**Exploratory Data Analysis (EDA) with Pandas on**

**Global Retail Sales Data**

The purpose of this project is to explore and analyze an reatail sale dataset using the Pandas framework to derive insights into customer behavior, product trends, and sales performance, sales trends.

**Goals of the Project:**

● Explore the global retail dataset using Pandas.

● Perform feature engineering to derive useful insights.

● Visualize data distributions and trends with various plot types.

● Summarize key findings that can aid in business decision-making.

**Materials and Methods :**

The data for this project is from a simulated global platform, containing information about Product Category, order location, international shipping, sales price, shipping information.

This dataset includes sales data, product categories, order dates, profit margins, and more. The analysis aims to understand sales performance, product trends, and sales trends.

**General Part :**

● Libraries Import: Numpy , Pandas , matplot , seaborn.

● Dataset Exploration: Initial exploration of the dataset, checking for missing values, duplicates, and generating summary statistics.

● Feature Engineering: Transformation of date columns and creation of new features like profit margin.

● Visualization in Pandas: Distribution analysis, relationships between variables, and time-based trends.

**Project Outcome & Insights :**

The project performs Exploratory Data Analysis (EDA) on an Global retail sale dataset to gain meaningful insights into sales performance, customer behavior, and sales trend.

**1. Sales Performance :**

● Customer Segment Wise Top Sales: The project groups sales based on different customer segments to identify the most profitable segments.

● Time Series Analysis: It shows sales trends over time, helping businesses identify seasonal fluctuations and peak sales periods.

● Top Performing Categories: Identifies the product categories with the highest sales and revenue.

**2. Profitability & Business Growth**

● Profit Margin Analysis: Helps understand profitability per order and identify areas for improving profit margins.

● Year-over-Year Sales Growth: Tracks annual sales growth percentages, enabling better financial planning.

**Feature Engineering:**

Created new columns such as:

● total sale amount according sale per unit.

● profit\_margin (Profit per order / Sales per order).

● order\_year, order\_month (Extracted from order\_date).

● actual price of product.

**Key Questions and Insights to be Addressed:**

1) Which product catergory have the highest sales ?

category\_wise\_prod\_quantity Product Category

Clothing 6097

Ornaments 3768

Other 2469

2) Difference between sales price and total sales price ?

0 100

1 100

2 49

3 9

4 10

...

7389 65

7390 158

7391 45

7392 23

7393 9

3) how do you calculate profit margin ?

profit margin

df['profit margin']=df['actual price']- df['Sales Price']

print(df['profit margin'])

0 100  
1 100  
2 9  
3 9  
4 10  
 ...   
7389 65  
7390 118  
7391 45  
7392 23  
7393 9

4) What is top 10 performing location ?

Order Location city wise top sale

Las Vegas 35234

Cleveland 35153

Detroit 34240

Austin 33655

Manchester 32426

Cardiff 32397

London 32104

Liverpool 30875

Dublin 30172

Glasgow 29000

5) Average rating wise product category

Product Category Average Rating

Other 3.581590

Clothing 3.499460

Ornaments 3.445922

6) month month\_wise\_sales\_trend

Month Sales

1 43417

2 42670

3 44946

4 45440

5 47531

6 38883

7 45115

8 40031

9 41466

10 41682

11 42324

12 42588

**Visualization:**

Several charts created to present inside including:









