DATA SCIENCE

-Eeshaan Dhanuka

TASK-2

2. Data Manipulation with Pandas-

Description:

This task involves using the Pandas library to manipulate data.

Responsibility:

Load a CSV file into a Pandas DataFrame. Perform operations like filtering data based on conditions, handling missing values, and calculating summary statistics.

LOADING CSV FILE:

```
[ ] import pandas as pd
```

```
data = pd.read_csv('/content/01.Data Cleaning and Preprocessing.csv')
```

+ Code + Text

```
data.head()
```

Performing Operations:

Handling missing values or null values:

```
data2 = data.fillna(value=0)
data2
```

```
data3 = data.fillna(method = 'pad')
data3
```

Statistics Summary:

```
[ ] data2.columns
```

```
Index(['Observation', 'Y-Kappa', 'ChipRate', 'BF-CMratio', 'BlowFlow',
       'ChipLevel4 ', 'T-upperExt-2 ', 'T-lowerExt-2  ', 'UCZAA',
       'WhiteFlow-4 ', 'AAWhiteSt-4 ', 'AA-Wood-4  ', 'ChipMoisture-4 ',
       'SteamFlow-4 ', 'Lower-HeatT-3', 'Upper-HeatT-3 ', 'ChipMass-4 ',
       'WeakLiquorF ', 'BlackFlow-2 ', 'WeakWashF ', 'SteamHeatF-3 ',
       'T-Top-Chips-4 ', 'SulphidityL-4 '],
      dtype='object')
```

```
data2.drop(['Observation'], axis = 1, inplace = True)
data2.columns
```

```
Index(['Y-Kappa', 'ChipRate', 'BF-CMratio', 'BlowFlow', 'ChipLevel4 ',
       'T-upperExt-2 ', 'T-lowerExt-2  ', 'UCZAA', 'WhiteFlow-4 ',
       'AAWhiteSt-4 ', 'AA-Wood-4  ', 'ChipMoisture-4 ', 'SteamFlow-4 ',
       'Lower-HeatT-3', 'Upper-HeatT-3 ', 'ChipMass-4 ', 'WeakLiquorF ',
       'BlackFlow-2 ', 'WeakWashF ', 'SteamHeatF-3 ', 'T-Top-Chips-4 ',
       'SulphidityL-4 '],
      dtype='object')
```

```
[ ] q1 = data2.quantile(0.25)
    q2 = data2.quantile(0.75)

    iqr = q2-q1
    print(iqr)
```

```
[ ] data2 = data2[~((data2<(q1-1.5*iqr))|(data2>(q2+1.5*iqr))).any(axis=1)]
    data2.shape
```

```
(241, 22)
```