jupyter ujjwal Last Checkpoint: 8 days ago

File Edit View Run Kernel Settings Help Not Trusted

Code JupyterLab Python 3 (ipykernel)

[6]:
```
import pandas as pd
df = pd.read_csv('placement-dataset.csv')

print(df)
```

```
    sr  cgpa     iq  placement
0    0   6.8  123.0        1.0
1    1   5.9  106.0        0.0
2    2   5.3  121.0        0.0
3    3   7.4  132.0        1.0
4    4   5.8  142.0        0.0
..  ..   ...    ...        ...
95  95   4.3  200.0        0.0
96  96   4.4   42.0        0.0
97  97   6.7  182.0        1.0
98  98   6.3  103.0        1.0
99  99   6.2  113.0        1.0

[100 rows x 4 columns]
```

df.isnull().sum()

[16]:
```
sr           0
cgpa         1
iq           0
placement    2
dtype: int64
```

[7]:
```
print(df.notnull().sum())
```

```
sr           100
cgpa          99
iq           100
placement     98
dtype: int64
```

[8]:
```


df_cleaned = df.dropna()
print(df_cleaned)
print(df_cleaned.notnull().sum())
```

```
    sr  cgpa     iq  placement
0    0   6.8  123.0        1.0
1    1   5.9  106.0        0.0
2    2   5.3  121.0        0.0
3    3   7.4  132.0        1.0
4    4   5.8  142.0        0.0
```

```
..  ..   ...    ...        ...
95  95   4.3  200.0        0.0
96  96   4.4   42.0        0.0
97  97   6.7  182.0        1.0
98  98   6.3  103.0        1.0
99  99   6.2  113.0        1.0

[97 rows x 4 columns]
sr           97
cgpa         97
iq           97
placement    97
dtype: int64
```

[10]:
```
df_filled = df.fillna(0)
print(df_filled)
print(df_filled.notnull().sum())
```

```
    sr  cgpa     iq  placement
0    0   6.8  123.0        1.0
1    1   5.9  106.0        0.0
2    2   5.3  121.0        0.0
3    3   7.4  132.0        1.0
4    4   5.8  142.0        0.0
..  ..   ...    ...        ...
95  95   4.3  200.0        0.0
96  96   4.4   42.0        0.0
97  97   6.7  182.0        1.0
98  98   6.3  103.0        1.0
99  99   6.2  113.0        1.0

[100 rows x 4 columns]
sr           100
cgpa         100
iq           100
placement    100
dtype: int64
```

[11]:
```
df_replaced = df.replace('0', '1')
print(df_replaced)
```

```
    sr  cgpa     iq  placement
0    0   6.8  123.0        1.0
1    1   5.9  106.0        0.0
2    2   5.3  121.0        0.0
3    3   7.4  132.0        1.0
4    4   5.8  142.0        0.0
..  ..   ...    ...        ...
95  95   4.3  200.0        0.0
96  96   4.4   42.0        0.0
97  97   6.7  182.0        1.0
98  98   6.3  103.0        1.0
99  99   6.2  113.0        1.0

[100 rows x 4 columns]
```

[12]:
```
df_interpolated = df.interpolate()
```

```
print(df_interpolated)
```

```
    sr  cgpa     iq  placement
0    0   6.8  123.0        1.0
1    1   5.9  106.0        0.0
2    2   5.3  121.0        0.0
3    3   7.4  132.0        1.0
4    4   5.8  142.0        0.0
..  ..   ...    ...        ...
95  95   4.3  200.0        0.0
96  96   4.4   42.0        0.0
97  97   6.7  182.0        1.0
98  98   6.3  103.0        1.0
99  99   6.2  113.0        1.0

[100 rows x 4 columns]
```

[14]:
```
nan_bool_series = pd.isna(df)
print(nan_bool_series)
```

```
       sr   cgpa     iq  placement
0   False  False  False      False
1   False  False  False      False
2   False  False  False      False
3   False  False  False      False
4   False  False  False      False
..    ...    ...    ...        ...
95  False  False  False      False
96  False  False  False      False
97  False  False  False      False
98  False  False  False      False
99  False  False  False      False

[100 rows x 4 columns]
```

[23]:
```
filtered_data = df[df['cgpa'] > 6]
print(filtered_data)
```

```
    sr  cgpa     iq  placement
0    0   6.8  123.0        1.0
3    3   7.4  132.0        1.0
5    5   7.1   48.0        1.0
8    8   6.1  156.0        0.0
11  11   6.9  138.0        1.0
13  13   6.4  116.0        1.0
14  14   6.1  103.0        0.0
20  20   6.6  120.0        1.0
21  21   7.1  151.0        1.0
26  26   7.0  199.0        1.0
29  29   7.0  112.0        1.0
30  30   7.6  128.0        1.0
32  32   7.0  139.0        1.0
35  35   6.8   90.0        1.0
37  37   8.1  149.0        1.0
38  38   6.5  160.0        1.0
42  42   7.6   89.0        1.0
43  43   6.8  141.0        1.0
44  44   7.5   61.0        1.0
48  48   6.6  138.0        1.0
```

```
52  52   7.0  175.0        1.0
53  53   8.3  168.0        1.0
54  54   6.4  141.0        1.0
55  55   7.8  114.0        1.0
56  56   6.1   65.0        0.0
57  57   6.5  130.0        1.0
58  58   8.0   79.0        1.0
60  60   6.9  139.0        1.0
61  61   7.3  137.0        1.0
63  63   6.3  128.0        1.0
64  64   7.0   64.0        1.0
65  65   8.1  166.0        1.0
66  66   6.9   96.0        1.0
69  69   8.5  120.0        1.0
70  70   6.3  127.0        1.0
71  71   6.1  132.0        1.0
72  72   7.3  116.0        1.0
74  74   6.7  154.0        1.0
77  77   7.3   50.0        1.0
78  78   6.1   81.0        0.0
79  79   6.5   90.0        1.0
82  82   6.5   37.0        1.0
83  83   7.5  130.0        1.0
90  90   7.3   86.0        1.0
91  91   7.5  158.0        1.0
93  93   6.8  112.0        1.0
97  97   6.7  182.0        1.0
98  98   6.3  103.0        1.0
99  99   6.2  113.0        1.0
```

[25]:
```
data_dict = {
    'sr': [1, 2, 3, 4, 5, 6],
    'cgpa': [6.8, 5.9, 5.3, 7.4, 5.8, 7.1],
    'iq': [123.0, 106.0, 121.0, 132.0, 142.0, 48.0],
    'placement': [1, 0, 0, 1, 0, 1]
}
print(data_dict)
```

```
{'sr': [1, 2, 3, 4, 5, 6], 'cgpa': [6.8, 5.9, 5.3, 7.4, 5.8, 7.1], 'iq': [123.0, 106.0, 121.0, 132.0, 142.0, 48.0], 'placement': [1, 0, 0, 1, 0, 1]}
```

[26]:
```
new_df = pd.DataFrame(data_dict)
print(new_df)
```

```
   sr  cgpa     iq  placement
0   1   6.8  123.0          1
1   2   5.9  106.0          0
2   3   5.3  121.0          0
3   4   7.4  132.0          1
4   5   5.8  142.0          0
5   6   7.1   48.0          1
```

[27]:
```
new_df_filled = new_df.fillna(0)
print(new_df_filled)
```

```
   sr  cgpa     iq  placement
0   1   6.8  123.0          1
1   2   5.9  106.0          0
2   3   5.3  121.0          0
3   4   7.4  132.0          1
```

```
4   5   5.8  142.0          0
5   6   7.1   48.0          1
```

[ ]: