**Project**

**Exploratory Data Analysis Project Report: Linear Regression on Order Data**

**Project Overview:**

The goal of this exploratory data analysis (EDA) project is to analyze and understand the relationships between order totals and various factors such as city, category, and product cost. We will leverage linear regression modeling to gain insights and make data-driven decisions.

**Data Description**

• **Dataset:**

Order data containing information about orders, customers, and products.

• **Columns of Interest:**

* Order ID: Unique identifier for each order.
* Order Date: Date when the order was placed.
* CustomerName: Name of the customer.
* State: State where the order was placed.
* City: City where the order was placed.
* order\_total: Total order amount.
* Product cost: Cost of the product.
* Category: Product category (e.g., Phones, Tables, Clothes).

**Key Insights and Visualizations:**

1. Order Total by City:

1. Create a bar chart or heatmap showing total order amounts for different cities.
2. Identify cities with the highest and lowest order totals.

* Order Total by Category:
* Generate a pie chart illustrating the distribution of order totals across product categories.

1. Highlight the most popular categories based on order amounts.

3. **Correlation Analysis:**

o Use scatter plots or heatmaps to explore correlations between order totals and other variables (e.g., product cost).

o Identify any significant relationships.

4. **Linear Regression Model:**

o Build a linear regression model with order\_total as the dependent variable.

o Select relevant features (predictors) such as product cost, city, and category.

o Evaluate model performance using metrics like Mean Absolute Error (MAE) or R-squared.

5. **Recommendations:**

o Based on the analysis, provide actionable recommendations to optimize order totals:

 Target specific cities or states with growth potential.

 Focus marketing efforts on popular product categories.

 Optimize product pricing based on category performance.

**Conclusion:**

This EDA project aims to uncover valuable insights from order data, enabling better decision-making and strategic planning. By combining visualizations and linear regression modeling, we can enhance our understanding of the factors influencing order totals.

Remember to customize the visual graphs and model parameters according to your specific dataset and business context. 📊🔍🚀