```python
#IMPORTING THE PANDAS

import pandas as pd

#IMPORTING THE DATASET CSV FILE.

df = pd.read_csv(r"smartphones.csv")

#GETTING A CONCISE SUMMARY OF THE DATAFRAME.
#USING .info()

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1020 entries, 0 to 1019
Data columns (total 11 columns):
 #    Column     Non-Null Count   Dtype
---   ------     --------------   -----
 0    model      1020 non-null    object
 1    price      1020 non-null    object
 2    rating     879 non-null     float64
 3    sim        1020 non-null    object
 4    processor  1020 non-null    object
 5    ram        1020 non-null    object
 6    battery    1020 non-null    object
 7    display    1020 non-null    object
 8    camera     1019 non-null    object
 9    card       1013 non-null    object
 10   os         1003 non-null    object
dtypes: float64(1), object(10)
memory usage: 87.8+ KB
```

```python
#GETTING THE STATISTICAL SUMMARY OF THE DATAFRAME.
#USING .describe()

df.describe()
```

```
           rating
count   879.000000
mean     78.258248
std       7.402854
min      60.000000
25%      74.000000
50%      80.000000
75%      84.000000
max      89.000000
```

```python
#TO RETURN THE FIRST 'n' ROWS OF A DATAFRAME (DEFAULT 5 ROWS)


df.head()
```

```
                        model     price   rating   \
0              OnePlus 11 5G   ₹54,999    89.0
1   OnePlus Nord CE 2 Lite 5G  ₹19,989    81.0
2         Samsung Galaxy A14 5G  ₹16,499  75.0
3         Motorola Moto G62 5G  ₹14,999   81.0
4           Realme 10 Pro Plus  ₹24,999   82.0


                                     sim   \
0  Dual Sim, 3G, 4G, 5G, VoLTE, Wi-Fi, NFC
1        Dual Sim, 3G, 4G, 5G, VoLTE, Wi-Fi
2        Dual Sim, 3G, 4G, 5G, VoLTE, Wi-Fi
3        Dual Sim, 3G, 4G, 5G, VoLTE, Wi-Fi
4        Dual Sim, 3G, 4G, 5G, VoLTE, Wi-Fi


                                     processor
ram   \
0  Snapdragon 8 Gen2, Octa Core, 3.2 GHz Processor  12 GB RAM, 256 GB
inbuilt
1      Snapdragon 695, Octa Core, 2.2 GHz Processor   6 GB RAM, 128 GB
inbuilt
2        Exynos 1330, Octa Core, 2.4 GHz Processor    4 GB RAM, 64 GB
inbuilt
3     Snapdragon  695, Octa Core, 2.2 GHz Processor   6 GB RAM, 128 GB
inbuilt
4     Dimensity 1080, Octa Core, 2.6 GHz Processor   6 GB RAM, 128 GB
inbuilt


                                     battery   \
0  5000 mAh Battery with 100W Fast Charging
1   5000 mAh Battery with 33W Fast Charging
2   5000 mAh Battery with 15W Fast Charging
3       5000 mAh Battery with Fast Charging
4   5000 mAh Battery with 67W Fast Charging


                                     display   \
0  6.7 inches, 1440 x 3216 px, 120 Hz Display wit...
1  6.59 inches, 1080 x 2412 px, 120 Hz Display wi...
2  6.6 inches, 1080 x 2408 px, 90 Hz Display with...
3  6.55 inches, 1080 x 2400 px, 120 Hz Display wi...
4  6.7 inches, 1080 x 2412 px, 120 Hz Display wit...


                                     camera   \
0  50 MP + 48 MP + 32 MP Triple Rear & 16 MP Fron...
1  64 MP + 2 MP + 2 MP Triple Rear & 16 MP Front ...
2  50 MP + 2 MP + 2 MP Triple Rear & 13 MP Front ...
3  50 MP + 8 MP + 2 MP Triple Rear & 16 MP Front ...
4  108 MP + 8 MP + 2 MP Triple Rear & 16 MP Front...


                          card            os
0        Memory Card Not Supported  Android v13
```

```
1    Memory Card (Hybrid), upto 1 TB   Android v12
2   Memory Card Supported, upto 1 TB   Android v13
3    Memory Card (Hybrid), upto 1 TB   Android v12
4         Memory Card Not Supported   Android v13
```

#TO RETURN THE LAST 'n' ROWS OF A DATAFRAME (DEFAULT 5 ROWS)

```
df.tail()
```

```
                               model      price   rating  \
1015      Motorola Moto Edge S30 Pro   ₹34,990     83.0
1016                    Honor X8 5G   ₹14,990     75.0
1017  POCO X4 GT 5G (8GB RAM + 256GB)  ₹28,990     85.0
1018           Motorola Moto G91 5G   ₹19,990     80.0
1019           Samsung Galaxy M52s 5G  ₹24,990     74.0


                                              sim  \
1015                Dual Sim, 3G, 4G, 5G, VoLTE, Wi-Fi
1016                Dual Sim, 3G, 4G, 5G, VoLTE, Wi-Fi
1017  Dual Sim, 3G, 4G, 5G, VoLTE, Wi-Fi, NFC, IR Bl...
1018          Dual Sim, 3G, 4G, 5G, VoLTE, Wi-Fi, NFC
1019                Dual Sim, 3G, 4G, 5G, VoLTE, Wi-Fi


                                        processor  \
1015   Snapdragon 8 Gen1, Octa Core, 3 GHz Processor
1016  Snapdragon  480+, Octa Core, 2.2 GHz Processor
1017   Dimensity 8100, Octa Core, 2.85 GHz Processor
1018    Snapdragon 695, Octa Core, 2.2 GHz Processor
1019                          Octa Core Processor


                              ram
battery  \
1015  8 GB RAM, 128 GB inbuilt  5000 mAh Battery with 68.2W Fast
Charging
1016  6 GB RAM, 128 GB inbuilt  5000 mAh Battery with 22.5W Fast
Charging
1017  8 GB RAM, 256 GB inbuilt   5080 mAh Battery with 67W Fast
Charging
1018  6 GB RAM, 128 GB inbuilt       5000 mAh Battery with Fast
Charging
1019  8 GB RAM, 128 GB inbuilt       5000 mAh Battery with Fast
Charging


                                         display  \
1015  6.67 inches, 1080 x 2460 px, 120 Hz Display wi...
1016  6.5 inches, 720 x 1600 px Display with Water D...
1017  6.6 inches, 1080 x 2460 px, 144 Hz Display wit...
1018  6.8 inches, 1080 x 2400 px Display with Punch ...
1019  6.5 inches, 1080 x 2400 px Display with Water ...
```

```
                                                          camera  \
1015   64 MP + 8 MP + 2 MP Triple Rear & 16 MP Front ...
1016   48 MP + 2 MP + Depth Sensor Triple Rear & 8 MP...
1017   64 MP + 8 MP + 2 MP Triple Rear & 16 MP Front ...
1018   108 MP + 8 MP + 2 MP Triple Rear & 32 MP Front...
1019   64 MP + 8 MP + 5 MP Triple Rear & 32 MP Front ...

                                        card           os
1015                             Android v12  No FM Radio
1016  Memory Card Supported, upto 1 TB  Android v11
1017          Memory Card Not Supported  Android v12
1018  Memory Card Supported, upto 1 TB  Android v12
1019  Memory Card Supported, upto 1 TB  Android v12
```

#CHECKING THE NUMBER OF ROWS AND COLOUMNS

```
df.shape
```

```
(1020, 11)
```

#CHECKING WHERE VALUES ARE MISSING (TRUE/FALSE)

```
df.isnull()
```

```
      model  price  rating    sim  processor    ram  battery  display
camera  \
0     False  False   False  False      False  False    False    False
False
1     False  False   False  False      False  False    False    False
False
2     False  False   False  False      False  False    False    False
False
3     False  False   False  False      False  False    False    False
False
4     False  False   False  False      False  False    False    False
False
...     ...    ...     ...    ...        ...    ...      ...      ...
...
1015  False  False   False  False      False  False    False    False
False
1016  False  False   False  False      False  False    False    False
False
1017  False  False   False  False      False  False    False    False
False
1018  False  False   False  False      False  False    False    False
False
1019  False  False   False  False      False  False    False    False
False
```

```
       card      os
0      False  False
1      False  False
2      False  False
3      False  False
4      False  False
...      ...    ...
1015   False  False
1016   False  False
1017   False  False
1018   False  False
1019   False  False

[1020 rows x 11 columns]
```

#COUNTING MISSING VALUES PER COLOUMN

```
df.isnull().sum()
```

```
model          0
price          0
rating       141
sim            0
processor      0
ram            0
battery        0
display        0
camera         1
card           7
os            17
dtype: int64
```

#TOTAL MISSING VALUES IN THE DATASET

```
df.isnull().sum().sum()
```

```
np.int64(166)
```

#CHECK DATA TYPES

```
df.dtypes
```

```
model        object
price        object
rating      float64
sim          object
processor    object
```

```
ram              object
battery          object
display          object
camera           object
card             object
os               object
dtype: object
```

#FILLING MISSING RATINGS WITH MEAN

```python
df['rating'] = df['rating'].fillna(df['rating'].mean())
```

#FILLING MISSING CATEGORICAL COLOUMNS WITH MODE

```python
for col in ['camera', 'card', 'os']:
    df[col] = df[col].fillna(df[col].mode()[0])
```

#VERIFYING

```python
print("\nMissing values after handling:\n", df.isnull().sum())
```

```
Missing values after handling:
 model          0
price          0
rating         0
sim            0
processor      0
ram            0
battery        0
display        0
camera         0
card           0
os             0
dtype: int64
```

#CHECKING NUMBER OF DUPLICATE ROWS

```python
print("Duplicate rows before:", df.duplicated().sum())
```

```
Duplicate rows before: 0
```

#REMOVING DUPLICATES

```python
df.drop_duplicates(inplace=True)
```

```python
#CHECKING AGAIN

print("Duplicate rows after:", df.duplicated().sum())
```
```
Duplicate rows after: 0
```
```python
#CLEAN AND CONVERT NUMERIC COLOUMNS STORED AS TEXT

#REMOVING ₹ AND COMMAS , CONVERTING PRICE TO FLOAT

df['price'] = df['price'].replace('[₹,]', '',
regex=True).astype(float)

#EXTRACTING NUMERIC RAM IN GB

df['ram'] = df['ram'].astype(str).str.extract(r'(\d+)').astype(float)

#EXTRACTING BATTERY CAPACITY IN MAH

df['battery'] = df['battery'].astype(str).str.extract(r'(\
d+)').astype(float)

#EXTRACTING DISPLAY SIZE IN INCHES

df['display'] = df['display'].astype(str).str.extract(r'(\d+\.\
d+)').astype(float)

#EXTRACTING MAIN CAMERA MEGAPIXELS

df['camera'] = df['camera'].astype(str).str.extract(r'(\
d+)').astype(float)

#VERIFYING DATATYPES

df.dtypes
```
```
model         object
price        float64
rating       float64
sim           object
processor     object
ram          float64
battery      float64
display      float64
camera       float64
card          object
```

```
os                object
dtype: object
```

#ENSURING THE DATASET IS CLEAN AND READY FOR ANALYSIS

#STANDARDISE COLOUMN NAMES

```
df.columns = df.columns.str.strip().str.lower()
```

#FINAL VALIDATION

```
print("Shape:", df.shape)
print("\nMissing values:\n", df.isnull().sum())
print("\nData types:\n", df.dtypes)
print("\nPreview cleaned data:\n", df.head())
```

```
Shape: (1020, 11)

Missing values:
 model         0
price         0
rating        0
sim           0
processor     0
ram           1
battery       1
display      49
camera       48
card          0
os            0
dtype: int64

Data types:
 model          object
price         float64
rating        float64
sim            object
processor      object
ram           float64
battery       float64
display       float64
camera        float64
card           object
os             object
dtype: object

Preview cleaned data:
                   model      price   rating  \
0            OnePlus 11 5G   54999.0     89.0
```

```
1   OnePlus Nord CE 2 Lite 5G   19989.0     81.0
2        Samsung Galaxy A14 5G   16499.0     75.0
3         Motorola Moto G62 5G   14999.0     81.0
4            Realme 10 Pro Plus   24999.0     82.0

                                                sim  \
0  Dual Sim, 3G, 4G, 5G, VoLTE, Wi-Fi, NFC
1        Dual Sim, 3G, 4G, 5G, VoLTE, Wi-Fi
2        Dual Sim, 3G, 4G, 5G, VoLTE, Wi-Fi
3        Dual Sim, 3G, 4G, 5G, VoLTE, Wi-Fi
4        Dual Sim, 3G, 4G, 5G, VoLTE, Wi-Fi

                                         processor   ram   battery
display  \
0  Snapdragon 8 Gen2, Octa Core, 3.2 GHz Processor   12.0    5000.0
6.70
1      Snapdragon 695, Octa Core, 2.2 GHz Processor    6.0    5000.0
6.59
2         Exynos 1330, Octa Core, 2.4 GHz Processor    4.0    5000.0
6.60
3     Snapdragon  695, Octa Core, 2.2 GHz Processor    6.0    5000.0
6.55
4     Dimensity 1080, Octa Core, 2.6 GHz Processor    6.0    5000.0
6.70

    camera                               card           os
0     50.0          Memory Card Not Supported   Android v13
1     64.0   Memory Card (Hybrid), upto 1 TB   Android v12
2     50.0   Memory Card Supported, upto 1 TB  Android v13
3     50.0    Memory Card (Hybrid), upto 1 TB  Android v12
4    108.0          Memory Card Not Supported   Android v13

#DATA SET IS READY FOR ANALYSIS

#FOR SAVING THE CLEANED DATA SET   ===>
                        #OPTIONAL


df.to_csv("smartphones_cleaned.csv", index=False)
```