**Open Visual Studio Code**

* Create a new folder.
* Inside the folder, create a file.
* Move your Hr-Data file.xlsx into the same folde**r**

**Panda Install**

|  |
| --- |
| pip install pandas openpyxl |

**Load and Inspect Data**

**import pandas as pd**

**# Load Excel File**

**file\_path = "Hr-Data file.xlsx" # Update the correct file name if needed**

**df = pd.read\_excel(file\_path)**

**# Display basic info**

**print("Data Overview:")**

**print(df.info()) # Check data types and missing values**

**print("\nFirst 5 Rows:")**

**print(df.head()) # Display first 5 rows**

**Check Data Structure**

**df.info() # Summary of columns and data types**

**df.describe() # Summary statistics for numerical columns**

**df.columns # List of column names**

**Data Cleaning**

**🔹 Check for Missing Values**

**df.isnull().sum() # Count missing values in each column**

**Data Visualization**

**Salary Distribution**

**plt.figure(figsize=(8, 5))**

**sns.histplot(df["Salary"], bins=20, kde=True, color='blue')**

**plt.title("Salary Distribution")**

**plt.xlabel("Salary")**

**plt.ylabel("Frequency")**

**plt.show()**

**Employee Count**

**plt.figure(figsize=(10, 5))**

**sns.countplot(y=df["Department"], order=df["Department"].value\_counts().index, palette="viridis")**

**plt.title("Number of Employees per Department")**

**plt.xlabel("Count")**

**plt.ylabel("Department")**

**plt.show()**

**Gender**

**plt.figure(figsize=(6, 4))**

**sns.countplot(x=df["Gender"], palette="pastel")**

**plt.title("Gender Distribution in HR Data")**

**plt.show()**