📊 Step-by-Step EDA Process for Shopify Sales Data

Step 1: Data Overview

- Shape of the dataset (rows, columns)
- Data types of each column
- Null/missing values
- Duplicate entries

Step 2: Data Cleaning

- Remove duplicates if any
- Handle missing values
- Convert data types (e.g., dates, numbers)
- Normalize text (e.g., city/country names)

Step 3: Univariate Analysis

Analyze each feature individually:

- **Categorical columns**: `Country`, `Product Type`, `Gateway`, etc.

  - Count plots, pie charts

- **Numerical columns**: `Quantity`, `Subtotal Price`, `Total Price Usd`, `Total Tax`

  - Histograms, boxplots, descriptive statistics (mean, median, std)

Step 4: Bivariate Analysis

Compare pairs of variables:

- Relationship between `Quantity` and `Total Price`
- Average `Total Tax` per `Country`
- Sales trends over time using `Invoice Date`
- Most popular products by quantity and revenue

Step 5: Time Series Analysis

- Convert `Invoice Date` to datetime
- Extract date features (day, month, year, weekday)
- Analyze sales trend over days/weeks/months

Step 6: Correlation Analysis

- Correlation matrix for numerical variables
- Heatmap to visualize relationships

Step 7: Grouped Aggregations

- Revenue per product
- Total orders per city/country
- Gateway-wise revenue
- Monthly/weekly trends

Step 8: Visualizations

Use plots to uncover insights:

- Bar plots, line plots, pie charts, boxplots, histograms
- Use `seaborn`, `matplotlib`, or `plotly`

Step 9: Insights & Summary

Summarize key findings:

- Which products perform best?
- Which locations generate most revenue?
- Are there seasonal trends?
- Which gateways are most used?