**Exploratory Data Analysis (EDA) Project - Diwali Sales Data**

**Introduction**

This project performs **Exploratory Data Analysis (EDA)** on a **Diwali Sales Dataset** to uncover insights into customer behavior, sales trends, and product performance. The analysis leverages **Pandas, NumPy, Matplotlib, and Seaborn** for data processing and visualization.

**Goals of the Project:**

* Explore and clean the dataset to ensure data integrity.
* Identify key trends in sales and customer demographics.
* Visualize distributions, correlations, and category-wise trends.
* Provide actionable insights for business decision-making.

**Materials and Methods**

**Libraries Used:**

* pandas: Data manipulation and processing.
* numpy: Numerical computations.
* seaborn & matplotlib: Data visualization.

**Dataset Overview:**

* The dataset contains **customer transactions from Diwali sales.**
* It includes features like **age, gender, marital status, occupation, state, product category, number of orders, and total sales amount.**

**Key Findings and Business Insights**

**1. Customer Demographics**

* **Male customers** make up the majority of sales, but female customers also contribute significantly.
* **Customers aged 26-35** are the biggest spenders.
* **Married individuals** tend to purchase more, indicating family-oriented buying patterns.

**2. Regional Insights**

* Some states have **higher sales volume**, suggesting targeted marketing strategies for underperforming regions.

**3. Product Trends**

* Specific **product categories perform exceptionally well** during sales, helping businesses prioritize inventory.
* **Top-selling products** should be promoted more aggressively during sales events.

**4. Recommendations for Business Growth**

✅ **Target Marketing:** Focus on the **26-35 age group** and **married individuals**. ✅ **Inventory Planning:** Stock more of **top-selling categories** before major sales. ✅ **Regional Promotions:** Invest in **state-wise marketing** to boost sales in underperforming regions. ✅ **Customer Engagement:** Implement loyalty programs for repeat customers.

**Feature Engineering:**

**Created new columns such as:**

* **Age Group** → Categorizing customers into age groups (e.g., 18-25, 26-35, etc.
* **Total Purchase Value per Customer** → Sum of all purchases per customer.
* **Top State Category** → Rank states by total sales.
* **Purchase Day** → Extract day of the week from the purchase date.

**Data Analysis & Insights**

1. **What was observed in the gender-wise sales analysis?**  
   **Answer:**
   * Male customers contributed a **higher percentage of total sales**.
   * Female customers also had significant spending.
   * Visualization used:

sns.barplot(x='Gender', y='Amount', data=sales\_gen)

1. **Which age group spent the most during the Diwali sale?**  
   **Answer:**  
   Customers **aged 26-35** had the highest total spending.
2. **Which states had the highest sales?**  
   **Answer:**
   * Certain states showed **significantly higher sales volume**.
   * Visualization:

sns.barplot(x='State', y='Amount', data=sales\_state)

1. **How does marital status affect sales?**  
   **Answer:**
   * **Married individuals** made more purchases.
   * This indicates **family-oriented buying behavior**.
2. **Which product categories had the highest sales?**  
   * **Electronics and fashion** dominated the sales.
   * Visualization:

sns.barplot(x='Product\_Category', y='Amount', data=sales\_category)

1. **Which products had the highest number of orders?**  
   **Answer:**
   * The **top 10 most sold products** were identified using:

df.groupby('Product\_ID')['Orders'].sum().nlargest(10).plot(kind='bar')

1. **What marketing strategy would you recommend based on this analysis?**  
   **Answer:**
   * **Target 26-35 age group** with promotions.
   * **Encourage loyalty programs** for repeat customers.
   * **Stock high-selling products** before major sales events.
2. **How can inventory management be improved?**  
   **Answer:**
   * Focus on **top-selling categories**.
   * **Reduce stock** for less popular products.
   * **Analyze regional demand** for optimized stocking.

**Conclusion:**

This **Diwali Sales EDA Project** provides valuable insights into customer behavior, sales patterns, and regional trends. By leveraging these findings, businesses can optimize their marketing strategies, inventory management, and customer engagement to drive higher sales and profitability.

Visualization:

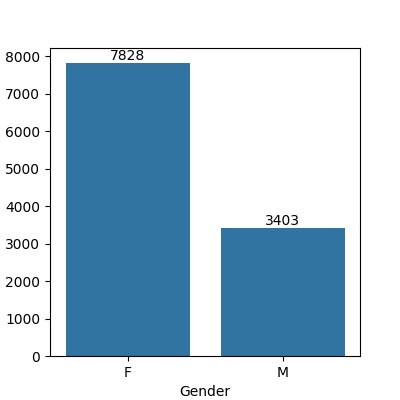


Fig.1 plotting a bar chart for Gender and it's count

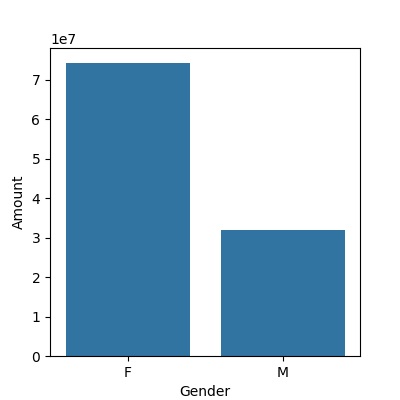


Fig.2 plotting a bar chart for gender vs total amount

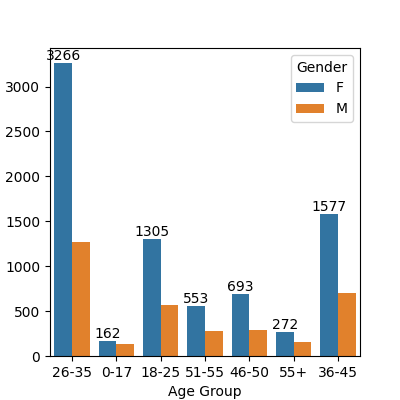


Fig.3 plotting a bar chart for gender vs total amount

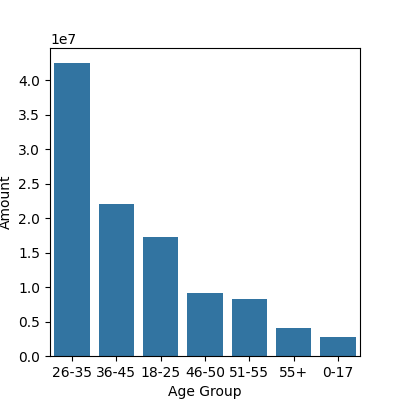


Fig 4. Age

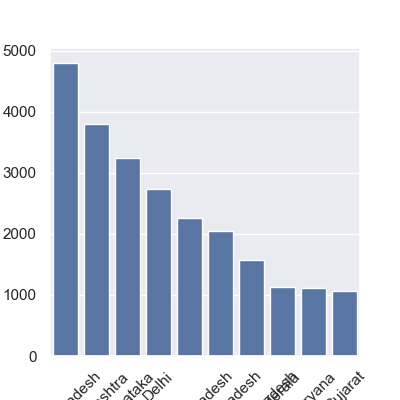


Fig 5. Total Amount vs Age Group

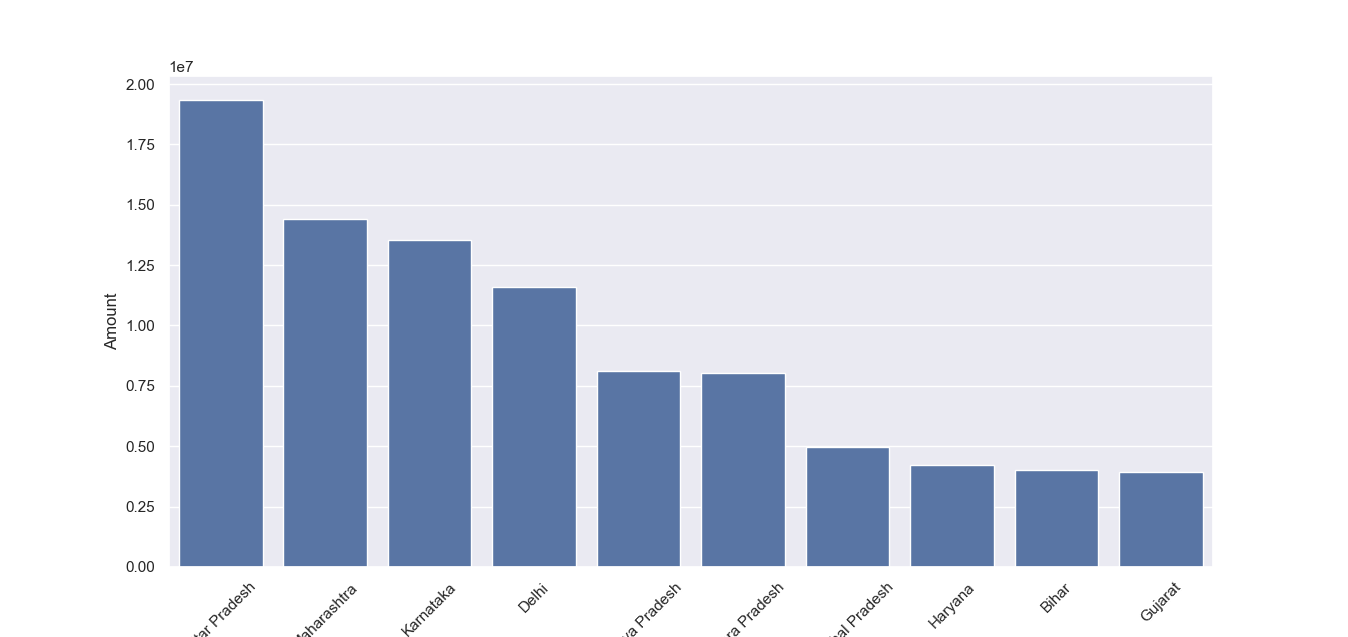


Fig 6. total number of orders from top 10 states

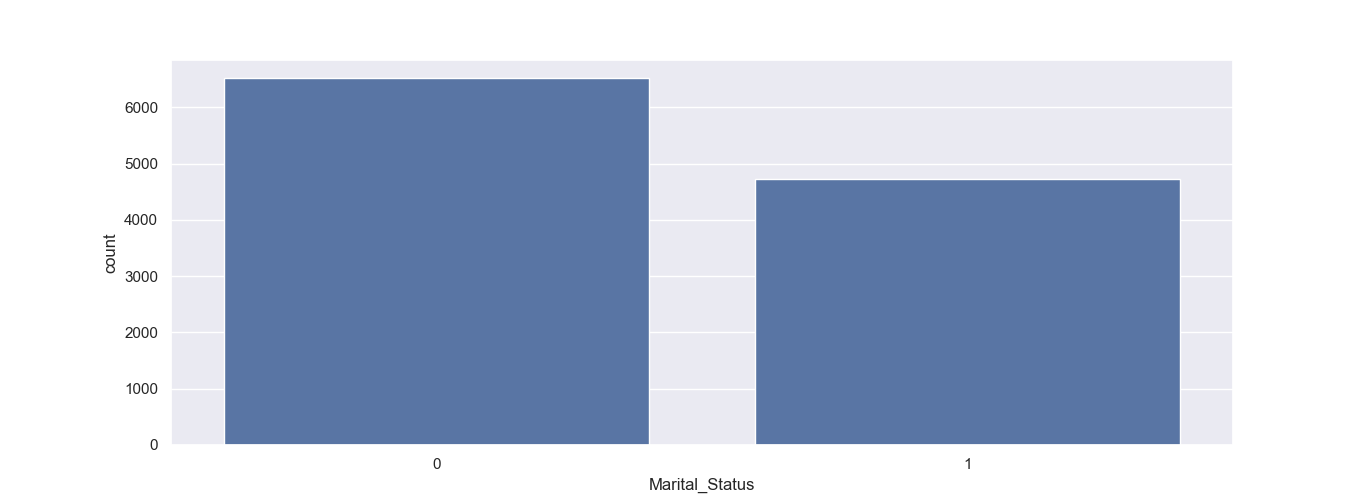


Fig 7. total amount/sales from top 10 states

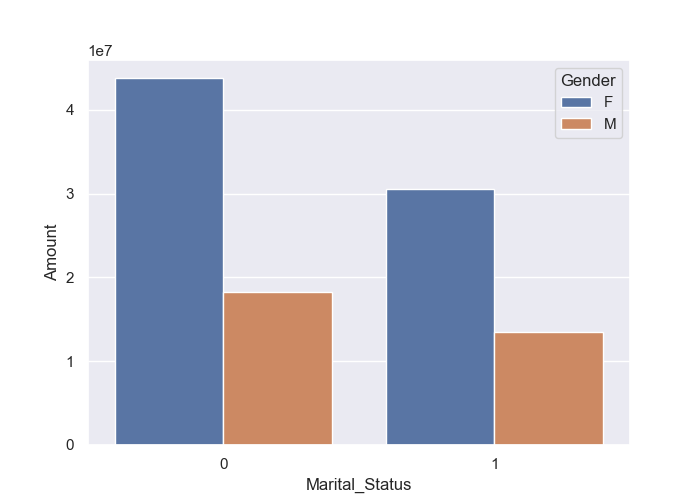


Fig 8. Marital Status

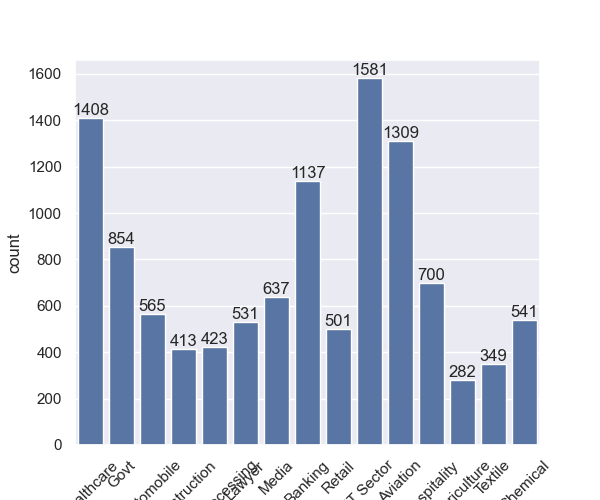


Fig 9. Marital Status By Gender

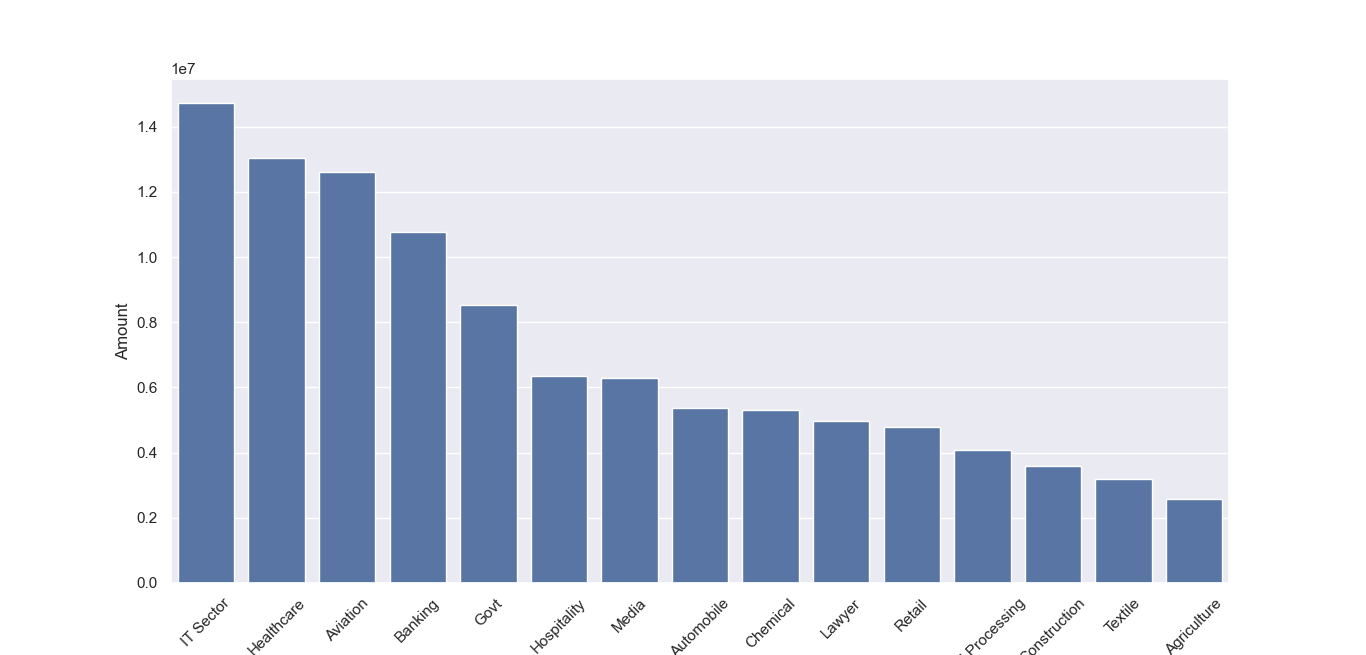


Fig 10. Occupation

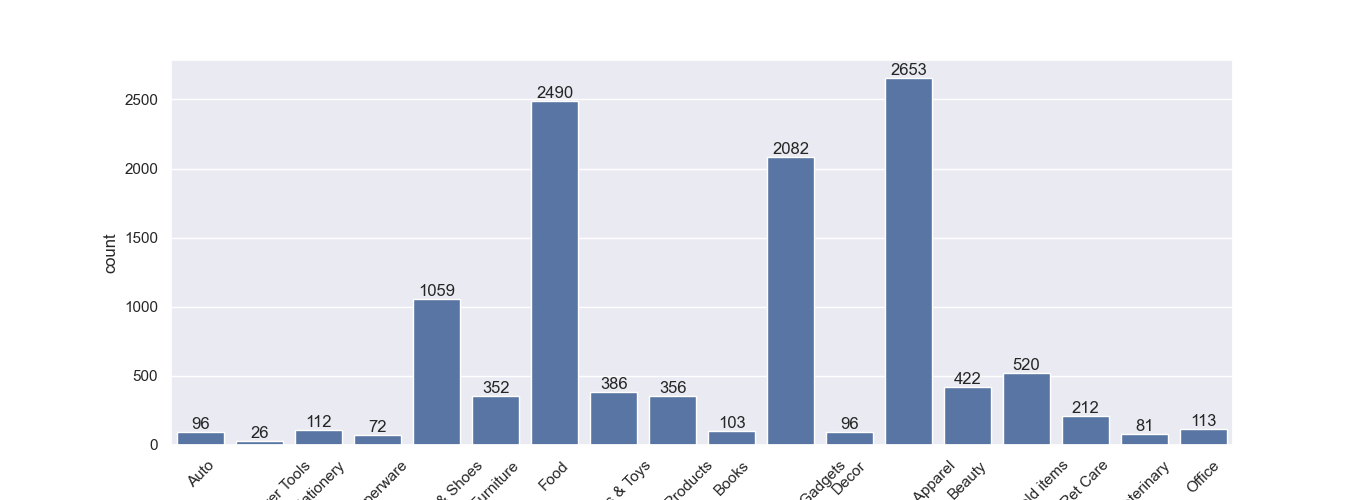


Fig 11. Occupation VS Amount

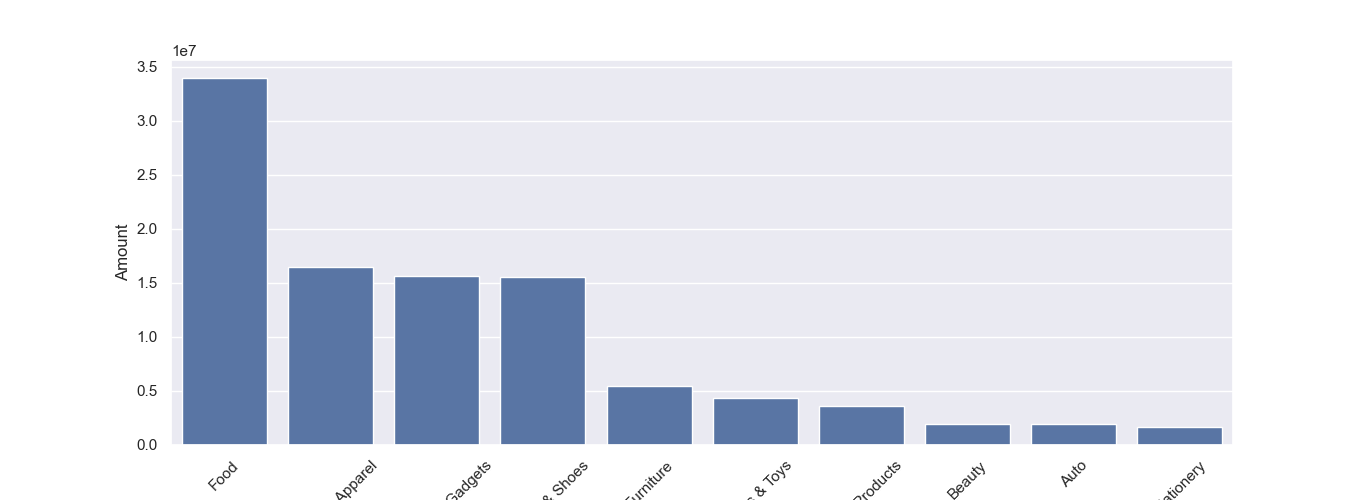


Fig 12. Product Category

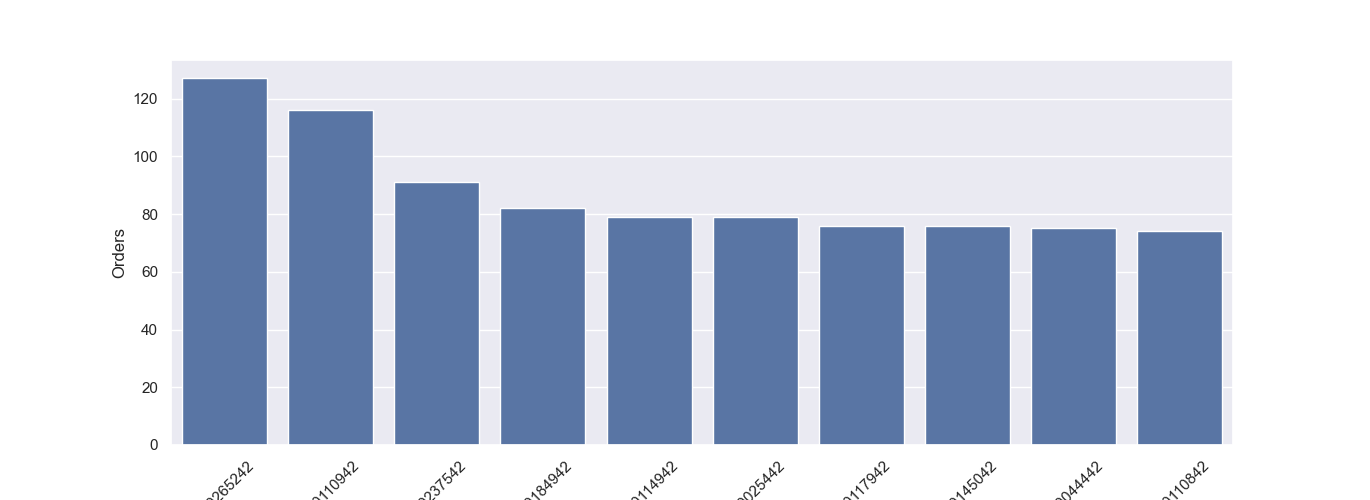


Fig 13. Product ID VS Order

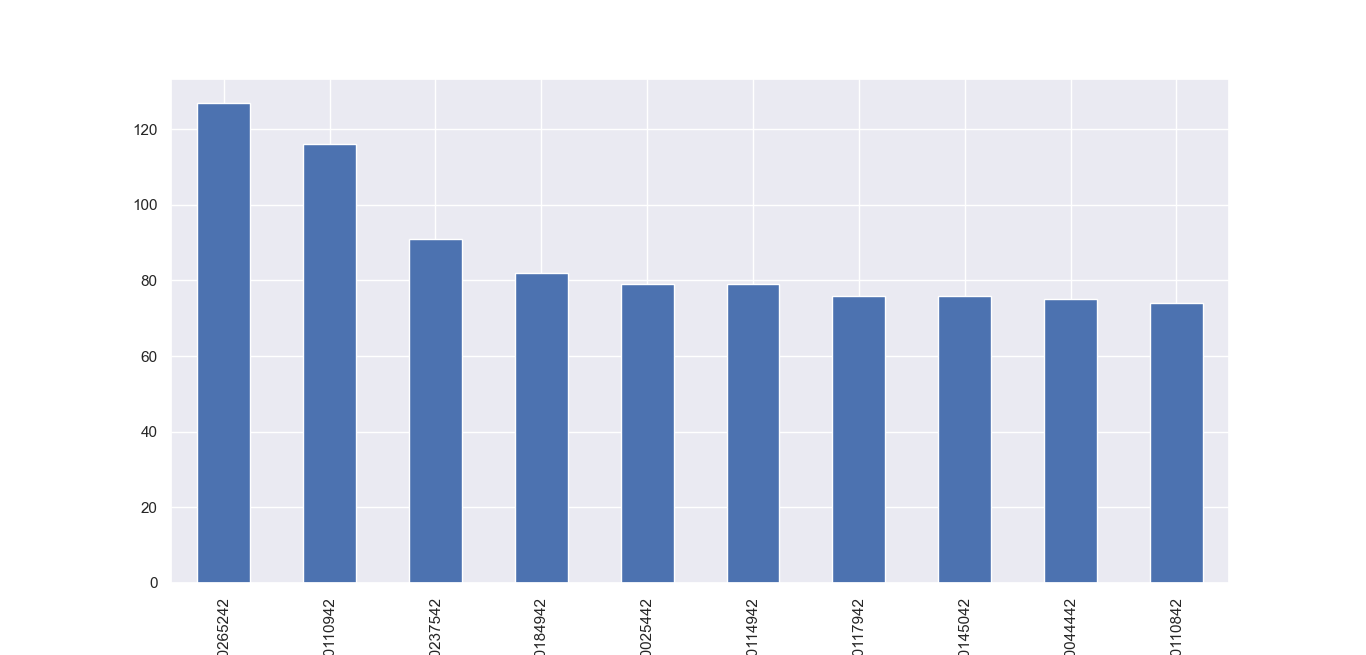


Fig 14.  top 10 most sold products