

Nikhil Sharma

Customer Shopping Data

17 July 2023

Customer Shopping Data Analysis Report

1. Introduction

This report presents findings from the analysis of the "Customer Shopping Data" dataset. The dataset, from Mehmet Tahir Aslan on Kaggle, includes shopping information from 10 Istanbul shopping malls from 2021 to 2023. The analysis aims to understand shopping trends in Istanbul, providing insights for researchers, data analysts, and machine learning enthusiasts.

2. Data Analysis

The analysis used Python, including the NumPy, Pandas, Matplotlib, Seaborn libraries. The Jupyter Notebook code was used for the analysis, which explored and understood the dataset. The key findings are as follows:

2.1 Data Shape and Columns

The dataset consists of 99,457 rows and 10 columns. The columns are: 'invoice_no', 'customer_id', 'gender', 'age', 'category', 'quantity', 'price', 'payment_method', 'invoice_date', and 'shopping_mall'.

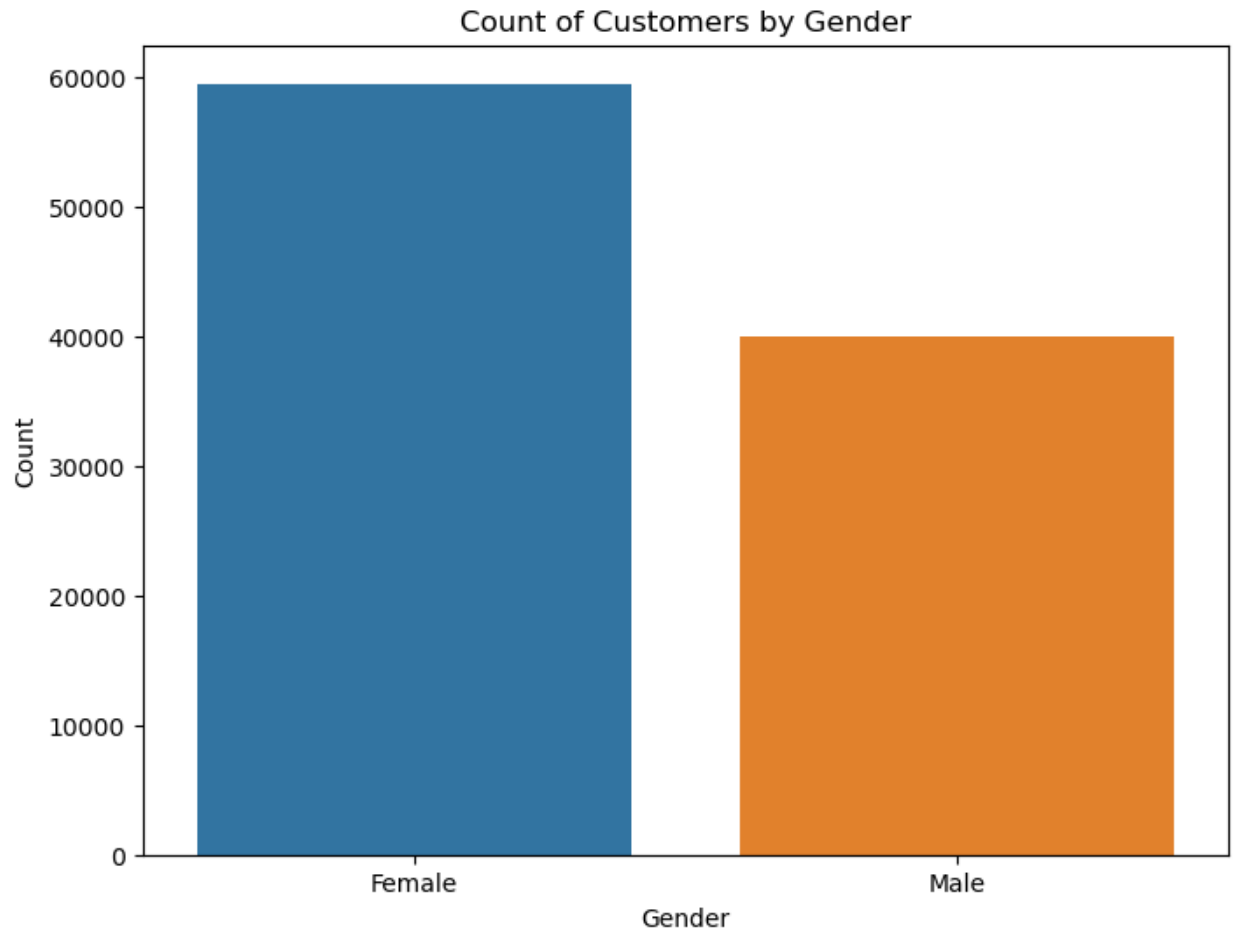
2.2 Data Types and Cleaning

The data types of the columns were checked and aligned with their respective attributes. The "invoice_date" column was converted to datetime for further analysis.

3. Findings

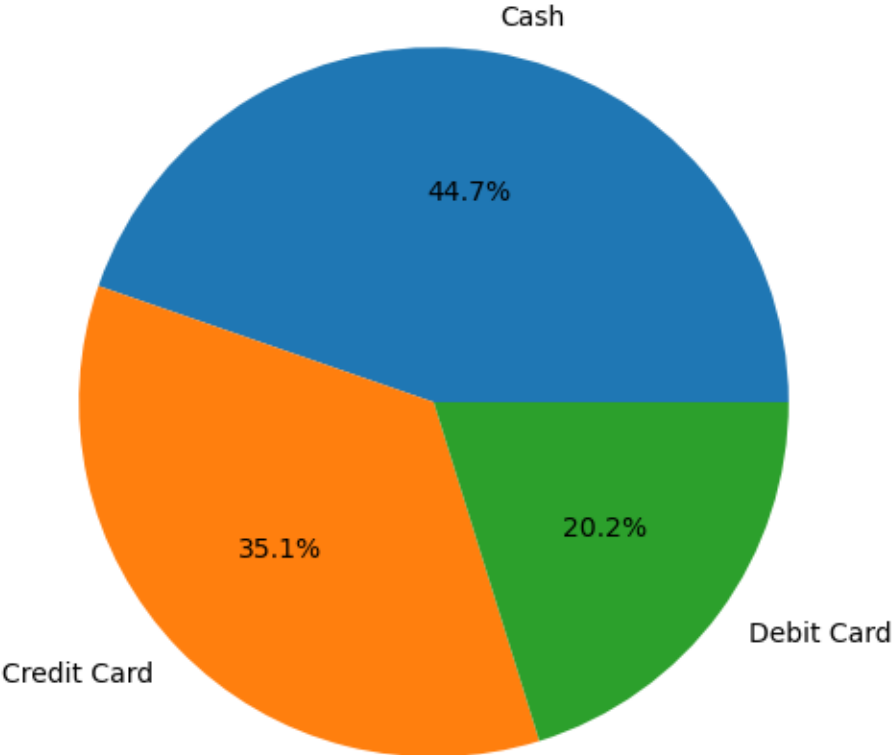
3.1 Customer Demographics and Shopping Habits

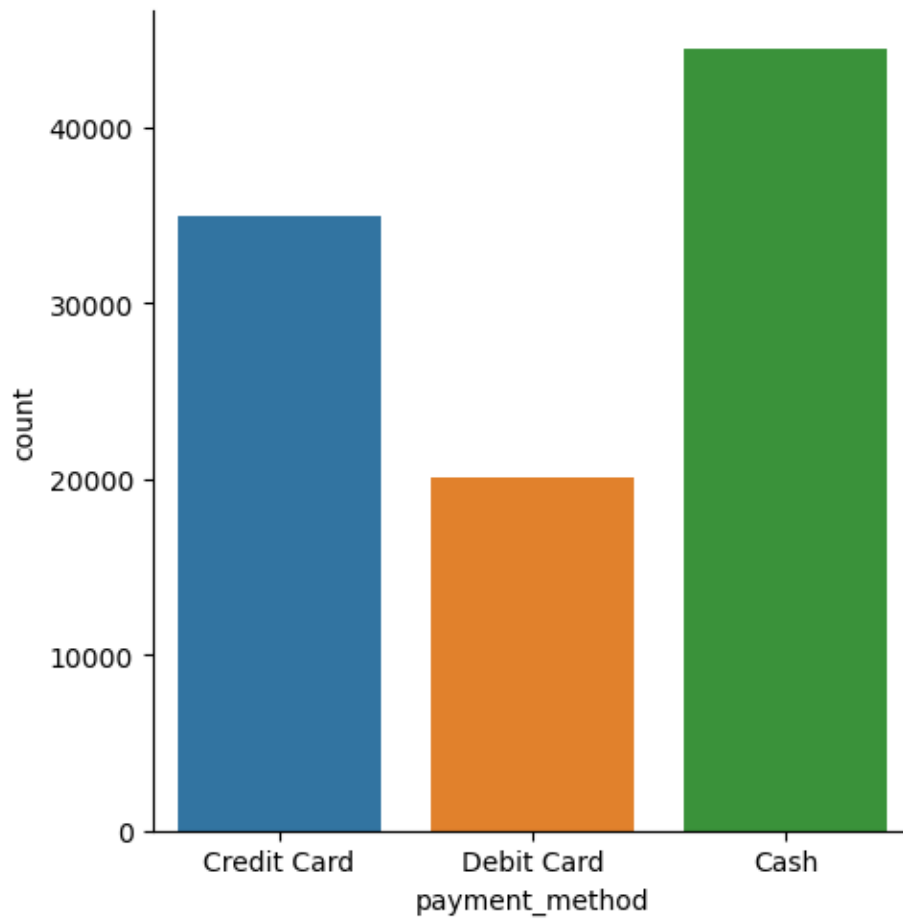
- **Gender Distribution:** The dataset indicates that the majority of customers are **female (59.8%)**, with males accounting for 40.19%.



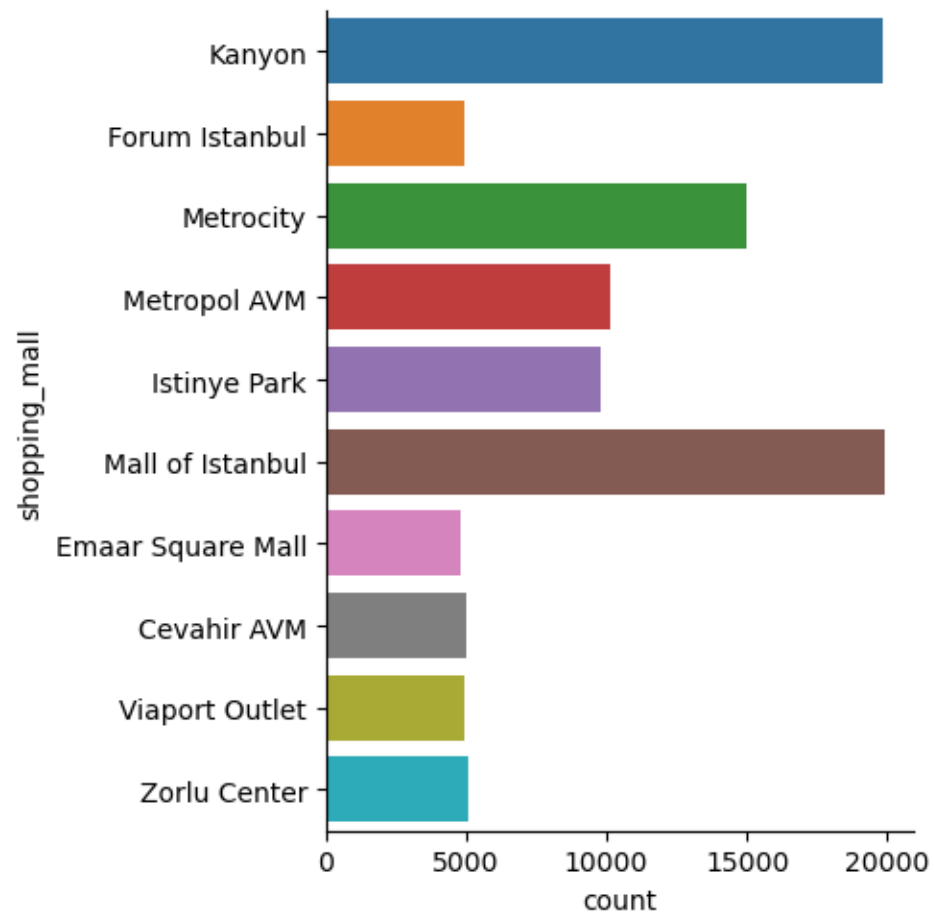
- **Age Distribution:** The largest age group of customers is **37**, accounting for **20.68%**.
- **Payment Methods:** **Cash** is the most common payment method, used in **44.68%** of transactions.

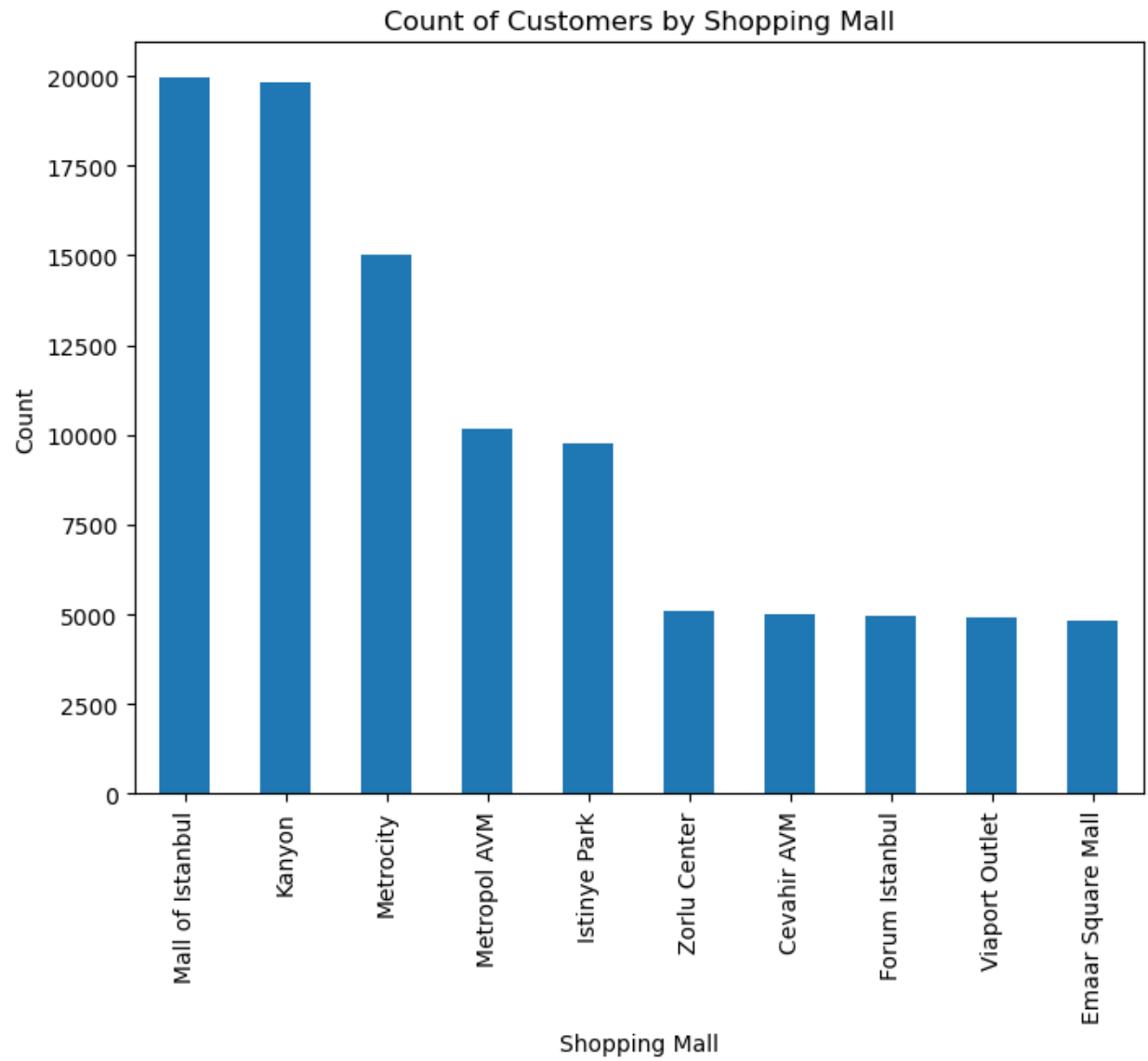
Proportional Distribution of Payment Methods



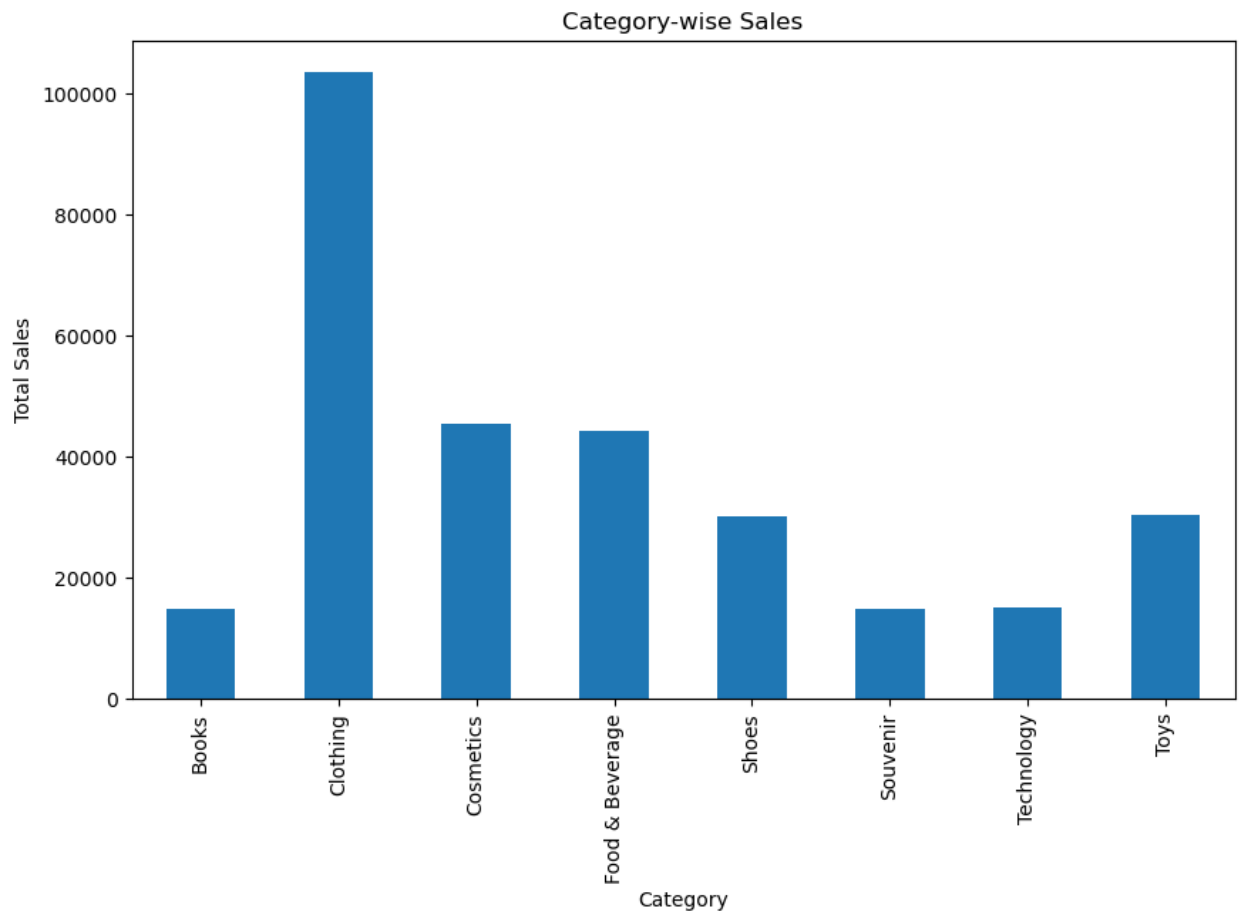


- **Shopping Mall Analysis:** The **Mall of Istanbul** had the highest number of customer visits, constituting **20%** of the total.



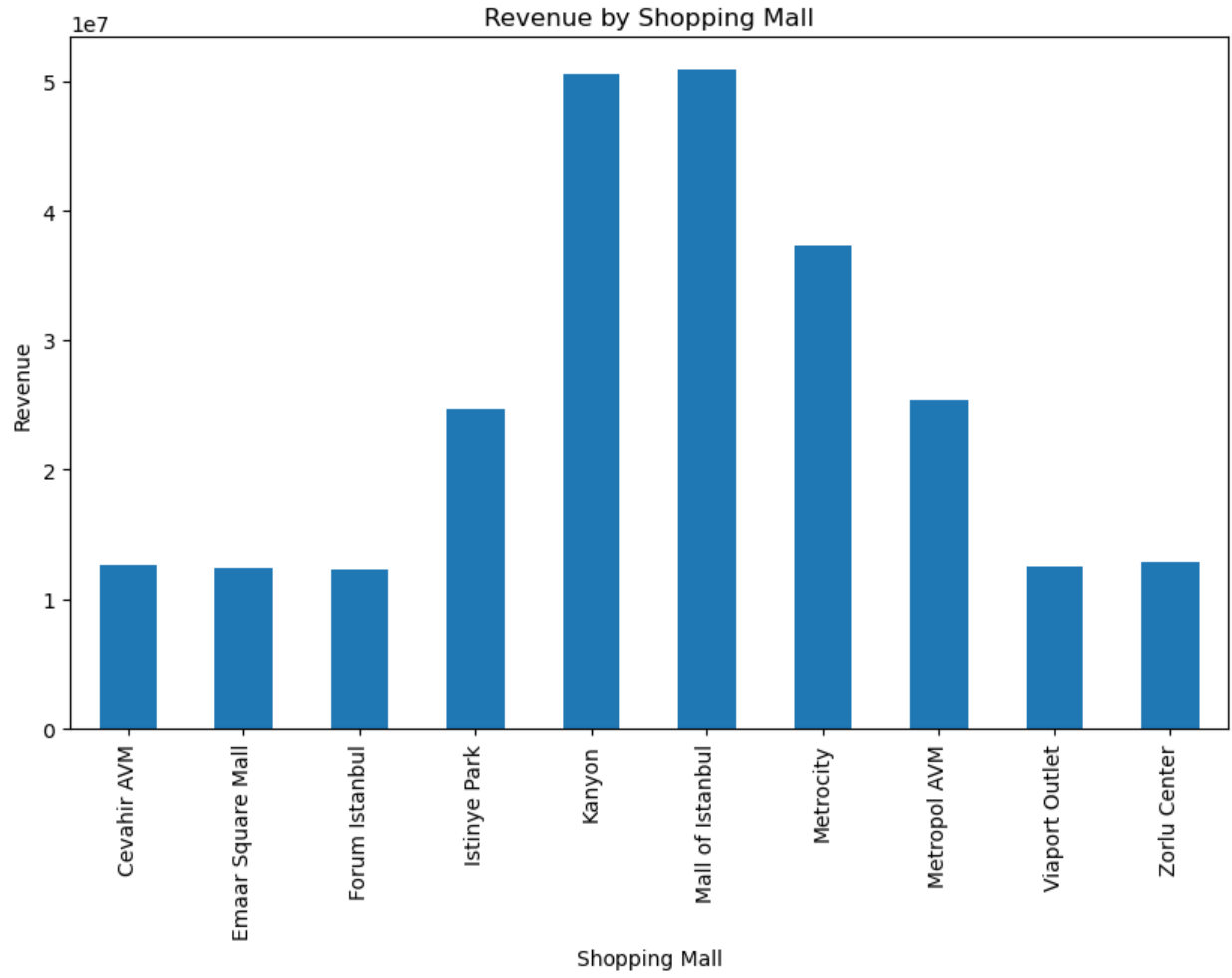


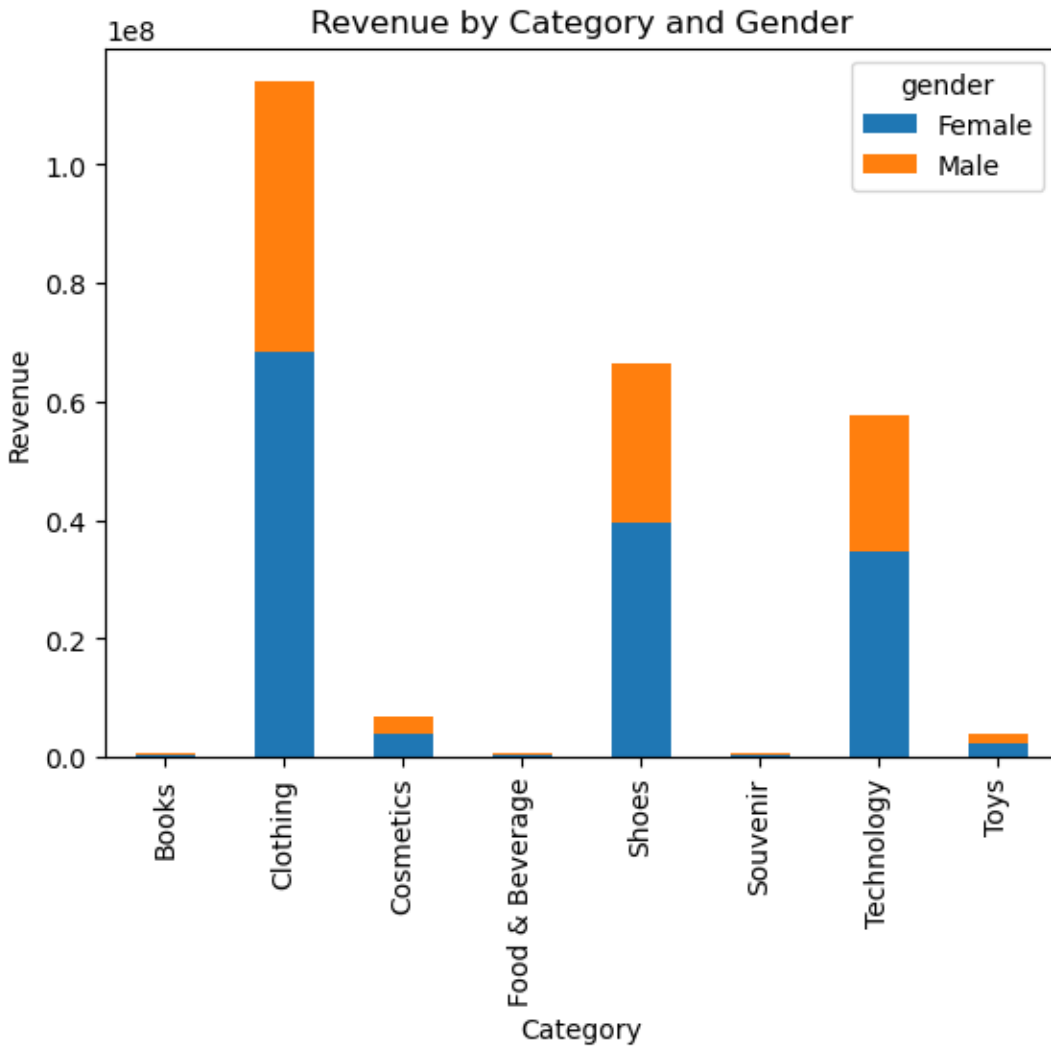
- **Category-wise Analysis:** The top-selling categories are **Clothing, Cosmetics, Food & Beverage, and Toys**. Technology had the widest price distribution.



3.2 Revenue Analysis

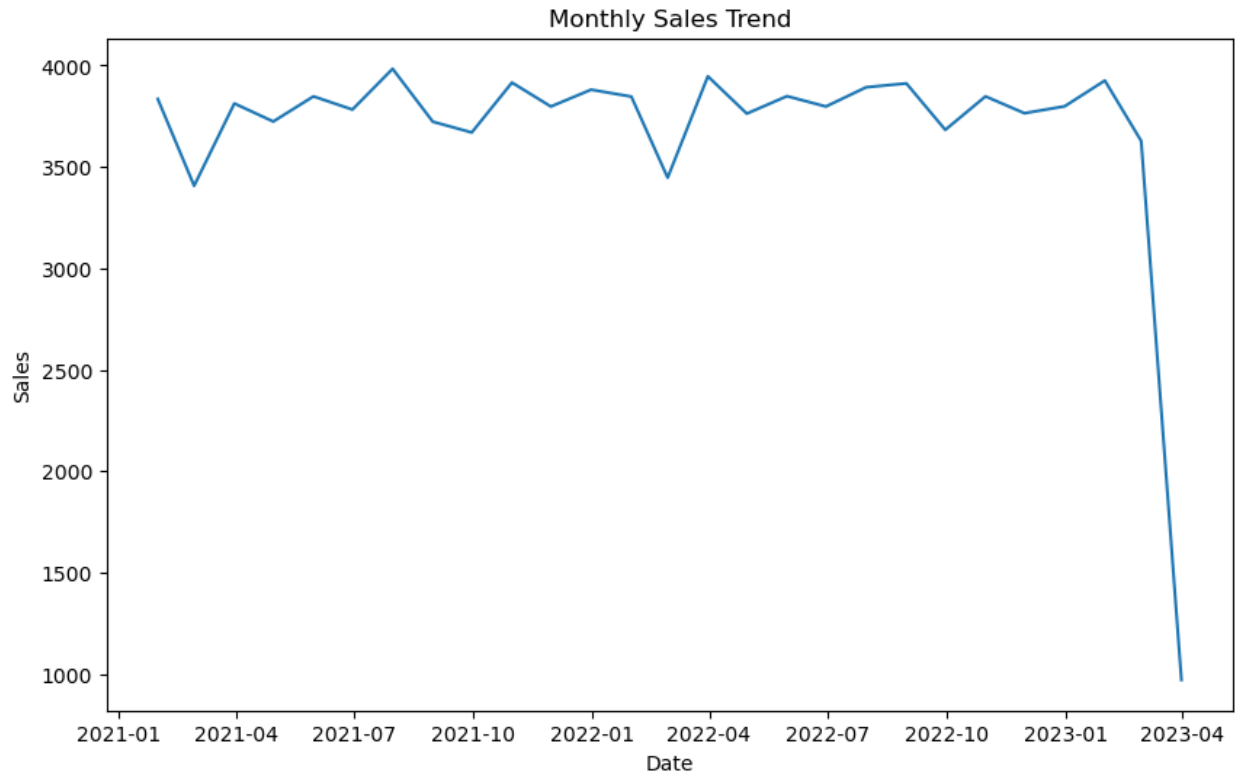
The analysis revealed the relationship between revenue and factors such as category, gender, and payment methods. Clothing, Shoes, and technology generated the most revenue, with males contributing more revenue in those categories than females. Higher quantity indicated higher revenue.





3.3 Monthly Trends

45,551 invoices were recorded in 2022, a slight increase from 45,382 in 2021. The first quarter of the year is the highest selling period, with the highest selling months being January, February, and March. The first week of the months are the best-selling days with Monday, Friday, and Tuesday leading the list.



4. Conclusion

The analysis provided valuable insights into Istanbul's shopping habits and trends. Most customers are female, aged 37, pay in cash, and frequent the Mall of Istanbul. The top-selling categories are Clothing, Cosmetics, Food & Beverage, and Toys, with Clothing generating the highest revenue.

5. Recommendations

Based on the analysis, the following is recommended:

1. Investigate the relationship between product categories and customer preferences.
2. Conduct market research to understand customer preferences and trends in different shopping malls.
3. Implement targeted marketing campaigns for specific customer segments.
4. Explore the impact of pricing strategies on customer purchasing behavior within different product categories.

6. Acknowledgments

We would like to express our gratitude to Mehmet Tahir Aslan for providing the "Customer Shopping Data" dataset on Kaggle.

7. References

The analysis used the following libraries and tools:

- NumPy: A library for numerical computing in Python.
- Pandas: A data manipulation and analysis library.
- Matplotlib: A plotting library.
- Seaborn: A data visualization library.
- Jupyter Notebook: An interactive coding environment.

For a detailed understanding of the analysis process, including the Jupyter Notebook code used, please refer to the original analysis documentation.

Best regards,

Nikhil Sharma