EDA_Checklist

1. Import necessary libraries
→ Pandas, NumPy, Matplotlib, Seaborn

2. Load the dataset
→ Use pd.read_excel() or pd.read_csv()

3. View first few rows
→ df.head()

4. Check shape and structure
→ df.shape, df.info()

5. Check for missing values
→ df.isnull().sum()

6. Identify and remove duplicates
→ df.duplicated().sum(), df.drop_duplicates()

7. Summary statistics
→ df.describe()

8. Analyze categorical columns
→ df['Column'].value_counts()

9. Visualize distributions
→ Histograms, boxplots, countplots

10. Correlation matrix
→ df.corr() and Seaborn heatmap

11. Time-based analysis
→ E.g., Monthly or weekly sales trends

12. Grouping and aggregation
→ Top-selling products, revenue by country, etc.

13. Save cleaned dataset
→ Use df.to_csv()

14. Document insights and observations

Recommended Visualizations

Histogram or KDE plots for distributions

Bar chart of top-selling products

Line chart for sales trends

Revenue by country

Heatmap for correlation matrix