Assignment No. 11

B. Load the dataset Employees.csv into a Pandas DataFrame.

i. Display the first 10 rows.
ii. The dataset contains columns Age, Salary, and Department. Some rows have missing values. Write a script to: Replace missing values in Age with the mean age. Replace missing values in Salary with the median salary. Drop rows where Department is missing.
iii. Create a new column in a DataFrame that categorizes employees based on their salary:

- Low (Salary < 50,000)
- Medium (50,000 ≤ Salary < 100,000)
- High (Salary > 100,000)

iv. Group the DataFrame by the Department column and calculate the average salary for each department.

Code :

[25]:
```
import pandas as pd
```

Load the dataset

[41]:
```
df = pd.read_csv('employees.csv')
df.head(10)
```

[41]:

	Name	Age	Salary	Department
0	Alice	25.0	48000.0	HR
1	Bob	34.0	55000.0	IT
2	Charlie	29.0	70000.0	Finance
3	David	NaN	52000.0	IT
4	Eve	28.0	NaN	NaN
5	Frank	45.0	62000.0	HR
6	Grace	38.0	88000.0	Finance
7	Hannah	27.0	47000.0	IT
8	Ian	NaN	53000.0	HR
9	Jane	32.0	60000.0	Finance

Replace missing values

[27]:
```
df['Age'] = df['Age'].fillna(df['Age'].mean()) # Age
```

[28]:
```
df['Salary'] = df['Salary'].fillna(df['Salary'].median())  # Salary
```

[29]:
```
df.sample(5)
```

[29]:

	Name	Age	Salary	Department
20	Uma	41.0	48000.0	HR
14	Oscar	32.5	120000.0	Finance
16	Quincy	31.0	55500.0	HR
15	Paul	33.0	58000.0	IT
3	David	32.5	52000.0	IT

Drop rows where 'Department' is missing

[30]:
```
df = df.dropna(subset=['Department'])
```

[31]:
```
df.sample(5)
```

[31]:

	Name	Age	Salary	Department
10	Karen	40.0	55500.0	IT
19	Tina	29.0	56000.0	HR
25	Zara	35.0	55500.0	Finance
2	Charlie	29.0	70000.0	Finance
8	Ian	32.5	53000.0	HR

Categorize employees based on salary

[39]:
```
def categorize_salary(salary):
    if salary < 50000:
        return 'Low'
    elif 50000 <= salary < 100000:
        return 'Medium'
    else:
        return 'High'

employees['Salary_Category'] = employees['Salary'].apply(categorize_salary)

print("\nDataset after handling missing values and categorization:")
print(employees)
```

```
Dataset after handling missing values and categorization:
       Name   Age    Salary Department Salary_Category
0     Alice  25.0   48000.0         HR             Low
1       Bob  34.0   55000.0         IT          Medium
2   Charlie  29.0   70000.0    Finance          Medium
3     David  32.5   52000.0         IT          Medium
5     Frank  45.0   62000.0         HR          Medium
6     Grace  38.0   88000.0    Finance          Medium
7    Hannah  27.0   47000.0         IT             Low
8       Ian  32.5   53000.0         HR          Medium
9      Jane  32.0   60000.0    Finance          Medium
10    Karen  40.0   55500.0         IT          Medium
11      Leo  32.5   78000.0    Finance          Medium
13     Nina  30.0   46000.0         HR             Low
14    Oscar  32.5  120000.0    Finance            High
15     Paul  33.0   58000.0         IT          Medium
16   Quincy  31.0   55500.0         HR          Medium
17   Rachel  32.5   72000.0    Finance          Medium
18    Steve  36.0   67000.0         IT          Medium
19     Tina  29.0   56000.0         HR          Medium
20      Uma  41.0   48000.0         HR             Low
21   Victor  39.0   54000.0    Finance          Medium
22    Wendy  24.0   51000.0         IT          Medium
24   Yvonne  32.5   49000.0         HR             Low
25     Zara  35.0   55500.0    Finance          Medium
```

Average salary by department

[40]:
```
average_salary_by_department = employees.groupby('Department')['Salary'].mean()

print("\nAverage salary by department:")
print(average_salary_by_department)
```

```
Average salary by department:
Department
Finance    74687.500000
HR         52187.500000
IT         55071.428571
Name: Salary, dtype: float64
```