

# TASK 2: EXPLORATORY DATA ANALYSIS (EDA)

## Introduction:

Exploratory Data Analysis (EDA) is the process of analyzing datasets to summarize their main characteristics using statistical methods and visualizations.

## Objectives of EDA:

- Understand data structure and variables
- Identify trends and patterns
- Detect outliers and missing values
- Validate assumptions before modeling

## Steps in EDA:

1. Load the dataset
2. Understand rows, columns, and data types
3. Perform summary statistics
4. Visualize data distributions
5. Detect anomalies and data issues

## Python Code for EDA:

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

# Load dataset
data = pd.read_csv("output.csv")

# Basic information
print(data.info())
print(data.describe())

# Check missing values
print(data.isnull().sum())

# Visualization
data['Title'].value_counts().head(10).plot(kind='bar')
plt.show()
```

## Outcome of EDA:

- Clear understanding of dataset quality
- Identification of key variables
- Dataset ready for further analysis or machine learning

## Conclusion:

Web Scraping helps in collecting data from the web, while EDA helps in understanding and preparing the data. Both are fundamental steps in the data science lifecycle.