

Software Documentation

Overview

This documentation outlines the steps, processes, and tools used in the data analysis project for a retail company. The project involves data cleaning, exploratory data analysis (EDA), visualization, and forecasting future sales.

Table of Contents

1. Introduction
 2. Tools and Technologies
 3. Data Cleaning Process
 4. Exploratory Data Analysis (EDA)
 5. Visualizations
 6. Insights and Recommendations
-

1. Introduction

The project analyzes a retail dataset to uncover trends and provide actionable insights for improving sales and customer retention. Tasks include data cleaning, segmentation, visualization, and forecasting.

2. Tools and Technologies

- **Programming Language:** Python
 - **Libraries:** pandas, NumPy, Matplotlib, Seaborn, statsmodels
 - **Visualization Tools:** Python-based libraries (e.g., Matplotlib, Seaborn)
-

3. Data Cleaning Process

Initial Data Inspection

- Previewed the dataset to understand its structure and identify potential issues.
- Checked for missing values, duplicate entries, and inconsistent data.

Steps Performed

- 1. Handling Missing Values:**
 - Removed rows with missing **Customer ID** values.
 - Filled missing **Age** values with the median age.
- 2. Duplicate Removal:**
 - Identified and removed duplicate rows.
- 3. Date Standardization:**
 - Standardized the **Date** column.
 - Removed invalid dates.
- 4. Derived Metrics:**
 - Created a **Revenue per Customer** column.
 - Aggregated **Total Amount** for total revenue calculations.

```
Initial Data Preview:
Transaction ID      Date Customer ID Gender Age Product Category \
0                1  2023-11-24  CUST001   Male  34         Beauty
1                2  2023-02-27  CUST002  Female  26         Clothing
2                3  2023-01-13  CUST003   Male  50        Electronics
3                4  2023-05-21  CUST004   Male  37         Clothing
4                5  2023-05-06  CUST005   Male  30         Beauty

Quantity  Price per Unit  Total Amount
0         3           50           150
1         2          500          1000
2         1           30           30
3         1          500           500
4         2           50           100

Missing Values:
Transaction ID      0
Date                0
Customer ID         0
Gender              0
Age                 0
Product Category    0
Quantity            0
Price per Unit      0
Total Amount        0
dtype: int64

Duplicate Rows:
0

Total Revenue: 456000
```

4. Exploratory Data Analysis (EDA)

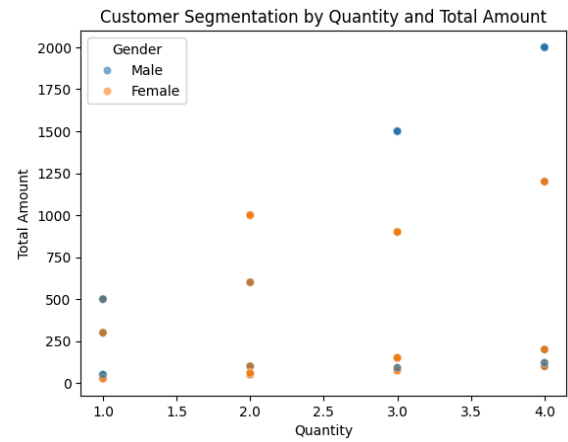
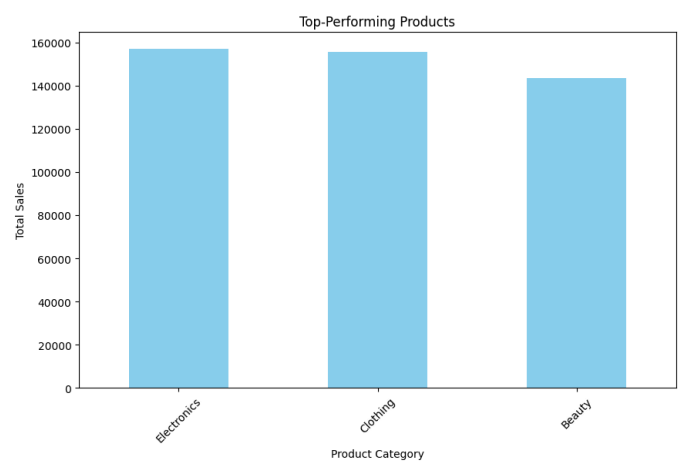
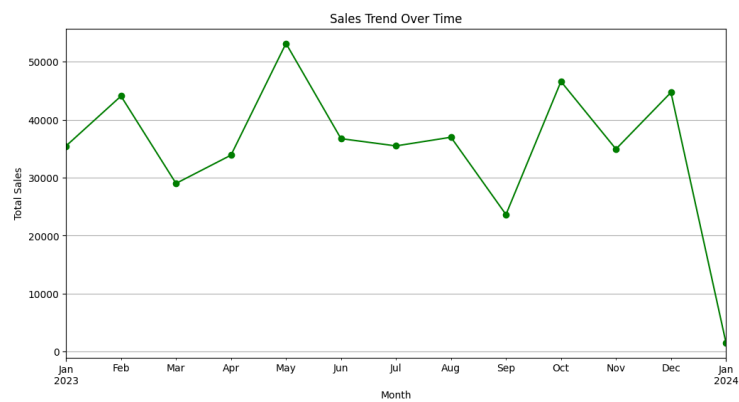
Objectives

- Analyze patterns in sales data.
- Segment customers based on purchasing behavior.
- Identify top-performing products, regions, and sales channels.

Findings

1. **Top-Performing Products:**
 - Identified [Category Name] as the highest revenue generator.
2. **Seasonal Trends:**
 - Observed peaks in sales during [specific months].
3. **Customer Segmentation:**
 - High-value customers often purchase [specific products].

```
Top-performing Products:  
Product Category  
Electronics    156905  
Clothing       155580  
Beauty         143515  
Name: Total Amount, dtype: int64
```



5. Visualizations

Sales Trend Analysis Chart

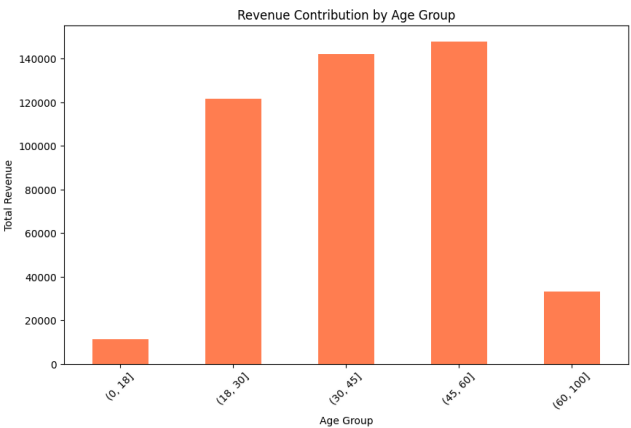
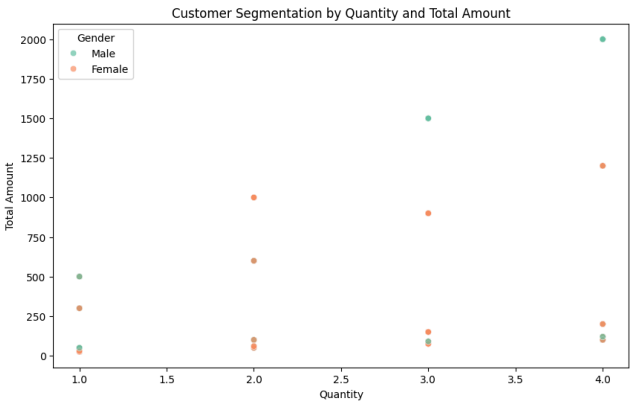
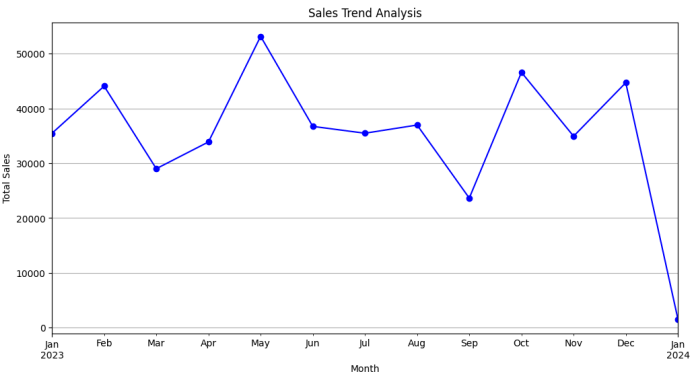
- Visualizes monthly sales trends, highlighting seasonal patterns.

Customer Segmentation

- Segments customers based on quantity purchased and total amount spent.

Revenue by Age Group

- Demonstrates revenue contributions by age groups, highlighting key demographics.



6. Insights and Recommendations

Insights

1. **Sales Trends:** Seasonal fluctuations and peak months identified.
2. **Customer Behavior:** High-value customers often prefer [specific products].
3. **Demographics:** Younger age groups contribute significantly to revenue.

Recommendations

1. **Increase Sales During Low-Performing Months:**
 - Offer targeted promotions.
 - Leverage marketing campaigns.
2. **Retain High-Value Customers:**
 - Introduce loyalty programs.
 - Personalize marketing strategies.
3. **Optimize Inventory:**
 - Expand stock for top-performing products.
 - Bundle complementary items to boost sales.