

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import requests
import scrapy
from bs4 import BeautifulSoup
```

paste which web URL you need to web scrapping

i am using flipkart top mobile phones

In [7]:

```
url='https://www.flipkart.com/search?q=top%20mobile%20phones&otracker=search&otracker1=se'
```

In [8]:

```
res=requests.get(url,headers=headers)
print(res)
```

```
<Response [200]>
```

In [10]:

```
data = BeautifulSoup(res.content, 'lxml')
data
```

Out[10]:

```
<!DOCTYPE html>
<html lang="en"><head><link href="https://rukminim2.flixcart.com" rel="p
reconnect"/><link href="//static-assets-web.flixcart.com/fk-p-linchpin-w
eb/fk-cp-zion/css/app_modules.chunk.905c37.css" rel="stylesheet"/><link
href="//static-assets-web.flixcart.com/fk-p-linchpin-web/fk-cp-zion/css/
app.chunk.d4bb4f.css" rel="stylesheet"/><meta content="text/html; charse
t=utf-8" http-equiv="Content-type"/><meta content="IE=Edge" http-equiv
="X-UA-Compatible"/><meta content="102988293558" property="fb:page_id"/>
<meta content="658873552,624500995,100000233612389" property="fb:admin
s"/><meta content="noodp" name="robots"/><link href="https://static-asse
ts-web.flixcart.com/www/promos/new/20150528-140547-favicon-retina.ico" r
el="shortcut icon"/><link href="/osdd.xml?v=2" rel="search" type="applic
ation/opensearchdescription+xml"/><meta content="website" property="og:t
ype"/><meta content="Flipkart.com" name="og_site_name" property="og:site
_name"/><link href="/apple-touch-icon-57x57.png" rel="apple-touch-icon"
sizes="57x57"/><link href="/apple-touch-icon-72x72.png" rel="apple-touch
-icon" sizes="72x72"/><link href="/apple-touch-icon-114x114.png" rel="ap
ple-touch-icon" sizes="114x114"/><link href="/apple-touch-icon-144x144.p
```

In [11]:

```
#<div class="_4rR01T">OPPO Reno10 5G (Ice Blue, 256 GB)</div>
```

In [14]:

```
w=data.find('div',attrs={'class':'_4rR01T'})
print(w)
print(w.text)
```

```
<div class="_4rR01T">OPPO Reno10 5G (Ice Blue, 256 GB)</div>
OPPO Reno10 5G (Ice Blue, 256 GB)
```

In [16]:

```
m=[]
for i in data.findAll('div',attrs={'class':'_4rR01T'}):
    m.append(i.text)
```

In [17]:

```
print(m)
print(len(m))
```

```
['OPPO Reno10 5G (Ice Blue, 256 GB)', 'OPPO Find N2 Flip (Moonlit Purple,
256 GB)', 'POCO C51 (Royal Blue, 64 GB)', 'POCO C51 (Power Black, 64 GB)',
'POCO C50 (Royal Blue, 32 GB)', 'POCO C55 (Cool Blue, 128 GB)', 'REDMI 12
(Moonstone Silver, 128 GB)', 'REDMI 12 (Pastel Blue, 128 GB)', 'POCO C55
(Power Black, 128 GB)', 'Infinix SMART 7 (Emerald Green, 64 GB)', 'Infinix
SMART 7 (Azure Blue, 64 GB)', 'Infinix SMART 7 (Night Black, 128 GB)', 'In
finix SMART 7 (Emerald Green, 128 GB)', 'Infinix SMART 7 (Azure Blue, 128
GB)', 'realme C53 (Champion Gold, 64 GB)', 'realme C53 (Champion Black, 64
GB)', 'Infinix SMART 7 (Night Black, 64 GB)', 'REDMI 12 (Pastel Blue, 128
GB)', 'MOTOROLA e13 (Aurora Green, 64 GB)', 'POCO C55 (Forest Green, 128 G
B)', 'SAMSUNG Galaxy F13 (Night sky Green, 64 GB)', 'realme C53 (Champion G
old, 128 GB)', 'vivo T2x 5G (Marine Blue, 128 GB)', 'vivo T2x 5G (Glimmer
Black, 128 GB)']
24
```

Get mobile brand names

In [18]:

```
brand=[]
for i in m:
    x=i.split("(")
    brand.append(x[0])
```

In [19]:

```
print(brand)
print(len(brand))
```

```
['OPPO Reno10 5G ', 'OPPO Find N2 Flip ', 'POCO C51 ', 'POCO C51 ', 'POCO
C50 ', 'POCO C55 ', 'REDMI 12 ', 'REDMI 12 ', 'POCO C55 ', 'Infinix SMART
7 ', 'Infinix SMART 7 ', 'Infinix SMART 7 ', 'Infinix SMART 7 ', 'Infinix
SMART 7 ', 'realme C53 ', 'realme C53 ', 'Infinix SMART 7 ', 'REDMI 12 ',
'MOTOROLA e13 ', 'POCO C55 ', 'SAMSUNG Galaxy F13 ', 'realme C53 ', 'vivo
T2x 5G ', 'vivo T2x 5G ']
24
```

In [20]:

```
#<div class="_3LWZLK">4.2<di
```

In [26]:

```
w1=data.find('span',attrs={'class': '_1lRcqv'})
print(w1)
print(w1.text)
print(len(w1))
```

```
<span class="_1lRcqv" id="productRating_LSTMOBGQP3UUYGFGYKDKTLXX3_MOBGQP3U
UYGFGYKD_"><div class="_3LWZLK">4.2</div></span>
4.2
1
```

Ratings

In []:

```
#<span class="_2_R_DZ"><span><span>12,825 Ratings&nbsp;</span><span class="_13vcmD">&
```

In [29]:

```
w2=data.find('span',attrs={'class': '_2_R_DZ'})
print(w2.text)
```

12,825 Ratings & 1,873 Reviews

In [32]:

```
n=[]
for i in data.findAll('span',attrs={'class':'_2_R_DZ'}):
    n.append(i.text)
print(n)
print(len(n))
```

```
['12,825 Ratings\xa0\xa01,873 Reviews', '269 Ratings\xa0\xa042 Reviews',
'56,272 Ratings\xa0\xa03,060 Reviews', '56,272 Ratings\xa0\xa03,060 Revi
ews', '53,052 Ratings\xa0\xa02,542 Reviews', '38,366 Ratings\xa0\xa02,22
5 Reviews', '2,665 Ratings\xa0\xa0228 Reviews', '2,665 Ratings\xa0\xa022
8 Reviews', '38,366 Ratings\xa0\xa02,225 Reviews', '63,638 Ratings\xa0\xa
a03,103 Reviews', '63,638 Ratings\xa0\xa03,103 Reviews', '63,638 Ratings
\xa0\xa03,103 Reviews', '63,638 Ratings\xa0\xa03,103 Reviews', '63,638 R
atings\xa0\xa03,103 Reviews', '17,973 Ratings\xa0\xa0888 Reviews', '17,9
73 Ratings\xa0\xa0888 Reviews', '63,638 Ratings\xa0\xa03,103 Reviews',
'8,985 Ratings\xa0\xa0690 Reviews', '45,314 Ratings\xa0\xa03,748 Review
s', '38,366 Ratings\xa0\xa02,225 Reviews', '1,67,538 Ratings\xa0\xa09,61
1 Reviews', '4,537 Ratings\xa0\xa0372 Reviews', '1,12,057 Ratings\xa0\xa
06,304 Reviews', '1,12,057 Ratings\xa0\xa06,304 Reviews']
24
```

In [36]:

```
rating=[]
for i in n:
    x=i.split('Ratings')
    rating.append(x[0])
print(rating)
```

```
['12,825 ', '269 ', '56,272 ', '56,272 ', '53,052 ', '38,366 ', '2,665 ',
'2,665 ', '38,366 ', '63,638 ', '63,638 ', '63,638 ', '63,638 ', '63,638
', '17,973 ', '17,973 ', '63,638 ', '8,985 ', '45,314 ', '38,366 ', '1,67,
538 ', '4,537 ', '1,12,057 ', '1,12,057 ']
```

RAM, ROM

In [65]:

```
#<div class="fMghEO"><ul class="_1xgFaf"><li class="rgWa7D">4 GB RAM | 64 GB ROM | Expanda
```

In [72]:

```
w3=data.find('div',attrs={'class':'fMghEO'})
print(w3.text)
```

```
8 GB RAM | 256 GB ROM17.02 cm (6.7 inch) Full HD+ Display64MP + 32MP + 8MP
| 32MP Front Camera5000 mAh BatteryDimensity 7050 Processor1 Year Manufact
urer Warranty for Phone and 6 Months Warranty for In the Box Accessories
```

In [73]:

```
ra=[]  
for i in data.findAll('div',attrs={'class':'fMghEO'}):  
    ra.append(i.text)  
print(ra)  
print(len(ra))
```

['8 GB RAM | 256 GB ROM17.02 cm (6.7 inch) Full HD+ Display64MP + 32MP + 8 MP | 32MP Front Camera5000 mAh BatteryDimensity 7050 Processor1 Year Manufacturer Warranty for Phone and 6 Months Warranty for In the Box Accessories', '8 GB RAM | 256 GB ROM17.27 cm (6.8 inch) Full HD+ Display50MP + 8MP | 32MP Front Camera4300 mAh BatteryMediatek Dimensity 9000+ Processor1 Year Manufacturer Warranty for Phone and 6 Months Warranty for In the Box Accessories', '4 GB RAM | 64 GB ROM | Expandable Upto 1 TB16.56 cm (6.52 inch) HD+ Display8MP Dual Rear Camera | 5MP Front Camera5000 mAh BatteryHelio G36 Processor1 Year Manufacturer Warranty for Phone and 6 Months Warranty for in the Box Accessories', '4 GB RAM | 64 GB ROM | Expandable Upto 1 TB16.56 cm (6.52 inch) HD+ Display8MP Dual Rear Camera | 5MP Front Camera5000 mAh BatteryHelio G36 Processor1 Year Manufacturer Warranty for Phone and 6 Months Warranty for in the Box Accessories', '2 GB RAM | 32 GB ROM | Expandable Upto 512 GB16.56 cm (6.52 inch) HD+ Display8MP Dual Camera | 5MP Front Camera5000 mAh Lithium Polymer BatteryMediatek Helio A22 Processor, Upto 2.0 GHz Processor1 Year Manufacturer Warranty for Phone and 6 Months Warranty for In the Box Accessories', '6 GB RAM | 128 GB ROM | Expandable Upto 1 TB17.04 cm (6.71 inch) HD+ Display50MP Dual Rear Camera | 5MP Front Camera5000 mAh BatteryMediatek Helio G85 Processor1 Year Manufacturer Warranty for Phone and 6 Months Warranty for In the Box Accessories', '6 GB RAM | 128 GB ROM | Expandable Upto 1 TB17.25 cm (6.79 inch) Full HD+ Display50MP + 8MP + 2MP | 8MP Front Camera5000 mAh BatteryHelio G88 Processor1 Year Manufacturer Warranty for Phone and 6 Months Warranty for in the Box Accessories', '6 GB RAM | 128 GB ROM | Expandable Upto 1 TB17.25 cm (6.79 inch) Full HD+ Display50MP + 8MP + 2MP | 8MP Front Camera5000 mAh BatteryHelio G88 Processor1 Year Manufacturer Warranty for Phone and 6 Months Warranty for in the Box Accessories', '6 GB RAM | 128 GB ROM | Expandable Upto 1 TB17.04 cm (6.71 inch) HD+ Display50MP Dual Rear Camera | 5MP Front Camera5000 mAh BatteryMediatek Helio G85 Processor1 Year Manufacturer Warranty for Phone and 6 Months Warranty for In the Box Accessories', '4 GB RAM | 64 GB ROM | Expandable Upto 2 TB16.76 cm (6.6 inch) HD+ Display13MP + AI Lens | 5MP Front Camera6000 mAh BatteryUnisoc Spreadtrum SC9863A1 Processor1 Year on Handset and 6 Months on Accessories', '4 GB RAM | 64 GB ROM | Expandable Upto 2 TB16.76 cm (6.6 inch) HD+ Display13MP + AI Lens | 5MP Front Camera6000 mAh BatteryUnisoc Spreadtrum SC9863A1 Processor1 Year on Handset and 6 Months on Accessories', '4 GB RAM | 128 GB ROM | Expandable Upto 512 GB16.76 cm (6.6 inch) HD+ Display13MP + AI Lens | 5MP Front Camera6000 mAh BatteryUnisoc Spreadtrum SC9863A1 Processor1 Year on Handset and 6 Months on Accessories', '4 GB RAM | 128 GB ROM | Expandable Upto 512 GB16.76 cm (6.6 inch) HD+ Display13MP + AI Lens | 5MP Front Camera6000 mAh BatteryUnisoc Spreadtrum SC9863A1 Processor1 Year on Handset and 6 Months on Accessories', '4 GB RAM | 128 GB ROM | Expandable Upto 512 GB16.76 cm (6.6 inch) HD+ Display13MP + AI Lens | 5MP Front Camera6000 mAh BatteryUnisoc Spreadtrum SC9863A1 Processor1 Year on Handset and 6 Months on Accessories', '6 GB RAM | 64 GB ROM | Expandable Upto 2 TB17.12 cm (6.74 inch) HD Display108MP + 2MP | 8MP Front Camera5000 mAh BatteryT612 Processor1 Year Manufacturer Warranty for Phone and 6 Months Warranty for In the Box Accessories', '6 GB RAM | 64 GB ROM | Expandable Upto 2 TB17.12 cm (6.74 inch) HD Display108MP + 2MP | 8MP Front Camera5000 mAh BatteryT612 Processor1 Year Manufacturer Warranty for Phone and 6 Months Warranty for In the Box Accessories', '4 GB RAM | 64 GB ROM | Expandable Upto 2 TB16.76 cm (6.6 inch) HD+ Display13MP + AI Lens | 5MP Front Camera6000 mAh BatteryUnisoc Spreadtrum SC9863A1 Processor1 Year on Handset and 6 Months on Accessories', '4 GB RAM | 128 GB ROM | Expandable Upto 1 TB17.25 cm (6.79 inch) Full HD+ Display50MP + 8MP + 2MP | 8MP Front Camera5000 mAh BatteryHelio G88 Processor1 Year Manufacturer Warranty for Phone and 6 Months Warranty for in the Box Accessories', '4 GB RAM | 64 GB ROM | Expandable Upto 1 TB16.51 cm (6.5 inch) HD+ Display13MP Rear Camera | 5MP Front Camera5000 mAh BatteryUnisoc T606 Processor1 Year on Handset and 6 Months on Accessories', '6 GB RAM | 128 GB ROM | Expandable Upto 1 TB17.04 cm (6.71 inch) HD+ Display50MP Dual Rear Camera | 5MP Front Camera5000 mAh BatteryMediatek Helio G85 Processor1 Year Manufact

```

urer Warranty for Phone and 6 Months Warranty for In the Box Accessories',
'4 GB RAM | 64 GB ROM | Expandable Upto 1 TB16.76 cm (6.6 inch) Full HD+ D
isplay50MP + 5MP + 2MP | 8MP Front Camera6000 mAh Lithium Ion BatteryExyno
s 850 Processor1 Year Warranty Provided By the Manufacturer from Date of P
urchase', '4 GB RAM | 128 GB ROM | Expandable Upto 2 TB17.12 cm (6.74 inc
h) HD Display108MP + 2MP | 8MP Front Camera5000 mAh BatteryT612 Processor1
Year Manufacturer Warranty for Phone and 6 Months Warranty for In the Box
Accessories', '6 GB RAM | 128 GB ROM16.71 cm (6.58 inch) Full HD+ Display5
0MP + 2MP | 8MP Front Camera5000 mAh BatteryDimensity 6020 Processor1 Year
of Device & 6 Months for Inbox Accessories', '6 GB RAM | 128 GB ROM16.71 c
m (6.58 inch) Full HD+ Display50MP + 2MP | 8MP Front Camera5000 mAh Batter
yDimensity 6020 Processor1 Year of Device & 6 Months for Inbox Accessorie
s']
24

```

In [77]:

```

ram = []
rom = []
front_camera= []
for i in ra:
    x = i.split('|')
    ram.append(x[0])
    rom.append(x[1])
    front_camera.append(x[2])

```


In [79]:

```
print(rom)
```

```
[' 256 GB ROM17.02 cm (6.7 inch) Full HD+ Display64MP + 32MP + 8MP ', ' 25
6 GB ROM17.27 cm (6.8 inch) Full HD+ Display50MP + 8MP ', ' 64 GB ROM ', '
64 GB ROM ', ' 32 GB ROM ', ' 128 GB ROM ', ' 128 GB ROM ', ' 128 GB ROM
', ' 128 GB ROM ', ' 64 GB ROM ', ' 64 GB ROM ', ' 128 GB ROM ', ' 128 GB
ROM ', ' 128 GB ROM ', ' 64 GB ROM ', ' 64 GB ROM ', ' 64 GB ROM ', ' 128
GB ROM ', ' 64 GB ROM ', ' 128 GB ROM ', ' 64 GB ROM ', ' 128 GB ROM ', '
128 GB ROM16.71 cm (6.58 inch) Full HD+ Display50MP + 2MP ', ' 128 GB ROM1
6.71 cm (6.58 inch) Full HD+ Display50MP + 2MP ']
```

RAM

In [82]:

```
RAM=[]
for i in ram:
    x=i.split(' RAM ')
    RAM.append(x[0])
print(RAM)
print(len(RAM))
```

```
['8 GB', '8 GB', '4 GB', '4 GB', '2 GB', '6 GB', '6 GB', '6 GB', '6 GB',
'4 GB', '4 GB', '4 GB', '4 GB', '4 GB', '6 GB', '6 GB', '4 GB', '4 GB', '4
GB', '6 GB', '4 GB', '4 GB', '6 GB', '6 GB']
24
```

ROM

In [83]:

```
ROM=[]
for i in rom:
    x=i.split("ROM")
    ROM.append(x[0])
print(ROM)
print(len(ROM))
```

```
[' 256 GB ', ' 256 GB ', ' 64 GB ', ' 64 GB ', ' 32 GB ', ' 128 GB ', ' 12
8 GB ', ' 128 GB ', ' 128 GB ', ' 64 GB ', ' 64 GB ', ' 128 GB ', ' 128 GB
', ' 128 GB ', ' 64 GB ', ' 64 GB ', ' 64 GB ', ' 128 GB ', ' 64 GB ', ' 1
28 GB ', ' 64 GB ', ' 128 GB ', ' 128 GB ', ' 128 GB ']
```

In [90]:

```
del(front_camera)
```

price,discount, and after discount price

In [91]:

```
#<div class="_3tbKJL"><div class="_25b18c"><div class="_30jeq3 _1_WHN1">₹6,499</div><div
```

In [92]:

```
w4=data.find('div',attrs={'class':'_3tbKJL'})
print(w4.text)
```

₹32,999₹38,99915% offFree delivery

In [94]:

```
p=[]
for i in data.findAll('div',attrs={'class':'_3tbKJL'}):
    p.append(i.text)
print(p)
print(len(p))
```

```
['₹32,999₹38,99915% offFree delivery', '₹89,999₹99,99910% offFree deliver
y', '₹6,499₹9,99935% offFree delivery by Today', '₹6,499₹9,99935% offFree
delivery by Today', '₹5,499₹8,99938% offFree delivery by Today', '₹8,499₹1
3,99939% offFree delivery by Today', '₹11,499₹15,99928% off', '₹11,499₹15,
99928% off', '₹8,499₹13,99939% offFree delivery by Today', '₹6,999₹9,9993
0% offFree delivery by Today', '₹6,999₹9,99930% offFree delivery by Toda
y', '₹7,599₹10,99930% offFree delivery by Today', '₹7,599₹10,99930% offFre
e delivery by Today', '₹7,599₹10,99930% offFree delivery by Today', '₹10,9
99₹12,99915% off', '₹10,999₹12,99915% offFree delivery by Today', '₹6,999
₹9,99930% off', '₹9,999₹14,99933% off', '₹7,299₹10,99933% offFree delivery
by Today', '₹8,499₹13,99939% offFree delivery by Today', '₹10,999₹14,9992
6% offFree delivery by Today', '₹9,999₹11,99916% offFree delivery by Toda
y', '₹13,999₹18,99926% offFree delivery by Today', '₹13,999₹18,99926% offF
ree delivery by Today']
24
```

In [97]:

```
p1=[]
d=[]
d1=[]
for i in p:
    x=i.split('₹')
    p1.append(x[0])
    d.append(x[1])
    d1.append(x[2])
```


In [103]:

```
print(d1)
```

```
['38,99915% offFree delivery', '99,99910% offFree delivery', '9,99935% off  
Free delivery by Today', '9,99935% offFree delivery by Today', '8,99938% o  
ffFree delivery by Today', '13,99939% offFree delivery by Today', '15,9992  
8% off', '15,99928% off', '13,99939% offFree delivery by Today', '9,99930%  
offFree delivery by Today', '9,99930% offFree delivery by Today', '10,9993  
0% offFree delivery by Today', '10,99930% offFree delivery by Today', '10,  
99930% offFree delivery by Today', '12,99915% off', '12,99915% offFree del  
ivery by Today', '9,99930% off', '14,99933% off', '10,99933% offFree deliv  
ery by Today', '13,99939% offFree delivery by Today', '14,99926% offFree d  
elivery by Today', '11,99916% offFree delivery by Today', '18,99926% offFr  
ee delivery by Today', '18,99926% offFree delivery by Today']
```

In [104]:

```
d2=[]  
for i in d1:  
    x=i.split(",999")  
    d2.append(x[1])
```

In [105]:

```
print(d2)
```

```
['15% offFree delivery', '10% offFree delivery', '35% offFree delivery by  
Today', '35% offFree delivery by Today', '38% offFree delivery by Today',  
'39% offFree delivery by Today', '28% off', '28% off', '39% offFree delive  
ry by Today', '30% offFree delivery by Today', '30% offFree delivery by To  
day', '30% offFree delivery by Today', '30% offFree delivery by Today', '3  
0% offFree delivery by Today', '15% off', '15% offFree delivery by Today',  
'30% off', '33% off', '33% offFree delivery by Today', '39% offFree delive  
ry by Today', '26% offFree delivery by Today', '16% offFree delivery by To  
day', '26% offFree delivery by Today', '26% offFree delivery by Today']
```

discount

In [122]:

```
discount=[]  
for i in d2:  
    x=i.split(' ')  
    discount.append(x[0])
```

In [123]:

```
print(discount)
```

```
['15%', '10%', '35%', '35%', '38%', '39%', '28%', '28%', '39%', '30%', '3  
0%', '30%', '30%', '30%', '15%', '15%', '30%', '33%', '33%', '39%', '26%',  
'16%', '26%', '26%']
```

In [112]:

```
d3=[]
for i in d1:
    x=i.split("%")
    d3.append(x[0])
print(d3)
print(len(d3))
```

```
['38,99915', '99,99910', '9,99935', '9,99935', '8,99938', '13,99939', '15,
99928', '15,99928', '13,99939', '9,99930', '9,99930', '10,99930', '10,9993
0', '10,99930', '12,99915', '12,99915', '9,99930', '14,99933', '10,99933',
'13,99939', '14,99926', '11,99916', '18,99926', '18,99926']
24
```

price

In [117]:

```
MRP=[x[:-2]for x in d3]
print(MRP)
```

```
['38,999', '99,999', '9,999', '9,999', '8,999', '13,999', '15,999', '15,99
9', '13,999', '9,999', '9,999', '10,999', '10,999', '10,999', '12,999', '1
2,999', '9,999', '14,999', '10,999', '13,999', '14,999', '11,999', '18,99
9', '18,999']
```

In [127]:

```
df=pd.DataFrame({"Mobile_Brand":brand,"Rating":rating,"RAM":RAM,"ROM":ROM,"Orignal_price"
"Price":dis_price})
```

In [128]:

```
df.head()
```

Out[128]:

	Mobile_Brand	Rating	RAM	ROM	Orignal_price	Discount	Price
0	OPPO Reno10 5G	12,825	8 GB	256 GB	38,999	15%	32,999
1	OPPO Find N2 Flip	269	8 GB	256 GB	99,999	10%	89,999
2	POCO C51	56,272	4 GB	64 GB	9,999	35%	6,499
3	POCO C51	56,272	4 GB	64 GB	9,999	35%	6,499
4	POCO C50	53,052	2 GB	32 GB	8,999	38%	5,499

In [129]:

```
df.to_csv('mobile.csv',encoding='utf-8')
```

In []:

