Importing Libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Reading Datasets

```
df=pd.read_csv('/content/Financial Analytics data.csv')
```

```
df
```

	S.No.	Name	Mar Cap - Crore	Sales Qtr - Crore	Unnamed: 4
0	1	Reliance Inds.	583436.72	99810.00	NaN
1	2	TCS	563709.84	30904.00	NaN
2	3	HDFC Bank	482953.59	20581.27	NaN
3	4	ITC	320985.27	9772.02	NaN
4	5	H D F C	289497.37	16840.51	NaN
...	...	...	...	...	...
483	496	Lak. Vilas Bank	3029.57	790.17	NaN
484	497	NOCIL	3026.26	249.27	NaN
485	498	Orient Cement	3024.32	511.53	NaN
486	499	Natl.Fertilizer	3017.07	2840.75	NaN
487	500	L T Foods	NaN	NaN	NaN

488 rows × 5 columns

Getting information about datasets

```
#shape of dataset
df.shape
```

```
(488, 5)
```

```
#Finding unique columns
df.columns
```

```
Index(['S.No.', 'Name', 'Mar Cap - Crore', 'Sales Qtr - Crore', 'Unnamed: 4'], dtype='object')
```

```
#Size of dataset
df.size
```

```
2440
```

```
#information about dataset
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 488 entries, 0 to 487
Data columns (total 5 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   S.No.              488 non-null    int64  
 1   Name               488 non-null    object 
 2   Mar Cap - Crore    479 non-null    float64
 3   Sales Qtr - Crore  365 non-null    float64
 4   Unnamed: 4         94 non-null     float64
dtypes: float64(3), int64(1), object(1)
memory usage: 19.2+ KB
```

```
#Descriptive statistics
df.describe()
```

	S.No.	Mar Cap - Crore	Sales Qtr - Crore	Unnamed: 4
count	488.000000	479.000000	365.000000	94.000000
mean	251.508197	28043.857119	4395.976849	1523.870106
std	145.884078	59464.615831	11092.206185	1800.008836
min	1.000000	3017.070000	47.240000	0.000000
25%	122.750000	4843.575000	593.740000	407.167500
50%	252.500000	9885.050000	1278.300000	702.325000
75%	378.250000	23549.900000	2840.750000	2234.815000
max	500.000000	583436.720000	110666.930000	7757.060000

```
#Dimenshions
df.ndim
```

```
2
```

```
#Finding number NULL values in dataset
df.isnull().sum()
```

```
S.No.                  0
Name                   0
Mar Cap - Crore        9
Sales Qtr - Crore    123
Unnamed: 4           394
dtype: int64
```

Cleaning Data

```
# Check for missing values
print("Missing values before cleaning:")
print(df.isnull().sum())
```

```
Missing values before cleaning:
S.No.                  0
Name                   0
Mar Cap - Crore        9
Sales Qtr - Crore    123
Unnamed: 4           394
dtype: int64
```

```
# Check for duplicates
print("\nDuplicates before cleaning:", df.duplicated().sum())
```

```
Duplicates before cleaning: 0
```

```
#Fillna Method Using Mean
col=df['Mar Cap - Crore']
col=col.fillna(col.mean(), inplace=True)
```

```
#After Fillna Method
df.isnull().sum()
```

```
S.No.                  0
Name                   0
Mar Cap - Crore        0
Sales Qtr - Crore    123
```

```
Unnamed: 4           394
dtype: int64
```

```
#Fillna Method Using Mean
col=df['Sales Qtr - Crore']
col=col.fillna(col.mean(), inplace=True)
```

```
#After Fillna Method
df.isnull().sum()
```

```
S.No.                  0
Name                   0
Mar Cap - Crore        0
Sales Qtr - Crore      0
Unnamed: 4           394
dtype: int64
```

```
# Drop unnamed columns
df = df.loc[:, ~df.columns.str.contains('^Unnamed')]
```

```
df
```

	S.No.	Name	Mar Cap - Crore	Sales Qtr - Crore
0	1	Reliance Inds.	583436.720000	99810.000000
1	2	TCS	563709.840000	30904.000000
2	3	HDFC Bank	482953.590000	20581.270000
3	4	ITC	320985.270000	9772.020000
4	5	H D F C	289497.370000	16840.510000
...	...	...	...	...
483	496	Lak. Vilas Bank	3029.570000	790.170000
484	497	NOCIL	3026.260000	249.270000
485	498	Orient Cement	3024.320000	511.530000
486	499	Natl.Fertilizer	3017.070000	2840.750000
487	500	L T Foods	28043.857119	4395.976849

488 rows × 4 columns

```
# Save the cleaned dataset to a new CSV file
df.to_csv('cleaned Financial Analytics data.csv', index=False)


# Display the cleaned dataset
print(df.head())
```

```
   S.No.            Name  Mar Cap - Crore  Sales Qtr - Crore
0      1  Reliance Inds.        583436.72           99810.00
1      2             TCS        563709.84           30904.00
2      3       HDFC Bank        482953.59           20581.27
3      4             ITC        320985.27            9772.02
4      5         H D F C        289497.37           16840.51
```