TASK-2

Load a csv file into a pandas dataframe.perform operation like filtering data based on conditions,handling missing values and calculating summery statistics

load CSV File by using pandas

```
import pandas as pd

df=pd.read_csv('HR_capstone_dataset.csv')

df
```

```
       satisfaction_level  last_evaluation  number_project  \
0                    0.38             0.53               2
1                    0.80             0.86               5
2                    0.11             0.88               7
3                    0.72             0.87               5
4                    0.37             0.52               2
...                   ...              ...             ...
14994                0.40             0.57               2
14995                0.37             0.48               2
14996                0.37             0.53               2
14997                0.11             0.96               6
14998                0.37             0.52               2

       average_montly_hours  time_spend_company  Work_accident
left  \
0                       157                   3              0     1

1                       262                   6              0     1

2                       272                   4              0     1

3                       223                   5              0     1

4                       159                   3              0     1

...                     ...                 ...            ...   ...

14994                   151                   3              0     1

14995                   160                   3              0     1

14996                   143                   3              0     1

14997                   280                   4              0     1

14998                   158                   3              0     1
```

```
       promotion_last_5years Department  salary
0                          0      sales     low
1                          0      sales  medium
2                          0      sales  medium
3                          0      sales     low
4                          0      sales     low
...                      ...        ...     ...
14994                      0    support     low
14995                      0    support     low
14996                      0    support     low
14997                      0    support     low
14998                      0    support     low

[14999 rows x 10 columns]
```

describe the dataset

```
df.describe().T
```

```
                           count        mean        std    min     25%  \
satisfaction_level     14999.0    0.612834   0.248631   0.09    0.44
0.64
last_evaluation        14999.0    0.716102   0.171169   0.36    0.56
0.72
number_project         14999.0    3.803054   1.232592   2.00    3.00
4.00
average_montly_hours   14999.0  201.050337  49.943099  96.00  156.00
200.00
time_spend_company     14999.0    3.498233   1.460136   2.00    3.00
3.00
Work_accident          14999.0    0.144610   0.351719   0.00    0.00
0.00
left                   14999.0    0.238083   0.425924   0.00    0.00
0.00
promotion_last_5years  14999.0    0.021268   0.144281   0.00    0.00
0.00

                          75%    max
satisfaction_level       0.82    1.0
last_evaluation          0.87    1.0
number_project           5.00    7.0
average_montly_hours   245.00  310.0
time_spend_company       4.00   10.0
Work_accident            0.00    1.0
left                     0.00    1.0
promotion_last_5years    0.00    1.0
```

identifying the missing values

```
missing_values=df.isnull().sum()

missing_values
```

```
satisfaction_level       0
last_evaluation          0
number_project           0
average_montly_hours     0
time_spend_company       0
Work_accident            0
left                     0
promotion_last_5years    0
Department               0
salary                   0
dtype: int64
```

```
df.dropna(inplace=True)

df.fillna(value=0, inplace=True)
```

```
       satisfaction_level  last_evaluation  number_project  \
0                    0.38             0.53               2
1                    0.80             0.86               5
2                    0.11             0.88               7
3                    0.72             0.87               5
4                    0.37             0.52               2
...                   ...              ...             ...
14994                0.40             0.57               2
14995                0.37             0.48               2
14996                0.37             0.53               2
14997                0.11             0.96               6
14998                0.37             0.52               2

       average_montly_hours  time_spend_company  Work_accident
left  \
0                       157                   3              0     1

1                       262                   6              0     1

2                       272                   4              0     1

3                       223                   5              0     1

4                       159                   3              0     1

...                     ...                 ...            ...   ...

14994                   151                   3              0     1
```

```
14995                    160                    3               0     1

14996                    143                    3               0     1

14997                    280                    4               0     1

14998                    158                    3               0     1


       promotion_last_5years Department  salary
0                          0      sales     low
1                          0      sales  medium
2                          0      sales  medium
3                          0      sales     low
4                          0      sales     low
...                      ...        ...     ...
14994                      0    support     low
14995                      0    support     low
14996                      0    support     low
14997                      0    support     low
14998                      0    support     low

[14999 rows x 10 columns]
```