# eBay

### eBay Project

#### **INDEX**

#### 1.WEB SCRAPING: EBAY SNEAKER CATEGORY

- 1. Scrape data for all products from the first 5 pages.
- 2. Extract details: product link, product title, price, reviews, likes.

#### 2.DATA HANDLING

- 1. Export scraped data into a CSV file.
- 2. Import data from the CSV file for further analysis.

#### 3.DATA CLEANING/WRANGLING

- 1. Handle missing values.
- 2. Perform feature engineering.
- 3. Replace incorrect or inconsistent values.
- 4. Correct data structure issues.

#### 4.EXPLORATORY DATA ANALYSIS (EDA)

- 1. Analyze the cleaned data to extract insights.
- 2. Visualize key patterns and trends.

### **IMPORT LIBRARIES**

```
import requests

from bs4 import BeautifulSoup

import pandas as pd

import numpy as np

from skimpy import skim

import seaborn as sns

import matplotlib.pyplot as plt

import warnings

warnings.filterwarnings("ignore")

headers = {"User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/91.0.4472.124 Safari/537.36"}
```

# Step-1 Ebay Web-scraping

- 1. Create a list to store product link which are scraped from the ebay website.
- 2. Requests-> to extract html data from website.
- 3. Soup -> to extract particular tags for the product links.
- 4. Using loop extract first 5-pages products link for scraping details like: title, price, review, likes

**NOTE:** each page show only 60 products

```
78]: # creating dictionary to store title and other details
     product_link=[]
     for page_num in range(0,5):
        link = "https://www.ebay.com/sch/i.html?_from=R40&_nkw=sneakers&_sacat=0&_pgn=" + str(page_num) + "&rt=nc"
         page = requests.get(link, "html", headers=headers)
         soup = BeautifulSoup(page.content, 'html.parser')
         # store links in product link list
         all_products = soup.find_all("ul", attrs={"class": "srp-results srp-grid clearfix"})
         for link in all_products:
             a_tags = link.find_all("a", attrs={"class": "s-item_link"})
            for a in a_tags:
                 product_link.append(a['href'])
         print("page",page_num,"link scraped and store in product_link")
     page 0 link scraped and store in product_link
     page 1 link scraped and store in product_link
     page 2 link scraped and store in product_link
     page 3 link scraped and store in product_link
     page 4 link scraped and store in product_link
```

#### Extracted product link and total links extracted

#### product\_link[:4]

['https://www.ebay.com/itm/196549827917?\_skw=sneakers&itmmeta=01JC2DXGFSER5NNVHSCZDCBQTN&hash=item2dc3485d4d%3Ag%3A9hAAAOSw00xmtRGr&itmprp=enc%3AAQAJAAA AwHoV3kP08IDx%2BKZ9MfhVJK1%2BgNn1EQ68tz9pldKmNr%2F1pBw11Ek%2FfyzPcfIiPz6UssIaOIyZzqccUWw2usFJgrvu1TfLbj%2BMfuc6E%2FYj3bBpGOeEHLUayd8g%2F3sbqEAS1UgzODKe0 y%2BPR%2FE996LyTGe3bt6LsOtNLRaU%2FP3OBhCeWmZ4EVpbIvSKqssCoOjhqFCmcOrUOHGWOLDF7E%2BqBfXd6Dkh%2BNK%2By%2BYoD%2Befk2q2WpSgrEPyHgd%2BQiMOUPTOMw%3D%3D%7Ctkp% 3AB1BMUICI9s3gZA&var=496318587386',

https://www.ebay.com/itm/176513863714?\_skw=sneakers&itmmeta=01JC2DXGFSN3N1ENND83FPND3H&hash=item29190bd022%3Ag%3ANfUAAOSwK81mtRMX&itmprp=enc%3AAQAJAAA AwHoV3kP08IDx%2BKZ9MfhVJKngBtFty1%2FsGFot8WTribDx9OtcC0AUTf902w7OmW59RvCiHtXs41Y6t7VoFPwb%2Fy1uMIXV0Zy6rC8dTnYyaINJsFHLY0alQ%2BBafiDZEMe4NXo8W10gsGHEAq% 2Fm8GCiNlqscdbQMfurn7Mjg0zV6KOMTSs3vZxB5q1%2B1kQ9FD2AILladAvWQXHc59KIxkFQ9n9AaqNuQW16kaWFln9r3D%2BVS1T4V1xHUmePQ%2BLf2SGiZA%3D%3D%7Ctkp%3AB1BMUICI9s3gZA &var=476111946700',

'https://www.ebay.com/itm/296765559816?\_skw=sneakers&itmmeta=01JC2DXGFSQVCYNVZVVSPD6ZAD&hash=item45189b1408%3Ag%3APpUAAOSwSyBnFfY-&itmprp=enc%3AAQAJAAA AwHoV3kP08IDx%2BKZ9MfhVJKmMRt4a1EwtEeibG5otvdBUiTVf5aSCwcLHADR3xO%2FkVDbIXTwzCA31aE1w1RNb37PkHM9AFQmk0MRyFEQno29%2F4XT3QW5JTDmdUGRb73%2FobWiXD4J%2FYmeDE JVqkLyaRy%2B11gZxhF1bTAtppw2eYn7fs1vib40gAUGsrJwE6FL7uHK1jbFQEwxHgnzoa9UDFYmD9hjyoyToZ6c%2BLSL14Xn1gvjEkKYoTytUSpYeRkVRkA%3D%3D%7Ctkp%3ABk9SR4CI9s3gZA&var=594573957545',

'https://www.ebay.com/itm/205089362973?\_skw=sneakers&epid=10064575364&itmmeta=01JC2DXGFS806PS5YS64C2AE0E&hash=item2fc047541d:g:y1oAAOSwD-ZnKvbt&itmprp=enc%3AAQAJAAAAwHoV3kP08IDx%2BKZ9MfhVJKkcMeqWZtpo%2F16RUjBg0h%2FfVBa0u1vW4Jm1DwFGzNNSGSXWbz96cX0JnX8SegwsRKhL4DiMfUdt68ToyV70cS4AME1afSp8M9TP7fggK3RQWRt%2FTpIYbZFpF%2F98pcLuyhMOdVw1oCUbfUTF%2BJSonPBoweTNC0Jj5ZvYtdUZDm0f%2FEUqstyDswnwnVy61fRKVvRWO8NqJBdbiYk6iTWZaQ2EpPuMoeGaRdy9Q0LC7tC5fQ%3D%3D%7Ctkp%3ABk9SR4CI9s3gZA']

#### len(product\_link)

300

### Create function to extract title from link through new variable single\_p\_soup

```
# creating function to extract product title from product page

def get_title(single_p_soup):
    try:
        product_title=single_p_soup.find("span",attrs={"class":"ux-textspans ux-textspans--BOLD"}).text.strip()
    except:
        product_title= ""
    return product_title
```

### Create function to extract Price from link through new variable single\_p\_soup

```
# creating function to extract price from product page (price will contain value review and price)

def get_price(single_p_soup):
    try:
        product_price=single_p_soup.find("div",attrs={"class":"x-price-primary"}).text
    except:
        product_price= ""

return product_price
```

### Create function to extract Review from link through new variable single\_p\_soup

```
# creating function to extract review from product page

def get_review(single_p_soup):
    try:
        product_review=single_p_soup.find("span",attrs={"class":"SECONDARY"}).text.strip("()")
    except:
        product_review= ""
    return product_review
```

### Create function to extract Likes from link through new variable single\_p\_soup

```
# creating function to extract review from product page

def get_likes(single_p_soup):
    try:
        product_likes=single_p_soup.find("span",attrs={"class":"x-watch-heart-btn-text"}).text.strip()
    except:
        product_likes= ""

    return product_likes
```

#### **Create dictionary to store scraped data**

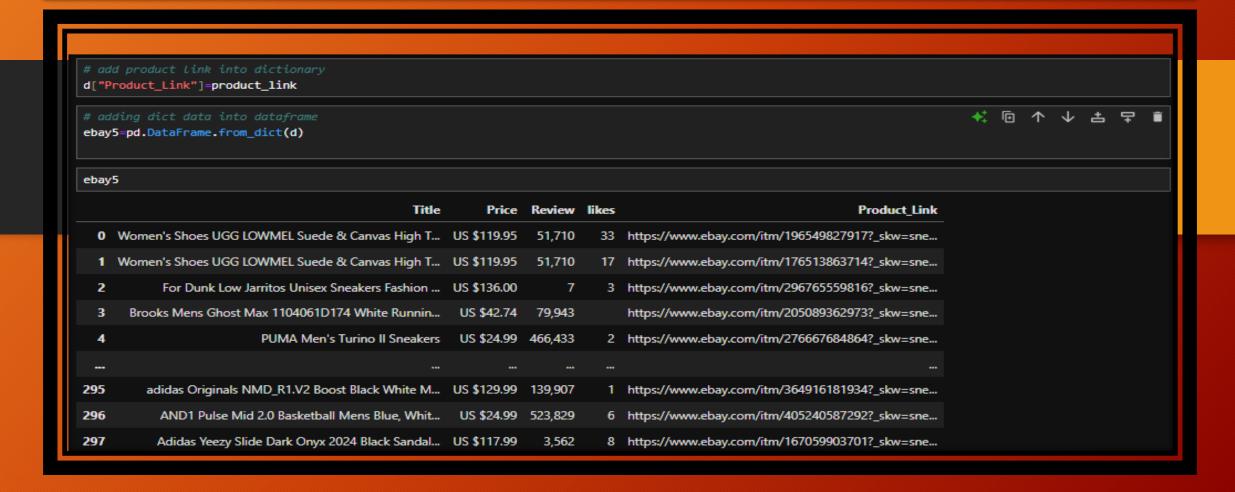
```
# open prodct link 1 by 1 and scrap data and store it in dictionary
d=("Title":[], "Price":[], "Review":[], "likes":[]}
for i in product_link:
    product_page = requests.get(i, headers=headers)
    single_p_soup=BeautifulSoup(product_page.content, "html.parser")

d["Title"].append(get_title(single_p_soup))
    d["Price"].append(get_price(single_p_soup))
    d["Review"].append(get_review(single_p_soup))
    d["likes"].append(get_likes(single_p_soup))
    print("Data successfully added in dictionary -> d")
Data successfully added in dictionary -> d
```

Create dictionary to store scraped data from product page from links which stored into product links list and total scraped data length

```
d={"Title":[],"Price":[], "Review":[],"likes":[]}
for i in product_link:
   product_page = requests.get(i, headers=headers)
   single_p_soup=BeautifulSoup(product_page.content, "html.parser")
   d["Title"].append(get_title(single_p_soup))
   d["Price"].append(get_price(single_p_soup))
   d["Review"].append(get_review(single_p_soup))
   d["likes"].append(get_likes(single_p_soup))
print("Data successfully added in dictionary -> d")
Data successfully added in dictionary -> d
print("Extracted Title : ",len(d["Title"]))
print("Extracted Price: ",len(d["Price"]))
print("Extracted Review: ",len(d["Review"]))
print("Extracted Likes: ",len(d["likes"]))
Extracted Title: 300
Extracted Price: 300
Extracted Review: 300
Extracted Likes: 300
```

### add prduct links into dictionary and create dataframe from dictionary with name ebay5



### **Step-2 Import & Export Data**

#### **Export scraped data into csv format and import the same**

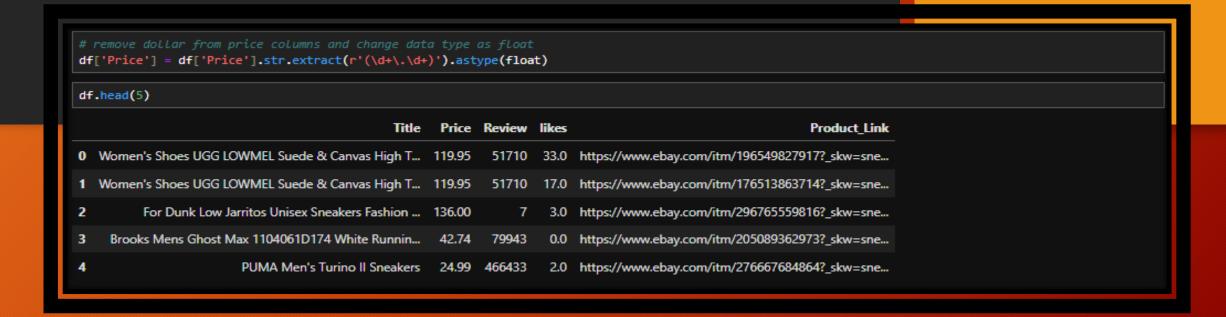


### Step-3 Data Analysis and Data Cleaing

#### Data summary to check dtype, missing values, memory consumption

```
[200]: df.info()
       <class 'pandas.core.frame.DataFrame'>
       RangeIndex: 300 entries, 0 to 299
       Data columns (total 5 columns):
                          Non-Null Count Dtype
            Column
            Title
                          300 non-null
                                         object
            Price
                          300 non-null
                                          object
                          290 non-null
                                          object
            Review
            likes
                          253 non-null
                                          float64
           Product_Link 300 non-null
                                         object
       dtypes: float64(1), object(4)
       memory usage: 11.8+ KB
```

#### Price column correction removed dolloar sign and change dtype as float



#### Data summary to check dtype, missing values, memory consumption

```
df.info()
[200]:
       <class 'pandas.core.frame.DataFrame'>
       RangeIndex: 300 entries, 0 to 299
       Data columns (total 5 columns):
                          Non-Null Count Dtype
            Column
            Title
                          300 non-null
                                          object
            Price
                          300 non-null
                                          object
                                          object
            Review
                          290 non-null
                                          float64
            likes
                          253 non-null
           Product_Link 300 non-null
                                          object
       dtypes: float64(1), object(4)
       memory usage: 11.8+ KB
```

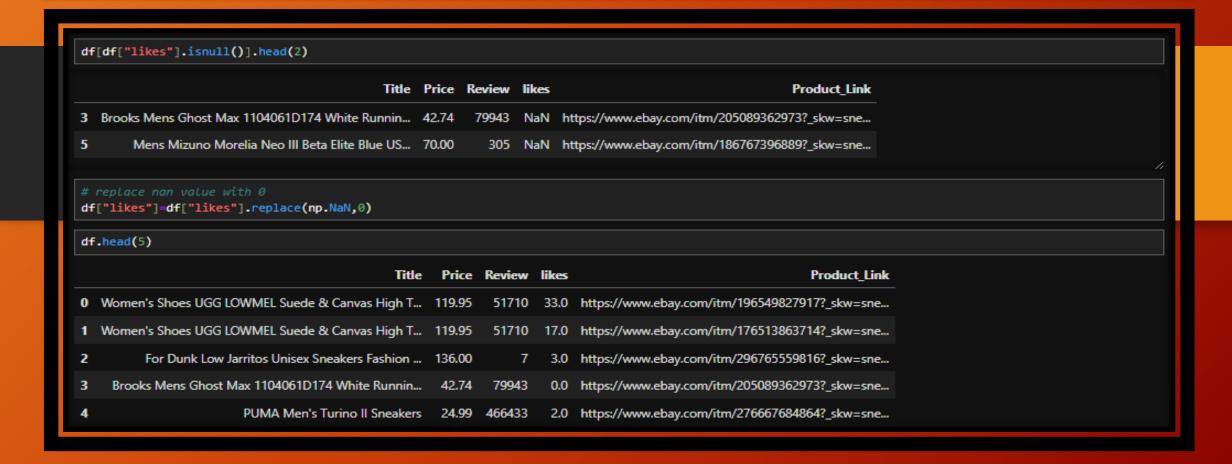
#### Replace "nan" value with "0" in review column before change data type

```
≮ 厄 ↑ ↓ 占 〒 🛢
df["Review"].to_list()
  و بيدونيد
 '223,529',
 '33,315',
 '33,315',
nan,
 '215',
 '4,789',
nan,
 '33,315',
 '33,315',
 '33,315',
 '33,315',
 '523,829',
 '223,529',
 nan,
 '2,159',
 '2,679',
 '523,829',
 '139,906',
df["Review"]=df["Review"].replace(np.nan,"0")
df["Review"]=df["Review"].str.replace(",","")
df["Review"].to_list()
 '139906',
 '33315',
 '33315',
 '33315',
```

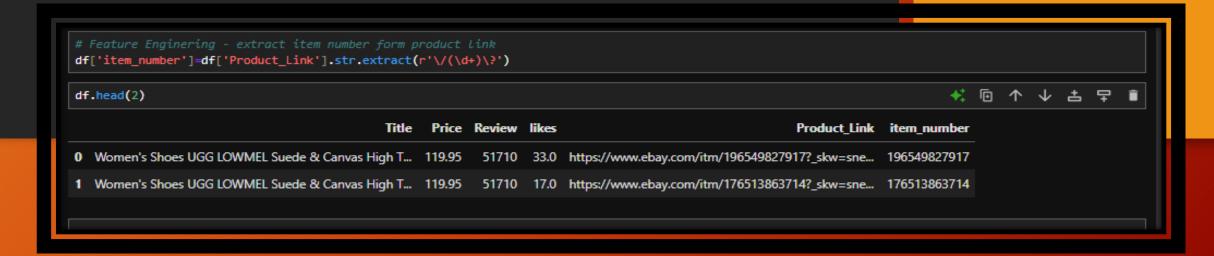
#### Change review column datatype object to int32

```
df["Review"]=df["Review"].astype("int")
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 300 entries, 0 to 299
Data columns (total 5 columns):
                  Non-Null Count Dtype
    Column
    Title
                  300 non-null
                                  object
    Price
                  300 non-null
                                  float64
     Review
                                  int32
                  300 non-null
                                  float64
    likes
                  253 non-null
    Product_Link 300 non-null
                                  object
dtypes: float64(2), int32(1), object(2)
memory usage: 10.7+ KB
```

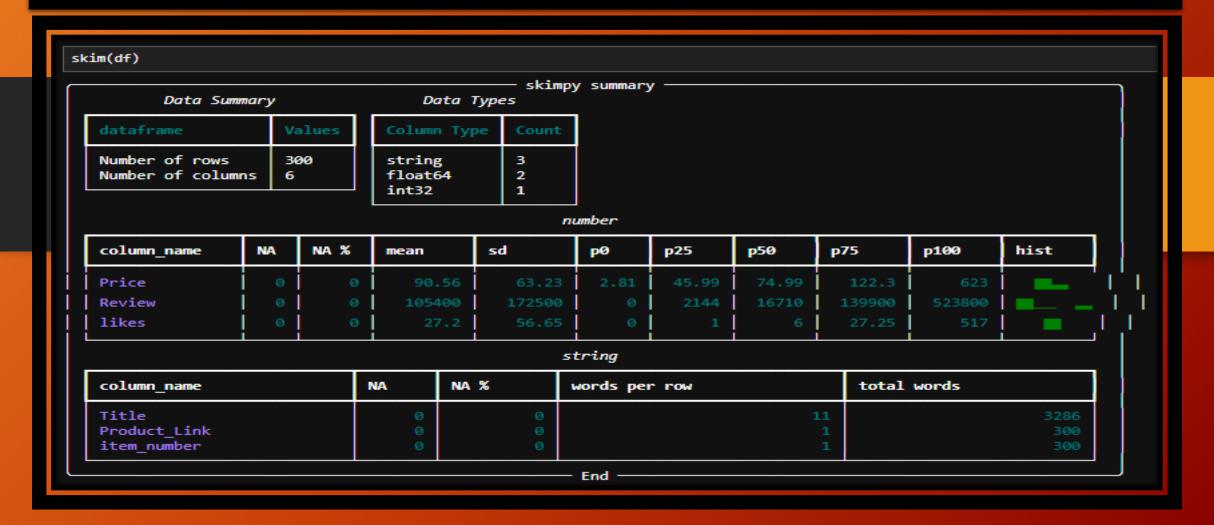
#### Fill missing values with 0 in likes columns using replace function



Extract item number form product link column and create new column with name item\_number



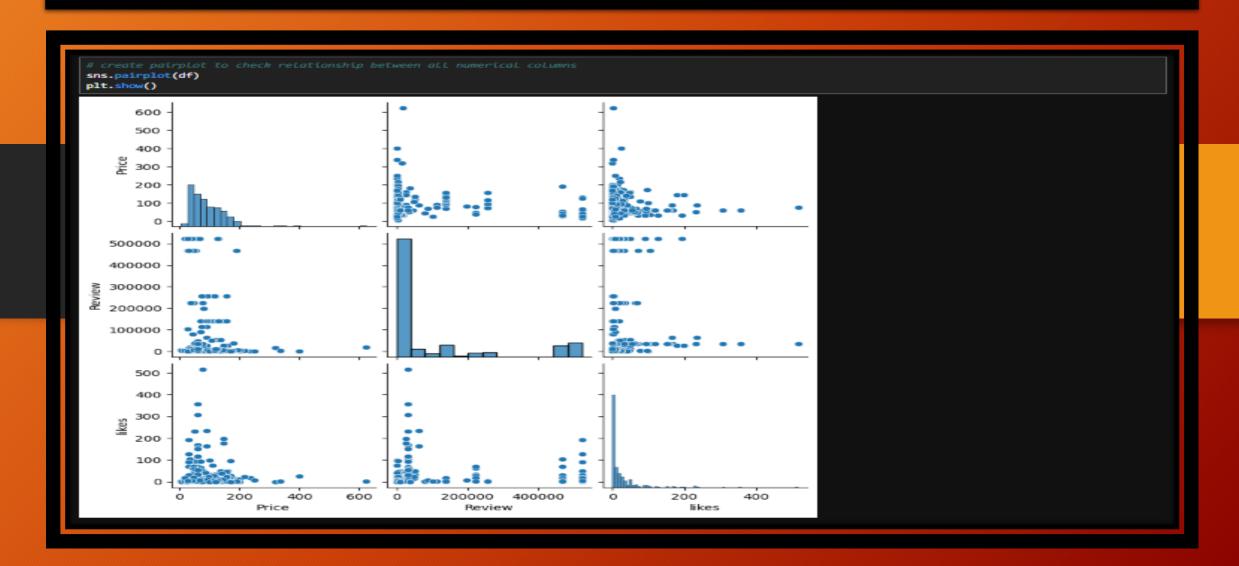
#### Check again data summary by importing skimmy for better understanding



#### Check again data summary by importing skimmy for better understanding

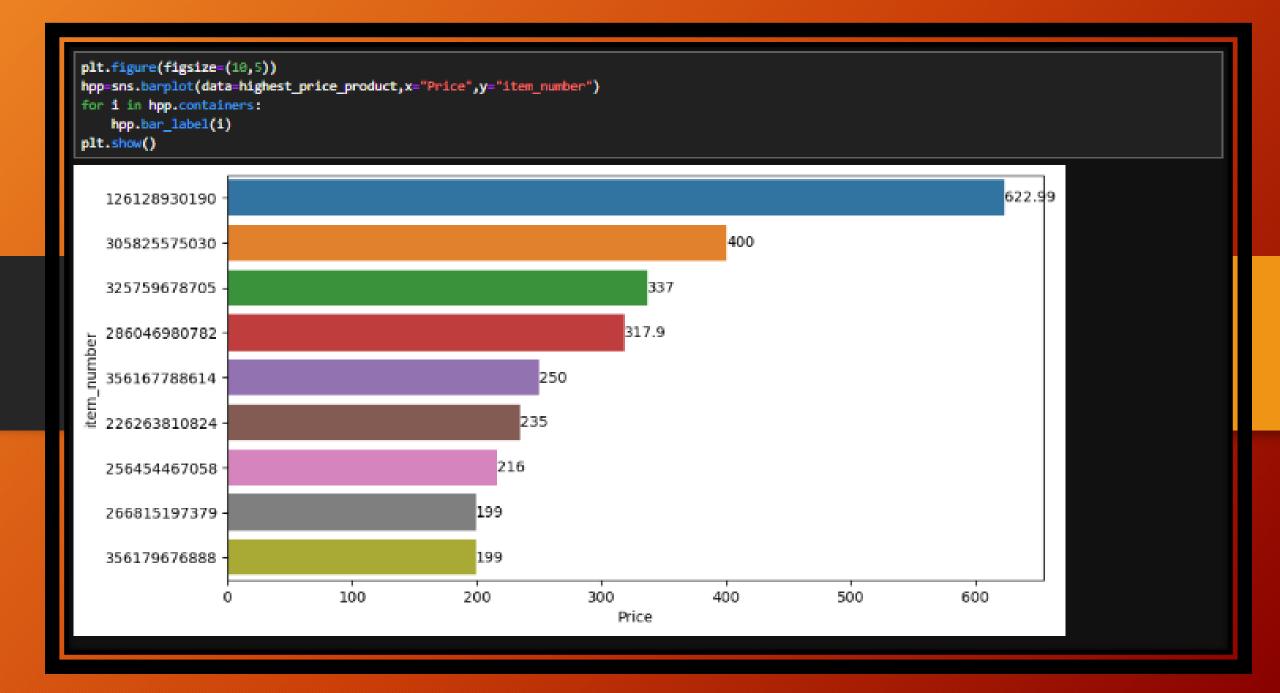
# Step-4 Exploratory Data Analysis (EDA)

#### **Check relation between all numerical columns**



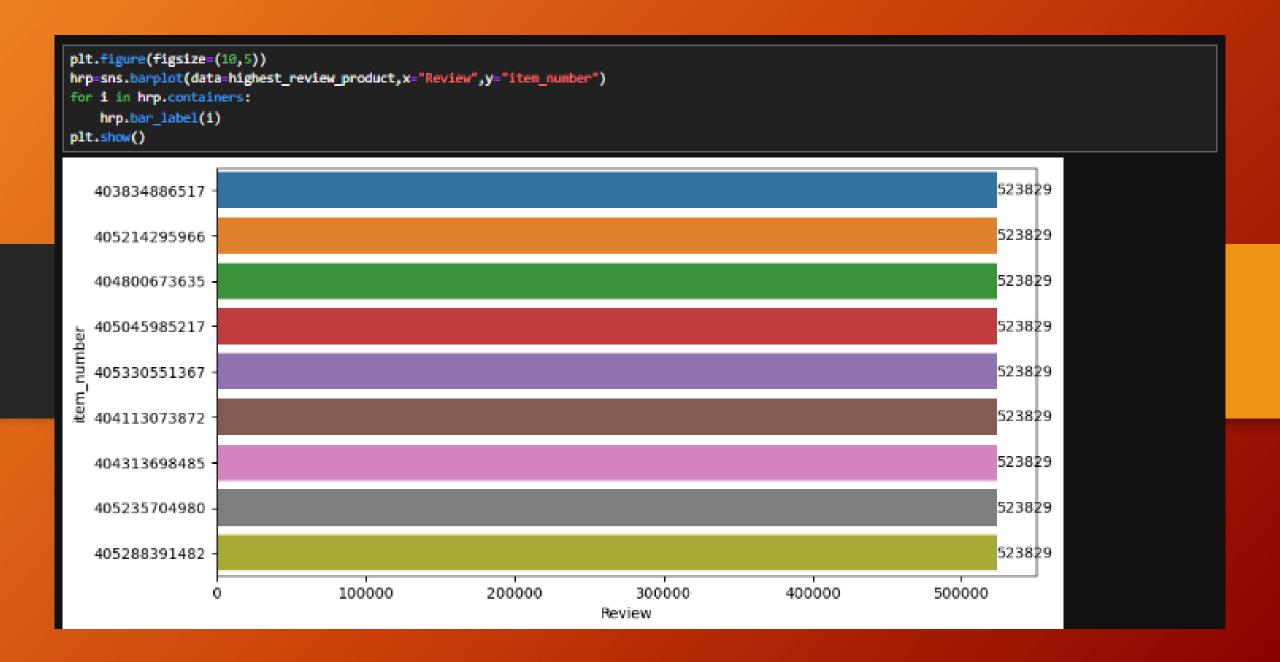
#### 10 Product which have highest price among all products

						1000A						
highe	<pre>18 highest price product in sneaker category st_price_product=df.sort_values(by="Price",as sst_price_product</pre>	**	0	<b></b>	Ψ	≛	₽	•				
	Title	Price	Review	likes	Product_Link	item_numbe	er					
270	Valentino Garavani Ready Go Runner Wool Sock S	622.99	17606	2.0	https://www.ebay.com/itm/126128930190?_skw=sne	12612893019	Ю					
132	FTP Osiris D3 2001 Size 12	400.00	85	25.0	https://www.ebay.com/itm/305825575030?_skw=sne	30582557503	10					
235	HOGAN men shoes Hogan-3R Recycle H590 sneaker	337.00	2098	1.0	https://www.ebay.com/itm/325759678705?_skw=sne	32575967870	)5					
208	Isaia White Leather Sneakers Shoes IDS53A Man	317.90	15612	0.0	https://www.ebay.com/itm/286046980782?_skw=sne	28604698078	2					
183	Adidas Yeezy Boost 700 V3 Dark US Size 9 Perfe	250.00	25	8.0	https://www.ebay.com/itm/356167788614?_skw=sne	35616778861	4					
123	Lanvin Paris Leather DDB0 Sneakers Size 44 New	235.00	98	19.0	https://www.ebay.com/itm/226263810824?_skw=sne	22626381082	4					
220	IF9280 TMNT Teenage Mutant Ninja Turtles adida	216.00	2679	22.0	https://www.ebay.com/itm/256454467058?_skw=sne	25645446705	8					
39	New Kiton Napoli Suede Leather Fashion Sneaker	199.00	129	12.0	https://www.ebay.com/itm/266815197379?_skw=sne	26681519737	9					
99	New Kiton Napoli Suede Leather Fashion Sneaker	199.00	129	12.0	https://www.ebay.com/itm/266815197379?_skw=sne	26681519737	9					
230	adidas Originals Country Japan IG4554 Preloved	199.00	5004	0.0	https://www.ebay.com/itm/356179676888?_skw=sne	35617967688	8					



#### 10 Product which have highest review among all products

high	# top 10 highest review product in sneaker category in ebay highest_review_product=df.sort_values(by="Review",ascending=False).head(10)										
high	est_review_product										
	Title	Price	Review	likes	Product_Link	item_number					
265	Puma Triple Mid Basketball Mens Red Sneakers	29.99	523829	91.0	https://www.ebay.com/itm/403834886517?_skw=sne	403834886517					
173	Puma Bmw Mms RCat Machina Lace Up Mens Black,	39.99	523829	2.0	https://www.ebay.com/itm/405214295966?_skw=sne	405214295966					
34	Puma Brww Mms RCat Machina Lace Up Mens Black,	39.99	523829	2.0	https://www.ebay.com/itm/405214295966?_skw=sne	405214295966					
281	Avia AviFactor 2.0 Running Mens Beige Sneaker	29.99	523829	31.0	https://www.ebay.com/itm/404800673635?_skw=sne	404800673635					
283	Puma Suede Triplex Lace Up Mens Red Sneakers	44.99	523829	21.0	https://www.ebay.com/itm/405045985217?_skw=sne	405045985217					
57	adidas Yeezy Foam Rnr Slip On Youth Boys Bro	59.99	523829	1.0	https://www.ebay.com/itm/405330551367?_skw=sne	405330551367					
260	Puma Viz Runner Repeat Running Mens Black Sne	49.99	523829	26.0	https://www.ebay.com/itm/404113073872?_skw=sne	404113073872					
50	Puma Ca Pro Embroidered Lace Up Mens White Sn	29.99	523829	127.0	https://www.ebay.com/itm/404313698485?_skw=sne	404313698485					
284	Puma Electrify Nitro 2 Running Mens Black Sne	54.99	523829	3.0	https://www.ebay.com/itm/405235704980?_skw=sne	405235704980					
206	Puma Mb.04 La France Basketball Mens Blue Sne	124.99	523829	0.0	https://www.ebay.com/itm/405288391482?_skw=sne	405288391482					

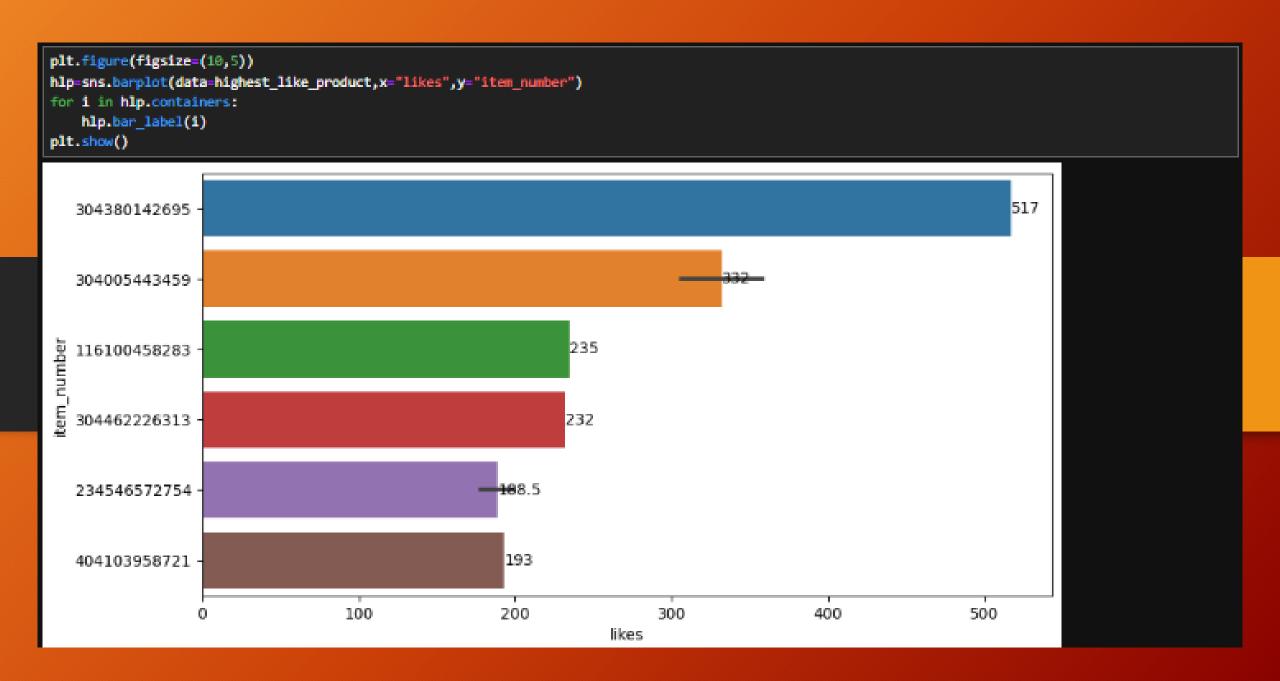


#### 10 Product which have highest likes among all products

# top 10 highest like product in sneaker category in ebay

highest\_like\_product=df.sort\_values(by="likes",ascending=False).head(10)
highest\_like\_product

					<del>-</del>	item_number
152 A	Adidas Adizero Ubersonic 4.1 CL Men's Tennis S	74.95	33315	517.0	https://www.ebay.com/itm/304380142695?_skw=sne	304380142695
6 Rec	ebok Classic Nylon Men's Running Shoe Black	59.95	33315	358.0	https://www.ebay.com/itm/304005443459?_skw=sne	304005443459
68 Re	ebok Classic Nylon Men's Running Shoe Black	59.95	33315	306.0	https://www.ebay.com/itm/304005443459?_skw=sne	304005443459
37 A	Air Force 1 '07 White/ Black CT2302-100 Fashio	88.35	63683	235.0	https://www.ebay.com/itm/116100458283?_skw=sne	116100458283
124 A	didas Vs Pace 2.0 Men's Skate Shoe White Athl	49.95	33315	232.0	https://www.ebay.com/itm/304462226313?_skw=sne	304462226313
24 A	didas Vs Pace 2.0 Men's Skate Shoe White Athl	49.95	33315	232.0	https://www.ebay.com/itm/304462226313?_skw=sne	304462226313
116 A	didas Vs Pace 2.0 Men's Skate Shoe White Athl	49.95	33315	232.0	https://www.ebay.com/itm/304462226313?_skw=sne	304462226313
<b>215</b> N	New NIKE Air Max 90 Men's classic Athletic Sne	144.99	26477	199.0	https://www.ebay.com/itm/234546572754?_skw=sne	234546572754
126 Po	uma Twitch Runner Trail Running Mens Black S	29.99	523829	193.0	https://www.ebay.com/itm/404103958721?_skw=sne	404103958721
<b>56</b> N	New NIKE Air Max 90 Men's classic Athletic Sne	144.99	26477	178.0	https://www.ebay.com/itm/234546572754?_skw=sne	234546572754



### **Thank You**

**linkedin:** www.linkedin.com/in/nishant-kumar-data-analyst

github: https://github.com/Nishant6020