

```
In [1]: import requests
from bs4 import BeautifulSoup as bs
import pandas as pd
import os
import numpy as np
from itertools import zip_longest
```

Data set from web scraping

```
In [2]: df=pd.read_csv(r"E:\2. MDS DEAKIN UNIVERSITY\TASK\Task8\mobile_recommendation_system_dataset.csv")
```

```
In [3]: df.head()
```

```
Out[3]:
```

	name	ratings	price	imgURL	corpus
0	REDMI Note 12 Pro 5G (Onyx Black, 128 GB)	4.2	23999	https://rukminim2.flixcart.com/image/312/312/x...	Storage128 GBRAM6 SystemAndroid 12Processor T...
1	OPPO F11 Pro (Aurora Green, 128 GB)	4.5	₹20,999	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 GBRAM6 GBExpandable Storage256GB S...
2	REDMI Note 11 (Starburst White, 64 GB)	4.2	13149	https://rukminim2.flixcart.com/image/312/312/x...	Storage64 GBRAM4 SystemAndroid 11Processor Sp...
3	OnePlus Nord CE 5G (Blue Void, 256 GB)	4.1	21999	https://rukminim2.flixcart.com/image/312/312/x...	Storage256 GBRAM12 SystemAndroid Q 11Processo...
4	APPLE iPhone 13 mini (Blue, 128 GB)	4.6	3537	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 SystemiOS 15Processor TypeA15 Bion...

```
In [6]: df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2546 entries, 0 to 2545
Data columns (total 5 columns):
 #   Column    Non-Null Count  Dtype  
---  -
 0   name      2546 non-null   object 
 1   ratings   2546 non-null   float64
 2   price     2546 non-null   object 
 3   imgURL    2546 non-null   object 
 4   corpus    2534 non-null   object 
dtypes: float64(1), object(4)
memory usage: 99.6+ KB

```

In [7]: `df.describe()`

Out[7]:

	ratings
count	2546.000000
mean	4.295797
std	0.214691
min	2.900000
25%	4.200000
50%	4.300000
75%	4.400000
max	5.000000

In [8]: `df.isnull().sum()`

Out[8]:

```

name      0
ratings   0
price     0
imgURL    0
corpus    12
dtype: int64

```

```
In [9]: df = df.dropna()
```

```
In [10]: df.head()
```

```
Out[10]:
```

	name	ratings	price	imgURL	corpus
0	REDMI Note 12 Pro 5G (Onyx Black, 128 GB)	4.2	23999	https://rukminim2.flixcart.com/image/312/312/x...	Storage128 GB RAM6 SystemAndroid 12Processor T...
1	OPPO F11 Pro (Aurora Green, 128 GB)	4.5	₹20,999	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 GB RAM6 GBExpandable Storage256GB S...
2	REDMI Note 11 (Starburst White, 64 GB)	4.2	13149	https://rukminim2.flixcart.com/image/312/312/x...	Storage64 GB RAM4 SystemAndroid 11Processor Sp...
3	OnePlus Nord CE 5G (Blue Void, 256 GB)	4.1	21999	https://rukminim2.flixcart.com/image/312/312/x...	Storage256 GB RAM12 SystemAndroid Q 11Processo...
4	APPLE iPhone 13 mini (Blue, 128 GB)	4.6	3537	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 SystemiOS 15Processor TypeA15 Bion...

```
In [11]: df.isnull().sum()
```

```
Out[11]: name      0
ratings    0
price      0
imgURL     0
corpus     0
dtype: int64
```

```
In [12]: df.isnull().any
```

```
Out[12]: <bound method DataFrame.any of      name ratings price imgURL corpus
0      False      False False  False False
1      False      False False  False False
2      False      False False  False False
3      False      False False  False False
4      False      False False  False False
...      ...      ...    ...    ...    ...
2540     False      False False  False False
2541     False      False False  False False
2542     False      False False  False False
2544     False      False False  False False
2545     False      False False  False False

[2534 rows x 5 columns]>
```

```
In [13]: df['price'].unique()
```

```
Out[13]: array(['23999', '₹20,999', '13149', '21999', '3537', '5537', '18996',  
              '9999', '16999', '12999', '15990', '6999', '₹10,990', '24463',  
              '7537', '15537', '31999', '10990', '17537', '₹13,990', '14463',  
              '₹22,990', '19087', '9499', '31637', '32537', '6499', '9490',  
              '7799', '₹9,499', '14999', '25937', '10280', '28537', '15999',  
              '14990', '73', '22000', '13999', '₹17,499', '19999', '13499',  
              '30537', '16200', '15790', '10999', '22537', '19990', '12990',  
              '10499', '12797', '28295', '8999', '23990', '18490', '1399',  
              '26927', '10537', '6099', '24927', '10073', '₹15,999', '11086',  
              '₹16,990', '11499', '7299', '8998', '15390', '14499', '9989',  
              '8537', '29960', '25000', '17999', '₹41,990', '9299', '32546',  
              '23537', '₹14,990', '8149', '₹12,990', '11999', '21800', '2899',  
              '23837', '6299', '18990', '4464', '8490', '₹15,799', '14890',  
              '13990', '17699', '27172', '23636', '12749', '₹14,999', '15499',  
              '31602', '27537', '21172', '5999', '27990', '20999', '19864',  
              '25856', '28999', '15337', '20537', '23463', '12498', '7490',  
              '24990', '₹13,999', '18499', '9261', '14885', '6927', '24999',  
              '7999', '13980', '14476', '18937', '1463', '3073', '6463', '20990',  
              '8499', '14775', '13828', '1172', '₹19,990', '16479', '₹16,980',  
              '₹27,990', '17691', '9049', '12280', '8597', '26073', '15526',  
              '12499', '29999', '12894', '25990', '24364', '19364', '26364',  
              '24537', '10490', '7773', '₹32,990', '22546', '28885', '31073',  
              '₹32,999', '28990', '₹39,999', '24546', '11990', '₹21,990',  
              '21990', '8864', '2890', '29073', '9537', '8399', '18999',  
              '₹39,990', '6464', '8464', '24900', '₹9,990', '₹17,990', '8828',  
              '8699', '28185', '16990', '29990', '7990', '17890', '27819',  
              '₹17,999', '12740', '28799', '21252', '₹54,999', '25546', '18884',  
              '28546', '14537', '18537', '28858', '17480', '2299', '15798',  
              '21490', '25989', '7199', '15490', '19890', '₹8,000', '29547',  
              '16537', '14299', '9000', '₹27,999', '₹18,990', '9990', '14989',  
              '11099', '15546', '22636', '18890', '₹15,499', '11249', '12790',  
              '23499', '1500', '5499', '21495', '26537', '17700', '30546',  
              '24472', '30780', '6990', '9799', '22999', '14900', '30163',  
              '8095', '15690', '24890', '20546', '7454', '7989', '22036',  
              '11490', '8697', '29463', '3828', '8541', '7479', '30999', '22221',  
              '₹20,990', '7399', '13536', '7760', '21000', '8888', '8977',  
              '1850', '15249', '7499', '₹13,490', '7295', '22840', '26999',  
              '32699', '14888', '₹23,990', '23900', '11073', '9772', '13940',  
              '₹10,490', '18609', '3463', '8601', '9575', '5073', '828', '29656',  
              '25999', '13402', '6199', '14111', '₹45,999', '17990', '8890',
```

'10300', '₹15,990', '₹23,999', '8899', '₹29,990', '32536', '14248',
'4463', '10649', '₹29,999', '14198', '13927', '8210', '₹89,999',
'11464', '3927', '7460', '10985', '537', '24373', '11890', '6172',
'30499', '9228', '14285', '₹35,990', '23068', '17636', '18927',
'7290', '₹49,990', '19927', '29775', '₹29,800', '17499', '14648',
'15200', '11640', '1537', '₹15,490', '9590', '7463', '6500',
'30063', '25980', '20846', '10796', '12840', '27599', '12900',
'8749', '6599', '8949', '14099', '11537', '9454', '14049', '13810',
'6489', '25537', '13995', '17676', '18290', '9239', '10397',
'11370', '8185', '12636', '30636', '28130', '15980', '28429',
'₹18,300', '13363', '21290', '17600', '849', '8878', '32037',
'6998', '14995', '₹16,490', '26463', '8400', '19000', '9448',
'₹24,999', '₹18,690', '9498', '17754', '₹28,990', '6750', '17490',
'₹9,900', '7364', '30490', '11290', '2149', '13199', '14699',
'₹9,490', '7790', '8788', '8482', '32495', '12879', '21070',
'17000', '22898', '₹21,999', '6660', '15699', '7450', '29499',
'13998', '2399', '8403', '8650', '31500', '₹11,990', '25900',
'23546', '17678', '8200', '5900', '11740', '₹11,309', '9080',
'6495', '30990', '5964', '29198', '12799', '536', '16795', '29992',
'2499', '15636', '₹13,989', '21563', '25036', '21249', '15485',
'13537', '27999', '27546', '13648', '6969', '29537', '29546',
'8991', '9670', '₹15,498', '31609', '24000', '7449', '3036',
'11975', '8150', '21546', '11637', '11200', '22980', '16499',
'25636', '31546', '21450', '25609', '12185', '13586', '12890',
'28746', '₹11,490', '25802', '22900', '8795', '1799', '14889',
'₹34,990', '11799', '31000', '27250', '₹60,990', '8799', '22970',
'26273', '6609', '4790', '30073', '19888', '17900', '27585',
'11172', '19900', '₹18,500', '463', '17996', '28346', '6490',
'24609', '24037', '29998', '17989', '₹24,990', '8299', '24989',
'7309', '16290', '10749', '9463', '15674', '₹29,489', '27499',
'20464', '22300', '₹8,990', '9779', '19979', '32556', '27636',
'23737', '12480', '27699', '21639', '11498', '26462', '16546',
'21499', '30633', '19949', '10636', '14640', '23137', '8288',
'23927', '19499', '11998', '9949', '20500', '₹7,999', '1662',
'1449', '22663', '19728', '₹10,900', '9689', '11899', '12830',
'₹15,800', '12980', '10470', '8049', '9980', '8990', '14964',
'7590', '₹16,970', '16295', '18600', '6877', '9849', '10538',
'6938', '12536', '8714', '10449', '12036', '2099', '8538', '13945',
'12490', '₹19,899', '9939', '20880', '₹9,779', '11199', '26990',
'17488', '31537', '21073', '24499', '11648', '10899', '26899',
'29730', '7316', '₹34,999', '14599', '22890', '27944', '27490',

```
'8290', '11699', '10190', '32557', '10500', '22490', '15248',
'18989', '10000', '9400', '16199', '12331', '12887', '12246',
'8805', '10700', '18998', '10546', '26499', '9450', '759', '30500',
'₹18,142', '12995', '21430', '21599', '12974', '13600', '₹12,200',
'11000', '12989', '11530', '₹23,789', '11599', '19795', '11535',
'₹18,049', '8195', '13699', '12899', '₹24,899', '8249', '19546',
'2879', '8690', '₹9,790', '19464', '11989', '10599', '21399',
'14137', '22990', '18799', '11400', '30900', '19463', '8449',
'₹31,990', '8598', '8450', '15964', '₹13,245', '16399', '14490',
'7189', '14120', '4999', '16490', '₹18,999', '31899', '12730',
'25037', '₹899', '₹20,880', '11195', '21199', '12398', '19977',
'₹30,990', '24490', '17964', '₹21,452', '30373', '6708', '20499',
'10849', '29208', '23899', '10890', '12520', '28036', '26037',
'6950', '6455', '31633', '18900', '21537', '7310', '₹12,499',
'₹8,590', '18790', '1999', '8390', '10699', '13430', '28700',
'20328', '19989', '28880', '12867', '9199', '23352', '16899',
'20909', '1499', '8487', '13550', '23036', '31551', '11500',
'6073', '14235', '25490', '12500', '32000', '14100', '19899',
'15995', '31172', '18546', '26464', '7329', '16390', '20538',
'15495', '3536', '6720', '7430', '10870', '9725', '25781', '12537',
'31037', '30862', '22444', '17464', '30658', '20349', '₹20,376',
'14083', '₹12,799', '24172', '31990', '16940', '5463', '7995',
'21637', '18299', '28864', '1573', '26546', '14298', '₹13,480',
'29220', '15300', '1299', '24046', '19955', '16989', '₹13,390',
'10390', '5646', '8875', '9945', '21072', '21299', '24325', '7791',
'19488', '27440', '30203', '20463'], dtype=object)
```

Removing special character in price of dataset

In [23]: *# Step 1: Ensure the column is of string type*

```
df['price'] = df['price'].astype(str)
```

Step 2: Remove ₹ symbol and commas

```
df['price'] = df['price'].str.replace('₹', '', regex=False) # Remove ₹
df['price'] = df['price'].str.replace(',', '', regex=False) # Remove commas
```

In [24]: *# Step 3: Strip any leading or trailing whitespaces*

```
df['price'] = df['price'].str.strip()
```

```
In [25]: # Step 4: Convert to numeric (this handles any errors and converts invalid values to NaN)
df['price'] = pd.to_numeric(df['price'], errors='coerce')
```

```
In [26]: # Step 5: Check if there are any remaining non-numeric values
print("Data type of price after conversion:", df['price'].dtype)
print("Missing prices after cleaning:", df['price'].isnull().sum())
```

Data type of price after conversion: int64
Missing prices after cleaning: 0

```
In [27]: # Step 6: Optional - drop rows with missing prices
df = df.dropna(subset=['price'])
```

```
In [28]: df.head()
```

```
Out[28]:
```

	name	ratings	price	imgURL	corpus
0	REDMI Note 12 Pro 5G (Onyx Black, 128 GB)	4.2	23999	https://rukminim2.flixcart.com/image/312/312/x...	Storage128 GB RAM6 SystemAndroid 12Processor T...
1	OPPO F11 Pro (Aurora Green, 128 GB)	4.5	20999	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 GB RAM6 GBExpandable Storage256GB S...
2	REDMI Note 11 (Starburst White, 64 GB)	4.2	13149	https://rukminim2.flixcart.com/image/312/312/x...	Storage64 GB RAM4 SystemAndroid 11Processor Sp...
3	OnePlus Nord CE 5G (Blue Void, 256 GB)	4.1	21999	https://rukminim2.flixcart.com/image/312/312/x...	Storage256 GB RAM12 SystemAndroid Q 11Processo...
4	APPLE iPhone 13 mini (Blue, 128 GB)	4.6	3537	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 SystemiOS 15Processor TypeA15 Bion...

```
In [29]: df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Index: 2534 entries, 0 to 2545
Data columns (total 5 columns):
 #   Column      Non-Null Count  Dtype
---  ---
 0   name        2534 non-null   object
 1   ratings     2534 non-null   float64
 2   price       2534 non-null   int64
 3   imgURL      2534 non-null   object
 4   corpus      2534 non-null   object
dtypes: float64(1), int64(1), object(3)
memory usage: 118.8+ KB

```

Overview 'name' column

```
In [40]: len(df['name'].unique())
```

```
Out[40]: 2128
```

```
In [39]: df[df['name']=='SAMSUNG Galaxy F13 (Waterfall Blue, 128 GB)']
```

```
Out[39]:
```

	name	ratings	price	imgURL	corpus
1403	SAMSUNG Galaxy F13 (Waterfall Blue, 128 GB)	4.3	10499	https://rukminim2.flixcart.com/image/312/312/x...	Storage128 GB RAM4 GBExpandable Storage1TB Sys...

```
In [41]: df['corpus'][0]
```

```
Out[41]: 'Storage128 GB RAM6 SystemAndroid 12Processor TypeMediatek Dimensity 1080Processor Speed2.6 50MP 8MP 2MP 50MP 16MP 5G Capacit
y5000 Display Size16.94 cm (6.67 inch)Resolution2400 x 1080 PixelsResolution TypeFull HD+ AMOLED DisplayGPUARM Mali-G68 MC4Di
splay TypeFull HD+ AMOLED DisplayOther Display FeaturesRefresh Rate: 120 Hz, Contrast: 5,000,000:1, 1920Hz PWM Dimming, Brigh
tness Level: 16000:1, Peak Brightness: 900 nits'
```

Extracting require informatio and perform feature slection

```

In [43]: import pandas as pd
import re

def extract_features(corpus):
    features = {}

    # Extract Storage
    features['storage'] = re.search(r'Storage(\d+GB)', corpus)
    features['storage'] = features['storage'].group(1) if features['storage'] else None

    # Extract RAM
    features['ram'] = re.search(r'RAM(\d+)', corpus)
    features['ram'] = features['ram'].group(1) + " GB" if features['ram'] else None

    # Extract OS
    features['os'] = re.search(r'System([A-Za-z0-9 ]+)', corpus)
    features['os'] = features['os'].group(1) if features['os'] else None

    # Extract Processor Type
    features['processor_type'] = re.search(r'Processor Type([A-Za-z0-9 ]+)', corpus)
    features['processor_type'] = features['processor_type'].group(1) if features['processor_type'] else None

    # Extract Processor Speed
    features['processor_speed'] = re.search(r'Processor Speed([0-9.]+)', corpus)
    features['processor_speed'] = features['processor_speed'].group(1) if features['processor_speed'] else None

    # Extract Rear Camera Primary
    features['rear_camera_primary'] = re.search(r'(\d+)MP \d+MP \d+MP (\d+)MP', corpus)
    features['rear_camera_primary'] = features['rear_camera_primary'].group(1) if features['rear_camera_primary'] else None

    # Extract Rear Camera Secondary
    features['rear_camera_secondary'] = re.search(r'(\d+)MP \d+MP (\d+)MP', corpus)
    features['rear_camera_secondary'] = features['rear_camera_secondary'].group(2) if features['rear_camera_secondary'] else None

    # Extract Front Camera Primary
    features['front_camera_primary'] = re.search(r'(\d+)MP \d+MP (\d+)MP (\d+)MP', corpus)
    features['front_camera_primary'] = features['front_camera_primary'].group(1) if features['front_camera_primary'] else None

    # Extract Front Camera Secondary

```

```

features['front_camera_secondary'] = re.search(r'(\d+)MP \d+MP (\d+)MP (\d+)MP', corpus)
features['front_camera_secondary'] = features['front_camera_secondary'].group(2) if features['front_camera_secondary'] else None

# Extract Display Size
features['display_size'] = re.search(r'Display Size([\d.]+ cm)', corpus)
features['display_size'] = features['display_size'].group(1) if features['display_size'] else None

# Extract Display Resolution
features['display_resolution'] = re.search(r'Resolution(\d+ x \d+)', corpus)
features['display_resolution'] = features['display_resolution'].group(1) if features['display_resolution'] else None

# Extract Display Type
features['display_type'] = re.search(r'Display Type([A-Za-z0-9 ]+)', corpus)
features['display_type'] = features['display_type'].group(1) if features['display_type'] else None

# Extract Refresh Rate
features['refresh_rate'] = re.search(r'Refresh Rate: (\d+ Hz)', corpus)
features['refresh_rate'] = features['refresh_rate'].group(1) if features['refresh_rate'] else None

# Extract Screen-to-Body Ratio
features['screen_to_body_ratio'] = re.search(r'Screen-to-Body Ratio: ([\d.]+%)', corpus)
features['screen_to_body_ratio'] = features['screen_to_body_ratio'].group(1) if features['screen_to_body_ratio'] else None

# Extract Aspect Ratio
features['aspect_ratio'] = re.search(r'Aspect Ratio: ([\d:]+)', corpus)
features['aspect_ratio'] = features['aspect_ratio'].group(1) if features['aspect_ratio'] else None

# Extract Network
features['network'] = re.search(r'(\d+G)', corpus)
features['network'] = features['network'].group(1) if features['network'] else None

# Extract Battery
features['battery'] = re.search(r'Capacity(\d+)', corpus)
features['battery'] = features['battery'].group(1) + " mAh" if features['battery'] else None

return features

# Apply the extraction function to each row in the corpus column
df_features = df['corpus'].apply(extract_features).apply(pd.Series)

```

```
# Add the extracted features as new columns to the original DataFrame
df = pd.concat([df, df_features], axis=1)
```

In [44]: df.head()

Out[44]:

	name	ratings	price	imgURL	corpus	storage	ram	os	processor_type
0	REDMI Note 12 Pro 5G (Onyx Black, 128 GB)	4.2	23999	https://rukminim2.flixcart.com/image/312/312/x...	Storage128 GBRAM6 SystemAndroid 12Processor T...	None	6 GB	Android 12Processor TypeMediatek Dimensity 108...	Mediatek Dimensity 1080Processor Speed2
1	OPPO F11 Pro (Aurora Green, 128 GB)	4.5	20999	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 GBRAM6 GBExpandable Storage256GB S...	256GB	6 GB	Android Pie 9	MediaTek Helio P70 Octa Core 2
2	REDMI Note 11 (Starburst White, 64 GB)	4.2	13149	https://rukminim2.flixcart.com/image/312/312/x...	Storage64 GBRAM4 SystemAndroid 11Processor Sp...	None	4 GB	Android 11Processor Speed2	None
3	OnePlus Nord CE 5G (Blue Void, 256 GB)	4.1	21999	https://rukminim2.flixcart.com/image/312/312/x...	Storage256 GBRAM12 SystemAndroid Q 11Processo...	None	12 GB	Android Q 11Processor TypeQualcomm Snapdragon ...	Qualcomm Snapdragon Octa Core 750G 5G Processo...
4	APPLE iPhone 13 mini (Blue, 128 GB)	4.6	3537	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 SystemiOS 15Processor TypeA15 Bion...	None	None	iOS 15Processor TypeA15 Bionic Chip 12MP 12MP ...	A15 Bionic Chip 12MP 12MP 12MP 63MP 12MP 12MP ...

5 rows × 22 columns



Feature selection dropping some variable which is not require

```
In [45]: df = df.drop(columns=['screen_to_body_ratio', 'aspect_ratio'])  
df.head()
```

Out[45]:

	name	ratings	price	imgURL	corpus	storage	ram	os	processor_type
0	REDMI Note 12 Pro 5G (Onyx Black, 128 GB)	4.2	23999	https://rukminim2.flixcart.com/image/312/312/x...	Storage128 GBRAM6 SystemAndroid 12Processor T...	None	6 GB	Android 12Processor TypeMediatek Dimensity 108...	Mediatek Dimensity 1080Processor Speed2
1	OPPO F11 Pro (Aurora Green, 128 GB)	4.5	20999	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 GBRAM6 GBExpandable Storage256GB S...	256GB	6 GB	Android Pie 9	MediaTek Helio P70 Octa Core 2
2	REDMI Note 11 (Starburst White, 64 GB)	4.2	13149	https://rukminim2.flixcart.com/image/312/312/x...	Storage64 GBRAM4 SystemAndroid 11Processor Sp...	None	4 GB	Android 11Processor Speed2	None
3	OnePlus Nord CE 5G (Blue Void, 256 GB)	4.1	21999	https://rukminim2.flixcart.com/image/312/312/x...	Storage256 GBRAM12 SystemAndroid Q 11Processo...	None	12 GB	Android Q 11Processor TypeQualcomm Snapdragon ...	Qualcomm Snapdragon Octa Core 750G 5G Processo...
4	APPLE iPhone 13 mini (Blue, 128 GB)	4.6	3537	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 SystemiOS 15Processor TypeA15 Bion...	None	None	iOS 15Processor TypeA15 Bionic Chip 12MP 12MP ...	A15 Bionic Chip 12MP 12MP 12MP 63MP 12MP 12MP ...

```
In [46]: df = df.drop(columns=['refresh_rate', 'display_type'])
df.head()
```

Out[46]:

	name	ratings	price	imgURL	corpus	storage	ram	os	processor_type
0	REDMI Note 12 Pro 5G (Onyx Black, 128 GB)	4.2	23999	https://rukminim2.flixcart.com/image/312/312/x...	Storage128 GBRAM6 SystemAndroid 12Processor T...	None	6 GB	Android 12Processor TypeMediatek Dimensity 108...	Mediatek Dimensity 1080Processor Speed2
1	OPPO F11 Pro (Aurora Green, 128 GB)	4.5	20999	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 GBRAM6 GBExpandable Storage256GB S...	256GB	6 GB	Android Pie 9	MediaTek Helio P70 Octa Core 2
2	REDMI Note 11 (Starburst White, 64 GB)	4.2	13149	https://rukminim2.flixcart.com/image/312/312/x...	Storage64 GBRAM4 SystemAndroid 11Processor Sp...	None	4 GB	Android 11Processor Speed2	None
3	OnePlus Nord CE 5G (Blue Void, 256 GB)	4.1	21999	https://rukminim2.flixcart.com/image/312/312/x...	Storage256 GBRAM12 SystemAndroid Q 11Processo...	None	12 GB	Android Q 11Processor TypeQualcomm Snapdragon ...	Qualcomm Snapdragon Octa Core 750G 5G Processo...
4	APPLE iPhone 13 mini (Blue, 128 GB)	4.6	3537	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 SystemiOS 15Processor TypeA15 Bion...	None	None	iOS 15Processor TypeA15 Bionic Chip 12MP 12MP ...	A15 Bionic Chip 12MP 12MP 12MP 63MP 12MP 12MP ...

In [49]: `df['corpus']`

Out[49]: 'Storage128 GBRAM6 SystemAndroid 12Processor TypeMediatek Dimensity 1080Processor Speed2.6 50MP 8MP 2MP 50MP 16MP 5G Capacit
y5000 Display Size16.94 cm (6.67 inch)Resolution2400 x 1080 PixelsResolution TypeFull HD+ AMOLED DisplayGPUARM Mali-G68 MC4Di
splay TypeFull HD+ AMOLED DisplayOther Display FeaturesRefresh Rate: 120 Hz, Contrast: 5,000,000:1, 1920Hz PWM Dimming, Brigh
tness Level: 16000:1, Peak Brightness: 900 nits'

```
In [47]: df.isnull().sum()
```

```
Out[47]: name                0
ratings                    0
price                     0
imgURL                    0
corpus                    0
storage                  1829
ram                      394
os                        7
processor_type           524
processor_speed          447
rear_camera_primary      994
rear_camera_secondary    774
front_camera_primary     994
front_camera_secondary   994
display_size             0
display_resolution       292
network                  0
battery                  328
dtype: int64
```

```
In [50]: # Use regex to extract the internal storage value after the word 'Storage'
df['internal_storage'] = df['corpus'].str.extract(r'Storage(\d+)\sGB')
```

```
In [51]: df.head()
```

Out[51]:

	name	ratings	price	imgURL	corpus	storage	ram	os	processor_type
0	REDMI Note 12 Pro 5G (Onyx Black, 128 GB)	4.2	23999	https://rukminim2.flixcart.com/image/312/312/x...	Storage128 GBRAM6 SystemAndroid 12Processor T...	None	6 GB	Android 12Processor TypeMediatek Dimensity 108...	Mediatek Dimensity 1080Processor Speed2
1	OPPO F11 Pro (Aurora Green, 128 GB)	4.5	20999	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 GBRAM6 GBExpandable Storage256GB S...	256GB	6 GB	Android Pie 9	MediaTek Helio P70 Octa Core 2
2	REDMI Note 11 (Starburst White, 64 GB)	4.2	13149	https://rukminim2.flixcart.com/image/312/312/x...	Storage64 GBRAM4 SystemAndroid 11Processor Sp...	None	4 GB	Android 11Processor Speed2	None
3	OnePlus Nord CE 5G (Blue Void, 256 GB)	4.1	21999	https://rukminim2.flixcart.com/image/312/312/x...	Storage256 GBRAM12 SystemAndroid Q 11Processo...	None	12 GB	Android Q 11Processor TypeQualcomm Snapdragon ...	Qualcomm Snapdragon Octa Core 750G 5G Processo...
4	APPLE iPhone 13 mini (Blue, 128 GB)	4.6	3537	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 SystemiOS 15Processor TypeA15 Bion...	None	None	iOS 15Processor TypeA15 Bionic Chip 12MP 12MP ...	A15 Bionic Chip 12MP 12MP 12MP 63MP 12MP 12MP ...

In [52]: df = df.drop(columns=['storage'])
df.head()

Out[52]:

	name	ratings	price	imgURL	corpus	ram	os	processor_type	processor
0	REDMI Note 12 Pro 5G (Onyx Black, 128 GB)	4.2	23999	https://rukminim2.flixcart.com/image/312/312/x...	Storage128 GBRAM6 SystemAndroid 12Processor T...	6 GB	Android 12Processor TypeMediatek Dimensity 108...	Mediatek Dimensity 1080Processor Speed2	
1	OPPO F11 Pro (Aurora Green, 128 GB)	4.5	20999	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 GBRAM6 GBExpandable Storage256GB S...	6 GB	Android Pie 9	MediaTek Helio P70 Octa Core 2	
2	REDMI Note 11 (Starburst White, 64 GB)	4.2	13149	https://rukminim2.flixcart.com/image/312/312/x...	Storage64 GBRAM4 SystemAndroid 11Processor Sp...	4 GB	Android 11Processor Speed2	None	
3	OnePlus Nord CE 5G (Blue Void, 256 GB)	4.1	21999	https://rukminim2.flixcart.com/image/312/312/x...	Storage256 GBRAM12 SystemAndroid Q 11Processo...	12 GB	Android Q 11Processor TypeQualcomm Snapdragon ...	Qualcomm Snapdragon Octa Core 750G 5G Processo...	
4	APPLE iPhone 13 mini (Blue, 128 GB)	4.6	3537	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 SystemiOS 15Processor TypeA15 Bion...	None	iOS 15Processor TypeA15 Bionic Chip 12MP 12MP ...	A15 Bionic Chip 12MP 12MP 12MP 63MP 12MP 12MP ...	

In [59]: df.info()

```

<class 'pandas.core.frame.DataFrame'>
Index: 2534 entries, 0 to 2545
Data columns (total 18 columns):
#   Column                Non-Null Count  Dtype
---  -
0   name                   2534 non-null   object
1   ratings                2534 non-null   float64
2   price                  2534 non-null   int64
3   imgURL                 2534 non-null   object
4   corpus                 2534 non-null   object
5   ram                    2140 non-null   object
6   os                     2527 non-null   object
7   processor_type         2010 non-null   object
8   processor_speed        2087 non-null   object
9   rear_camera_primary    1540 non-null   object
10  rear_camera_secondary  1760 non-null   object
11  front_camera_primary   1540 non-null   object
12  front_camera_secondary 1540 non-null   object
13  display_size           2534 non-null   object
14  display_resolution     2242 non-null   object
15  network                2534 non-null   object
16  battery                2206 non-null   object
17  internal_storage       2118 non-null   object
dtypes: float64(1), int64(1), object(16)
memory usage: 440.7+ KB

```

```
In [53]: df.isnull().sum()
```

```
Out[53]: name          0
ratings          0
price            0
imgURL           0
corpus           0
ram              394
os               7
processor_type    524
processor_speed    447
rear_camera_primary 994
rear_camera_secondary 774
front_camera_primary 994
front_camera_secondary 994
display_size      0
display_resolution 292
network           0
battery           328
internal_storage  416
dtype: int64
```

```
In [61]: # Remove non-digit characters and convert to integer
df['internal_storage'] = df['internal_storage'].str.extract(r'(\d+)').astype(float)
```

```
In [64]: df['ram'] = df['ram'].str.replace('GB', '', regex=False).astype(float)
# Processor Speed: remove any non-numeric characters (e.g., "GHz") and convert to float
df['processor_speed'] = df['processor_speed'].str.extract(r'([\d.]+)').astype(float)

# Rear Camera Primary
df['rear_camera_primary'] = df['rear_camera_primary'].str.extract(r'(\d+)').astype(float)

# Rear Camera Secondary
df['rear_camera_secondary'] = df['rear_camera_secondary'].str.extract(r'(\d+)').astype(float)

# Front Camera Primary
df['front_camera_primary'] = df['front_camera_primary'].str.extract(r'(\d+)').astype(float)

# Front Camera Secondary
df['front_camera_secondary'] = df['front_camera_secondary'].str.extract(r'(\d+)').astype(float)
```

```
In [65]: # Display Resolution: extract width and height separately
resolution_split = df['display_resolution'].str.extract(r'(\d+)\s*[xX]\s*(\d+)')
df['display_width'] = resolution_split[0].astype(float)
df['display_height'] = resolution_split[1].astype(float)
```

```
In [68]: df = df.drop(columns=['display_resolution'])
df.head()
```

Out[68]:

	name	ratings	price	imgURL	corpus	ram	os	processor_type	processor_
0	REDMI Note 12 Pro 5G (Onyx Black, 128 GB)	4.2	23999	https://rukminim2.flixcart.com/image/312/312/x...	Storage128 GBRAM6 SystemAndroid 12Processor T...	6.0	Android 12Processor TypeMediatek Dimensity 108...	Mediatek Dimensity 1080Processor Speed2	
1	OPPO F11 Pro (Aurora Green, 128 GB)	4.5	20999	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 GBRAM6 GBExpandable Storage256GB S...	6.0	Android Pie 9	MediaTek Helio P70 Octa Core 2	
2	REDMI Note 11 (Starburst White, 64 GB)	4.2	13149	https://rukminim2.flixcart.com/image/312/312/x...	Storage64 GBRAM4 SystemAndroid 11Processor Sp...	4.0	Android 11Processor Speed2	None	
3	OnePlus Nord CE 5G (Blue Void, 256 GB)	4.1	21999	https://rukminim2.flixcart.com/image/312/312/x...	Storage256 GBRAM12 SystemAndroid Q 11Processo...	12.0	Android Q 11Processor TypeQualcomm Snapdragon ...	Qualcomm Snapdragon Octa Core 750G 5G Processo...	
4	APPLE iPhone 13 mini (Blue, 128 GB)	4.6	3537	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 SystemiOS 15Processor TypeA15 Bion...	NaN	iOS 15Processor TypeA15 Bionic Chip 12MP 12MP ...	A15 Bionic Chip 12MP 12MP 12MP 63MP 12MP 12MP ...	

In []:

```
In [70]: # Battery: remove non-numeric characters (e.g., "mAh") and convert to numeric
df['battery'] = df['battery'].str.extract(r'(\d+)').astype(float)
```

```
In [73]: # Extract numeric values from display_size and convert to float  
df['display_size'] = df['display_size'].str.extract(r'(\d+(\.\d+)?)')[0].astype(float)
```

```
In [78]: import re  
  
# Function to extract network type from the corpus text  
def extract_network(corpus):  
    match = re.search(r'\b(2G|3G|4G|5G)\b', corpus)  
    return match.group(0) if match else None  
  
# Apply the function to the corpus column to create the network column  
df['network'] = df['corpus'].apply(extract_network)
```

```
In [82]: # Remove the 'G' from the network column and convert it to a numeric value  
df['network'] = df['network'].str.replace('G', '').astype(float)
```

```
In [83]: df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Index: 2534 entries, 0 to 2545
Data columns (total 19 columns):
#   Column                Non-Null Count  Dtype
---  -
0   name                   2534 non-null   object
1   ratings                2534 non-null   float64
2   price                  2534 non-null   int64
3   imgURL                 2534 non-null   object
4   corpus                 2534 non-null   object
5   ram                    2140 non-null   float64
6   os                     2527 non-null   object
7   processor_type         2010 non-null   object
8   processor_speed        2087 non-null   float64
9   rear_camera_primary    1540 non-null   float64
10  rear_camera_secondary  1760 non-null   float64
11  front_camera_primary   1540 non-null   float64
12  front_camera_secondary 1540 non-null   float64
13  display_size           2534 non-null   float64
14  network                2534 non-null   float64
15  battery                2206 non-null   float64
16  internal_storage       2118 non-null   float64
17  display_width          2242 non-null   float64
18  display_height         2242 non-null   float64
dtypes: float64(13), int64(1), object(5)
memory usage: 460.5+ KB

```

In [84]:

Out[84]:

	name	ratings	price	imgURL	corpus	ram	os	processor_type	processor_
0	REDMI Note 12 Pro 5G (Onyx Black, 128 GB)	4.2	23999	https://rukminim2.flixcart.com/image/312/312/x...	Storage128 GBRAM6 SystemAndroid 12Processor T...	6.0	Android 12Processor TypeMediatek Dimensity 108...	Mediatek Dimensity 1080Processor Speed2	
1	OPPO F11 Pro (Aurora Green, 128 GB)	4.5	20999	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 GBRAM6 GBExpandable Storage256GB S...	6.0	Android Pie 9	MediaTek Helio P70 Octa Core 2	
2	REDMI Note 11 (Starburst White, 64 GB)	4.2	13149	https://rukminim2.flixcart.com/image/312/312/x...	Storage64 GBRAM4 SystemAndroid 11Processor Sp...	4.0	Android 11Processor Speed2	None	
3	OnePlus Nord CE 5G (Blue Void, 256 GB)	4.1	21999	https://rukminim2.flixcart.com/image/312/312/x...	Storage256 GBRAM12 SystemAndroid Q 11Processo...	12.0	Android Q 11Processor TypeQualcomm Snapdragon ...	Qualcomm Snapdragon Octa Core 750G 5G Processo...	
4	APPLE iPhone 13 mini (Blue, 128 GB)	4.6	3537	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 SystemiOS 15Processor TypeA15 Bion...	NaN	iOS 15Processor TypeA15 Bionic Chip 12MP 12MP ...	A15 Bionic Chip 12MP 12MP 12MP 63MP 12MP 12MP ...	

In []:

In []:

Missing value treatment

```
In [88]: import pandas as pd
import matplotlib.pyplot as plt

# Assuming the DataFrame is already loaded as df
# Calculate skewness for numeric columns
numeric_columns = df.select_dtypes(include=['float64', 'int64']).columns
skewness_values = df[numeric_columns].skew()

# Display skewness values
print("Skewness Values:")
print(skewness_values)

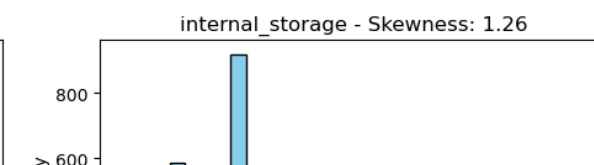
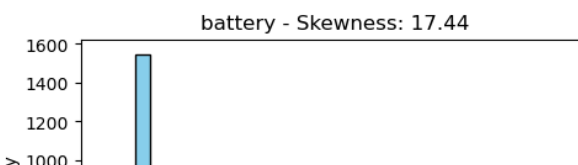
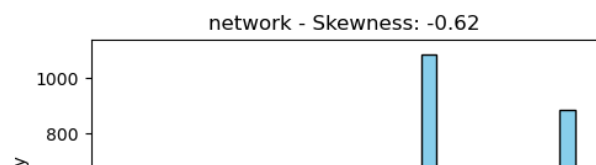
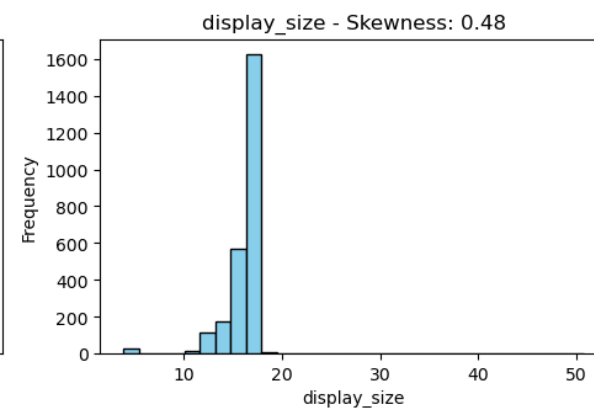
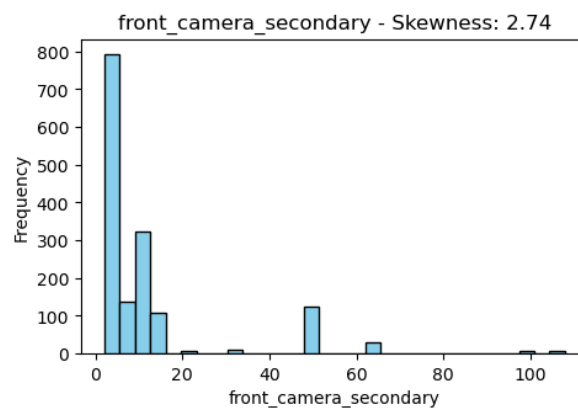
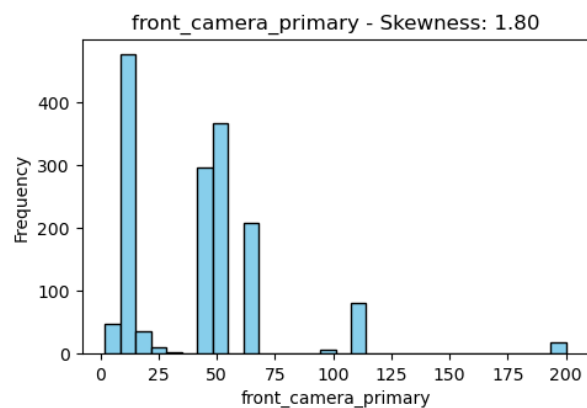
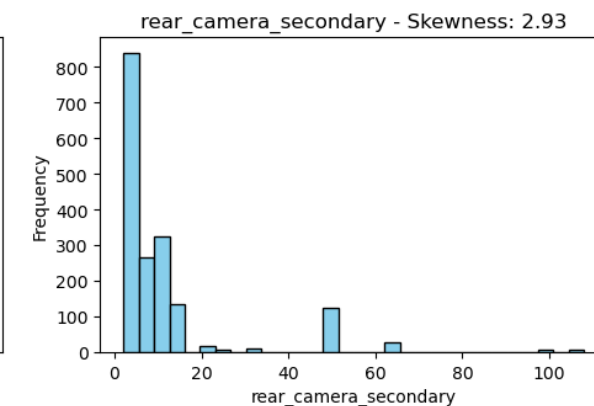
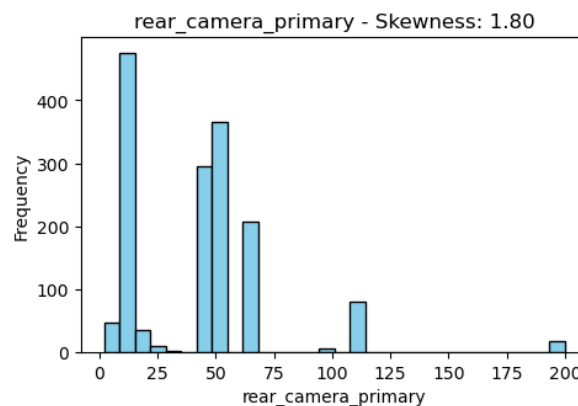
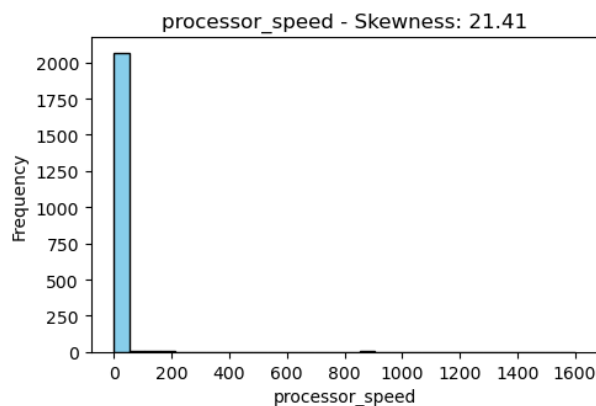
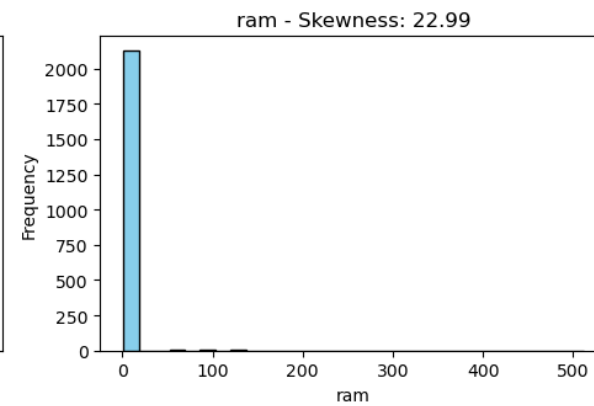
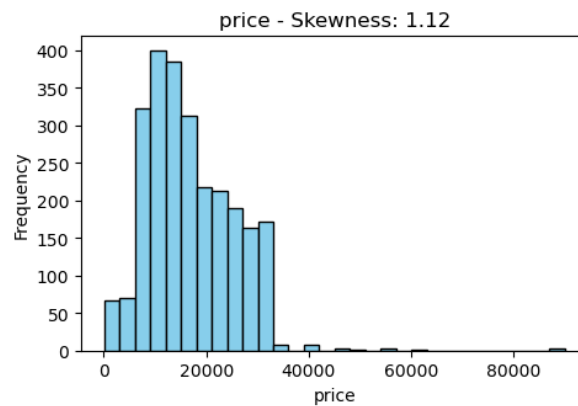
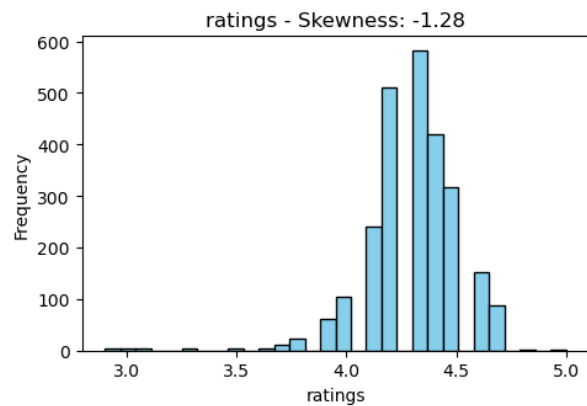
# Plot histograms for each numeric column in one row of three graphs
fig, axes = plt.subplots(6, 3, figsize=(15, 20)) # Adjust the number of rows and columns based on your data
axes = axes.flatten() # Flatten axes to easily iterate over them

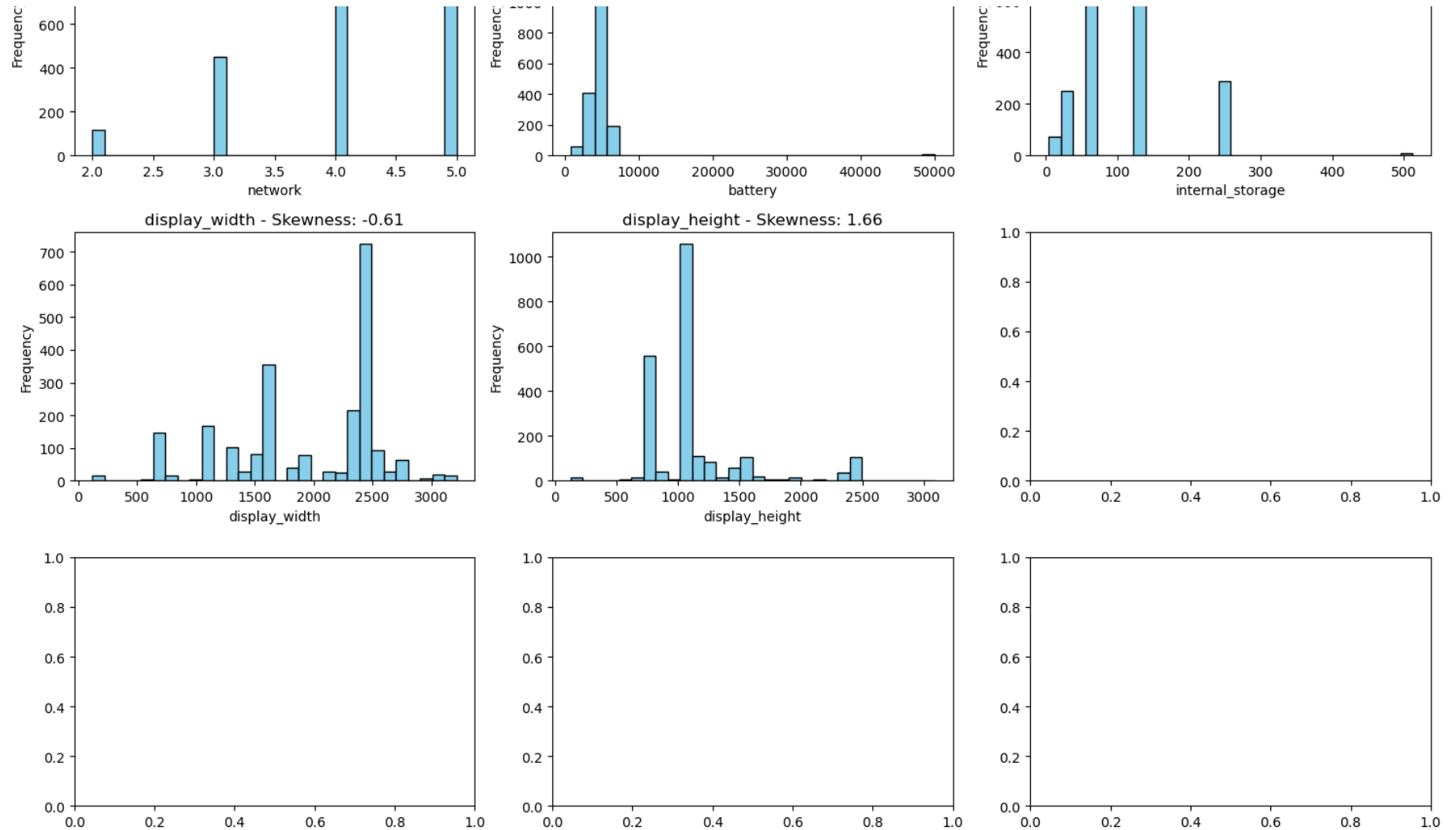
# Plot histogram for each numeric column
for i, col in enumerate(numeric_columns):
    axes[i].hist(df[col].dropna(), bins=30, color='skyblue', edgecolor='black') # Drop NA values before plotting
    axes[i].set_title(f'{col} - Skewness: {skewness_values[col]:.2f}')
    axes[i].set_xlabel(col)
    axes[i].set_ylabel('Frequency')

# Adjust layout for better spacing
plt.tight_layout()
plt.show()
```

Skewness Values:

ratings	-1.283340
price	1.123912
ram	22.993727
processor_speed	21.409509
rear_camera_primary	1.803058
rear_camera_secondary	2.927778
front_camera_primary	1.803058
front_camera_secondary	2.744070
display_size	0.483405
network	-0.615387
battery	17.442931
internal_storage	1.262516
display_width	-0.613568
display_height	1.663414
dtype:	float64





Handling missing values based on skewness: Skewness > 1 or < -1 : This indicates a highly skewed distribution, so it's better to use the median as it's more robust for skewed distributions.

Skewness between -1 and 1 : This indicates a more normal distribution, so using the mean is appropriate.

For categorical data (like network, os, processor_type): Use the mode to handle missing values.

Replacing missing values: We will:

Replace missing values with median or mean for numeric columns based on their skewness.

For categorical columns (like network, os, processor_type), replace missing values with the mode.

Handling value based on skewness

```
In [97]: # Handle missing values based on skewness
for column in df.columns:
    if df[column].dtype in ['float64', 'int64']: # Numeric columns
        skewness_value = skewness_values[column]

        if abs(skewness_value) > 1: # Highly skewed
            # Replace NaN with median
            df[column] = df[column].fillna(df[column].median())
        else: # Less skewed (approx normal distribution)
            # Replace NaN with mean
            df[column] = df[column].fillna(df[column].mean())

        # Round the column to 2 decimal places
        df[column] = df[column].round(2)

    elif df[column].dtype == 'object': # Categorical columns
        # Replace NaN with mode
        df[column] = df[column].fillna(df[column].mode()[0])
```

```
In [98]: df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Index: 2534 entries, 0 to 2545
Data columns (total 19 columns):
#   Column                Non-Null Count  Dtype
---  -
0   name                   2534 non-null   object
1   ratings                2534 non-null   float64
2   price                  2534 non-null   int64
3   imgURL                 2534 non-null   object
4   corpus                 2534 non-null   object
5   ram                    2534 non-null   float64
6   os                     2534 non-null   object
7   processor_type         2534 non-null   object
8   processor_speed        2534 non-null   float64
9   rear_camera_primary    2534 non-null   float64
10  rear_camera_secondary  2534 non-null   float64
11  front_camera_primary   2534 non-null   float64
12  front_camera_secondary 2534 non-null   float64
13  display_size           2534 non-null   float64
14  network                2534 non-null   float64
15  battery                2534 non-null   float64
16  internal_storage       2534 non-null   float64
17  display_width          2534 non-null   float64
18  display_height         2534 non-null   float64
dtypes: float64(13), int64(1), object(5)
memory usage: 460.5+ KB

```

```
In [99]: df.describe().T
```

Out[99]:

	count	mean	std	min	25%	50%	75%	max
ratings	2534.0	4.296251	0.214859	2.90	4.2	4.30	4.40	5.0
price	2534.0	16974.611681	8605.435236	73.00	10499.0	15499.00	23463.00	89999.0
ram	2534.0	6.694159	13.335359	1.00	4.0	6.00	8.00	512.0
processor_speed	2534.0	5.015852	49.157275	0.10	2.0	2.20	2.40	1600.0
rear_camera_primary	2534.0	44.713496	24.049922	2.00	48.0	48.00	50.00	200.0
rear_camera_secondary	2534.0	10.114049	13.288859	2.00	5.0	7.00	10.00	108.0
front_camera_primary	2534.0	44.713496	24.049922	2.00	48.0	48.00	50.00	200.0
front_camera_secondary	2534.0	9.150355	13.438426	2.00	5.0	5.00	8.00	108.0
display_size	2534.0	16.055047	1.931632	3.81	16.0	16.54	16.76	50.8
network	2534.0	4.078927	0.838808	2.00	4.0	4.00	5.00	5.0
battery	2534.0	4745.157064	1985.599324	800.00	4500.0	5000.00	5000.00	50000.0
internal_storage	2534.0	116.014207	65.886884	4.00	64.0	128.00	128.00	512.0
display_width	2534.0	1940.299369	585.399524	120.00	1600.0	1940.30	2400.00	3216.0
display_height	2534.0	1114.245067	397.757532	128.00	960.0	1080.00	1080.00	3088.0

In [100...

```
df.head()
```

Out[100...

	name	ratings	price	imgURL	corpus	ram	os	processor_type	processor_
0	REDMI Note 12 Pro 5G (Onyx Black, 128 GB)	4.2	23999	https://rukminim2.flixcart.com/image/312/312/x...	Storage128 GBRAM6 SystemAndroid 12Processor T...	6.0	Android 12Processor TypeMediatek Dimensity 108...	Mediatek Dimensity 1080Processor Speed2	
1	OPPO F11 Pro (Aurora Green, 128 GB)	4.5	20999	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 GBRAM6 GBExpandable Storage256GB S...	6.0	Android Pie 9	MediaTek Helio P70 Octa Core 2	
2	REDMI Note 11 (Starburst White, 64 GB)	4.2	13149	https://rukminim2.flixcart.com/image/312/312/x...	Storage64 GBRAM4 SystemAndroid 11Processor Sp...	4.0	Android 11Processor Speed2	Qualcomm Snapdragon 680Processor Speed2	
3	OnePlus Nord CE 5G (Blue Void, 256 GB)	4.1	21999	https://rukminim2.flixcart.com/image/312/312/x...	Storage256 GBRAM12 SystemAndroid Q 11Processo...	12.0	Android Q 11Processor TypeQualcomm Snapdragon ...	Qualcomm Snapdragon Octa Core 750G 5G Processo...	
4	APPLE iPhone 13 mini (Blue, 128 GB)	4.6	3537	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 SystemiOS 15Processor TypeA15 Bion...	6.0	iOS 15Processor TypeA15 Bionic Chip 12MP 12MP ...	A15 Bionic Chip 12MP 12MP 12MP 63MP 12MP 12MP ...	

In [101...

df.isnull().sum()

```
Out[101...  name                0
 ratings             0
 price               0
 imgURL              0
 corpus              0
 ram                 0
 os                  0
 processor_type      0
 processor_speed     0
 rear_camera_primary 0
 rear_camera_secondary 0
 front_camera_primary 0
 front_camera_secondary 0
 display_size        0
 network             0
 battery             0
 internal_storage    0
 display_width       0
 display_height      0
 dtype: int64
```

renaming dataset

```
In [103... # Rename columns
df.rename(columns={
    'ram': 'ram_in_GB',
    'price': 'price_in_rupees',
    'processor_speed': 'processor_speed_in_GHz',
    'rear_camera_primary': 'rear_camera_primary_megapixel',
    'rear_camera_secondary': 'rear_camera_secondary_megapixel',
    'front_camera_primary': 'front_camera_primary_megapixel',
    'front_camera_secondary': 'front_camera_secondary_megapixel',
    'display_size': 'display_size_in_cm',
    'network': 'network_G',
    'battery': 'battery_in_mAH',
    'internal_storage': 'internal_storage_in_GB'
}, inplace=True)
```

In [104...

df.head()

Out[104...

	name	ratings	price_in_rupees	imgURL	corpus	ram_in_GB	os	processor_t
0	REDMI Note 12 Pro 5G (Onyx Black, 128 GB)	4.2	23999	https://rukminim2.flixcart.com/image/312/312/x...	Storage128 GBRAM6 SystemAndroid 12Processor T...	6.0	Android 12Processor TypeMediatek Dimensity 108...	Media Dimen 1080Proce: Spe
1	OPPO F11 Pro (Aurora Green, 128 GB)	4.5	20999	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 GBRAM6 GBExpandable Storage256GB S...	6.0	Android Pie 9	MediaTek H P70 Octa C
2	REDMI Note 11 (Starburst White, 64 GB)	4.2	13149	https://rukminim2.flixcart.com/image/312/312/x...	Storage64 GBRAM4 SystemAndroid 11Processor Sp...	4.0	Android 11Processor Speed2	Qualco Snapdra 680Proce: Spe
3	OnePlus Nord CE 5G (Blue Void, 256 GB)	4.1	21999	https://rukminim2.flixcart.com/image/312/312/x...	Storage256 GBRAM12 SystemAndroid Q 11Processo...	12.0	Android Q 11Processor TypeQualcomm Snapdragon ...	Qualco Snapdra Octa Core 7: 5G Proces
4	APPLE iPhone 13 mini (Blue, 128 GB)	4.6	3537	https://rukminim2.flixcart.com/image/312/312/k...	Storage128 SystemiOS 15Processor TypeA15 Bion...	6.0	iOS 15Processor TypeA15 Bionic Chip 12MP 12MP ...	A15 Bionic C 12MP 12 12MP 63 12MP 12M

In [105...

df.info()

```

<class 'pandas.core.frame.DataFrame'>
Index: 2534 entries, 0 to 2545
Data columns (total 19 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   name                                  2534 non-null   object
1   ratings                              2534 non-null   float64
2   price_in_rupees                      2534 non-null   int64
3   imgURL                               2534 non-null   object
4   corpus                              2534 non-null   object
5   ram_in_GB                           2534 non-null   float64
6   os                                   2534 non-null   object
7   processor_type                      2534 non-null   object
8   processor_speed_in_GHz              2534 non-null   float64
9   rear_camera_primary_megapixel       2534 non-null   float64
10  rear_camera_secondary_megapixel     2534 non-null   float64
11  front_camera_primary_megapixel      2534 non-null   float64
12  front_camera_secondary_megapixel    2534 non-null   float64
13  display_size_in_cm                  2534 non-null   float64
14  network_G                           2534 non-null   float64
15  battery_in_mAH                      2534 non-null   float64
16  internal_storage_in_GB               2534 non-null   float64
17  display_width                       2534 non-null   float64
18  display_height                      2534 non-null   float64
dtypes: float64(13), int64(1), object(5)
memory usage: 460.5+ KB

```

```
In [110... df.to_csv(r"E:\2. MDS DEAKIN UNIVERSITY\TASK\Task8\notebook\final_data_for recommendation system\final_data_for recommendation
```

===== End of value
=====

```
In [ ]:
```