EDA (EXPLORATORY DATA ANALYSIS) PROJECT

EDA ON SALES DATA

- Tools Used:
- Libraries: pandas, numpy, matplotlib, seaborn, streamlit, openpyxl
- Data Source: Sales Data (Excel)
- Packages Installed: pandas, numpy, matplotlib, seaborn, streamlit, openpyxl
- Data Import: Loaded sales data using pd.read_excel().
- Initial Data Check:
- df.head() to preview the data
- Checked for missing values, duplicates, and data types.

- **Missing Values:**
- Removed row with missing 'Ship Code'.
- Filled missing 'Order Release Date' with 'Order Date'.
- **Duplicates:**
- Removed duplicate rows using drop_duplicates().
- Data Types: Verified correct data types for columns.
- Shape of Data: 23,194 records, 12,873 unique orders.
- **Revenue and Order Count:**
- Total Revenue: $24,485,214.76
- Total Orders: 12,873

KEY INSIGHTS FROM DATA

- **Top 10 Items by Revenue**: Identified highest-grossing products like Item I0021, I0016.
- Average Order Value: Calculated mean order value for insights on sales behavior.
- **Monthly Sales Trend:**
- Fluctuating sales across months.
- Peak months show higher revenue, indicating seasonal demand.
- High and Low Sales Months: Identified months with the highest and lowest sales.

- **Top 10 Customers:** Identified top customers by order count.
- Order Size Distribution: Analyzed common order sizes and quantities.
- **Sales by Warehouse:**
- Top warehouses for sales distribution: DWN and LPI.
- Visualized warehouse contribution with bar charts.
- **Shipping and Delivery:**
- 88.5% delayed shipments, indicating a need for improved logistics and delivery time optimization.
- Unit Price vs Total Sales: Analyzed pricing impact on total sales.

RECOMMENDATIONS

- **Inventory Management:**
- Increase inventory for high-revenue items (I0021, I0016, I0051).
- Focus marketing efforts on high-sales products.

- **Shipment Optimization:**
- Analyze causes for delayed shipments and optimize logistics.
- Work with logistics partners to reduce delays.

- **Customer Engagement:**
- Offer personalized promotions for top customers based on order history.
- Implement loyalty programs and email campaigns.

- **Seasonal Sales Strategy:**
- Plan marketing campaigns during peak sales months.
- Offer discounts or incentives during slow months.

- **Pricing & Profitability:**
- Focus on high-margin products despite some low-priced items generating higher sales.

BI_ASSISTANT

- Open terminal - run streamlit file
- python -m streamlit run BI_assistant_EDAfile.py

(using anaconda hence used python - m)

- It will open in browser