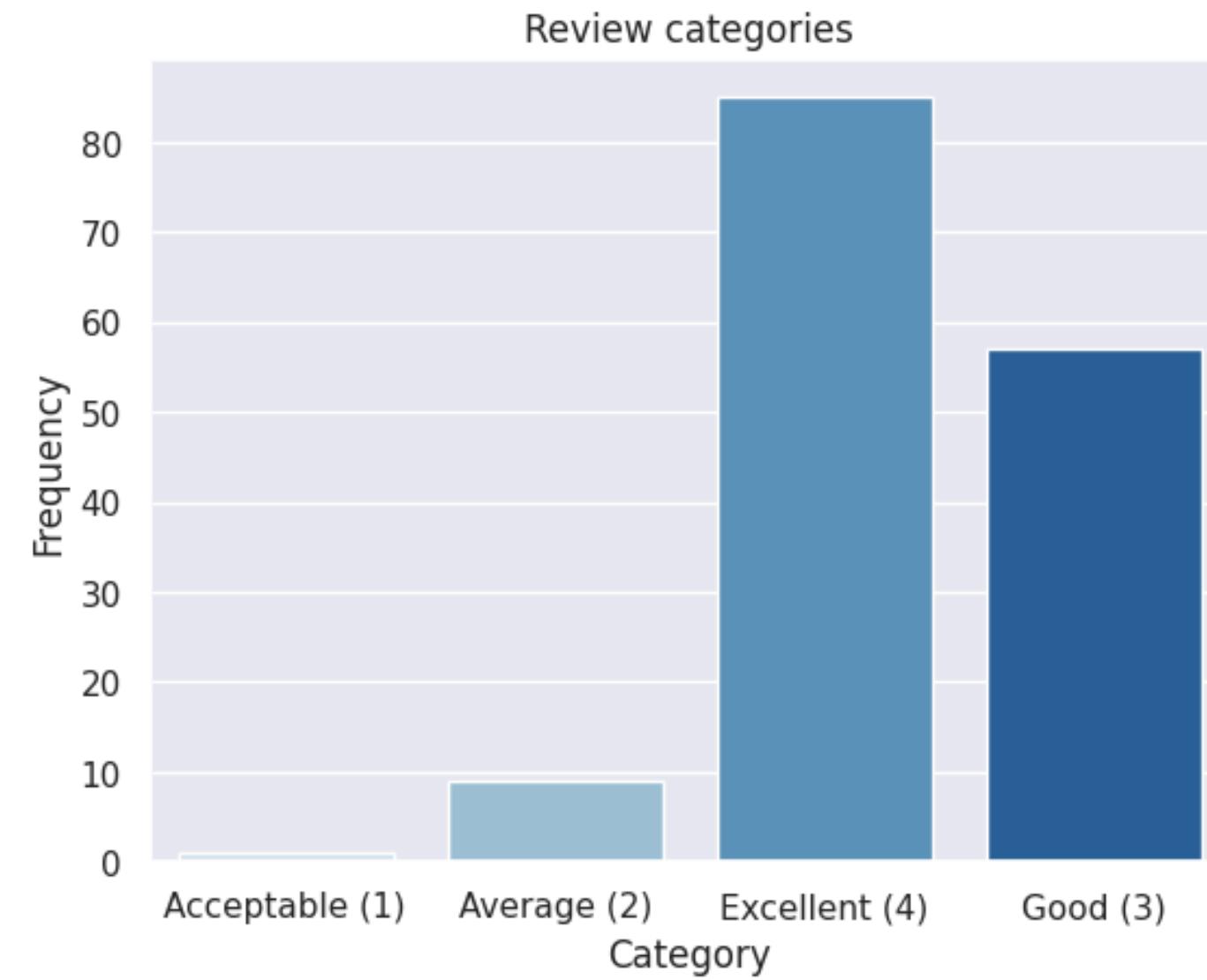
- Majority of hotels have received positive reviews
- There are not significant relationships within the data

# Graphical analysis:



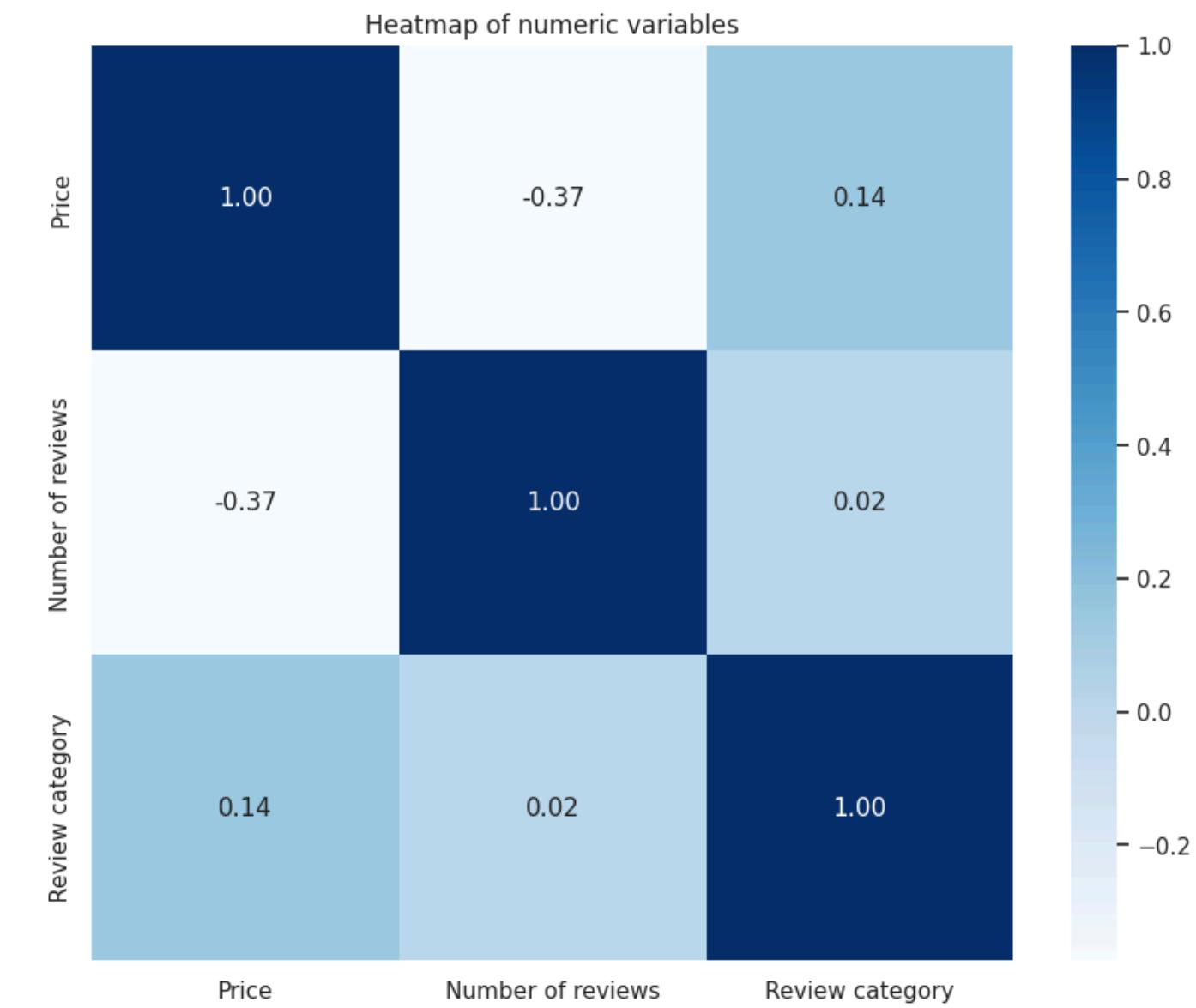
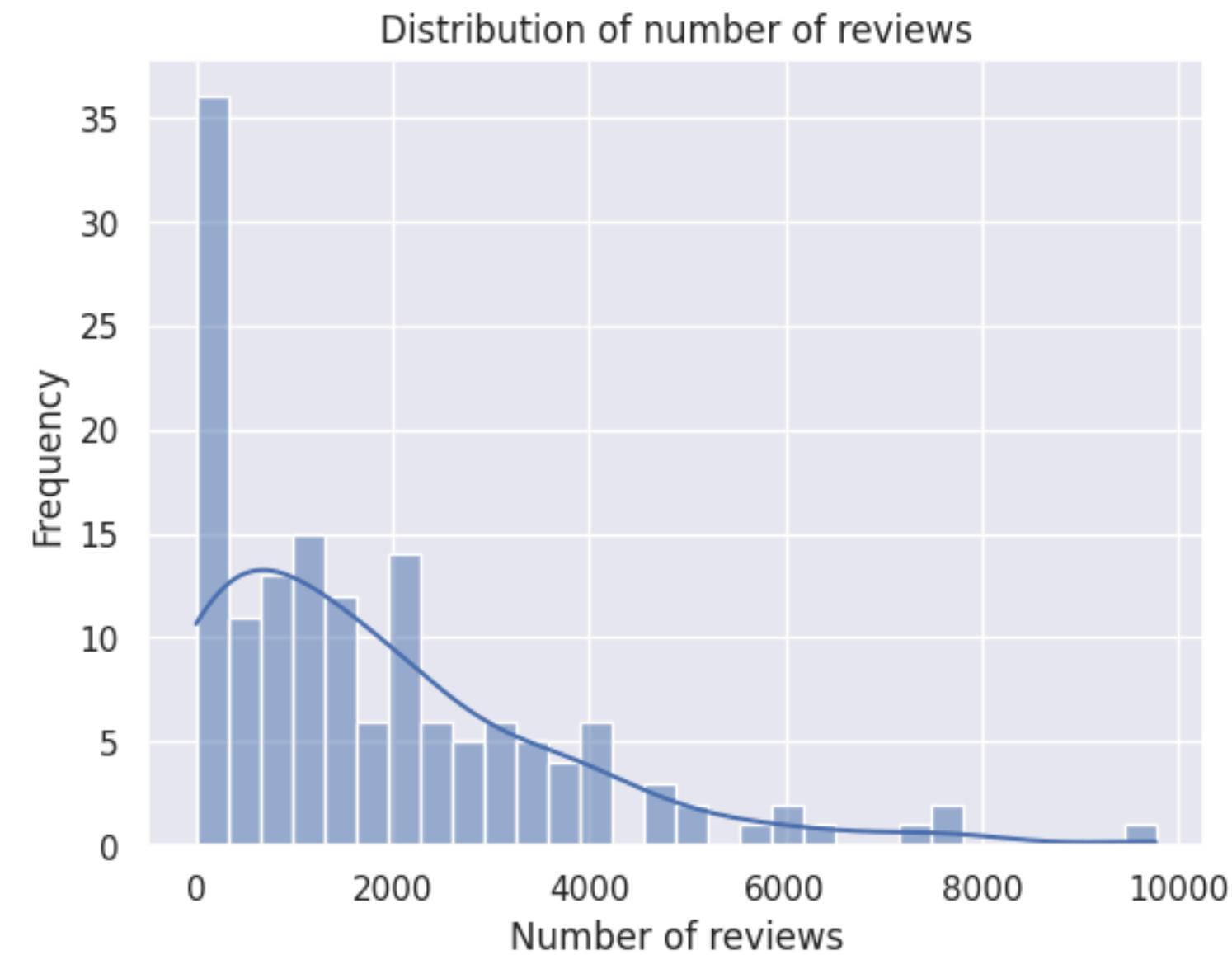
- Majority of hotel prices are clustered in the lower price range (right skewed distribution)

- Several outliers in the higher price range



- Majority of hotels fall into high review categories (3-4 stars), and fewer hotels were found in the lowest review categories

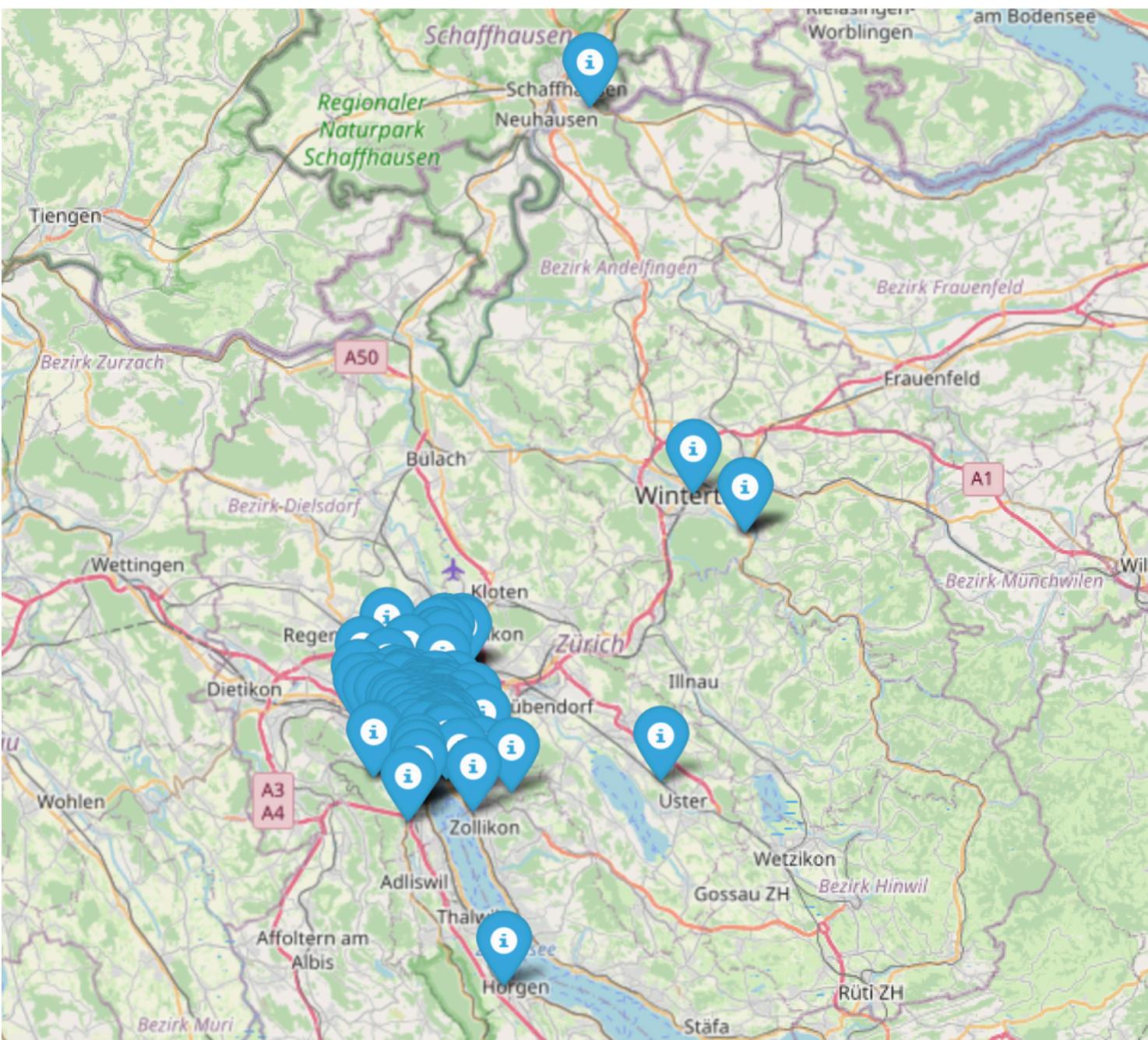
- Hotels with higher review categories generally have higher prices



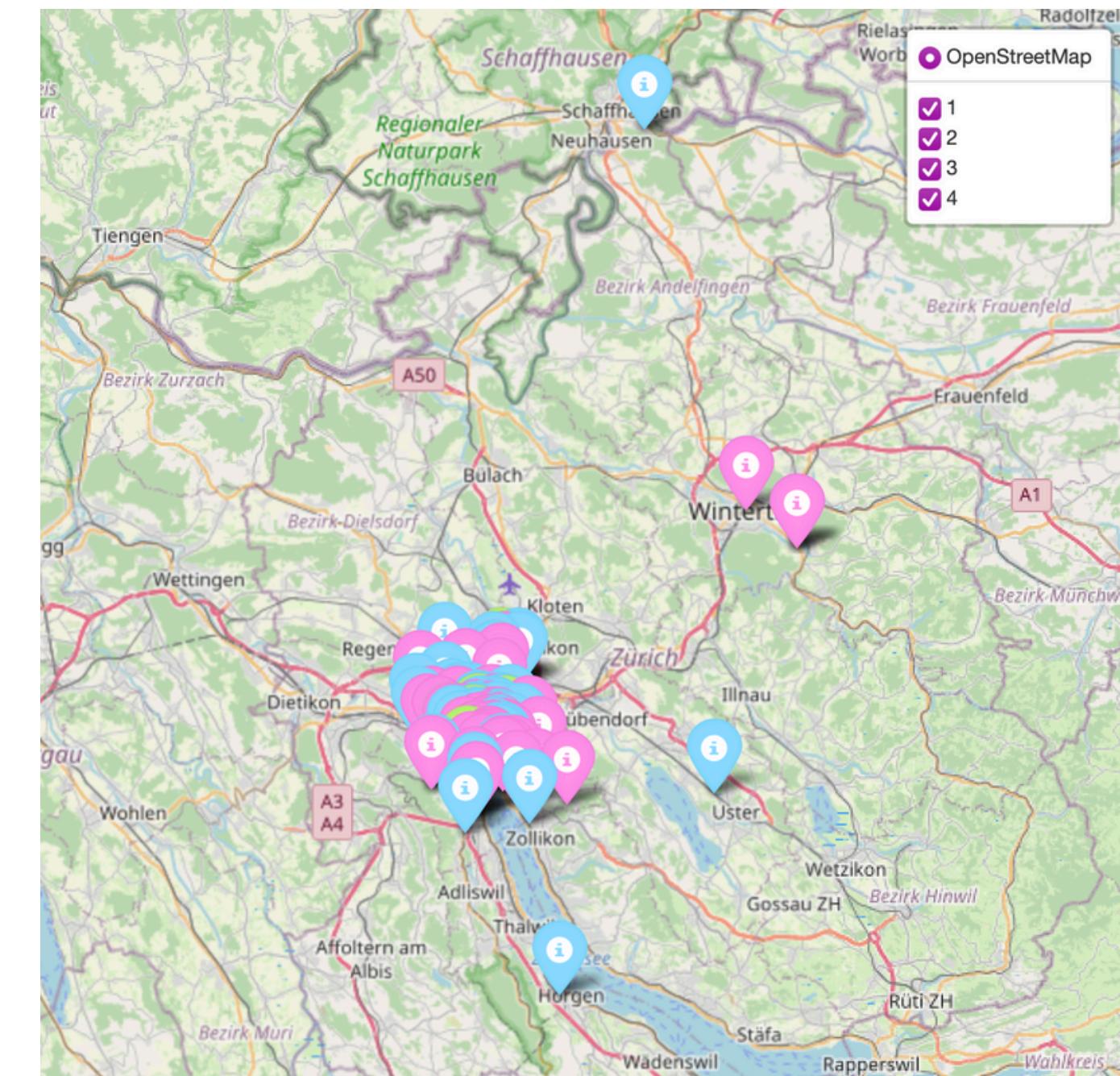
- Most hotels have a relatively low number of reviews, with a sharp decline in frequency as the number of reviews increases (right skewed distribution)
- Correlation heatmap with no meaningful relationships

# Geocoding geopy library

Location of hotels



Location of hotels by category of review



Geographical data (1pt.)

# Modelling methods

## ANOVA

To examine whether there are statistically significant differences in the average hotel price across four distinct review categories

### **Hypothesis:**

***H<sub>0</sub>*:** There is no significant difference in the average prices between the four review categories.

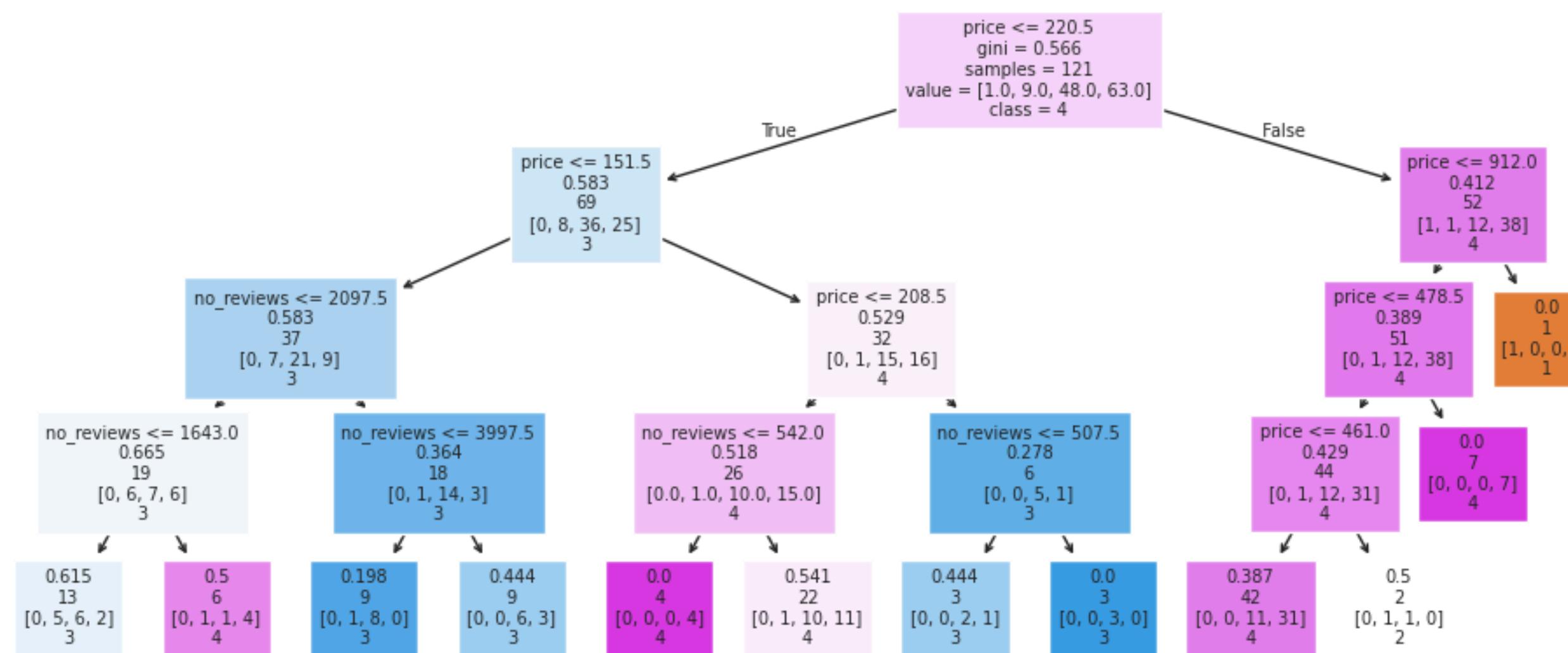
***H<sub>1</sub>*:** At least one review category has a significantly different average price compared to the others.

Result: p-value < 0.05

Classification models (1 pt.)

# Decision tree classifier

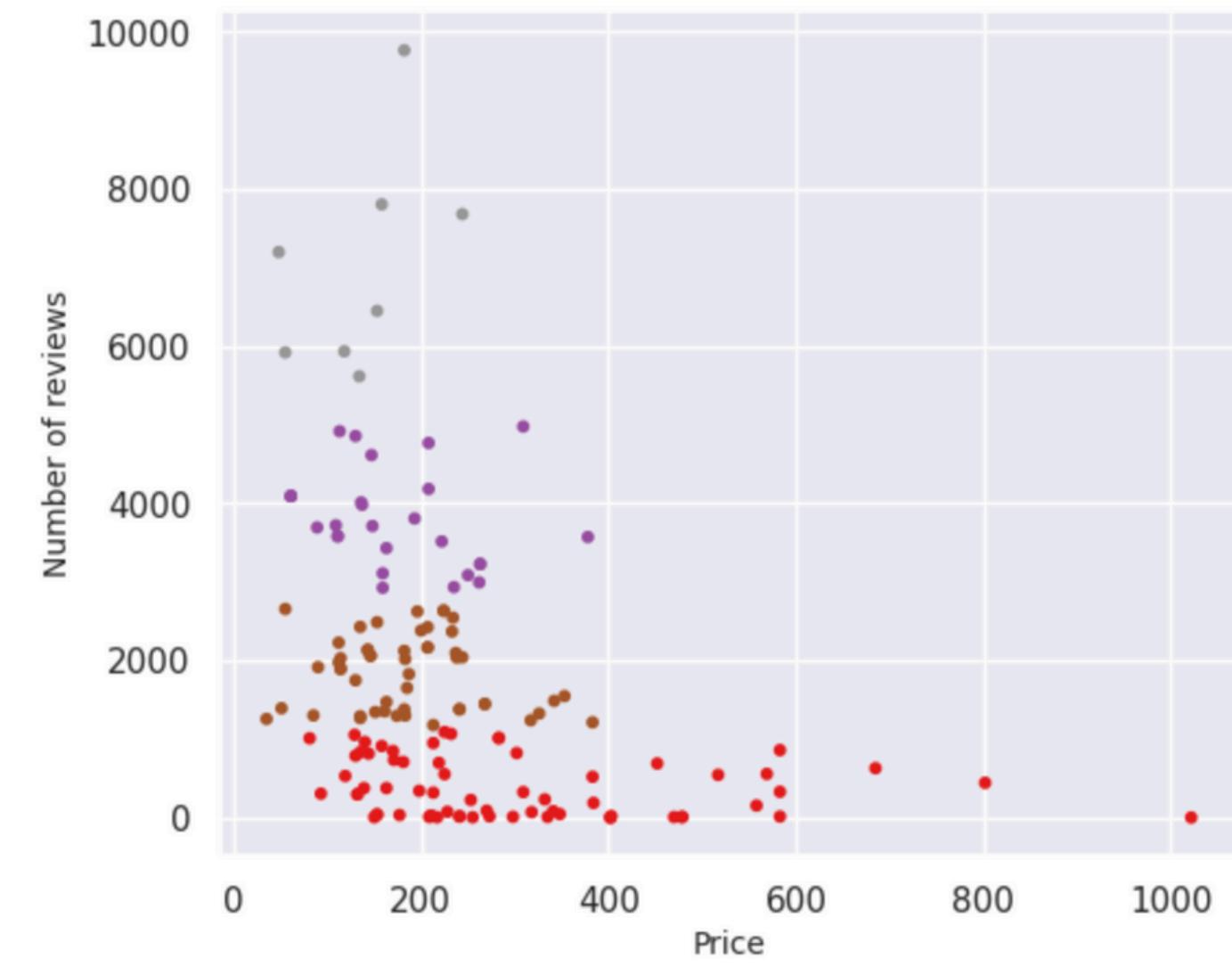
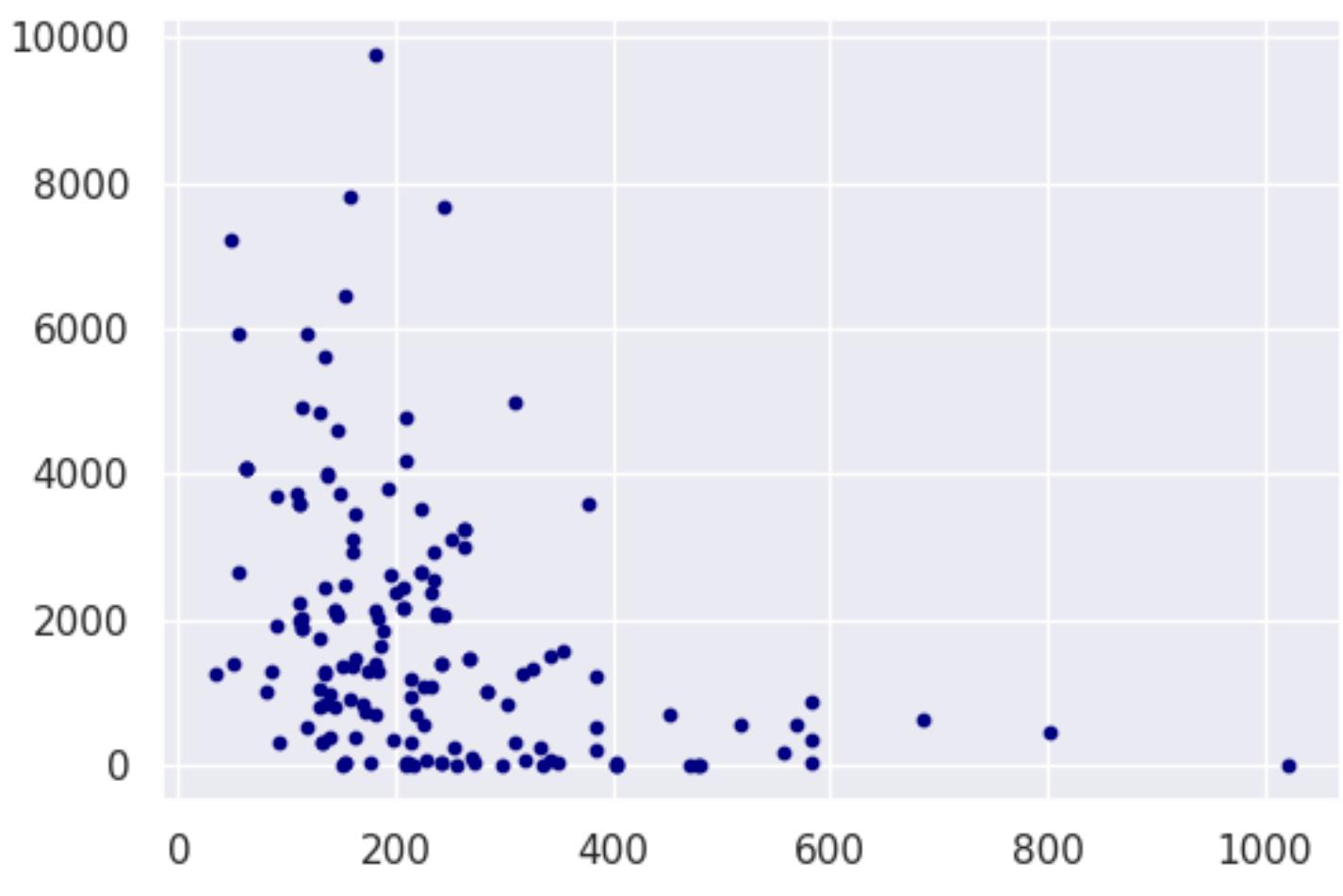
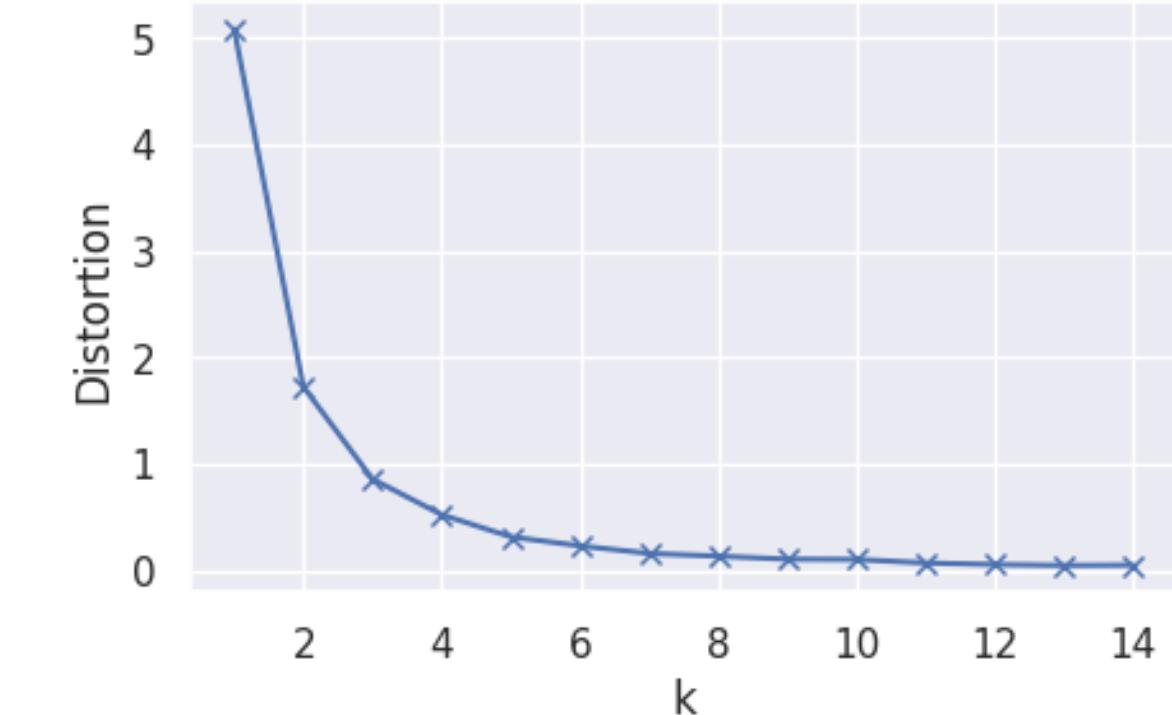
To classify data into categories based on input features. **Number of reviews** and **price** variables were used to predict the categorical target variable: **category of review**



# k-means classifier

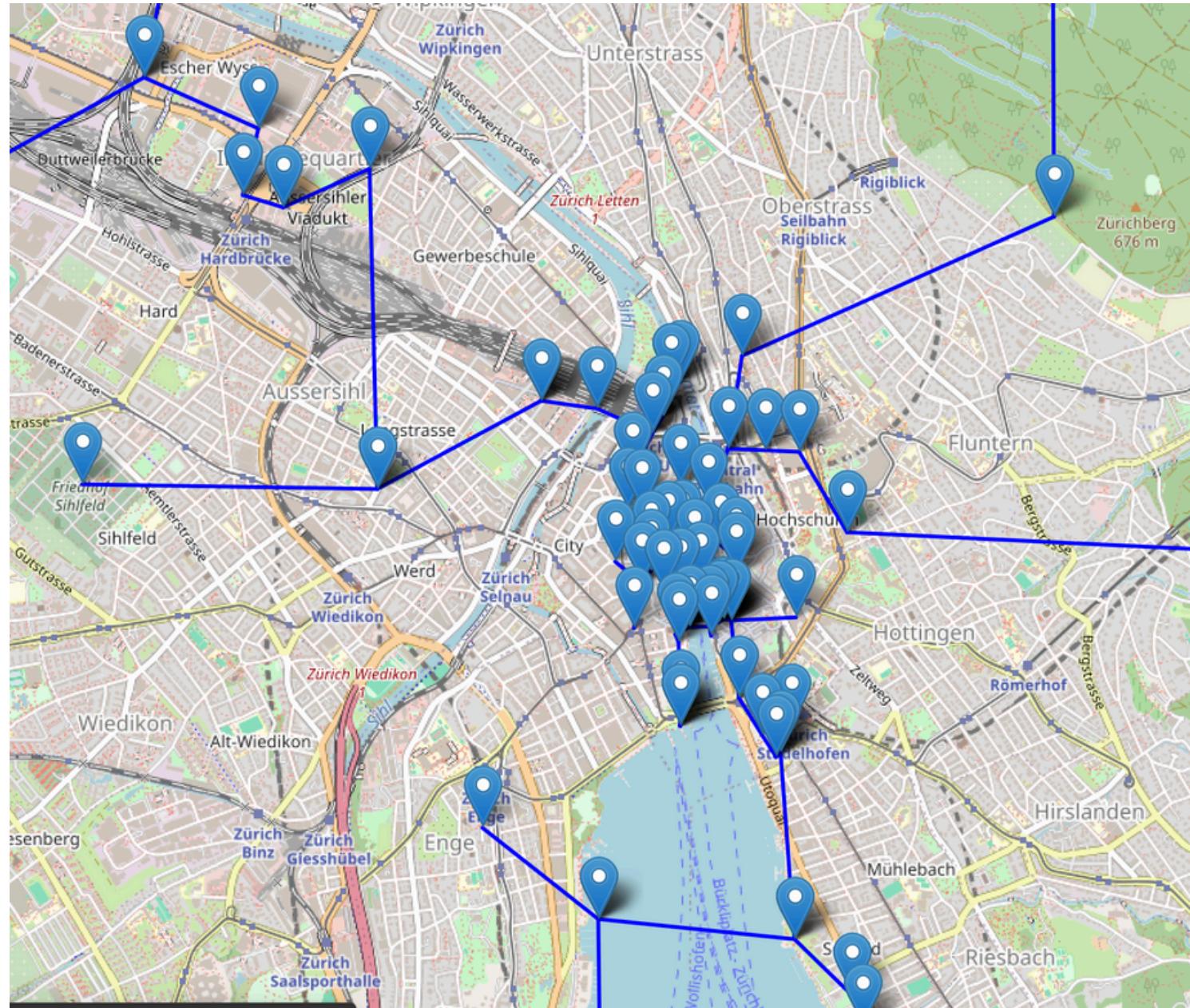
Parts the set of data points into K clusters, where each point belongs to the cluster with the nearest mean.

Elbow method

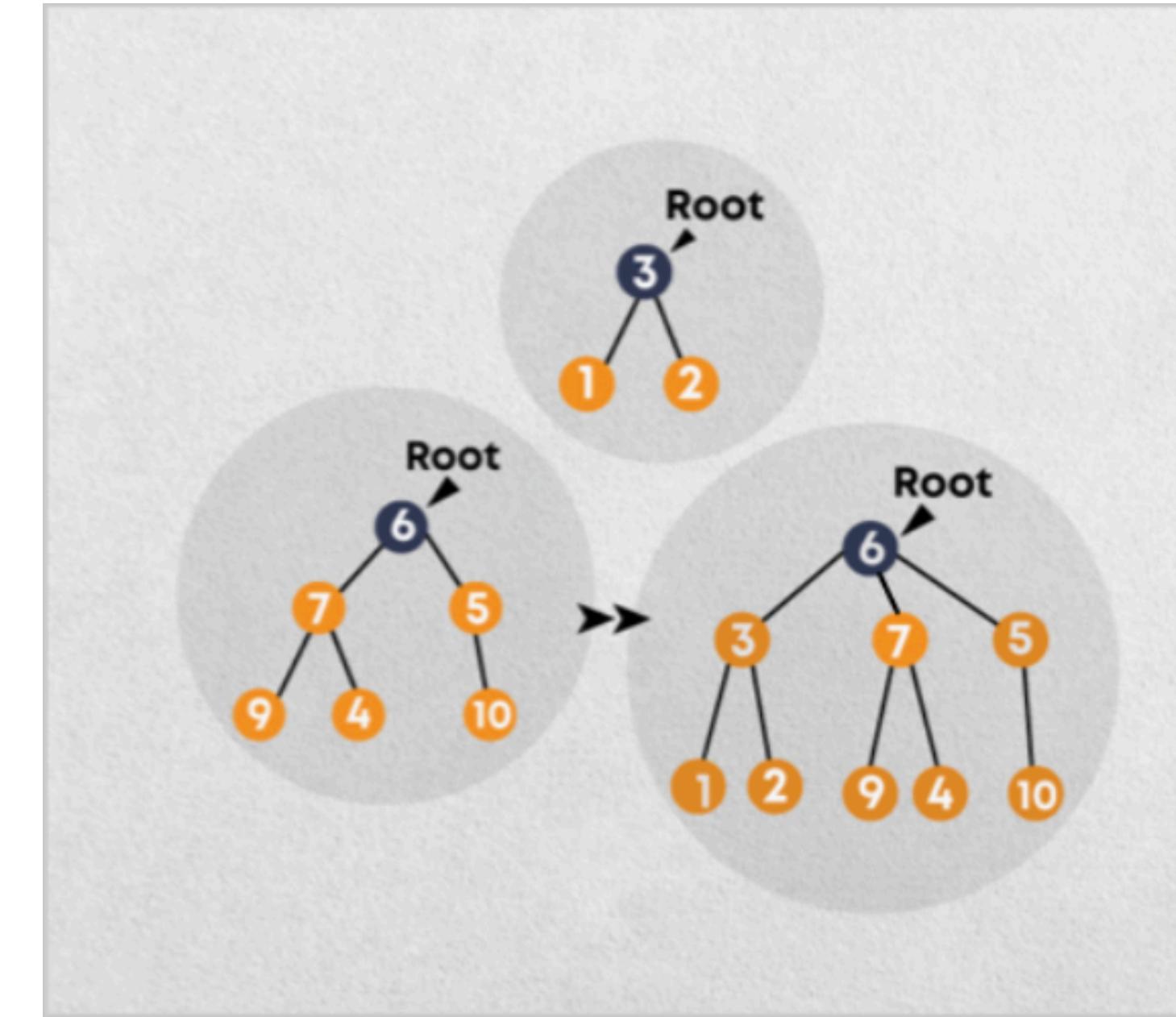


# Distance metrics

## Minimum spanning tree



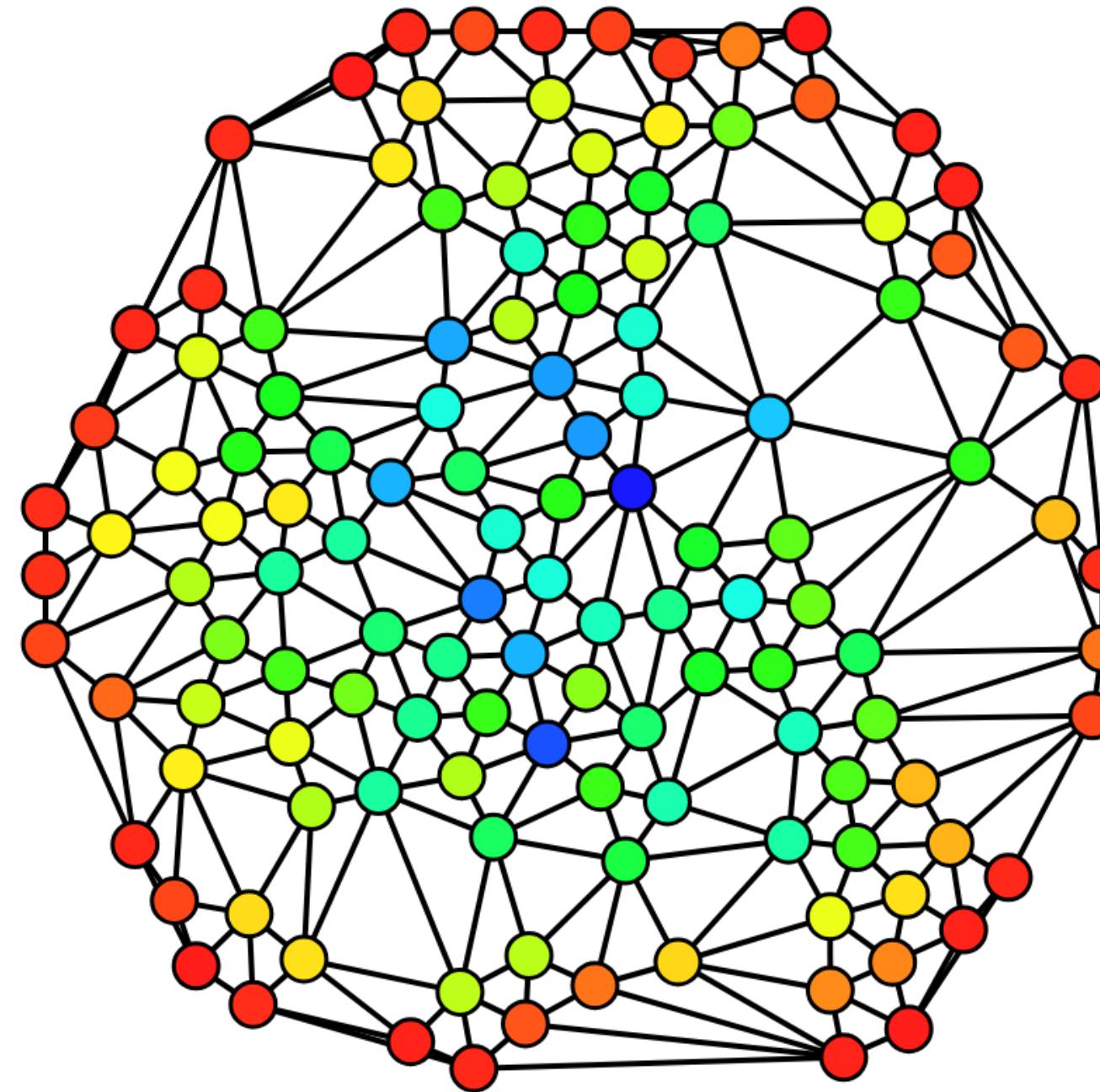
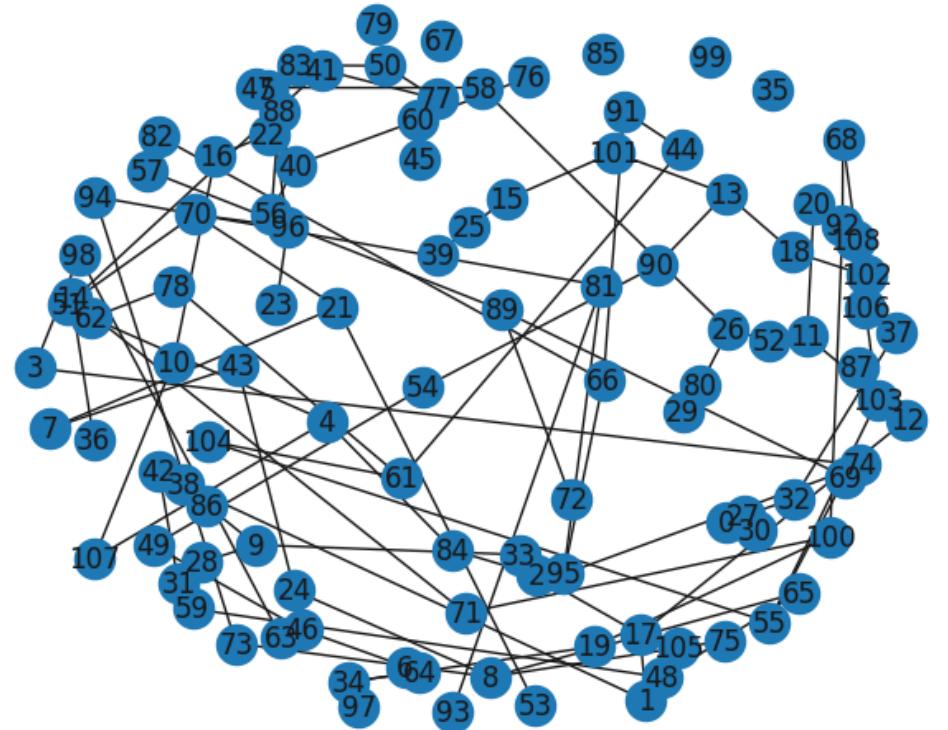
Minimum spanning tree of attractions



# Distance metrics

## Betweenness centrality

- proportion of smallest paths through a point
- Useful in social networks
- Bottlenecks and influential points
- Usually between 0 - 1
- Categorical 10 - 0



# **Results and discussion**

# Decision tree classifier

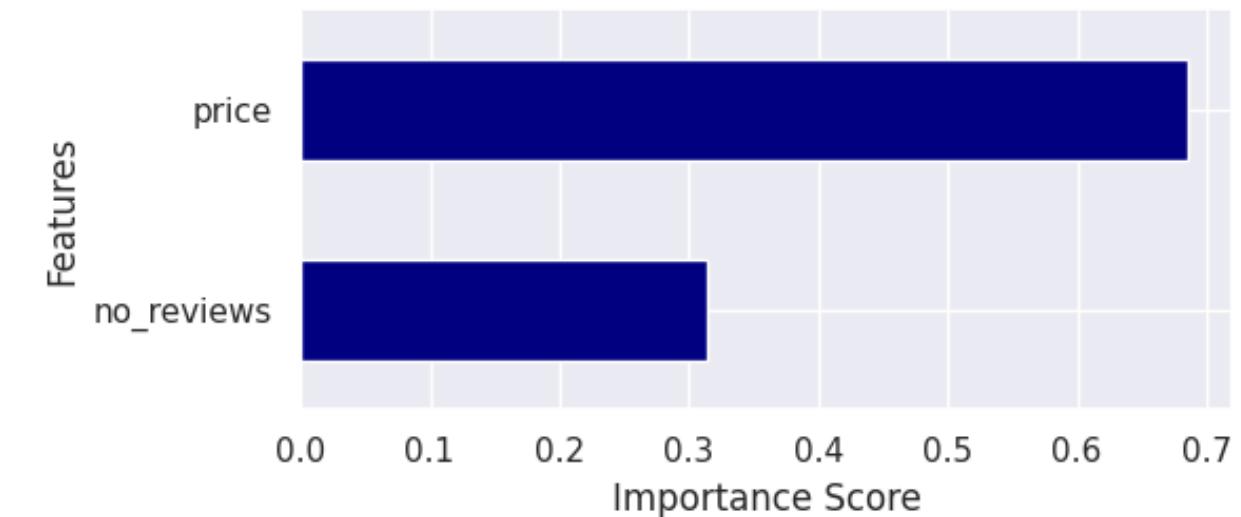
**Classification report:**

	precision	recall	f1-score	support
3	0.64	0.78	0.70	9
4	0.90	0.82	0.86	22
accuracy				31
macro avg	0.77	0.80	0.78	31
weighted avg	0.82	0.81	0.81	31

**Classification report:**

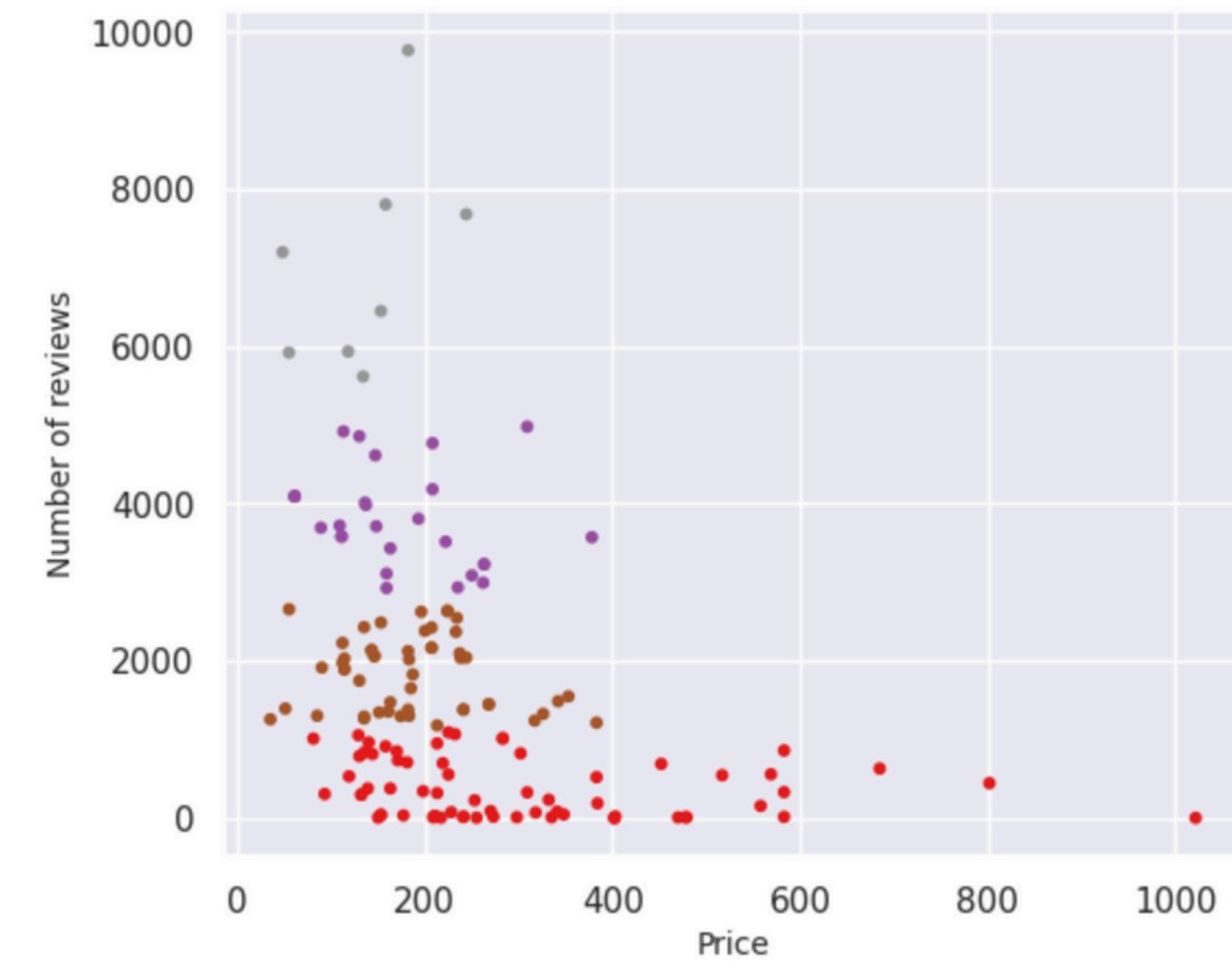
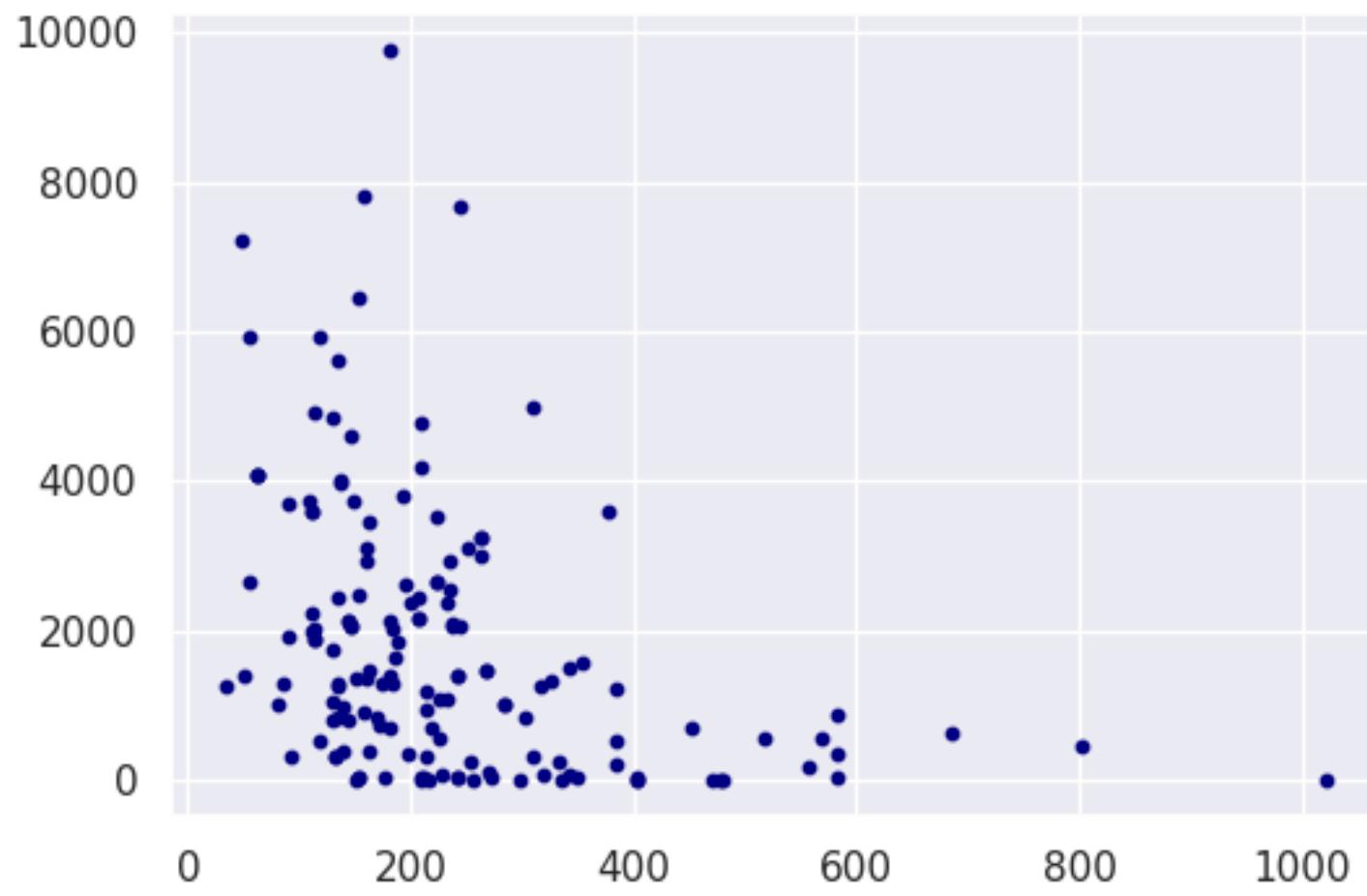
```
[[ 7  2]
 [ 4 18]]
```

**Feature importance:**



- The model correctly predicted 18 instances as class 4 and 7 instances as class 3.
- 64% of the time the prediction was correct for class 3, and for class 4, 90% of the time.
- In an overall performance, the model correctly predicted the class for 81% of the instances

# k-means classifier



**Silhouette score: 0.5726**

- The clusters are moderately well-defined
- The data points have good cohesion
- A higher score closer to 1 would suggest better separation and cohesion

# General improvements

## **Hotels and data analysis:**

- Adjust the models or using techniques such as class balancing to handle underrepresented classes.
- Investigate other features and including additional variables. Including interactions between external datasets could provide more predictive power.
- We are working with less than 20 hotels, so include more data, in order for it to be more meaningful and not having imbalanced classes.

## **Distance metrics:**

- Take into account travel times, or creating a specific route for a visitor through the city for the minimum spanning tree.
- Include big public transportation hubs and include the use of public transportation in the created routes.

# Interactive application

## Streamlit

streamlit app

Fork · · ·

Hotel data analysis

Attraction data analysis

Distance Metrics

### Data Analytics Project:

#### Exploration of Zurich Hotels and Tourist Attractions

Authors:

Adalía Fernanda Aneiros

Diego Arturo González Juárez

Stanislaw Zapala

	name	price	centrality
	Statthotel	182 CHF	★★★★★
	Marktgasse Hotel	163 CHF	★★★★★
	Zuerich Niederdorf - Grossmuenster	228 CHF	★★★★★
	Storchen Zuerich - Lifestyle Boutique Hotel	583 CHF	★★★★★
	Boutique Hotel Wellenberg	263 CHF	★★★★★
	Boutique Hotel Wellenberg	263 CHF	★★★★★
	Hotel Adler Zuerich	207 CHF	★★★★★
	Hotel Adler Zuerich	207 CHF	★★★★★
	Aparthotel Splendid City Center	403 CHF	★★★★★



Creative presentation (1pt.)

# Conclusion

This project provides a comprehensive analysis of Zurich's hotels and tourist attractions using different techniques and analysis methods. By leveraging them, this study uncovered meaningful patterns and insights relevant to Zurich's tourism industry.

To have a real application of the analysis, the geocoding content learned during class, in combination with the minimum spanning tree analysis and betweenness centrality, helped us in the creation of connections between hotels and attractions, to end with a multi-dimensional understanding of Zurich's hotel market and tourist attractions, which demonstrates how advanced data analytics can drive informed decision-making, contributing to a more efficient and customer-centric tourist industry in Zurich

# Bibliography

- Zhao, B. (2022). Web Scraping. In L. A. Schintler & C. L. McNeely (Eds.), Encyclopedia of Big Data (Chapter 483). Cham: Springer. [https://doi.org/10.1007/978-3-319-32010-6\\_483](https://doi.org/10.1007/978-3-319-32010-6_483)
- HomeToGo. (n.d.). Home To Go. Retrieved January 13, 2025, from <https://www.hometogo.ch>
- ZurichTourism. (2022). Sights in the Zurich region (Zurich Tourism). Retrieved from <https://opendata.swiss/de/dataset/sehenswurdigkeiten-in-der-region-zurich-zurich-tourismus>

# Appendix

# Code-snippets

Web scraping:

The screenshot shows a table titled '\_root / Main\_selector' with a 'Data preview' button at the top right. The table has columns: ID, Selector, type, Multiple, Parent selectors, and Actions. There are six rows, each representing a selector with its corresponding details and action buttons (Element preview, Data preview, Edit, Delete). An 'Add new selector' button is located at the bottom left.

Main_selector					
ID	Selector	type	Multiple	Parent selectors	Actions
name	a.hotel__title-url	SelectorText	no	Main_selector	<button>Element preview</button> <button>Data preview</button> <button>Edit</button> <button>Delete</button>
address	p	SelectorText	no	Main_selector	<button>Element preview</button> <button>Data preview</button> <button>Edit</button> <button>Delete</button>
description	div.short-info	SelectorText	no	Main_selector	<button>Element preview</button> <button>Data preview</button> <button>Edit</button> <button>Delete</button>
review_cat	div.hotel__review-text	SelectorText	no	Main_selector	<button>Element preview</button> <button>Data preview</button> <button>Edit</button> <button>Delete</button>
price	div.hotel__price-number	SelectorText	no	Main_selector	<button>Element preview</button> <button>Data preview</button> <button>Edit</button> <button>Delete</button>
no_reviews	div.hotel__review-number	SelectorText	no	Main_selector	<button>Element preview</button> <button>Data preview</button> <button>Edit</button> <button>Delete</button>

Data preparation:

```
# Change 'no_reviews' to numeric first by unwanted characters and words using regular expressions and
# then changing it from a string to a numeric data type

df_hotels['no_reviews'] = df_hotels['no_reviews'].str.replace(r'\(|\)|reviews|review', '', regex=True)
df_hotels['no_reviews'] = pd.to_numeric(df_hotels['no_reviews'])
df_hotels.head()
```

Python

```
# Encoding of categorical variable using a dictionary
dicc_review = {'Acceptable' : 1, 'Average' : 2, 'Good' : 3, 'Excellent' : 4}
df_hotels['review_cat'] = df_hotels['review_cat'].map(dicc_review)
df_hotels.head()
```

Python

```
# Apply the function to our data using a lambda function
df_hotels[['Latitude', 'Longitude']] = df_hotels['address'].apply(lambda loc: pd.Series(get_coordinates(loc)))
df_hotels.head()
```

Python

## Data storage Hotels df (sql):

DA-Project\da-project\hotels\_v1 - HeidiSQL 12.6.0.6765

Archivo Editar Buscar Consulta Herramientas Ir a Ayuda

Filtro de base Filtro de tabla Host: 127.0.0.1 Base de datos: da-project Tabla: hotels\_v1 Datos Consulta\* Consulta

DA-Project

case\_study\_1  
case\_study\_2  
da-project 1.7 MiB  
data\_attract... 1.5 MiB  
hotels\_v1 160.0 KiB  
exam2  
indexes\_transa..  
information\_s...  
mysql  
performance\_...  
sqlintro  
sys

da-project.hotels\_v1: 294 filas en total (exact)

#	-	web_scraper_order	web_scraper_start_url	name	address	description	review_cat	price
1	0	1732027449-1	https://www.hotels-zurich.com/search/?sort=...	Zurich Youth Hostel	Mutschellenstrasse 114A, Zurich, Switzerland	Offering 421 rooms, the 2-star Zurich Youth Ho...	Excellent	61 CHF
2	1	1732027449-2	https://www.hotels-zurich.com/search/?sort=...	Saint Georges Hotel	Weberstrasse 11, Zurich, Switzerland	Located 15 km from Kloten airport, the 2-star ...	Excellent	114 CHF
3	2	1732027449-3	https://www.hotels-zurich.com/search/?sort=...	Ibis Zuerich City West	Schiffbaustrasse 11, Zurich, Switzerland	A 10-minute drive from Old Town, the 3-star I...	Good	143 CHF
4	3	1732027449-4	https://www.hotels-zurich.com/search/?sort=...	Hotel Seehof	Seehofstrasse 11, Zurich, Switzerland	Featuring a currency exchange and a lift, the 3...	Good	213 CHF
5	4	1732027449-5	https://www.hotels-zurich.com/search/?sort=...	Inside Five	Motorenstrasse 11, Zurich, Switzerland	Located about a 5-minute walk from Markthall...	Good	112 CHF
6	5	1732027449-6	https://www.hotels-zurich.com/search/?sort=...	Swiss Star District 11 - Self Check-In	Friesstrasse 33, Zurich, Switzerland	Situated just 13 minutes' walk from Messe Zuri...	Average	132 CHF
7	6	1732027449-7	https://www.hotels-zurich.com/search/?sort=...	Cozy Urban Studio In The City Be-11	Bertastrasse 15, Zurich, Switzerland	12 m² Cozy Urban Studio In The City Be-11 is...	Good	150 CHF
8	7	1732027452-8	https://www.hotels-zurich.com/search/?sort=...	Hotel Adler Zuerich	Rosengasse 10, Zurich, Switzerland	Offering a restaurant and a storage room, the ...	Excellent	207 CHF
9	8	1732027452-9	https://www.hotels-zurich.com/search/?sort=...	Sheraton Zuerich Hotel	Pfingstweidstrasse 100, Zurich, Switzerland	The 4-star Sheraton Zurich Hotel is located in t...	Good	283 CHF
10	9	1732027452-10	https://www.hotels-zurich.com/search/?sort=...	Boutique Hotel Wellenberg	Niederdorfstrasse 10, Zurich, Switzerland	Boutique Hotel Wellenberg Zurich is located in ...	Excellent	263 CHF
11	10	1732027452-11	https://www.hotels-zurich.com/search/?sort=...	Sorell Hotel St. Peter	In Gassen 10, Zurich, Switzerland	Located within 5 minutes' walk of the Romanes...	Excellent	268 CHF
12	13	1732027452-14	https://www.hotels-zurich.com/search/?sort=...	Leonardo Boutique Hotel Righihof Zurich	Universitatstrasse 101, Zurich, Switzerland	Located within 25 minutes' walk of Limmat Riv...	Good	137 CHF
13	14	1732027452-15	https://www.hotels-zurich.com/search/?sort=...	Hotel Hottingen	Hottingerstrasse 31, Zurich, Switzerland	Offering a bureau de change and a vending m...	Excellent	140 CHF
14	15	1732027452-16	https://www.hotels-zurich.com/search/?sort=...	Swiss Night By Fassbind	Steinwiesstrasse 8-10, Zurich, Switzerland	The 3-star Swiss Night By Fassbind Hotel Zuric...	Excellent	146 CHF
15	17	1732027452-18	https://www.hotels-zurich.com/search/?sort=...	Gaestehaus Hunziker	Hagenholzstrasse 104B, Zurich, Ch, Zurich, Swi...	Located 4.4 km from Lindenhof hill, which offer...	Good	181 CHF
16	19	1732027454-20	https://www.hotels-zurich.com/search/?sort=...	Hotel Seidenhof	Sihlstrasse 9, Zurich, Switzerland	The Hotel Seidenhof is part of the Sorell Hotels...	Excellent	237 CHF
17	20	1732027454-21	https://www.hotels-zurich.com/search/?sort=...	Sorell Hotel Rex	Weinbergstrasse 92, Zurich, Switzerland	Sorell Hotel Rex Zurich is located in the Unterst...	Good	153 CHF
18	21	1732027454-22	https://www.hotels-zurich.com/search/?sort=...	Sorell Hotel Righblick - Studios & Spa Suites	Germaniastrasse 99, Zurich, Switzerland	The 4-star Sorell Hotel Righblick - Studios & Sp...	Good	253 CHF
19	22	1732027454-23	https://www.hotels-zurich.com/search/?sort=...	Locke Am Platz Zurich	Tessinerplatz 9, Zurich, Switzerland	Offering location approximately a 5-minute wal...	Excellent	232 CHF
20	23	1732027454-24	https://www.hotels-zurich.com/search/?sort=...	Numa I Craft Apartments	9 Baeckerstrasse, Zurich, Switzerland	Numa I Craft Apartments Zurich hotel has 18 r...	Excellent	225 CHF
21	25	1732027456-26	https://www.hotels-zurich.com/search/?sort=...	Zurich Youth Hostel	Mutschellenstrasse 114A, Zurich, Switzerland	Offering 421 rooms, the 2-star Zurich Youth Ho...	Excellent	61 CHF
22	27	1732027456-28	https://www.hotels-zurich.com/search/?sort=...	Swiss Night By Fassbind	Steinwiesstrasse 8-10, Zurich, Switzerland	The 3-star Swiss Night By Fassbind Hotel Zuric...	Excellent	146 CHF
23	28	1732027456-29	https://www.hotels-zurich.com/search/?sort=...	Statthotel	Schneggengasse 8, Zurich, Switzerland	Statthotel Zurich hotel offers 13 rooms in the v...	Good	182 CHF
24	29	1732027456-30	https://www.hotels-zurich.com/search/?sort=...	Swiss Star Tower - Self Check-In	Friesstrasse 8, Zurich, Switzerland	Swiss Star Tower - Self Check-In Apartment f...	Average	112 CHF
25	30	1732027456-31	https://www.hotels-zurich.com/search/?sort=...	Swiss Star Zurich University - Self Check-In	Universitaetstrasse 80/82, Zurich, Switzerland	Swiss Star Zurich University - Self Check-In, lo...	Average	158 CHF
26	31	1732027456-32	https://www.hotels-zurich.com/search/?sort=...	Cozy Urban Studio In The City Be31	Bertastrasse 15, Zurich, Switzerland	Situated in the Wiedikon district, 1-room Cozy ...	Good	149 CHF

Filtro: Expresión regular

# Data storage Attractions df (sql):

DA-Project\da-project\data\_attractions\_v5 - HeidiSQL 12.6.0.6765

Archivo Editar Buscar Consulta Herramientas Ir a Ayuda

Filtro de base Filtro de tabla ★ Host: 127.0.0.1 Base de datos: da-project Tabla: data\_attractions\_v5 Datos Consulta\* ⚡

DA-Project

da-project.data\_attractions\_v5: 432 filas en total (exact)

#	_type	identifier	copyrightholder	license	category	name	description	titleteaser
1 0	LocalBusiness	1,011,454	Zurich Tourism www.zuerich.com	CC BY-SA		Jungfraujoch with the Zürich Card	<p>Smart travelers can save 15% on a day tri...	Jungfraujoch with the Zürich Card
2 1	LocalBusiness	1,011,189	Zurich Tourism www.zuerich.com	CC BY-SA		Urban Art at Zurich Airport	<h3>Onur Dinc<p>The artist, who grew up in ...	Urban Art at Zurich Airport
3 2	SportsActivityLocation	1,010,518	Zurich Tourism www.zuerich.com	CC BY-SA		Orelliweg	<p>The Rigiblick funicular takes you up to the ...	Orelliweg
4 3	CivicStructure	1,001,309	Zurich Tourism www.zuerich.com	CC BY-SA		Bronze City Model on Stadthausquai	<p>Thanks to this model, both visually impaire...	Bronze City Model
5 4	TouristAttraction	1,001,235	Zurich Tourism www.zuerich.com	CC BY-SA		Polybahn	<p>The &ldquo;Polyb&auml;hnli&rdquo;, as t...	Polybahn
6 5	Church	552,721	Zurich Tourism www.zuerich.com	CC BY-SA		Enge Church	<p>The Reformed Enge Church is well worth ...	Enge Church
7 5	TouristAttraction	552,715	Zurich Tourism www.zuerich.com	CC BY-SA		Sihlfeld Cemetery	<p>Originally planned as a &ldquo;central ce...	Sihlfeld Cemetery
8 7	TouristAttraction	543,214	Zurich Tourism www.zuerich.com	CC BY-SA		Archaeological Windows	<p>In an inconspicuous alley, in an undergrou...	Archaeological Windows
9 8	Church	538,186	Zurich Tourism www.zuerich.com	CC BY-SA		Cloister (Grossmünster)	<p>Everyone who visits Zurich is familiar with ...	Cloister (Grossmünster)
10 9	TouristAttraction	391,814	Zurich Tourism www.zuerich.com	CC BY-SA		Winterthur	<p>The former industrial city of Winterthur is ...	Winterthur

#	titleteaser	zurichcard	osm_id	image	photo	datemodified	address	geocc
1	Jungfraujoch with the Zürich Card	TRUE		{'url': 'https://www.zuerich.com/sites/default/f...	[{'url': 'https://www.zuerich.com/sites/default/...]	2024-10-14T15:56	CH Grindelwald 3818 Eigergletscher-Jungfraujo...	
2	Urban Art at Zurich Airport	FALSE		{'url': 'https://www.zuerich.com/sites/default/f...	[{'url': 'https://www.zuerich.com/sites/default/...]	2024-10-14T15:56	CH Zürich-Flughafen 8058	
3	Orelliweg	FALSE	54939277	{'url': 'https://www.zuerich.com/sites/default/f...	[{'url': 'https://www.zuerich.com/sites/default/...]	2024-11-20T16:27	CH Zürich 8044 Orelliweg	
4	Bronze City Model	FALSE		{'url': 'https://www.zuerich.com/sites/default/f...	[{'url': 'https://www.zuerich.com/sites/default/...]	2024-10-14T15:56	CH Zürich 8001 Stadthausquai 17	
5	Polybahn	TRUE		{'url': 'https://www.zuerich.com/sites/default/f...	[{'url': 'https://www.zuerich.com/sites/default/...]	2024-10-14T15:56	CH Zürich 8001 Limmatquai 144	
6	Enge Church	FALSE		{'url': 'https://www.zuerich.com/sites/default/f...		2024-10-14T15:56	CH Zürich 8002 Bluntschlistieg	
7	Sihlfeld Cemetery	FALSE	10634343	{'url': 'https://www.zuerich.com/sites/default/f...		2024-10-14T15:56	CH Zürich 8003 Aemtlerstrasse 151	
8	Archaeological Windows	FALSE		{'url': 'https://www.zuerich.com/sites/default/f...	[{'url': 'https://www.zuerich.com/sites/default/...]	2024-10-14T15:56	CH Zürich 8001 Schillerstrasse 5	
9	Cloister (Grossmünster)	FALSE		{'url': 'https://www.zuerich.com/sites/default/f...	[{'url': 'https://www.zuerich.com/sites/default/...]	2024-10-14T15:56	CH Zürich 8001 Zwingliplatz	
10	Winterthur	FALSE	1682243	{'url': 'https://www.zuerich.com/sites/default/f...	[{'url': 'https://www.zuerich.com/sites/default/...]	2024-10-14T15:56	CH Winterthur 8400	

#	postalcode	streetaddress	telephone	email	url	latitude	longitude	centrality	betweenness_centrality
1	3818	Eigergletscher-Jungfraujoch	+41 33 828 72 33	info@jungfrau.ch	https://www.jungfrau.ch/de-ch/jungfraujoch-t...	46.5475395	7.98517047	0.009259259	0
2	8058				https://www.flughafen-zuerich.ch/de/passagie...	47.450098	8.561898	0.018518519	2
3	8044	Orelliweg				47.38525	8.5623059	0.018518519	3
4	8001	Stadthausquai 17				47.369305	8.54144	0.018518519	0
5	8001	Limmatquai 144			https://www.polybahn.ch/	47.37658	8.544195	0.027777778	4
6	8002	Bluntschlistieg			https://musik-kirche-enge.ch/	47.362322	8.530517	0.009259259	0
7	8003	Aemtlerstrasse 151	+41 44 412 06 40		https://www.stadt-zuerich.ch/prd/de/index/be...	47.375206	8.508275	0.009259259	0
8	8001	Schillerstrasse 5		sekretariat@theol.uzh.ch	https://www.stadt-zuerich.ch/hbd/de/index/st...	47.3649628	8.54683	0.018518519	5
9	8001	Zwingliplatz			https://www.stadt-zuerich.ch/hbd/de/index/im...	47.370129	8.544402	0.037037037	6
10	8400					47.499172	8.72915	0.027777778	0

## Non-graphical and graphical exploratory analysis:

```
# Summary statistics of numerical variables  
df_hotels[['price', 'no_reviews']].describe()
```

Python

```
# Histogram of the 'price' to with a density curve to highlighting their frequency and distribution  
  
sns.histplot(df_hotels['price'], bins=20, color="navy", kde=True)  
plt.xlabel('Price')  
plt.ylabel('Frequency')  
plt.title('Histogram of prices with density line')  
plt.show()
```

Python

## Classification model and evaluation metrics:

```
# Initialize the classification tree model  
clf = DecisionTreeClassifier(random_state=20,  
| | | | | | | max_depth=4)  
  
# Train the classification tree model  
clf = clf.fit(X_train, y_train)  
  
# Make model predictions  
y_pred = clf.predict(X_test)  
y_pred
```

Python

```
# Classification report (metrics)
print('Classification report')
print(classification_report(y_test, y_pred))
```

Python

Creativity (creation of an interactive application using streamlit):

```
# Setup streamlit header
st.set_page_config(page_title='Data Analytics Project')
st.write('# Data Analytics Project')
st.write('Authors:')
st.write('Adalía Fernanda Aneiros')
st.write('Diego Arturo González Juárez')
st.write('Stanislaw Zapala')

# Read Data
hotels_df = pd.read_csv('data\\hotels_v1.csv')
attractions_df = pd.read_csv('data\\data_attractions_v5.csv')

hotels_df['centrality'] = hotels_df.apply(lambda row: '★'*row.distance_score, axis= 1)
hotels_df = hotels_df.sort_values(by= ['total_distance'], ascending=True, ignore_index= True)
display_hotels = hotels_df[['name', 'price', 'centrality']]
```

## Integration and visualization of geographical data:

```
# Initialize the geolocator object using the Nominatim geocoding service
geolocator = Nominatim(user_agent="exercise", timeout=10)

# Function to get coordinates based on location
def get_coordinates(location):
    try:
        geo = geolocator.geocode(location)
        if geo:
            return geo.latitude, geo.longitude
        else:
            return None, None
    except GeocoderTimedOut:
        return None, None
```

Python

```
# Visualization of hotel locations using an interactive Folium map
# https://msticpy.readthedocs.io/en/latest/visualization/FoliumMap.html

mp.init_notebook()
df_hotels.mp_plot.folium_map(
    lat_column="Latitude", long_column="Longitude", zoom_start=10
)
```

Python

## ANOVA:

```
# One-way Anova to determine if there are statistically significant differences in the average price across the four review categories.

# Create subsets (groups)
c1 = df_hotels.loc[df_hotels['review_cat'] == 1]
c2 = df_hotels.loc[df_hotels['review_cat'] == 2]
c3 = df_hotels.loc[df_hotels['review_cat'] == 3]
c4 = df_hotels.loc[df_hotels['review_cat'] == 4]

# Create ANOVA
fvalue, pvalue = stats.f_oneway(
    c1['price'],
    c2['price'],
    c3['price'],
    c4['price'])

print('F-value:', fvalue.round(3), 'p-value', pvalue.round(4))
```

Python

## k-means clustering:

```
# Number of clusters
k = 4

# Perform k-means clustering
kmeans_apmts = KMeans(n_clusters=k, random_state=42).fit(X)

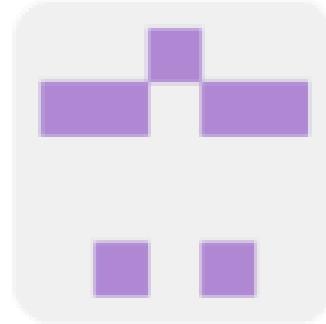
# Add the clusters to data frame
X['cluster'] = kmeans_apmts.predict(X)

# Show number of apartments per cluster
X['cluster'].value_counts().sort_values(ascending=False)
```

Python

Source code of interactive app available in:

**DiegoGlez28/  
Zurich\_Hotels\_App**



1 Contributor 0 Issues 0 Stars 0 Forks

[DiegoGlez28/Zurich\\_Hotels\\_App](#)

Contribute to DiegoGlez28/Zurich\_Hotels\_App development by creating an account on GitHub.

 GitHub