# **Project documentation**

**Laptop Products** 

Hamza Lamrabet
Master Wirtschaftsinformatik
lamrabet@uni-hildesheim.de

# **Objective**

 Identify the most suitable laptop brand by considering both price and customer reviews.



# **Agenda**



**Data Cleaning and Preparation** 

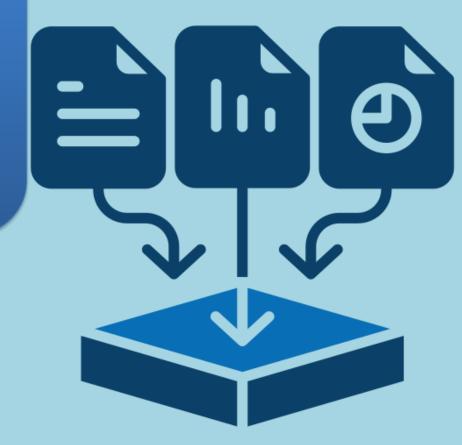
Analysis and Visualizations

Findings and Insights



## **Data Collection**

collecting data from Amazon by scraping information about laptops, which included the laptop name, price, and customer reviews.



### **Tools Used**

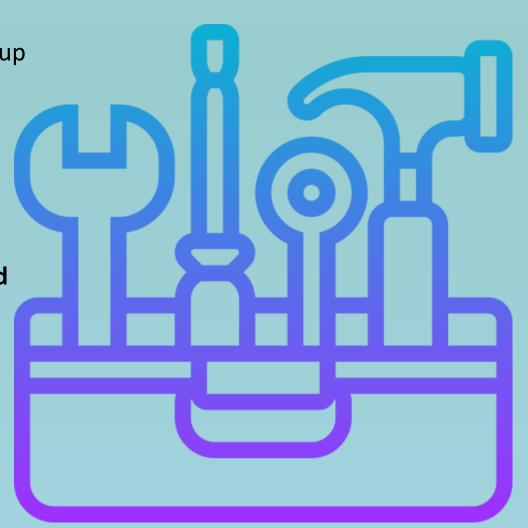
Python libraries such as BeautifulSoup for web scraping.

Pandas for data manipulation.

Requests for making HTTP requests.

CSV for saving the data.

After scraping, the data is organized and saved into a CSV file for easier analysis.



## **Data Cleaning and Preparation**

steps in the data cleaning and preparation phase, which are essential for making the dataset ready for analysis and visualization:

Remove Duplicates: Ensured each record is unique.

 Find and Replace Values: Corrected inconsistencies in data (e.g "Male" vs "male").

 Categorize Data: Grouped values like income into categories (e.g low, medium, high).

Other important cleaning tasks to focus on:

- Handling Missing Data.
- Converting Data Types.



# **Analysis and Visualizations**

The final Power BI report contains:

 A Brand Distribution Bar Chart showing product availability by brand.

 A Rating Gauge that displays the average customer rating.

 A Price Comparison Bar Chart highlighting the average prices for each brand.

 A Summary of Total Reviews to demonstrate the strength of the data.



# **Findings and Insights**

#### **HP**

**33 products**, **4,151 reviews**, **4.41 rating**, **\$1,000+** (Premium, high satisfaction).

#### **Acer**

25 products, 71,000 reviews, 4.31 rating, \$700–\$800 (Affordable, well-rated).

#### Lenovo

19 products, 7,067 reviews, 4.18 rating, \$800–\$900 (Balanced pricing & rating).

#### Dell

14 products, 79 reviews, 3.51 rating, ~\$2,000 (High-end, lower satisfaction).

#### Samsung

10 products, 7,025 reviews, 4.4 rating, ~\$1,000 (Best rating, mid-range).



### **Attachment**

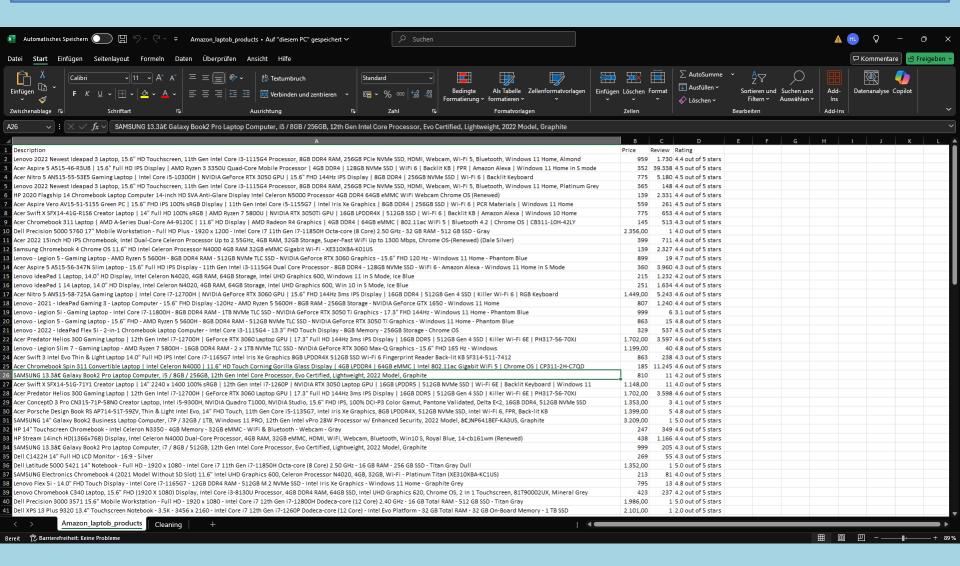
#### The code

```
from bs4 import BeautifulSoup
import pandas as pd
import requests
import csv
URL =
"https://www.amazon.com/s?k=laptop&i=electronics&bbn=172282&rh=n%3A172282%2Cp_89
%3AAcer%7CApple%7CDell%7CHP%7CLenovo%7CSAMSUNG%2Cp_6%3AATVPDKIKX0DER%2Cp_
72%3A1248882011&dc&page=10&crid=1MYVLR12FLK7W&qid=1674856773&rnid=1248877011&
sprefix=lactob%2Caps%2C275&ref=sr pg 10"
HEADERS = ({"User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36
(KHTML, like Gecko) Chrome/109.0.0.0 Safari/537.36", "Accept-Language": "en-US, en;q=0.5"})
webpage =requests.get(URL, headers=HEADERS)
print(webpage)
soup = BeautifulSoup(webpage.content , "html.parser")
print(soup)
durl = soup.find('a',{"class": "a-link-normal s-underline-text s-underline-link-text s-link-style a-
text-normal"})
link = durl.get("href")
link product = "https://amazon.com"+link
print(link_product)
des = soup.find_all("span", {"class": "a-size-medium a-color-base a-text-normal"})
pr = soup.find_all("span", {"class":"a-price-whole"})
rv = soup.find_all("span", {"class" : "a-size-base s-underline-text"})
rt = soup.find all("span", {"class": "a-icon-alt"})
```

```
Description = []
for i in range(len(des)):
  Description.append(des[i].text)
print (Description)
Price = []
for i in range(len(pr)):
  Price.append(pr[i].text)
print(Price)
Review = []
for i in range(len(rv)):
  Review.append(rv[i].text)
print(Review)
Rating = []
for i in range(len(rt)):
  Rating.append(rt[i].text)
print(Rating)
a = {
  "Description": Description,
  "Price": Price,
  "Review": Review.
  "Rating": Rating
df = pd.DataFrame.from_dict(a, orient='index')
df= df.transpose()
df.to_csv("Amazon_laptob_products2")
print ("saved to file")
```

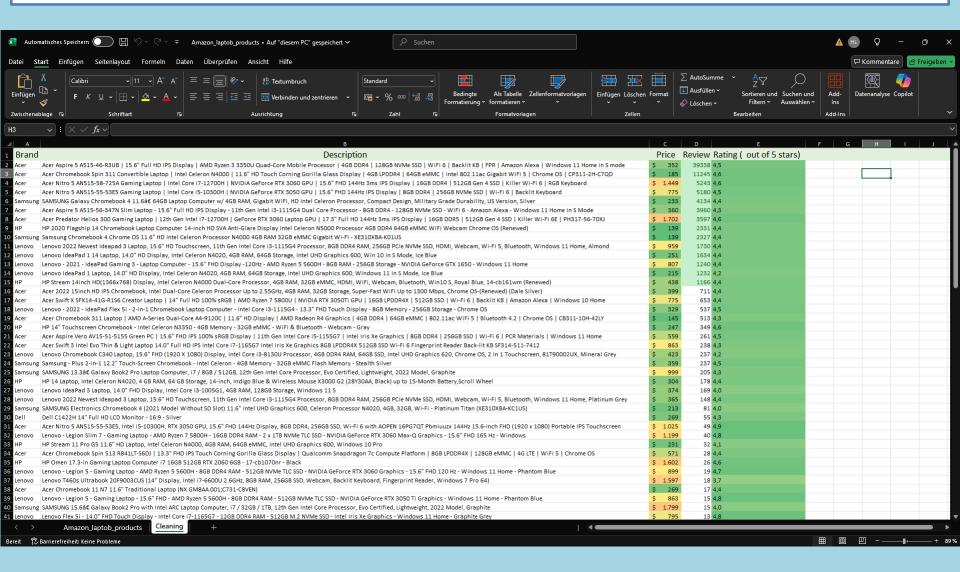
### **Attachment**

#### **Row Dataset**



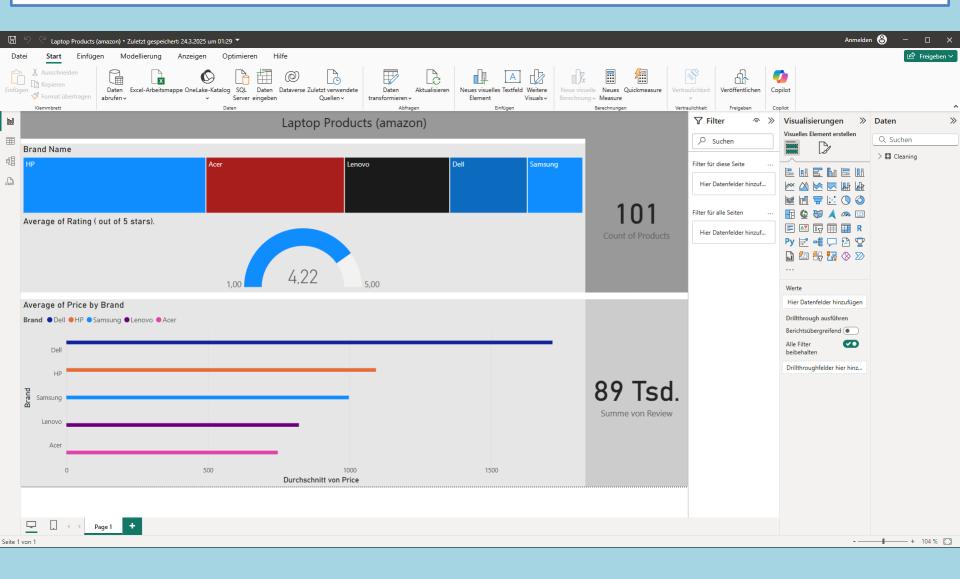
### **Attachment**

### **Data Cleaning and Preparation**



## **Attachment**

### **Analysis and Visualizations**



# Thank you for your attention