Financial Data Analytics

July 3, 2024

[1]:
```
import pandas as pd
df=pd.read_csv("Financial Analytics UncleanData.csv")
#display(df)
df.info()
print(df.head(200))
```

```
<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
     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
..     …                 …              …                 …          …
195    202  Shri.City Union.         12995.31            1338.09         NaN
196    203    L&T Technology         12942.25                NaN       969.1
197    204   Rain Industries         12655.17            3050.81         NaN
198    205     Torrent Power         12599.37            2754.64         NaN
199    206    Dilip Buildcon         12526.06            1942.12         NaN

[200 rows x 5 columns]
```

[2]:
```
df['Unnamed: 4'] = df['Unnamed: 4'].fillna(0)
df['Sales Qtr - Crore'] = df['Sales Qtr - Crore'].fillna(0)
df['Sales Qtr - Crore'] = df['Sales Qtr - Crore'].astype(float) + df['Unnamed:␣
 ↪4'].astype(float)
```

[3]:
```
print(df.head(200))
```

```
     S.No.              Name  Mar Cap - Crore  Sales Qtr - Crore  Unnamed: 4
0        1    Reliance Inds.        583436.72           99810.00         0.0
1        2               TCS        563709.84           30904.00         0.0
2        3         HDFC Bank        482953.59           20581.27         0.0
3        4               ITC        320985.27            9772.02         0.0
4        5           H D F C        289497.37           16840.51         0.0
..     …                 …              …                 …          …
195    202  Shri.City Union.         12995.31            1338.09         0.0
196    203    L&T Technology         12942.25             969.10       969.1
197    204   Rain Industries         12655.17            3050.81         0.0
198    205     Torrent Power         12599.37            2754.64         0.0
199    206    Dilip Buildcon         12526.06            1942.12         0.0

[200 rows x 5 columns]
```

[4]:
```
df = df.drop(columns=['Unnamed: 4'])
```

[5]:
```
print(df.head(200))
```

```
     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
..     …                 …              …                 …
195    202  Shri.City Union.         12995.31            1338.09
196    203    L&T Technology         12942.25             969.10
197    204   Rain Industries         12655.17            3050.81
198    205     Torrent Power         12599.37            2754.64
199    206    Dilip Buildcon         12526.06            1942.12

[200 rows x 4 columns]
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 488 entries, 0 to 487
Data columns (total 4 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  488 non-null    float64
dtypes: float64(2), int64(1), object(1)
memory usage: 15.4+ KB
```

[7]:
```
df['Mar Cap - Crore'] = df['Mar Cap - Crore'].fillna('Not Available')
df['Sales Qtr - Crore'] = df['Sales Qtr - Crore'].fillna('Not Available')
```

[8]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 488 entries, 0 to 487
Data columns (total 4 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   S.No.              488 non-null    int64  
 1   Name               488 non-null    object 
 2   Mar Cap - Crore    488 non-null    object 
 3   Sales Qtr - Crore  488 non-null    float64
dtypes: float64(1), int64(1), object(2)
memory usage: 15.4+ KB
```

[9]:
```
duplicates = df[df['Name'].duplicated()]
if not duplicates.empty:
    print("Duplicate Names found:")
    print(duplicates['Name'].unique())  # Display unique duplicated names
else:
    print("No Duplicate Names found.")
```

```
No Duplicate Names found.
```

[10]:
```
not_availablemc_rows = df[df['Mar Cap - Crore'] == 'Not Available']
if not not_availablemc_rows.empty:
    print("Rows where 'Mar Cap - Crore' is 'Not Available':")
    print(not_availablemc_rows)
    not_availablemc_rows.to_excel('Mar_Cap_Unavailable.xlsx', index=False)
else:
    print("No rows found where 'Mar Cap - Crore' is 'Not Available'.")

print("======================================================================")


not_available_sales_rows = df[df['Sales Qtr - Crore'] == 0]
if not not_available_sales_rows.empty:
    print("Rows where 'Sales Qtr' is '0':")
    print(not_available_sales_rows)
    not_available_sales_rows.to_excel('Sales_Qtr_Unavailable.xlsx', index=False)
else:
    print("No rows found where 'Sales Qtr' is '0'.")
```

```
Rows where 'Mar Cap - Crore' is 'Not Available':
     S.No.             Name Mar Cap - Crore  Sales Qtr - Crore
99     100    Colgate-Palm.   Not Available                0.0
147    150  Endurance Tech.   Not Available                0.0
```

```
193    200  Natl. Aluminium   Not Available                0.0
243    250    Mahanagar Gas   Not Available                0.0
287    300       Bajaj Corp   Not Available                0.0
337    350      ISGEC Heavy   Not Available                0.0
387    400              BSE   Not Available                0.0
437    450     Force Motors   Not Available                0.0
487    500        L T Foods   Not Available                0.0
```

```
==========================================================================
Rows where 'Sales Qtr' is '0':
     S.No.              Name Mar Cap - Crore  Sales Qtr - Crore
49      50     Bharti Infra.        61776.92                0.0
99     100     Colgate-Palm.   Not Available                0.0
147    150   Endurance Tech.   Not Available                0.0
171    176  Info Edg.(India)        14845.05                0.0
185    192     Max Financial        13401.76                0.0
193    200   Natl. Aluminium   Not Available                0.0
224    231     Bombay Burmah        10864.53                0.0
241    248  Sundaram Clayton        10074.36                0.0
243    250     Mahanagar Gas   Not Available                0.0
258    271      Mahindra CIE         8587.04                0.0
287    300        Bajaj Corp   Not Available                0.0
314    327      Prism Cement         6176.23                0.0
332    345          GE Power          5497.4                0.0
337    350       ISGEC Heavy   Not Available                0.0
338    351              MMTC         5300.0                0.0
370    383       Swan Energy         4721.49                0.0
374    387      Shoppers St.         4558.06                0.0
379    392   Stand.Chart.PLC         4487.31                0.0
387    400              BSE   Not Available                0.0
393    406  Ujjivan Fin.Ser.         4293.42                0.0
396    409        Jindal Saw         4278.31                0.0
398    411       Linde India         4198.33                0.0
409    422     JP Associates         4074.37                0.0
418    431               HMT          3973.5                0.0
424    437  Gayatri Projects         3835.73                0.0
437    450     Force Motors   Not Available                0.0
446    459     JP Power Ven.          3597.6                0.0
451    464     Amber Enterp.         3529.87                0.0
459    472   Hind.Construct.         3452.57                0.0
487    500         L T Foods   Not Available                0.0
```

[11]:
```
df.to_excel('output.xlsx', index=False)
```