"Exploratory Sales Transaction Data Analysis: Data-Driven Insights for Business Growth"

Project Overview

This project analyzes sales transaction data to uncover key business insights. Using Python libraries like Pandas, NumPy, Matplotlib, and Seaborn, we clean and preprocess the data, handle missing values, and perform exploratory data analysis (EDA). The project identifies top-performing products, customer segments, and regional sales trends. Additionally, it examines the impact of shipping delays and profit margins. The insights gained help optimize sales strategies and improve customer satisfaction.

Goals of the Project

✓ Analyze sales transaction data to identify key trends and patterns.
✓ Clean and preprocess the dataset for accurate analysis.
✓ Perform exploratory data analysis (EDA) using visualizations.
✓ Identify top-performing products, customer segments, and sales regions.
✓ Understand the impact of shipping delays on customer satisfaction.
✓ Provide data-driven insights to improve sales strategies.

Materials and Methods

- **Dataset**: The data used for this project is sourced from a **sales transaction dataset** containing details about various sales transactions.
- **Data Points**:
    - **Sales Data**: Includes transaction details like sales per order, order quantity, and total sales.

    - **Product Categories**: Information on the product types sold (e.g., electronics, apparel, etc.).
    - **Order Dates and Ship Dates**: Contains the order and shipping dates, allowing us to evaluate shipping delays.
    - **Profit Margins**: The difference between sales per order and cost per order, used for profitability analysis.
    - **Customer Details**: Includes customer IDs, customer segments, and geographical regions.
    - **Shipping Information**: Shipping type, delays, and status for each order.
- **Analysis Goals**:
    - To analyze **sales performance**, identifying top-performing products, categories, and regions.
    - To explore **customer behavior** through metrics like order frequency and customer segments.
    - To understand **shipping delays** and their impact on customer satisfaction and sales.

Project Outcome & Insights

- **Top-Performing Products**:
  The analysis revealed the best-selling products, helping to identify which items contribute most to the overall revenue. These insights guide inventory management and marketing strategies.
- **Customer Segments**:
  A deeper understanding of customer segments was gained, identifying which groups (e.g., age, region) generate the highest sales. This helps target specific customer bases for promotions and campaigns.
- **Sales Trends**:
  The project identified sales trends over time, including seasonality

effects and key spikes or drops. These trends assist in forecasting future sales and planning marketing efforts accordingly.

- **Shipping Delays and Customer Satisfaction**:
  Analyzing shipping delays showed their negative impact on sales, as delayed shipments often lead to lower customer satisfaction and fewer repeat purchases. Timely deliveries are key to maintaining customer loyalty.
- **Profitability Insights**:
  By calculating profit margins across different products and regions, the project highlighted areas where profitability could be improved by adjusting pricing or reducing costs.

Feature Engineering

- **Date Features**:
  - **Order Year, Month, Weekday**: Extracted from the order date to analyze sales trends and seasonality.
- **Shipping Delays**:
  - **Actual Shipping Delay**: Calculated as the difference between order and ship dates.
- **Profit Margin**:
  - **Profit Margin**: Ratio of profit per order to sales per order.
- **Customer Behavior**:
  - **Returning Customer**: Identified based on repeated customer IDs.
- **Shipping Categories**:
  - **Shipping Category**: Categorical variable based on shipping delays (Same Day, Fast, Moderate, Delayed).

Key Questions and Insights

1.What are the top-performing products and categories?

```
top_products = df.groupby('Product_name')['Sales_per_order'].sum().sort_values(ascending=False).head(10)

print("\nTop-Performing Products:\n", top_products)
```

Answer: Top performing categories:

Category_of_product

Home & Kitchen 143082.86000

Beauty 142189.26832

Clothing 126500.30000

Sports 121050.20000

Electronics 115145.65000

2. What are sales by region?

```
sales_by_region = df.groupby('Customer_region')['Sales_per_order'].sum().sort_values(ascending=False)

print("\nSales by Region:\n", sales_by_region)
```

Answer: Sales by Region:

Customer_region

South 179555.63832

North 166451.40000

East 155483.10000

West 146478.14000

3.Which product have the highest sales?

```
top_products = df.groupby('Product_name')['Sales_per_order'].sum().sort_values(ascending=False).head(10)

print("\nTop-Performing Products:\n", top_products)
```

Answer: Top-Performing Products:

Product_name

Perfume 42936.63832

Blender 34043.22000

T-Shirt 33479.24000

Jacket 30947.41000

Basketball 30453.13000

Toaster 30423.40000

Vacuum Cleaner 30176.36000

Headphones 28297.31000

Lipstick 27750.04000

Coffee Maker 27062.82000

4. How do sales trends vary over time?

```
monthly_revenue = df.groupby(['Order_year', 'Order_month'])['Sales_per_order'].sum().reset_index()

print("\nMonthly Revenue:\n", monthly_revenue)
```

Answer : Monthly Revenue:

	Order_year	Order_month	Sales_per_order
0	2024	1	4082.93000
1	2024	2	58953.03832
2	2024	3	63559.85000
3	2024	4	48043.75000
4	2024	5	47306.32000
5	2024	6	44281.91000
6	2024	7	61574.46000
7	2024	8	46180.65000
8	2024	9	54746.63000
9	2024	10	45772.54000
10	2024	11	56949.02000
11	2024	12	52235.64000
12	2025	1	64281.54000

5.What are Customer Segment Wise Top Sales?

```
customer_segment_sales = df.groupby('Customer_segment')['Sales_per_order'].sum().sort_values(ascending=False)

print("\nCustomer Segment Wise Top Sales:\n", customer_segment_sales)
```

Answer: Customer Segment Wise Top Sales:

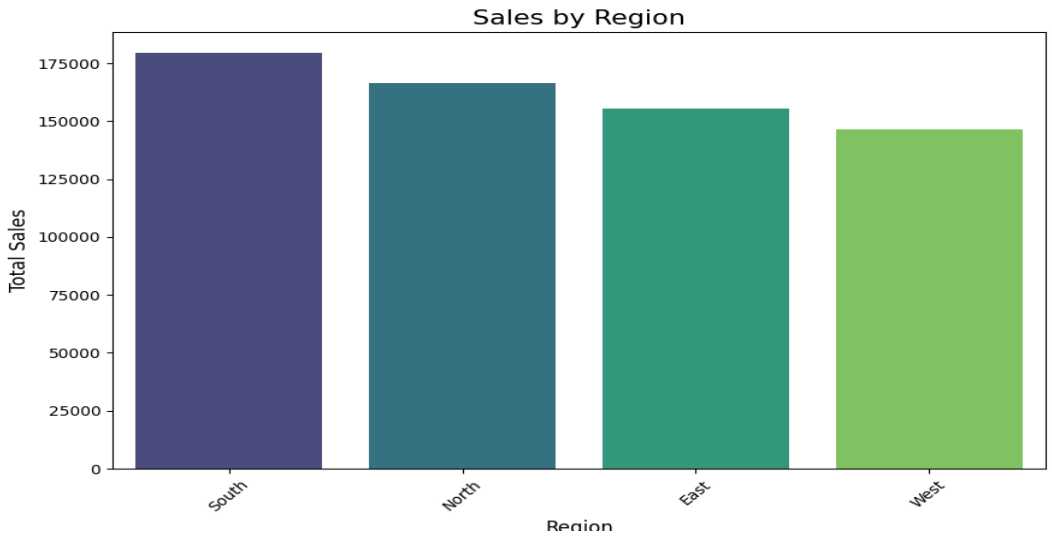
Customer_segment	
Consumer	335219.50944
Corporate	211062.68944

Home Office 101686.07944

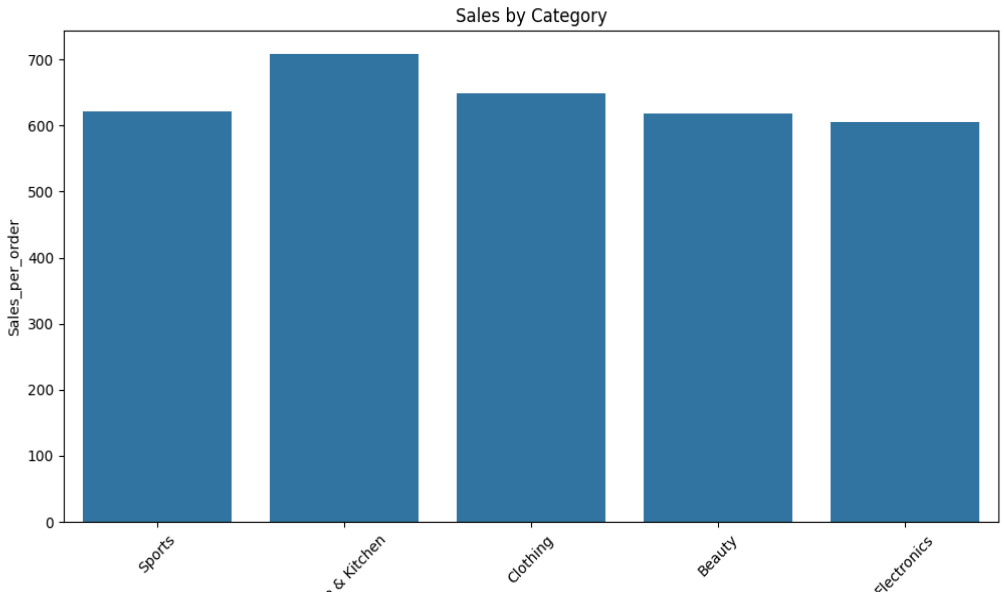
Visualization:

Several charts created to present inside including:

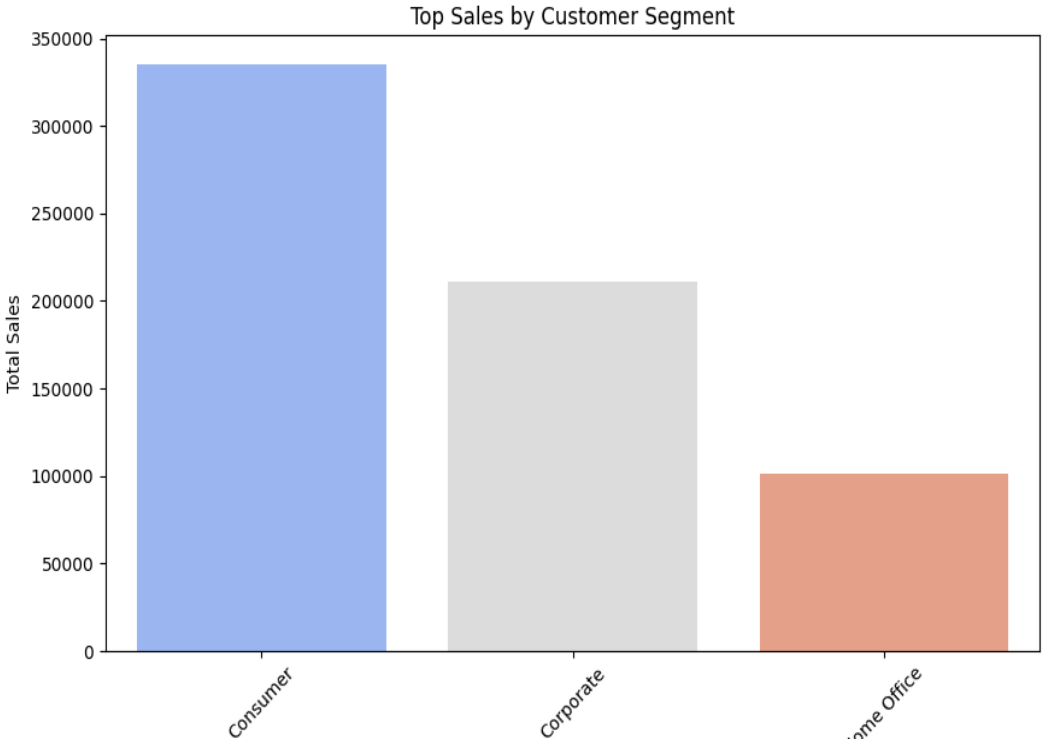
- Sales by Region (Bar chart)



- Sales by Category (Bar chart)



- Top sales by customer segment (Bar chart)



- Sales trend over time (Line chart)