**Project Title: Sales Data Analysis using Python**

Project Report  
Prepared by: Ritika Roy  
Date: June 2025

# 1. Project Overview

To perform an in-depth Exploratory Data Analysis (EDA) on customer transactions to uncover patterns and meaningful insights related to gender, age group, geography, profession, and product category. This project helps businesses identify and understand their most valuable customer segments, optimize marketing strategies, and tailor product recommendations for better business decisions.

# 2. Data Overview

* **Dataset Name:** Sales\_Data.csv
* **Data Source:** Internal transaction logs containing both demographic and transactional information.
* **Features:**
  + **Demographics:** Gender, Age, Profession
  + **Transactions:** Orders (Quantity), Amount (Total spent)
  + **Products:** Product Category, Product ID
  + **Geography:** State (Customer Location)

The dataset had minor issues such as blank columns and missing values, which were cleaned using pandas.

# 3. Tools and Libraries Used

* **Python** (v3.9+): Primary language for data processing and visualization.
* **Pandas:** Used extensively for data manipulation, cleaning, aggregation.
* **NumPy:** Supported statistical operations and numerical computations.
* **Matplotlib & Seaborn:** For static and appealing visualizations using professional color palettes.
* **Jupyter Notebook:** Used as the IDE to execute, document, and narrate the project journey.
* **Color Palette:** A deep blue-based theme was applied using Seaborn’s dark\_palette and custom RGB settings to maintain visual harmony and professionalism.

# 4. Step-by-Step Exploratory Data Analysis (EDA)

**1. Gender-wise Transactions Count (Bar Chart)**

* Used sns.countplot() to visualize transaction frequency by gender.
* Result: Female users dominate in number of transactions.

**2. Gender Distribution (Pie Chart)**

* Used plt.pie() to show proportion of male vs. female customers.
* Result: Females account for more than 60% of total transactions.

**3. Gender-wise Sales Amount (Bar Chart)**

* Grouped by Gender and calculated total Amount spent.
* Result: Females contribute significantly more to the revenue than males.

**4. Age Group-wise Transactions Count (Bar Chart)**

* Visualized count of transactions by different age groups.
* Result: 26–35 is the most active and profitable segment.

**5. Age Group-wise Total Amount (Bar Chart)**

* Used groupby with Amount to calculate total purchase value.
* Result: Again, the 26–35 age group leads, confirming high purchasing power.

**6. Top 5 States by Orders (Bar Chart)**

* Grouped by State and used Orders to rank top performers.
* Top 3 States: **Uttar Pradesh, Maharashtra, Karnataka**.

**7. Top 5 States by Sales Amount (Bar Chart)**

* Similar to above, but based on revenue instead of order count.
* Result: Same top 3 states dominate in both metrics.

**8. Product Category-wise Transaction Count (Bar Chart)**

* Displayed frequency of orders by each product category.
* Top Categories: **Beauty, Sports, Electronics**.

**9. Product Category-wise Sales Amount (Bar Chart)**

* Sorted by total amount spent on each category.
* Beauty leads, followed by Sports and Electronics.

**10. Product Category and Gender-wise Transactions**

* Used hue='Gender' in Seaborn to split category-wise data by gender.
* Result: Females outperform males across all categories.

**11. Profession-wise Transactions Count**

* Used sns.countplot() to show which professions shop the most.
* Common Professions: **IT, Healthcare, Aviation**.

**12. Amount-wise Top Professions (Bar Chart)**

* Ranked professions by revenue generated.
* Result: IT sector dominates, followed by Healthcare and Aviation.

**13. Top 10 Products by Quantity Ordered**

* Grouped by Product\_ID, sorted by Orders.
* Result: Revealed best-sellers which can be key for inventory and promotions.

All charts included:

* Titles
* Font and label adjustments
* Custom dark blue color schemes for a polished, cohesive look.

# 5. Key Insights

* **Customer Demographics:**
  + Most customers are **female**, aged **26–35**.
  + This group represents the most engaged and highest-spending demographic.
* **Top Locations:**
  + **Uttar Pradesh**, **Maharashtra**, and **Karnataka** are crucial markets.
  + Marketing efforts should be localized for these high-performing states.
* **Product Preferences:**
  + Focus areas should include **Beauty**, **Sports**, and **Electronics**.
  + Targeted offers in these segments could increase conversions.
* **Professional Insights:**
  + Professionals in **IT**, **Healthcare**, and **Aviation** are the biggest spenders.
  + These groups may benefit from personalized loyalty programs or curated recommendations.
* **Product-Level Opportunities:**
  + The Top 10 products by order volume should be prioritized for promotions and bundling strategies.

# 6. Customer Summary

Female customer aged 26–35, located in Uttar Pradesh, Maharashtra, or Karnataka, working in IT, Healthcare, or Aviation, with a preference for Beauty, Sports, or Electronics products. She is highly engaged and contributes significantly to company revenue.

# 7. Conclusion

This comprehensive analysis provides valuable insights into customer behavior, preferences, and geography. By identifying high-value customer segments and product trends, businesses can drive targeted growth and improve customer satisfaction. Continuous use of analytics and segmentation will lead to more effective decision-making and revenue optimization.