Exploratory Data Analysis (EDA) Checklist

Objective

To guide students through the step-by-step process of exploring a dataset using Python.

EDA Checklist

1. Import necessary libraries (Pandas, Numpy, Matplotlib, Seaborn)
2. Load the dataset using Pandas (pd.read_excel or pd.read_csv)
3. View first few rows of the dataset (df.head())
4. Understand the shape and structure of the dataset (df.shape, df.info())
5. Check for missing values (df.isnull().sum())
6. Identify and remove duplicates if any (df.duplicated().sum())
7. Summary statistics of numerical features (df.describe())
8. Analyze categorical columns (df['Column'].value_counts())
9. Visualize distributions (histograms, boxplots, countplots)
10. Correlation matrix (df.corr(), heatmap)
11. Time-based analysis (e.g., sales trends over time)
12. Grouping and aggregation (e.g., top-selling products, revenue by country)
13. Save cleaned data if necessary
14. Document insights and observations

Recommended Visualizations

- Distribution of numerical columns using histograms or KDE plots
- Count of top products using bar charts
- Sales trend over time using line plots
- Revenue by country using bar charts
- Heatmap of correlations between numeric variables

Deliverables

Exploratory Data Analysis (EDA) Checklist

1. Cleaned and analyzed Jupyter Notebook (*.ipynb)
2. Summary Report (PDF or Word) including key insights and screenshots.