Problem Statement

- With the growing number of smartphones released each month, consumers face difficulty comparing prices, features, and specifications across online platforms. Manually checking each website is time-consuming and inefficient.
- By using its features
- Collect data from platforms like flipkart, amazon, samsung

```python
In [1]: # By using its features
        # Collect data from platforms like flipkart, amazon, samsung
```

```python
In [8]: # import the packages
        import re
        import requests
        from bs4 import BeautifulSoup
        import pandas as pd
```

```python
In [4]: # Give the url of the website
        url = "https://www.flipkart.com/search?q=mobiles&otracker=search&otracker1=sea
        web_page = requests.get(url)
        print(web_page)
        # web_page.text ==> It will give you code of the web page in the text format
        # ==> To convert that text into HTML format we have to use BeautifuSoup
```

<Response [200]>

```python
In [5]: html_code = BeautifulSoup(web_page.text) # It Gives data in the html code form
```

```python
In [6]: mob_price = html_code.find_all("div",class_="Nx9bqj _4b5DiR")
        mob_price
```

```
Out[6]: [<div class="Nx9bqj _4b5DiR">₹44,999</div>,
         <div class="Nx9bqj _4b5DiR">₹44,999</div>,
         <div class="Nx9bqj _4b5DiR">₹6,799</div>,
         <div class="Nx9bqj _4b5DiR">₹7,499</div>,
         <div class="Nx9bqj _4b5DiR">₹44,999</div>,
         <div class="Nx9bqj _4b5DiR">₹79,999</div>,
         <div class="Nx9bqj _4b5DiR">₹8,499</div>,
         <div class="Nx9bqj _4b5DiR">₹11,499</div>,
         <div class="Nx9bqj _4b5DiR">₹8,999</div>,
         <div class="Nx9bqj _4b5DiR">₹7,849</div>,
         <div class="Nx9bqj _4b5DiR">₹79,999</div>,
         <div class="Nx9bqj _4b5DiR">₹79,999</div>,
         <div class="Nx9bqj _4b5DiR">₹6,199</div>,
         <div class="Nx9bqj _4b5DiR">₹7,999</div>,
         <div class="Nx9bqj _4b5DiR">₹12,999</div>,
         <div class="Nx9bqj _4b5DiR">₹11,999</div>,
         <div class="Nx9bqj _4b5DiR">₹79,999</div>,
         <div class="Nx9bqj _4b5DiR">₹11,999</div>,
         <div class="Nx9bqj _4b5DiR">₹8,499</div>,
         <div class="Nx9bqj _4b5DiR">₹7,499</div>,
         <div class="Nx9bqj _4b5DiR">₹6,199</div>,
         <div class="Nx9bqj _4b5DiR">₹9,499</div>,
         <div class="Nx9bqj _4b5DiR">₹13,499</div>,
         <div class="Nx9bqj _4b5DiR">₹6,799</div>]
```

Extracting prices and kept into a list

```
In [9]: price = [] # Creating list object to store the mobile prices
        for i in mob_price:
            #prices.append(i.text) # Adding only text data into the list
            p = re.findall("\d+\,\d+",i.text)
            price.append(p[0])
        price
```

```
Out[9]:  ['44,999',
          '44,999',
          '6,799',
          '7,499',
          '44,999',
          '79,999',
          '8,499',
          '11,499',
          '8,999',
          '7,849',
          '79,999',
          '79,999',
          '6,199',
          '7,999',
          '12,999',
          '11,999',
          '79,999',
          '11,999',
          '8,499',
          '7,499',
          '6,199',
          '9,499',
          '13,499',
          '6,799']
```

```python
mob_features = html_code.find_all("div",class_="KzDlHZ")
mob_features # Give all mobile name and brand in the HTML format in list
```

```
Out[113…  [<div class="KzDlHZ">POCO C71 (Cool Blue, 128 GB)</div>,
          <div class="KzDlHZ">Samsung Galaxy F06 5G (Bahama Blue, 64 GB)</div>,
          <div class="KzDlHZ">POCO M7 Pro 5G (Lunar Dust, 128 GB)</div>,
          <div class="KzDlHZ">POCO C71 (Desert Gold, 64 GB)</div>,
          <div class="KzDlHZ">POCO M7 5G (Ocean Blue, 128 GB)</div>,
          <div class="KzDlHZ">Samsung Galaxy F06 5G (Bahama Blue, 128 GB)</div>,
          <div class="KzDlHZ">Samsung Galaxy F06 5G (Bahama Blue, 128 GB)</div>,
          <div class="KzDlHZ">POCO M7 5G  - Locked with Airtel Prepaid (Satin Black, 1
          28 GB)</div>,
          <div class="KzDlHZ">POCO M7 5G (Mint Green, 128 GB)</div>,
          <div class="KzDlHZ">OPPO K13x 5G 6000mAh and 45W SUPERVOOC Charger &amp; AI
          (Breeze Blue, 128 GB)</div>,
          <div class="KzDlHZ">Motorola g45 5G (Brilliant Green, 128 GB)</div>,
          <div class="KzDlHZ">POCO M7 5G (Mint Green, 128 GB)</div>,
          <div class="KzDlHZ">OPPO K13x 5G 6000mAh and 45W SUPERVOOC Charger &amp; AI
          (Midnight Violet, 128 GB)</div>,
          <div class="KzDlHZ">POCO C71 (Power Black, 128 GB)</div>,
          <div class="KzDlHZ">OPPO K13x 5G 6000mAh and 45W SUPERVOOC Charger &amp; AI
          (Sunset Peach, 128 GB)</div>,
          <div class="KzDlHZ">Samsung Galaxy F06 5G (Lit Violet, 64 GB)</div>,
          <div class="KzDlHZ">POCO M7 Pro 5G (Lavender Frost, 128 GB)</div>,
          <div class="KzDlHZ">MOTOROLA g35 5G (Guava Red, 128 GB)</div>,
          <div class="KzDlHZ">POCO C71 (Power Black, 64 GB)</div>,
          <div class="KzDlHZ">POCO C75 5G (Enchanted Green, 64 GB)</div>,
          <div class="KzDlHZ">POCO M7 Pro 5G (Lunar Dust, 256 GB)</div>,
          <div class="KzDlHZ">POCO X7 5G (Yellow, 128 GB)</div>,
          <div class="KzDlHZ">MOTOROLA G96 5G (Pantone Greener Pastures, 128 GB)</di
          v>,
          <div class="KzDlHZ">Ai+ Pulse (Blue, 64 GB)</div>]
```

Extracting brand names and kept into a list

```python
In [114…  brand = [] # Creating list object to store brand names of the mobiles
          for i in mob_features:
              brand.append(i.text.split()[0]) # By split() and [0] we can extract only b
          brand
```

```
Out[114…  ['POCO',
           'Samsung',
           'POCO',
           'POCO',
           'POCO',
           'Samsung',
           'Samsung',
           'POCO',
           'POCO',
           'OPPO',
           'Motorola',
           'POCO',
           'OPPO',
           'POCO',
           'OPPO',
           'Samsung',
           'POCO',
           'MOTOROLA',
           'POCO',
           'POCO',
           'POCO',
           'POCO',
           'MOTOROLA',
           'Ai+']
```

Extracting the model names and kept into a list

```
In [115…  model = []
          for i in mob_features:
              model.append(i.text.split("(")[0]) # Extract the model name by split() and
          model
```

```
Out[115…  ['POCO C71 ',
           'Samsung Galaxy F06 5G ',
           'POCO M7 Pro 5G ',
           'POCO C71 ',
           'POCO M7 5G ',
           'Samsung Galaxy F06 5G ',
           'Samsung Galaxy F06 5G ',
           'POCO M7 5G  - Locked with Airtel Prepaid ',
           'POCO M7 5G ',
           'OPPO K13x 5G 6000mAh and 45W SUPERVOOC Charger & AI ',
           'Motorola g45 5G ',
           'POCO M7 5G ',
           'OPPO K13x 5G 6000mAh and 45W SUPERVOOC Charger & AI ',
           'POCO C71 ',
           'OPPO K13x 5G 6000mAh and 45W SUPERVOOC Charger & AI ',
           'Samsung Galaxy F06 5G ',
           'POCO M7 Pro 5G ',
           'MOTOROLA g35 5G ',
           'POCO C71 ',
           'POCO C75 5G ',
           'POCO M7 Pro 5G ',
           'POCO X7 5G ',
           'MOTOROLA G96 5G ',
           'Ai+ Pulse ']
```

Extracting color names and kept into a list

```python
In [116…  # Extracting colors
          color = []
          for i in mob_features:
              color.append(i.text.split("(")[1].split(",")[0])
          color
```

```
Out[116…   ['Cool Blue',
            'Bahama Blue',
            'Lunar Dust',
            'Desert Gold',
            'Ocean Blue',
            'Bahama Blue',
            'Bahama Blue',
            'Satin Black',
            'Mint Green',
            'Breeze Blue',
            'Brilliant Green',
            'Mint Green',
            'Midnight Violet',
            'Power Black',
            'Sunset Peach',
            'Lit Violet',
            'Lavender Frost',
            'Guava Red',
            'Power Black',
            'Enchanted Green',
            'Lunar Dust',
            'Yellow',
            'Pantone Greener Pastures',
            'Blue']
```

Extracting GB and kept into list

```
In [117…   GB = []
           for i in mob_features:
               GB.append(i.text.split(",")[1].split(" ")[1])
           GB
```

```
Out[117…  ['128',
           '64',
           '128',
           '64',
           '128',
           '128',
           '128',
           '128',
           '128',
           '128',
           '128',
           '128',
           '128',
           '128',
           '128',
           '64',
           '128',
           '128',
           '64',
           '64',
           '256',
           '128',
           '128',
           '64']
```

Extracting features

```
In [12]:  feature = html_code.find_all("li",class_="J+igdf")
```

From features extract the ROM and kept into list

```
In [119…  ROM = []
          for i in feature:
              a=(re.findall(r"(\d+)\sGB\sROM",i.text))
              if len(a)>0:
                  ROM.append(a[0])
          ROM
```

```
Out[119…  ['128',
          '64',
          '128',
          '64',
          '128',
          '128',
          '128',
          '128',
          '128',
          '128',
          '128',
          '128',
          '128',
          '128',
          '128',
          '64',
          '128',
          '128',
          '64',
          '64',
          '256',
          '128',
          '128',
          '64']
```

From features extract the display size and kept into list

```python
In [120…  display = []
          for i in feature:
              a = re.findall(r"(\d+\.\d+)\sinch",i.text)
              if len(a)>0:
                  display.append(a[0])
          display
```

```
Out[120…   ['6.88',
            '6.7',
            '6.67',
            '6.88',
            '6.88',
            '6.7',
            '6.7',
            '6.88',
            '6.88',
            '6.67',
            '6.5',
            '6.88',
            '6.67',
            '6.88',
            '6.67',
            '6.7',
            '6.67',
            '6.72',
            '6.88',
            '6.88',
            '6.67',
            '6.67',
            '6.67',
            '6.745']
```

From features extract the battery capacity and kept into list

```python
In [121…   battery = []
           for i in feature:
               a = re.findall(r"\d+\smAh",i.text)
               if len(a)>0:
                   battery.append(a[0])
           battery
```

```
Out[121…  ['5200 mAh',
           '5000 mAh',
           '5110 mAh',
           '5200 mAh',
           '5160 mAh',
           '5000 mAh',
           '5000 mAh',
           '5160 mAh',
           '5160 mAh',
           '6000 mAh',
           '5000 mAh',
           '5160 mAh',
           '6000 mAh',
           '5200 mAh',
           '6000 mAh',
           '5000 mAh',
           '5110 mAh',
           '5000 mAh',
           '5200 mAh',
           '5160 mAh',
           '5110 mAh',
           '5500 mAh',
           '5500 mAh',
           '5000 mAh']
```

Creating DataFrame

```python
import pandas as pd
dic = {"Brand":brand,
       "Model":model,
       "Color":color,
       "GB":GB,
       "ROM":ROM,
       "Battery":battery,
       "Price":price
      }
data_frame = pd.DataFrame(dic)
data_frame
```

	Brand	Model	Color	GB	ROM	Battery	Price
0	POCO	POCO C71	Cool Blue	128	128	5200 mAh	6,799
1	Samsung	Samsung Galaxy F06 5G	Bahama Blue	64	64	5000 mAh	7,499
2	POCO	POCO M7 Pro 5G	Lunar Dust	128	128	5110 mAh	11,499
3	POCO	POCO C71	Desert Gold	64	64	5200 mAh	6,199
4	POCO	POCO M7 5G	Ocean Blue	128	128	5160 mAh	8,499
5	Samsung	Samsung Galaxy F06 5G	Bahama Blue	128	128	5000 mAh	7,999
6	Samsung	Samsung Galaxy F06 5G	Bahama Blue	128	128	5000 mAh	8,999
7	POCO	POCO M7 5G - Locked with Airtel Prepaid	Satin Black	128	128	5160 mAh	7,849
8	POCO	POCO M7 5G	Mint Green	128	128	5160 mAh	9,499
9	OPPO	OPPO K13x 5G 6000mAh and 45W SUPERVOOC Charger...	Breeze Blue	128	128	6000 mAh	12,999
10	Motorola	Motorola g45 5G	Brilliant Green	128	128	5000 mAh	11,999
11	POCO	POCO M7 5G	Mint Green	128	128	5160 mAh	8,499
12	OPPO	OPPO K13x 5G 6000mAh and 45W SUPERVOOC Charger...	Midnight Violet	128	128	6000 mAh	12,999
13	POCO	POCO C71	Power Black	128	128	5200 mAh	6,799
14	OPPO	OPPO K13x 5G 6000mAh and 45W SUPERVOOC Charger...	Sunset Peach	128	128	6000 mAh	11,999
15	Samsung	Samsung Galaxy F06 5G	Lit Violet	64	64	5000 mAh	7,499
16	POCO	POCO M7 Pro 5G	Lavender Frost	128	128	5110 mAh	11,499
17	MOTOROLA	MOTOROLA g35 5G	Guava Red	128	128	5000 mAh	9,999
18	POCO	POCO C71	Power Black	64	64	5200 mAh	6,199

	Brand	Model	Color	GB	ROM	Battery	Price
19	POCO	POCO C75 5G	Enchanted Green	64	64	5160 mAh	7,299
20	POCO	POCO M7 Pro 5G	Lunar Dust	256	256	5110 mAh	13,499
21	POCO	POCO X7 5G	Yellow	128	128	5500 mAh	14,999
22	MOTOROLA	MOTOROLA G96 5G	Pantone Greener Pastures	128	128	5500 mAh	17,999
23	Ai+	Ai+ Pulse	Blue	64	64	5000 mAh	5,499

Convert the dataframe into CSV file

```
In [126…  data_frame.to_csv("Mobiles Data.csv")
```

```
In [ ]:  # Con
```

```
In [131…  data_frame.to_excel("Mobiles Data Excel.xlsx")
```

```
In [ ]:
```