```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns


df = pd.read_csv("/content/hr_clean_for_bi.csv")

print("Shape:", df.shape)
print(df.head(2))
```

```
Shape: (1470, 33)
   Age Attrition     BusinessTravel  DailyRate Department  DistanceFromHome  \
0   19       Yes  Travel_Frequently        602      Sales                 1
1   50       Yes  Travel_Frequently        959      Sales                 1

   Education    EducationField  EnvironmentSatisfaction  Gender  ...  \
0          1  Technical Degree                        3  Female  ...
1          4             Other                        4    Male  ...

   StockOptionLevel  TotalWorkingYears  TrainingTimesLastYear WorkLifeBalance  \
0                 0                  1                      5               4
1                 0                  5                      4               3

   YearsAtCompany YearsInCurrentRole  YearsSinceLastPromotion  \
0               0                  0                        0
1               0                  0                        0

   YearsWithCurrManager  Tenure_Bucket Attrition_Flag
0                     0          <=1y              1
1                     0          <=1y              1

[2 rows x 33 columns]
```

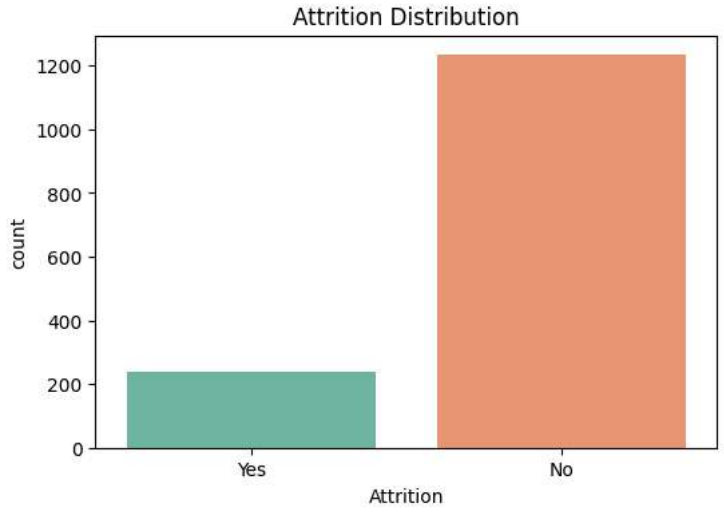
```
print("\n--- Missing values ---\n", df.isna().sum().sort_values(ascending=False))
print("\n--- Data types ---\n", df.dtypes)
print("\n--- Attrition Counts ---\n", df['Attrition'].value_counts())
```

```
--- Missing values ---
 Age                         0
Attrition                    0
BusinessTravel               0
DailyRate                    0
Department                   0
DistanceFromHome             0
Education                    0
EducationField               0
EnvironmentSatisfaction      0
Gender                       0
HourlyRate                   0
JobInvolvement               0
JobLevel                     0
JobRole                      0
JobSatisfaction              0
MaritalStatus                0
MonthlyIncome                0
MonthlyRate                  0
NumCompaniesWorked           0
OverTime                     0
PercentSalaryHike            0
PerformanceRating            0
RelationshipSatisfaction     0
StockOptionLevel             0
TotalWorkingYears            0
TrainingTimesLastYear        0
WorkLifeBalance              0
YearsAtCompany               0
YearsInCurrentRole           0
YearsSinceLastPromotion      0
YearsWithCurrManager         0
Tenure_Bucket                0
Attrition_Flag               0
dtype: int64

--- Data types ---
 Age                          int64
Attrition                    object
BusinessTravel               object
DailyRate                     int64
Department                   object
DistanceFromHome              int64
```

```
Education                 int64
EducationField           object
EnvironmentSatisfaction   int64
Gender                   object
HourlyRate                int64
JobInvolvement            int64
JobLevel                  int64
JobRole                  object
JobSatisfaction           int64
MaritalStatus            object
MonthlyIncome             int64
MonthlyRate               int64
NumCompaniesWorked        int64
OverTime                 object
```

```
plt.figure(figsize=(6,4))
sns.countplot(data=df, x="Attrition", palette="Set2")
plt.title("Attrition Distribution")
plt.show()
```

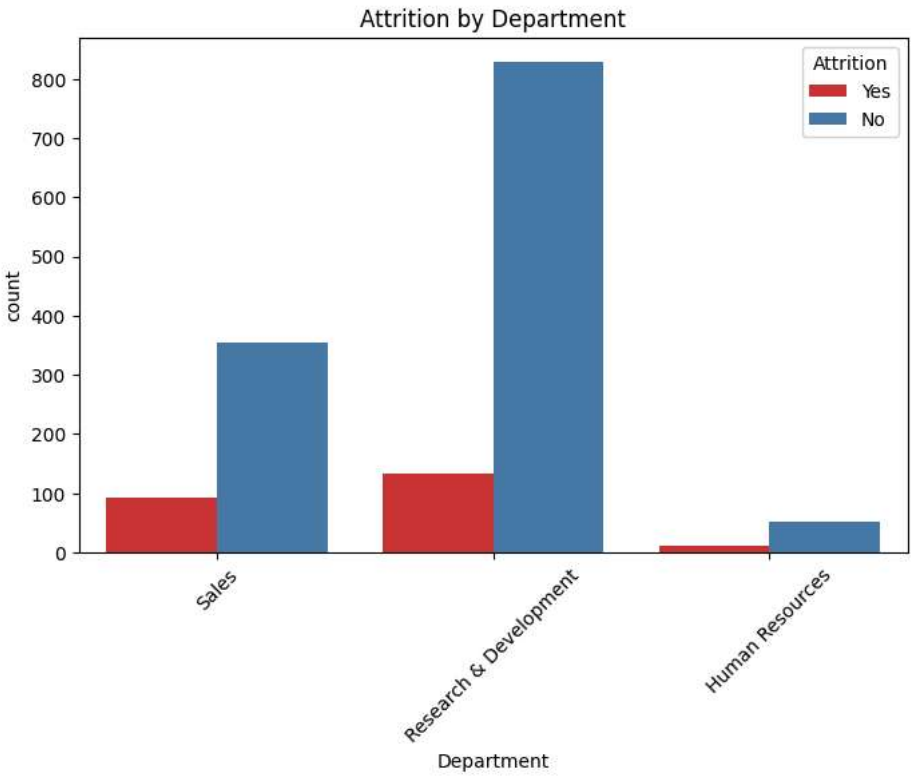
```
/tmp/ipython-input-1436007917.py:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend

  sns.countplot(data=df, x="Attrition", palette="Set2")
```



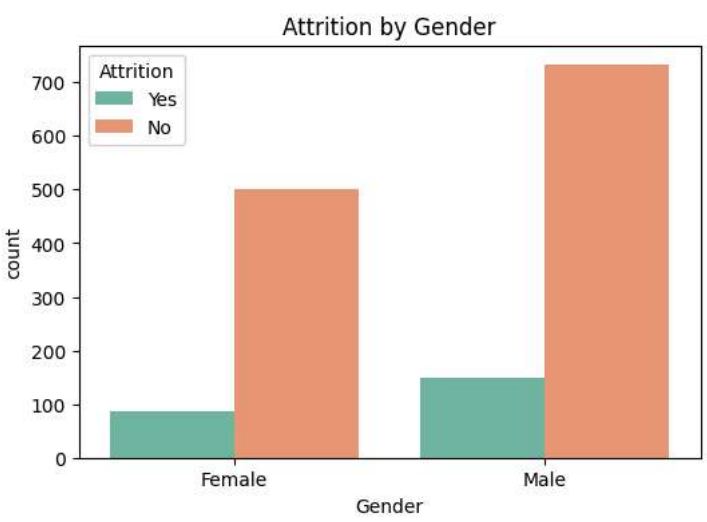

```
print("\nAttrition %:\n", df["Attrition"].value_counts(normalize=True) * 100)
```

```
Attrition %:
 Attrition
No     83.877551
Yes    16.122449
Name: proportion, dtype: float64
```

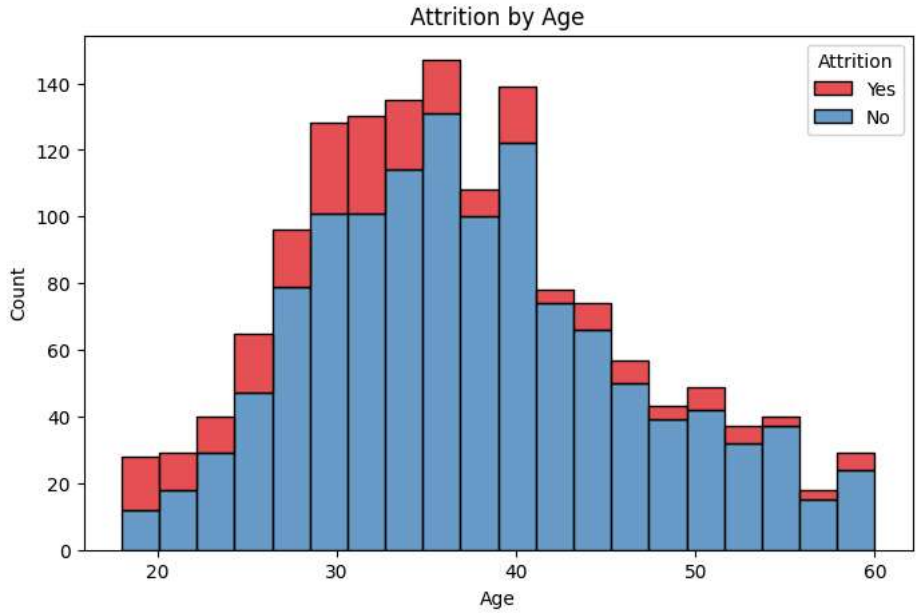
```
plt.figure(figsize=(8,5))
sns.countplot(data=df, x="Department", hue="Attrition", palette="Set1")
plt.title("Attrition by Department")
plt.xticks(rotation=45)
plt.show()
```




```
plt.figure(figsize=(6,4))
sns.countplot(data=df, x="Gender", hue="Attrition", palette="Set2")
plt.title("Attrition by Gender")
plt.show()
```

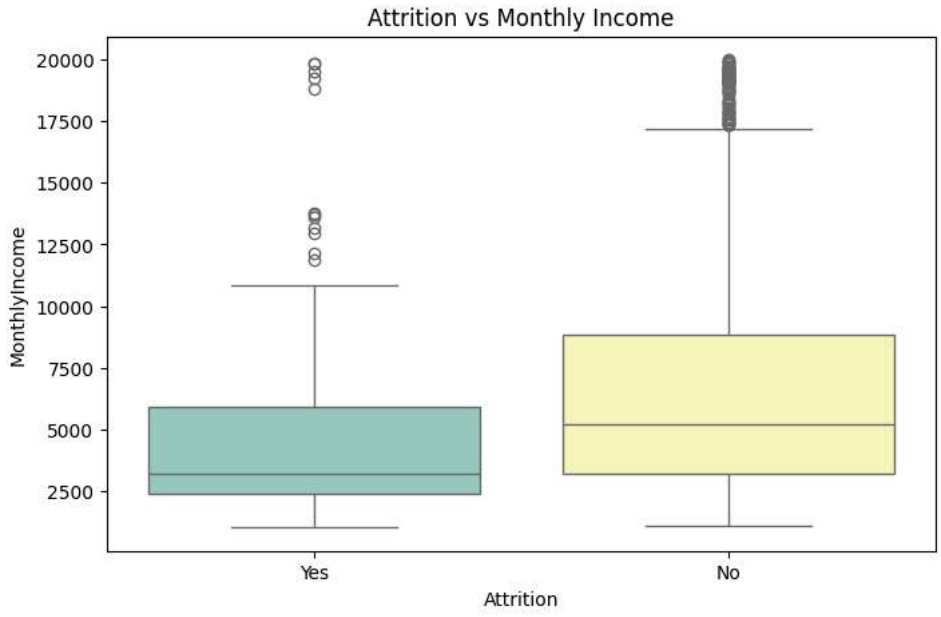



```
plt.figure(figsize=(8,5))
sns.histplot(data=df, x="Age", hue="Attrition", multiple="stack", bins=20, palette="Set1")
plt.title("Attrition by Age")
plt.show()
```

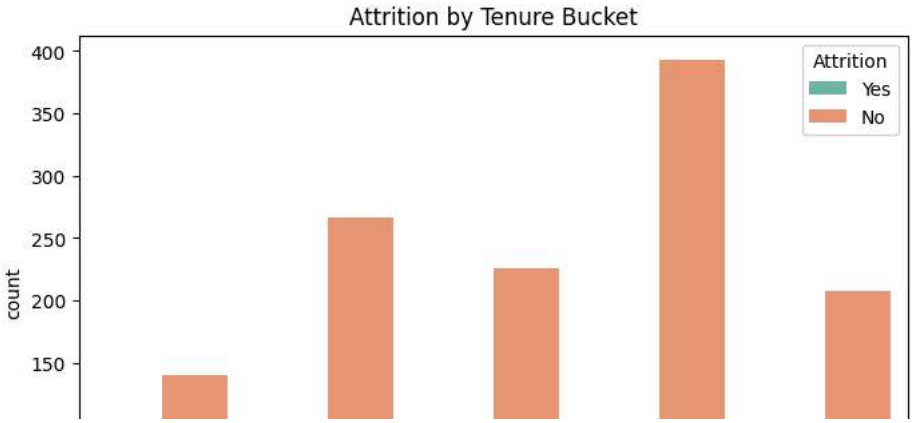



```
plt.figure(figsize=(8,5))
sns.boxplot(data=df, x="Attrition", y="MonthlyIncome", palette="Set3")
plt.title("Attrition vs Monthly Income")
plt.show()
```

```
/tmp/ipython-input-412261594.py:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legenc

  sns.boxplot(data=df, x="Attrition", y="MonthlyIncome", palette="Set3")
```




```
if "Tenure_Bucket" in df.columns:
    plt.figure(figsize=(8,5))
    sns.countplot(data=df, x="Tenure_Bucket", hue="Attrition", palette="Set2")
    plt.title("Attrition by Tenure Bucket")
    plt.show()
```




```
plt.figure(figsize=(12,8))
sns.heatmap(df.corr(numeric_only=True), cmap="coolwarm", cbar=True)
plt.title("Correlation Heatmap (Numerical Features)")
plt.show()
```