Final Project : Submission date: Jan 2 2026

EDA Project

Task: Choose any dataset (CSV or Excel) and perform a complete **Exploratory Data Analysis (EDA) and basic modeling workflow** covering the following steps:

Steps to Follow:

1. **Import Data**

   o   Load the dataset from CSV, Excel, or a SQL database into Python.

   o   Display the first few rows and check the structure of the dataset.

2. **Export Data**

   o   Save a copy of the dataset (after any preprocessing) to CSV or Excel.

3. **Data Cleaning**

   o   Handle missing values (impute or remove).

   o   Remove duplicates if present.

   o   Identify and handle outliers using visualization or statistical methods.

4. **Data Transformation**

   o   Apply **normalization or standardization** to numeric features if required.

5. **Descriptive Statistics**

   o   Calculate **mean, median, mode, standard deviation**, and other summary statistics.

   o   Provide a brief interpretation of key statistics.

6. **Basic Visualization**

   o   Create **line plots, bar charts, and histograms**.

   o   Customize titles, labels, and legends.

7. **Advanced Visualization**

   o   Create **pair plots, heatmaps, and violin plots**.

   o   Analyze **correlations and covariance** between numeric features.

8. **Interactive Visualization**

   o   Use **Plotly or Plotly Dash** to create at least one interactive visualization or simple dashboard.

9. **Probability Analysis**

   o   Visualize **probability distributions** for numeric features.

10. **Modeling – Classification (k-NN)**

   o   Implement **k-Nearest Neighbors (k-NN)** on a numeric target variable or class (if available).

   o   Split data into training and testing sets, train the model, and evaluate accuracy.

11. **Modeling – Clustering (k-Means)**

   o   Apply **k-Means clustering** to group numeric data points.

   o   Visualize clusters using scatter plots or pair plots.

   o   Interpret the clusters.

12. **Summary & Insights**

   o   Write a **brief report** summarizing your findings, patterns, and insights from the data

```
EDA_Project/
|
├── data/                # CSV, Excel files
|
├── modules/             # Python modules
|   ├── data_import.py
|   ├── data_cleaning.py
|   ├── transformation.py
|   ├── stats_analysis.py
|   ├── visualization.py
|   └── modeling.py
|
├── main.py              # Main Python script to run the EDA project
|
├── outputs/             # Plots, reports, exported datasets
|
└── README.md
```