**Project Title**

**Exploratory Data Analysis (EDA) on the Global Superstore**

**Objective**

To perform an in-depth exploratory data analysis on the Global Superstore dataset to uncover insights into sales, profitability, customer behavior, product performance, regional trends, and shipping efficiency. The goal is to identify actionable patterns that can help improve decision-making in sales, marketing, and operations.

**Problem Statement**

**1. Sales Trends and Patterns**

* Identify seasonal fluctuations and yearly trends in sales across **categories, regions, and customer segments**.
* Analyze monthly and quarterly sales growth patterns.

**2. Profitability Analysis**

* Compare **profit margins** across product categories, customer segments, and regions.
* Identify high-margin and low-margin products.

**3. Customer Segmentation**

* Segment customers based on **purchase frequency, spending amount, and product preferences**.
* Identify the most valuable customer segments.

**4. Product Performance**

* Determine **top-selling** and **underperforming** products.
* Analyze which categories and sub-categories drive the most revenue and profit.

**5. Regional Analysis**

* Compare sales and profitability across **countries, regions, and states**.
* Identify growth opportunities and regional buying preferences.

**6. Shipping Dynamics**

* Analyze preferred **shipping methods** and their impact on delivery times.
* Evaluate shipping costs versus profitability.

**7. Promotional Effectiveness**

* Assess the impact of **discounts and promotions** on sales and profit.
* Determine if higher discounts always lead to increased profitability.

**8. Return Analysis**

* Investigate **return rates** by product category, region, and customer segment.
* Identify reasons for high returns and potential improvements.

**Tasks & Deliverables**

**A. Data Understanding & Cleaning**

1. Load the dataset into a Python environment (Pandas, NumPy).
2. Check for missing values, duplicates, and outliers.
3. Convert data types where necessary (e.g., dates to datetime).
4. Handle incorrect or inconsistent entries (e.g., inconsistent region names).

**B. Exploratory Data Analysis**

1. **Univariate Analysis** – Distributions of sales, profit, discount, and quantity.
2. **Bivariate Analysis** – Relationships between sales and profit, sales and discount, category vs. profit, etc.
3. **Multivariate Analysis** – Interaction of multiple factors (e.g., segment + region + category).

**C. Visualization**

* Use **Matplotlib** and **Seaborn** for:
  + Sales trends over time
  + Profit by category and region
  + Customer segmentation patterns
  + Shipping performance metrics
  + Discount vs. profit scatterplots
  + Regional sales heatmaps

**D. Insights & Recommendations**

* Summarize key patterns found in the analysis.
* Provide **data-backed recommendations** to improve sales, profitability, and operational efficiency.

**Tools & Libraries**

* **Python**: Pandas, NumPy, Matplotlib, Seaborn (optional for interactive visuals)
* **SQL** (optional): For querying specific insights
* **Jupyter Notebook**: For combining code, visuals, and explanations

**Expected Output**

* A **Jupyter Notebook** (or Python script) containing:
  + Data loading and cleaning
  + EDA with visualizations
  + Insightful commentary for each analysis
* A **summary report** (PDF/Word) highlighting:
  + Key findings
  + Actionable recommendations
  + Supporting charts/graphs