**Project: Exploratory Data Analysis on the Global Superstore Database**

**Objective**

Conduct a detailed analysis of the Global Superstore Database to uncover valuable insights across multiple business dimensions, such as sales trends, profitability, customer behavior, product performance, and logistics. This project aims to enable data-driven decisions that improve profitability, customer satisfaction, and operational efficiency.

**Scope of Analysis**

1. **Sales Trends and Patterns**
   * Objective: Identify seasonal fluctuations and observe long-term trends in sales across different product categories and regions.
   * Tasks:
     + Perform time-series analysis to detect seasonal trends and monthly or yearly sales fluctuations.
     + Identify top-performing categories and regions based on consistent growth trends.
     + Visualization: Line plots and seasonal decomposition charts to highlight trends.
2. **Profitability Analysis**
   * Objective: Determine profit margins across product categories, customer segments, and regions to pinpoint high-margin areas.
   * Tasks:
     + Calculate profit margins by category, segment, and region.
     + Identify product categories and regions with the highest and lowest profitability.
3. **Customer Segmentation**
   * Objective: Segment customers based on purchasing behavior to determine high-value customer groups and preferences.
   * Tasks:
     + Identify customer segments that generate the highest revenue.
4. **Product Performance**
   * Objective: Assess the performance of products and categories to identify top-sellers and items with underperformance.
   * Tasks:
     + Rank products by sales and profitability.
     + Compare the performance of product categories and sub-categories.
5. **Regional Analysis**
   * Objective: Evaluate sales and profitability by region to identify growth opportunities and regional customer preferences.
   * Tasks:
     + Perform a comparative analysis of sales across regions.
     + Identify regional strengths and areas for growth based on profitability metrics.
6. **Shipping Dynamics**
   * Objective: Analyze shipping preferences, delivery times, and costs to optimize logistics and enhance customer satisfaction.
   * Tasks:
     + Calculate average shipping time and cost for each region and product category.
     + Assess the impact of shipping on profitability.
     + Visualization: bar charts for shipping costs by region.
7. **Promotional Effectiveness**
   * Objective: Evaluate the impact of promotions and discounts on sales and profitability.
   * Tasks:
     + Measure sales uplift during promotional periods compared to regular sales.
     + Analyze the effect of discounts on profit margins.
     + Visualization: Sales vs. discount bar plots.
8. **Return Analysis**
   * Objective: Investigate return rates and underlying reasons to minimize returns and improve product quality.
   * Tasks:
     + Calculate return rates by product category and region.
     + Identify common reasons for returns and high-return products.
     + Visualization: Return rate charts by category and reason distribution pie charts.

**Expected Outcomes**

* Clear identification of seasonal sales trends and top-performing products.
* Insights into high-margin segments and underperforming regions or categories.
* Segmented view of valuable customer groups based on purchasing patterns.
* Actionable insights on shipping dynamics to reduce costs and improve customer satisfaction.
* Analysis of promotions and discounts to strategize future marketing initiatives.
* Data-backed strategies for minimizing return rates by improving product quality or clarifying product descriptions.

**Deliverables**

1. **Codebase**: Provide a well-documented Jupyter Notebook or Python script with code for each analysis step.
2. **Dashboard**: Optionally, present the findings in a Power BI or Tableau dashboard for interactive exploration.
3. **Report**: Summarize the key insights, charts, and recommendations in a project report for stakeholders.

Contents

[Problem Statement & Objective 5](#_Toc169250064)

[Problem Statement: 5](#_Toc169250065)

[Objectives 5](#_Toc169250066)

[Introduction 6](#_Toc169250067)

[Exploratory Data Analysis (EDA) 6](#_Toc169250068)

[EDA Pipeline 6](#_Toc169250069)

[About the company 6](#_Toc169250070)

[Tools and platforms used in project 7](#_Toc169250071)

[Why Python and MYSQL? 7](#_Toc169250072)

[Platforms used 7](#_Toc169250073)

[Versions of platform 7](#_Toc169250074)

[Chapter 1: Data Loading/reading 8](#_Toc169250075)

[Load Data in MySQL 8](#_Toc169250076)

[Import Necessary Library 8](#_Toc169250077)

[Establish a connection to the Euro Mart database using MySQL Connector 8](#_Toc169250078)

[Load Data (Jupyter Notebook or VS Code) from MySQL 10](#_Toc169250079)

[Retrieve data from a MySQL database table and load it into a pandas DataFrame for further analysis in Jupyter Notebook 10](#_Toc169250080)

[Chapter 2: Familiarize with Data & Identifying the Target Variable 11](#_Toc169250081)

[Explore the provided data (column names, data types) 11](#_Toc169250082)

[Overview of data 11](#_Toc169250083)

[Interpretation 11](#_Toc169250084)

[Chapter 3: Data Preparation & Transformation 12](#_Toc169250085)

[Data Cleaning 12](#_Toc169250086)

[Handle missing values (imputation or deletion) 12](#_Toc169250087)

[Data Reduction: Remove unwanted columns or rows 12](#_Toc169250088)

[Format data types (numerical & categorical) 12](#_Toc169250089)

[Rename of columns 13](#_Toc169250090)

[Change data types of columns and Memory optimization 13](#_Toc169250091)

[Feature Engineering (Create new features/variables) 13](#_Toc169250092)

[Create new features 13](#_Toc169250093)

[Chapter 4: Data Analysis & Visualization 15](#_Toc169250094)

[Overview of data before analysis 15](#_Toc169250095)

[Univariate analysis 15](#_Toc169250096)

[Summary statistics 15](#_Toc169250097)

[Unique elements in column 16](#_Toc169250098)

[Univariate Analysis of Categorical Columns 16](#_Toc169250099)

[Multivariate analysis 23](#_Toc169250100)

[Identifying Top Performers 23](#_Toc169250101)

[Address operational inefficiencies and customer engagement issues in underperforming areas 32](#_Toc169250102)

[Optimizing Discounts & Shipping Modes 33](#_Toc169250103)

[Feedback Analysis 35](#_Toc169250104)

[Improving Product Mix 37](#_Toc169250105)

[Chapter 5: Summary and Suggestions 42](#_Toc169250106)

[Identifying top performers 42](#_Toc169250107)

[Understanding challenges 42](#_Toc169250108)

[Optimizing discounts & shipping modes 43](#_Toc169250109)

[Leveraging customer feedback 43](#_Toc169250110)

[Improving product mix 43](#_Toc169250111)

[Additional recommendations 44](#_Toc169250112)

[Annexure 45](#_Toc169250113)

[Git hub link 45](#_Toc169250114)

[Code snippets 45](#_Toc169250115)

[Plotting Bar Graphs 45](#_Toc169250116)

[Plotting Pie chart 45](#_Toc169250117)

[Plotting box and whisker chart 46](#_Toc169250118)

[Plotting heat maps 46](#_Toc169250119)

[Plotting Line chat 46](#_Toc169250120)

[Adding gradient to table 47](#_Toc169250121)