

✓ Step 1 :- Import Libraries

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.preprocessing import LabelEncoder, StandardScaler
```

✓ Step 2 :- Import Dataset

```
Data = {
    'Name': ['Alice', 'Bob', 'Charlie', 'David', None],
    'Age': [25, None, 30, 22, 29],
    'Gender': ['Female', 'Male', 'Male', 'Male', 'Female'],
    'Salary': [50000, 60000, None, 52000, 58000],
    'Department': ['HR', 'IT', 'IT', 'HR', 'Finance']
}
```

✓ Step 3 :- Create a DataFrame

[+ Code](#)
[+ Text](#)

```
df = pd.DataFrame(Data)
```

```
df
```



	Name	Age	Gender	Salary	Department
0	Alice	25.0	Female	50000.0	HR
1	Bob	NaN	Male	60000.0	IT
2	Charlie	30.0	Male	NaN	IT
3	David	22.0	Male	52000.0	HR
4	None	29.0	Female	58000.0	Finance

✓ Step 4 :- Identify the Null Values in the DataFrame

```
df['Age'].fillna(df['Age'].mean()) #Filling Missing Values inside the Age Column
```



```
0    25.0
1    26.5
2    30.0
3    22.0
4    29.0
Name: Age, dtype: float64
```

```
df
```



	Name	Age	Gender	Salary	Department
0	Alice	25.0	Female	50000.0	HR
1	Bob	NaN	Male	60000.0	IT
2	Charlie	30.0	Male	NaN	IT
3	David	22.0	Male	52000.0	HR
4	None	29.0	Female	58000.0	Finance


```
df['Salary'].fillna(df['Salary'].mean())
```



```
0    50000.0
1    60000.0
2    55000.0
3    52000.0
```

```
4      58000.0
Name: Salary, dtype: float64
```

df




	Name	Age	Gender	Salary	Department
0	Alice	25.0	Female	50000.0	HR
1	Bob	NaN	Male	60000.0	IT
2	Charlie	30.0	Male	NaN	IT
3	David	22.0	Male	52000.0	HR
4	None	29.0	Female	58000.0	Finance

✓ Step 5 :- Handling Categorical Data

```
Encoder = LabelEncoder()
```

```
df['Gender'] = Encoder.fit_transform(df['Gender'])
```

df




	Name	Age	Gender	Salary	Department
0	Alice	25.0	0	50000.0	HR
1	Bob	NaN	1	60000.0	IT
2	Charlie	30.0	1	NaN	IT
3	David	22.0	1	52000.0	HR
4	None	29.0	0	58000.0	Finance

✓ Step 6 :- Remove Duplicates

```
df = df.drop_duplicates()
```

df



	Name	Age	Gender	Salary	Department
0	Alice	25.0	0	50000.0	HR
1	Bob	NaN	1	60000.0	IT
2	Charlie	30.0	1	NaN	IT
3	David	22.0	1	52000.0	HR
4	None	29.0	0	58000.0	Finance


✓ Step 7 :- Outlier Treatment

```
Q1 = df[['Age', 'Salary']].quantile(0.25)
```

```
Q3 = df[['Age', 'Salary']].quantile(0.75)
```

```
IQR = Q3 - Q1
```

IQR



```
Age      5.0
Salary   7000.0
dtype: float64
```

```
Lower_Bound = Q1 - 1.5 * IQR  
Upper_Bound = Q3 + 1.5 * IQR
```

```
for Col in ['Age' , 'Salary']:  
    df = df[(df[Col] >= Lower_Bound[Col]) & (df[Col] <= Upper_Bound[Col])]
```

df



	Name	Age	Gender	Salary	Department
0	Alice	25.0	0	50000.0	HR
3	David	22.0	1	52000.0	HR
4	None	29.0	0	58000.0	Finance

✓ Step 8:- Feature Scaling

```
Scaler = StandardScaler()
```

```
df[['Age','Salary']] = Scaler.fit_transform(df[['Age','Salary']])
```

df



	Name	Age	Gender	Salary	Department
0	Alice	-0.116248	0	-0.980581	HR
3	David	-1.162476	1	-0.392232	HR
4	None	1.278724	0	1.372813	Finance

Start coding or [generate](#) with AI.

Start coding or [generate](#) with AI.

Start coding or [generate](#) with AI.