



DATA ANALYSIS

DATA ANALYSIS OF VARIOUS SMARTPHONES BY WEBSCRAPPING

PROBLEM STATEMENT

The main aim of this project is to analyse the various smartphones which are taken randomly from the flipkart .By this project we can analyse the features of the smartphones.

1.Product Name

2.Price

3.Ratings

4.Features

5.Discount

6.Brand



ESSENTIAL TOOLS AND LIBRARIES

1.Numpy

2.Pandas

3.Matplotlib.pyplot

4.Seaborn

5.Requests

6.Beautiful Soup



IMPORT PACKAGES

Life cycle of the data science is:

1.Business Understanding

2.Data Requirement

3.Data Collection

4.Data Cleaning

5.Data Analysing

```
import warnings
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import requests
from bs4 import BeautifulSoup

import re
import time

from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"
```

UNDERSTANDING THE DATA

```
: url="https://www.flipkart.com/search?q=smartphones&otracker=search&otracker1=search&marketplace=FLIPKART&as-show=on&as=off"  
:  
: page= requests.get(url)  
:  
: page.status_code  
:  
: 200
```

So here I took a URL from the flipkart about various smartphones randomly.

DATA REQUIREMENT

We need to collect the required data for the further process. Here we need the div classes of product names, prices, ratings, discount features.

```
soup = BeautifulSoup(pagecontent)
soup
```

...

```
print(soup.prettify())
```

...

- We use BeautifulSoup to collect this required data

DATA COLLECTION

```
for i in range(1,31):
    url= "https://www.flipkart.com/search?q=smartphones&otracker=search&otracker1=search&marketplace=FLIPKART&as-show=on&as=off{}".format(i)
    page= requests.get(url)
    pagecontent= page.text
    soup= BeautifulSoup(pagecontent)

    for x in soup.find_all('div',attrs={'class':'_3ply-c row'}):
        pname= x.find('div',attrs={'class':'_4rR01T'})
        cost= x.find('div',attrs={'class':'_30jeq3 _1_WHN1'})
        rat= x.find('div',attrs={'class':'_3LWZ1K'})
        spec= x.find('div',attrs={'class':'fMghEO'})
        dis= x.find('div',attrs={'class':'_3Ay6Sb'})

        if pname is None:
            product_name.append(np.NaN)

        else:
            product_name.append(pname.text)

        if cost is None:
            price.append(np.NaN)

        else:
            price.append(cost.text)

        if rat is None:
            rating.append(np.NaN)

        else:
            rating.append(rat.text)

        if spec is None:
            features.append(np.NaN)

        else:
            features.append(spec.text)

        if dis is None:
            discount.append(np.NaN)

        else:
            discount.append(dis.text)

    pagenum.append(i)
```

smartphones_df= pd.DataFrame({'ProductName':product_name, 'Price':price, 'Rating':rating, 'Features':features,'Discount':discount,'PageNo':pagenum})								
smartphones_df								
	ProductName	Price	Rating	Features	Discount	PageNo		
0	vivo T2x 5G (Aurora Gold, 128 GB)	₹14,999	4.3	8 GB RAM 128 GB ROM16.71 cm (6.58 inch) Full...	28% off	1		
1	vivo T2x 5G (Marine Blue, 128 GB)	₹12,999	4.4	6 GB RAM 128 GB ROM16.71 cm (6.58 inch) Full...	31% off	1		
2	vivo T2x 5G (Aurora Gold, 128 GB)	₹12,999	4.4	6 GB RAM 128 GB ROM16.71 cm (6.58 inch) Full...	31% off	1		
3	vivo T2x 5G (Glimmer Black, 128 GB)	₹11,999	4.4	4 GB RAM 128 GB ROM16.71 cm (6.58 inch) Full...	33% off	1		
4	vivo T2x 5G (Aurora Gold, 128 GB)	₹11,999	4.4	4 GB RAM 128 GB ROM16.71 cm (6.58 inch) Full...	33% off	1		
...		
667	SAMSUNG Galaxy F13 (Nightsky Green, 64 GB)	₹10,999	4.3	4 GB RAM 64 GB ROM Expandable Upto 1 TB16...	26% off	30		
668	Motorola G34 5G (Ocean Green, 128 GB)	₹11,999	4.2	8 GB RAM 128 GB ROM16.51 cm (6.5 inch) HD+ D...	20% off	30		
669	SAMSUNG Galaxy F13 (Waterfall Blue, 64 GB)	₹10,999	4.3	4 GB RAM 64 GB ROM Expandable Upto 1 TB16...	26% off	30		
670	realme 11x 5G (Midnight Black, 128 GB)	₹15,999	4.3	8 GB RAM 128 GB ROM Expandable Upto 2 TB17...	15% off	30		
671	realme 11x 5G (Midnight Black, 128 GB)	₹14,999	4.4	6 GB RAM 128 GB ROM Expandable Upto 2 TB17...	11% off	30		

Smartphones_df is the dataframe of the data

DATA CLEANING

```
smartphones_df["Brand"] =smartphones_df["ProductName"].apply(lambda x: " ".join(re.findall("^\w+",x)))
smartphones_df
```

	ProductName	Price	Rating	Features	Discount	PageNo	ROM	Brand
0	vivo T2x 5G (Aurora Gold, 128 GB)	₹14,999	4.3	8 GB RAM 128 GB ROM16.71 cm (6.58 inch) Full...	28% off	1	128 GB ROM	vivo
1	vivo T2x 5G (Marine Blue, 128 GB)	₹12,999	4.4	6 GB RAM 128 GB ROM16.71 cm (6.58 inch) Full...	31% off	1	128 GB ROM	vivo
2	vivo T2x 5G (Aurora Gold, 128 GB)	₹12,999	4.4	6 GB RAM 128 GB ROM16.71 cm (6.58 inch) Full...	31% off	1	128 GB ROM	vivo
3	vivo T2x 5G (Glimmer Black, 128 GB)	₹11,999	4.4	4 GB RAM 128 GB ROM16.71 cm (6.58 inch) Full...	33% off	1	128 GB ROM	vivo
4	vivo T2x 5G (Aurora Gold, 128 GB)	₹11,999	4.4	4 GB RAM 128 GB ROM16.71 cm (6.58 inch) Full...	33% off	1	128 GB ROM	vivo
...
667	SAMSUNG Galaxy F13 (Night sky Green, 64 GB)	₹10,999	4.3	4 GB RAM 64 GB ROM Expandable Upto 1 TB16...	26% off	30	64 GB ROM	SAMSUNG
668	Motorola G34 5G (Ocean Green, 128 GB)	₹11,999	4.2	8 GB RAM 128 GB ROM16.51 cm (6.5 inch) HD+ D...	20% off	30	128 GB ROM	Motorola
669	SAMSUNG Galaxy F13 (Waterfall Blue, 64 GB)	₹10,999	4.3	4 GB RAM 64 GB ROM Expandable Upto 1 TB16...	26% off	30	64 GB ROM	SAMSUNG
670	realme 11x 5G (Midnight Black, 128 GB)	₹15,999	4.3	8 GB RAM 128 GB ROM Expandable Upto 2 TB17...	15% off	30	128 GB ROM	realme
671	realme 11x 5G (Midnight Black, 128 GB)	₹14,999	4.4	6 GB RAM 128 GB ROM Expandable Upto 2 TB17...	11% off	30	128 GB ROM	realme

```
smartphones_df["Price"]=smartphones_df["Price"].apply(lambda x: str(x).replace('₹','').replace(',','')).astype(float)
```

```
smartphones_df
```

	ProductName	Price	Rating	Features	Discount	PageNo	ROM	Brand
0	vivo T2x 5G (Aurora Gold, 128 GB)	14999.0	4.3	8 GB RAM 128 GB ROM16.71 cm (6.58 inch) Full...	28% off	1	128 GB ROM	vivo
1	vivo T2x 5G (Marine Blue, 128 GB)	12999.0	4.4	6 GB RAM 128 GB ROM16.71 cm (6.58 inch) Full...	31% off	1	128 GB ROM	vivo
2	vivo T2x 5G (Aurora Gold, 128 GB)	12999.0	4.4	6 GB RAM 128 GB ROM16.71 cm (6.58 inch) Full...	31% off	1	128 GB ROM	vivo
3	vivo T2x 5G (Glimmer Black, 128 GB)	11999.0	4.4	4 GB RAM 128 GB ROM16.71 cm (6.58 inch) Full...	33% off	1	128 GB ROM	vivo
4	vivo T2x 5G (Aurora Gold, 128 GB)	11999.0	4.4	4 GB RAM 128 GB ROM16.71 cm (6.58 inch) Full...	33% off	1	128 GB ROM	vivo
...
667	SAMSUNG Galaxy F13 (Night sky Green, 64 GB)	10999.0	4.3	4 GB RAM 64 GB ROM Expandable Upto 1 TB16...	26% off	30	64 GB ROM	SAMSUNG
668	Motorola G34 5G (Ocean Green, 128 GB)	11999.0	4.2	8 GB RAM 128 GB ROM16.51 cm (6.5 inch) HD+ D...	20% off	30	128 GB ROM	Motorola
669	SAMSUNG Galaxy F13 (Waterfall Blue, 64 GB)	10999.0	4.3	4 GB RAM 64 GB ROM Expandable Upto 1 TB16...	26% off	30	64 GB ROM	SAMSUNG
670	realme 11x 5G (Midnight Black, 128 GB)	15999.0	4.3	8 GB RAM 128 GB ROM Expandable Upto 2 TB17...	15% off	30	128 GB ROM	realme
671	realme 11x 5G (Midnight Black, 128 GB)	14999.0	4.4	6 GB RAM 128 GB ROM Expandable Upto 2 TB17...	11% off	30	128 GB ROM	realme

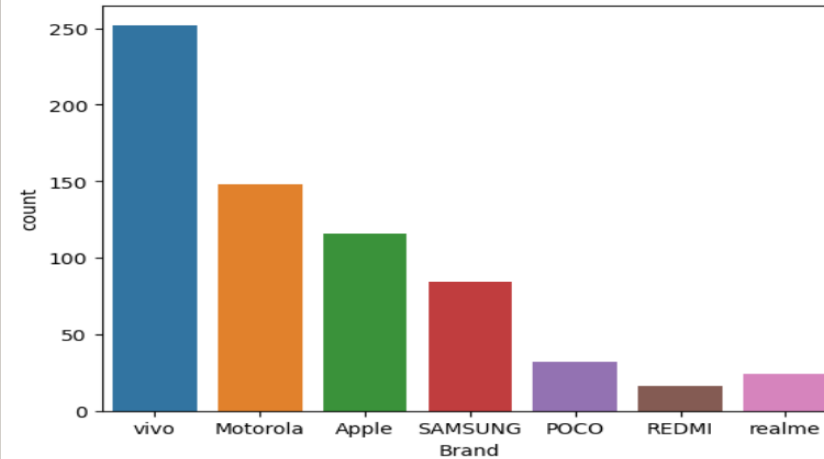
672 rows × 8 columns

Missing data is nothing but data cleaning

DATA VISUALIZATION

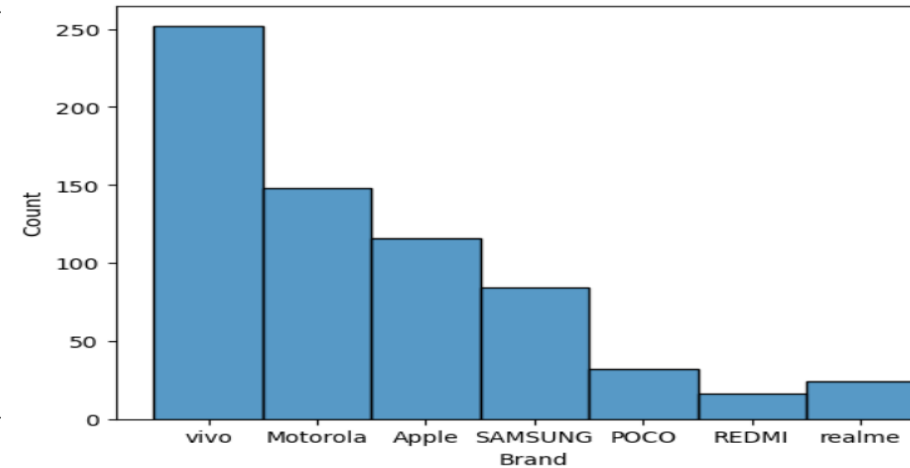
```
sns.countplot(smartphones_df, x="Brand")
```

<Axes: xlabel='Brand', ylabel='count'>



```
sns.histplot(smartphones_df, x="Brand")
```

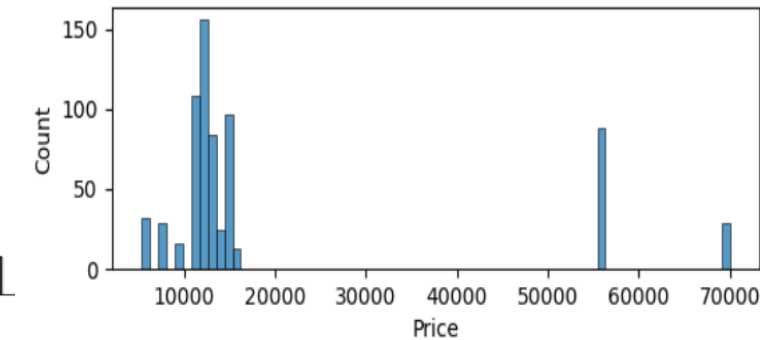
<Axes: xlabel='Brand', ylabel='Count'>



```
plt.figure(figsize=(6,2))  
sns.histplot(smartphones_df.Price)
```

<Figure size 600x200 with 0 Axes>

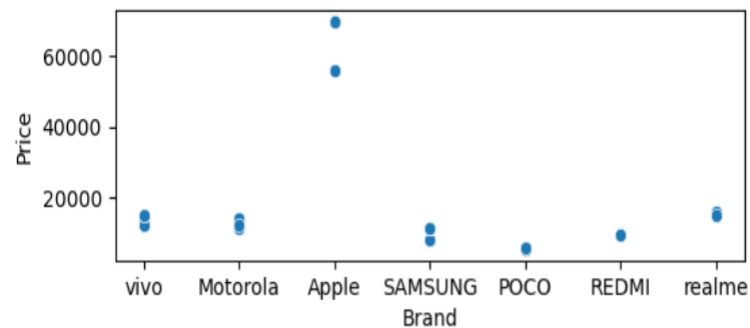
<Axes: xlabel='Price', ylabel='Count'>



```
plt.figure(figsize=(6,2))  
sns.scatterplot(data=smartphones_df,x=smartphones_df["Brand"],y=smartphones_df["Price"])
```

<Figure size 600x200 with 0 Axes>

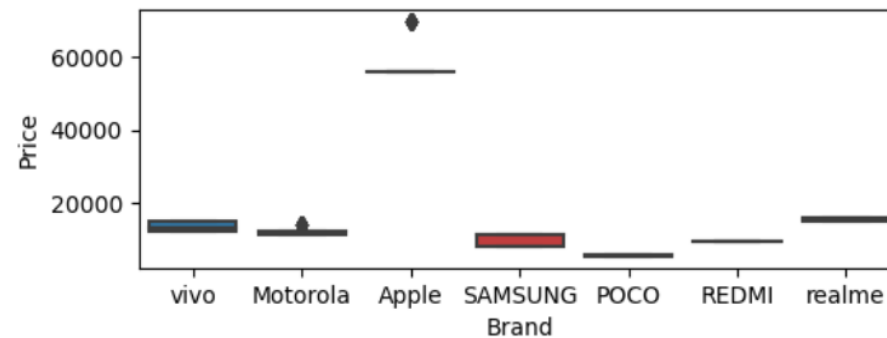
<Axes: xlabel='Brand', ylabel='Price'>



```
plt.figure(figsize=(6,2))  
sns.boxplot(data=smartphones_df,x=smartphones_df["Brand"],y=smartphones_df["Price"])
```

<Figure size 600x200 with 0 Axes>

<Axes: xlabel='Brand', ylabel='Price'>



These are the data visualizations of brand and prices of the smartphones

THANK YOU!