**Title:** *Exploratory Data Analysis (EDA) on train.csv*

**1. Objective**

The goal of this work is to explore and understand the dataset train.csv by performing basic statistical analysis and visualizations. This helps in identifying patterns, relationships, and potential data quality issues.

**2. Tools & Libraries Used**

* **Python 3** – Programming language for analysis
* **Pandas** – Data manipulation and analysis
* **Seaborn** – Statistical data visualization
* **Matplotlib** – Plotting and graphing

**3. Installation Steps**

In the Jupyter Notebook, we installed the required libraries using:

python

Copy code

!pip install pandas --quiet

!pip install seaborn --quiet

!pip install matplotlib --quiet

This ensures that the libraries are available before running the code.

**4. Code Overview**

1. **Library Installation & Import**  
   Installed and imported Pandas, Seaborn, and Matplotlib.
2. **Data Loading**  
   Loaded the dataset using:

python

Copy code

df = pd.read\_csv('train.csv')

1. **Basic Overview**
   * Displayed first few rows (head())
   * Displayed column data types and non-null counts (info())
   * Generated descriptive statistics (describe(include='all'))
   * Checked for missing values (isnull().sum())
2. **Visual Exploration**
   * **Pairplot** to see pairwise relationships:

python

Copy code

sns.pairplot(df.sample(min(300, len(df)), random\_state=42))

* + **Correlation Heatmap** for numerical relationships:

python

Copy code

sns.heatmap(df.corr(), annot=True, cmap='coolwarm')

* + **Histograms** for distribution of values:

python

Copy code

df.hist(figsize=(10,6), bins=20, edgecolor='black')

**5. Key Benefits of This Approach**

* **Step-by-step installation** ensures smooth execution even on new systems.
* **Basic statistics** help quickly understand the dataset’s structure.
* **Visualizations** make trends, correlations, and outliers more visible.

**6. Next Steps**

* Perform deeper statistical analysis.
* Handle missing values and outliers.
* Create more targeted visualizations for specific columns.