

Customer Shopping

Exploratory Data Analysis
using Python



by Chyntia Novrianti Kudadiri

01

Understanding
Data

02

Bussiness
Problem

03

Data
Preparation

04

Data
Visualization

05

Recommendation
for Business

Understanding Data

Costumer Shopping Dataset is a dataset containing shopping information in Istanbul.

This dataset comes from 10 different shopping malls starting from Jan 1, 2021 to March 8, 2023.

The dataset can be downloaded on [Kaggle](#)



Attribute Information:


- **invoice_no**: Invoice number.
- **customer_id**: Customer number.
- **gender**: customer's gender.
- **age**: customer's age.
- **category**: category of the purchased product.
- **quantity**: The quantities per transaction.
- **price**: Product price per unit in Turkish Liras (TL).
- **payment_method**: payment method used for the transaction.
- **invoice_date**: The day when a transaction was generated.
- **shopping_mall**: shopping mall where the transaction was made.

Business Problem

Shopping is one of the factors that contribute to economic growth. Shopping can increase a country's income. Therefore, by analyzing customer behavior using Customer Shopping Dataset, companies can determine business strategies to attract customers to shop.

In this data set, I'll explore some of the following through visualization :

- What month is the most and the least total spending?
- Which gender spend more money?
- What payment method is the most used by customers?
- Which category the customer spend most money?

- 
- Which shopping mall is most visited by customers?
 - Distribution by customer's age
 - Distribution by price
 - In 2023, Which shopping mall is most visited by customers?
 - In 2023, Which Shopping mall has the highest total revenue?
 - In 2023, What day in a week have the most transaction?
 - In 2023, What day in a week have the most total spending?
 - In 2023, Which category the customer spend most money?

Data Preparation

This project use Google Colab. If you want to see the whole project, you can find it [here](#).

I did data preparation before explore the data :

1. Add column `total_price` = `quantity * price`
2. Add column `invoice_month` (month of `invoice_date`)
3. Change type of `invoice_date` and `invoice_month` into `datetime`
4. Add column `invoice_date` (day of week of `invoice_date`)
5. Add column `age_group` (range of age)
6. Add column `price_group` (range of price)

Data Visualization

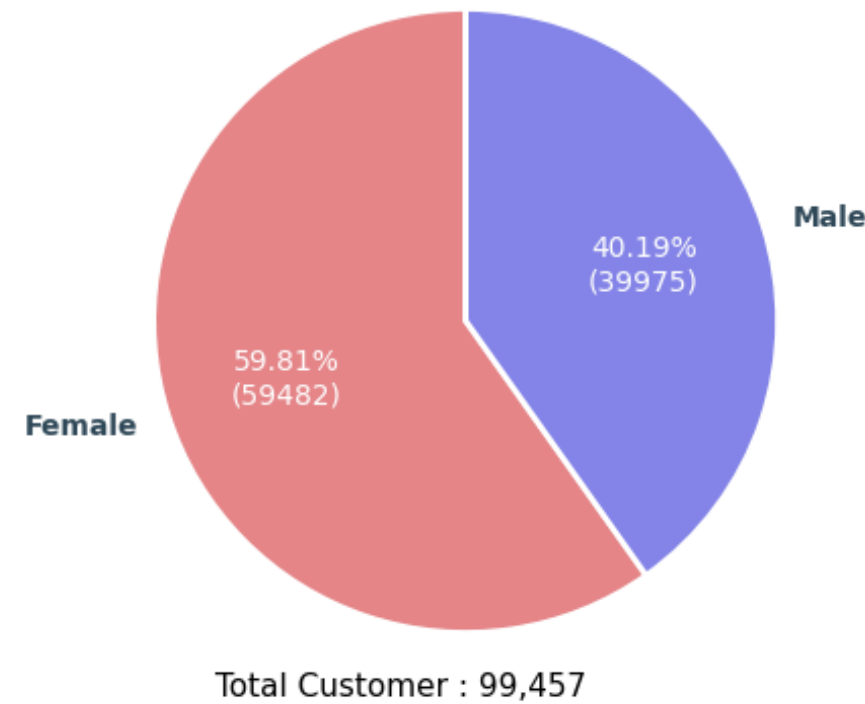
Total Spending per Month



Customers spent around **10.31** million in **July 2021** which is the **highest** total spending and the **lowest** total spending in **February 2022** around **8.34** million.

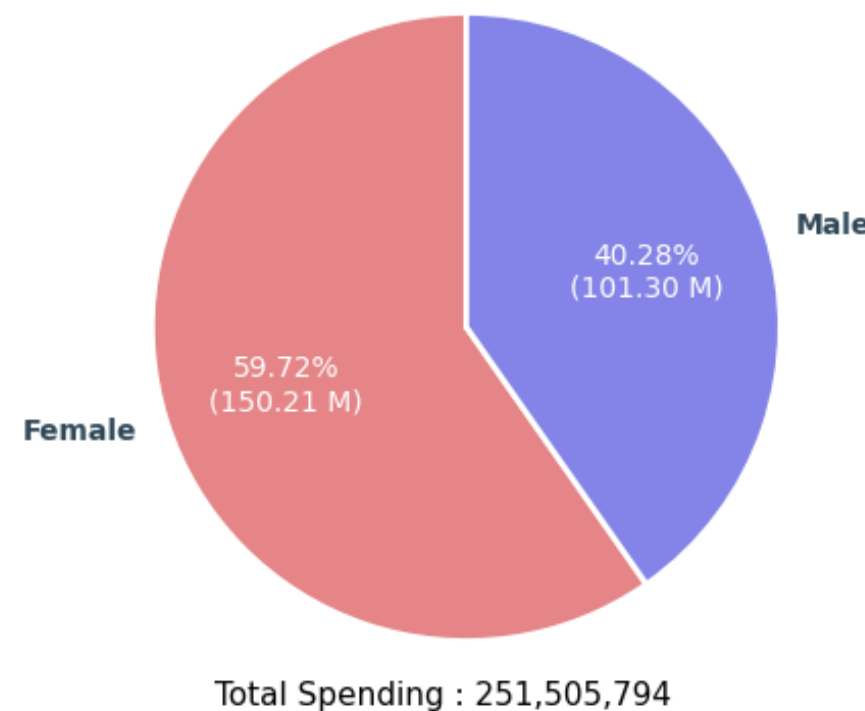
Number of Customers and Total Spending by Gender

Percentage Total Customer per Gender

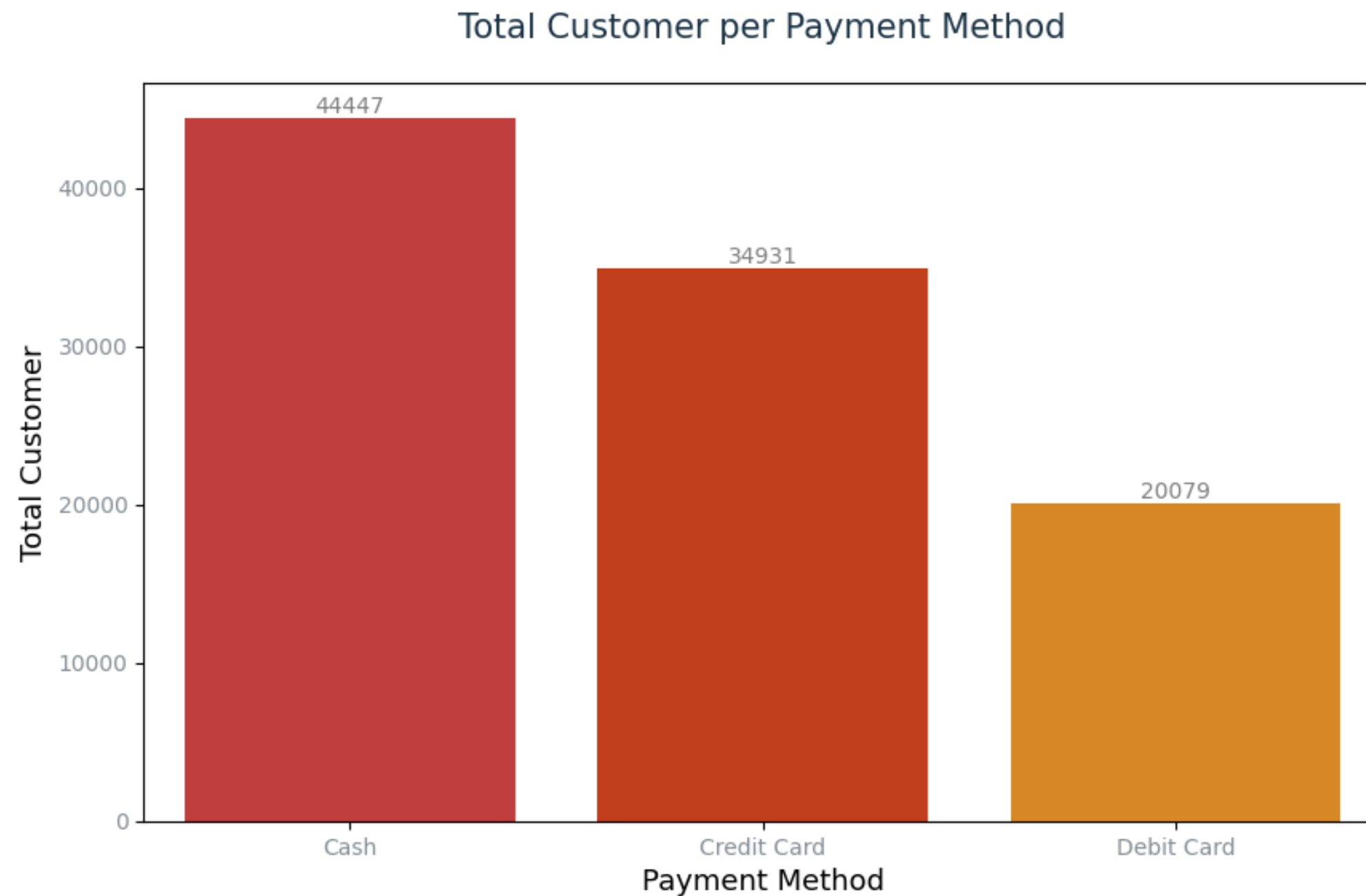


From Jan 2021 to March 2023, the number of customers is **99,457** customers, with total spending of **251,505,794 TL** (Turkish Liras).

Percentage Total Spending per Gender



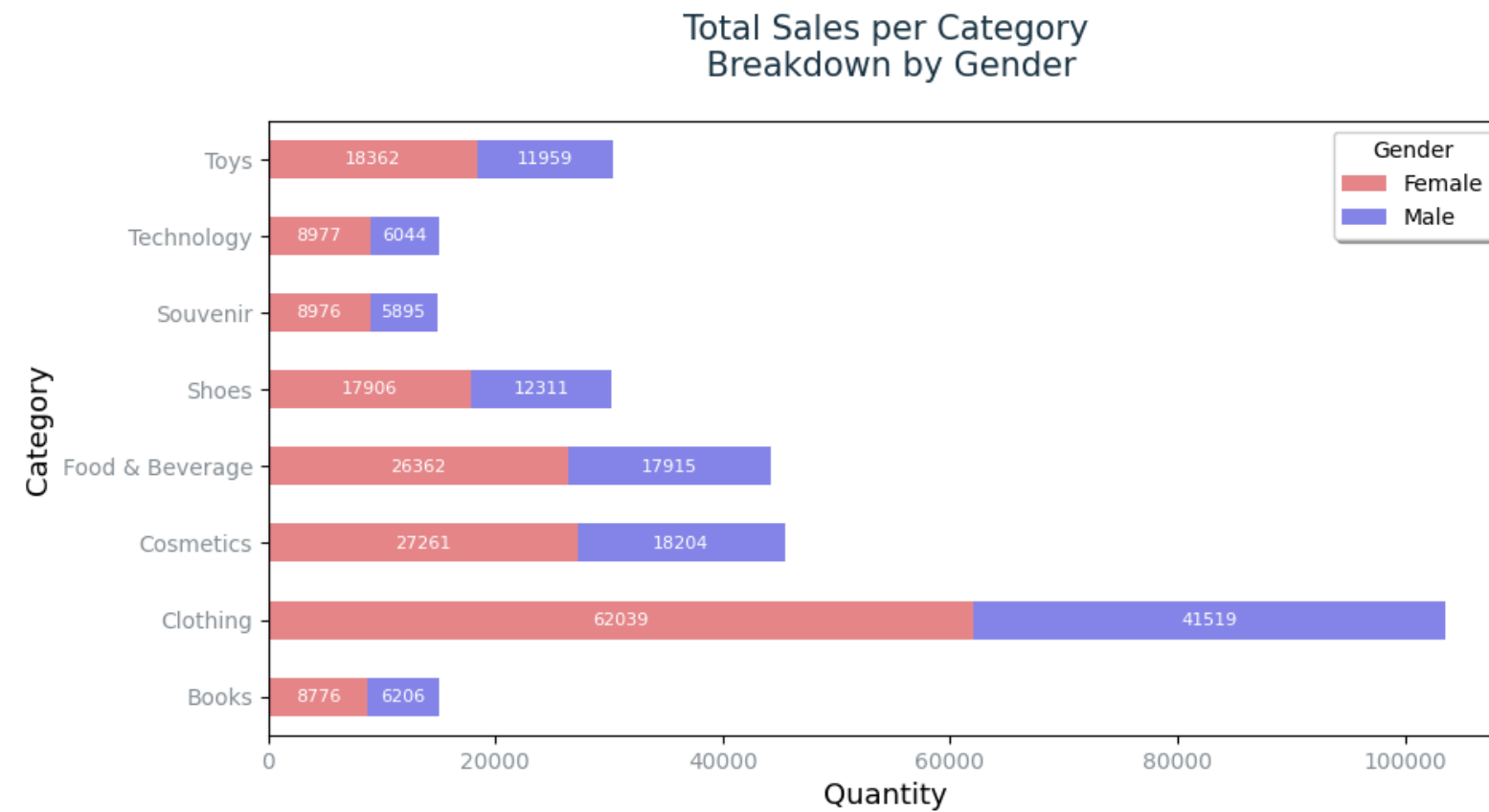
Among of them are dominated by female around **59.81%** or **59482** customers with total spending of around **150.21** million TL. Male customers are around **40.19%** or **39,975** customers with total spending of around **101.30** million TL.



44,447

of the total customers paid by Cash, the most payment method used.

Payment method by Credit card was used by 34,931 customer and Debit Card was the least used by customers.

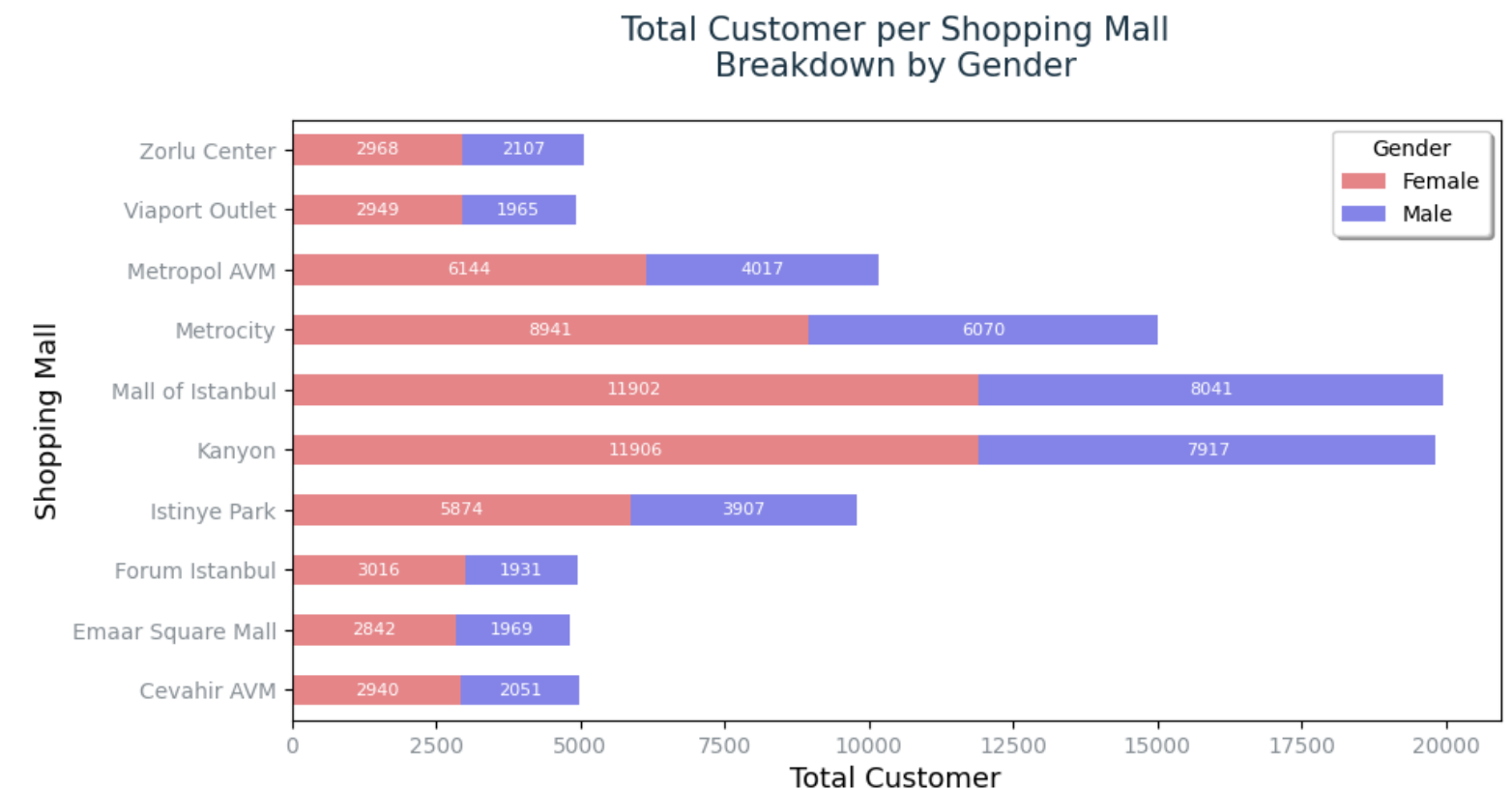


Top Category

The most purchased by customers is Clothing and the least purchased is Souvenir.

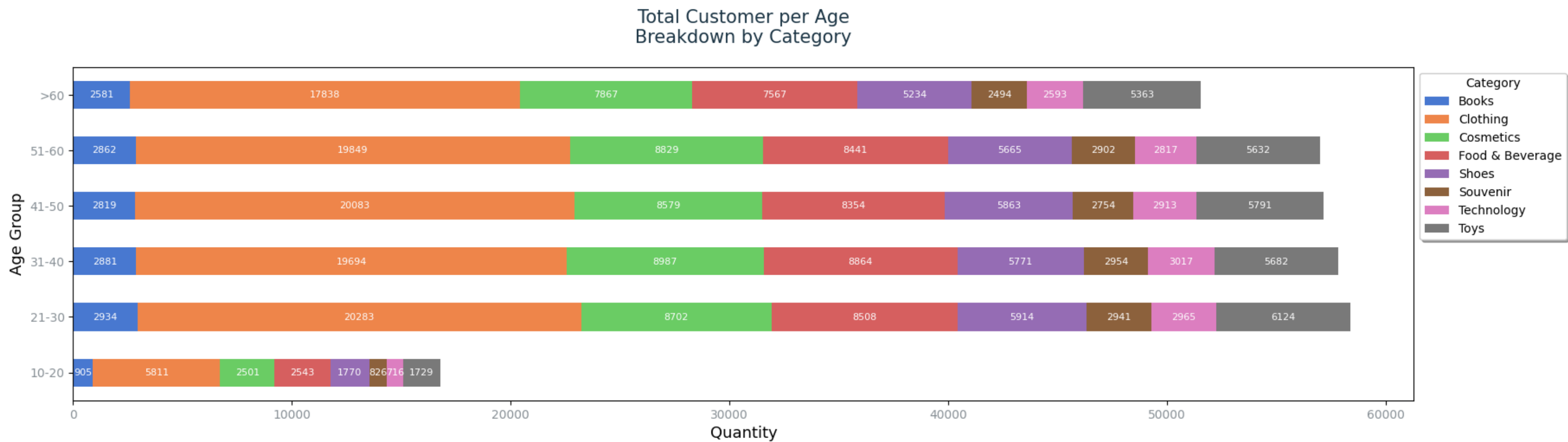
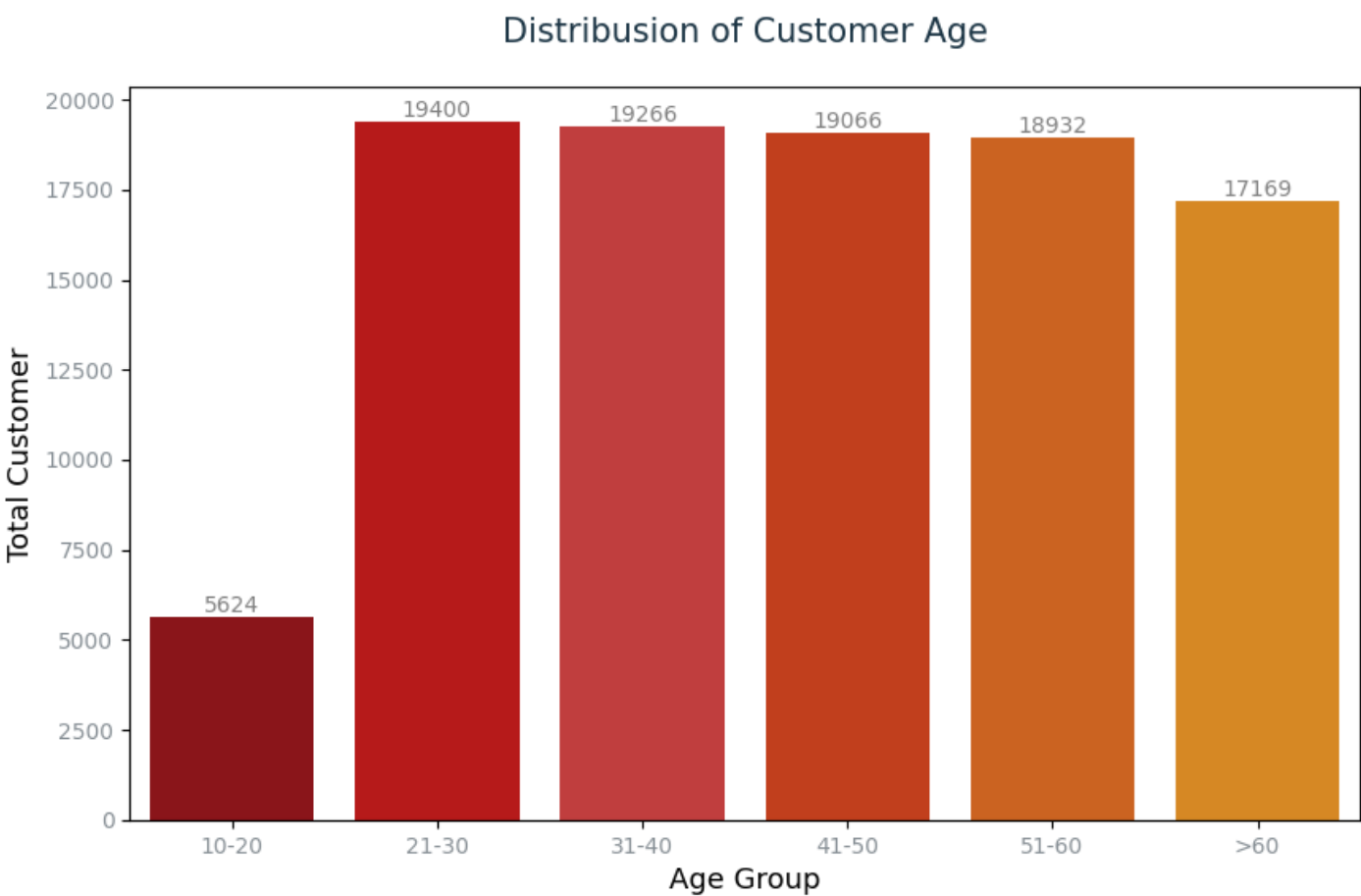
Top Shopping Mall

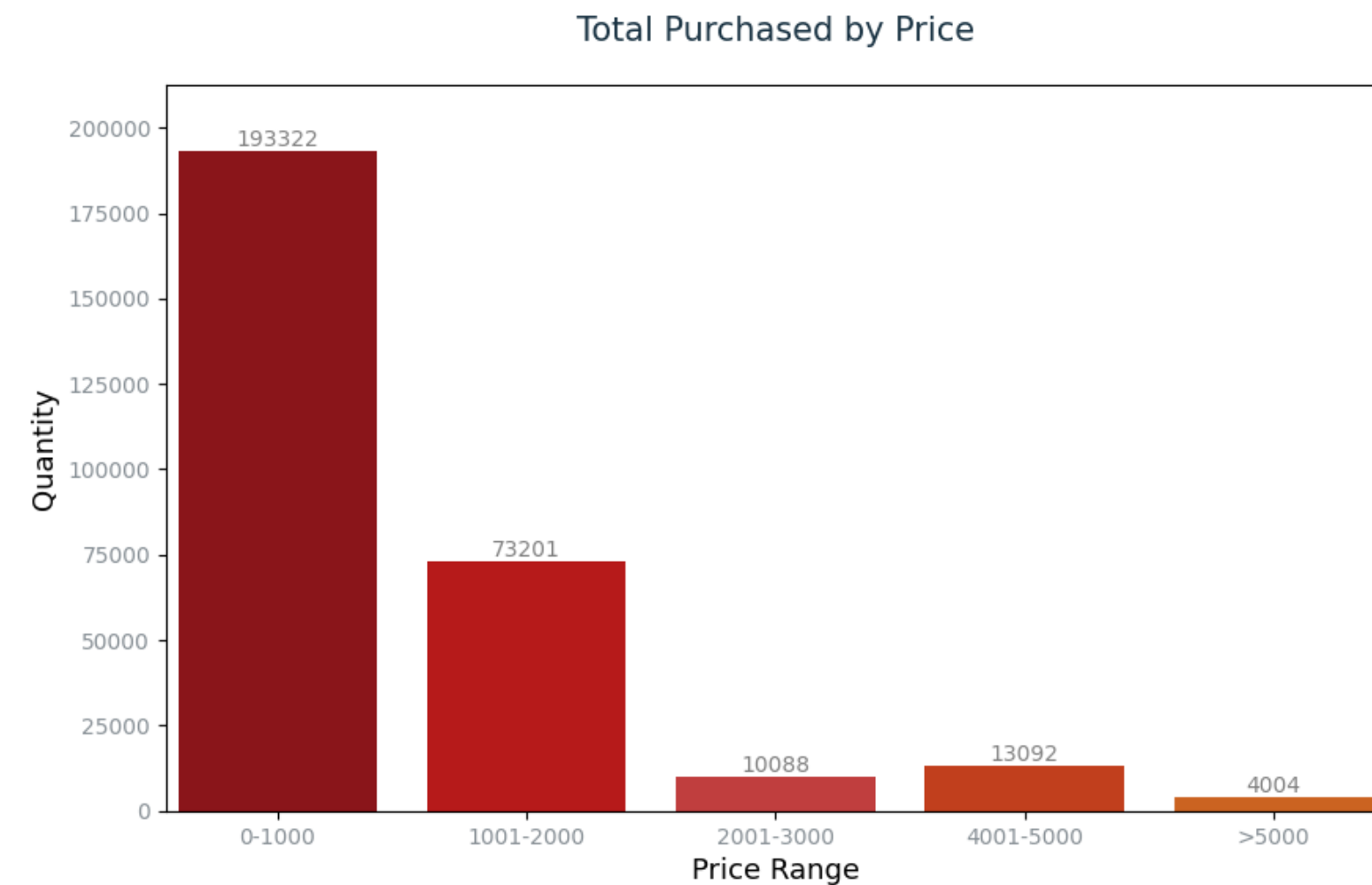
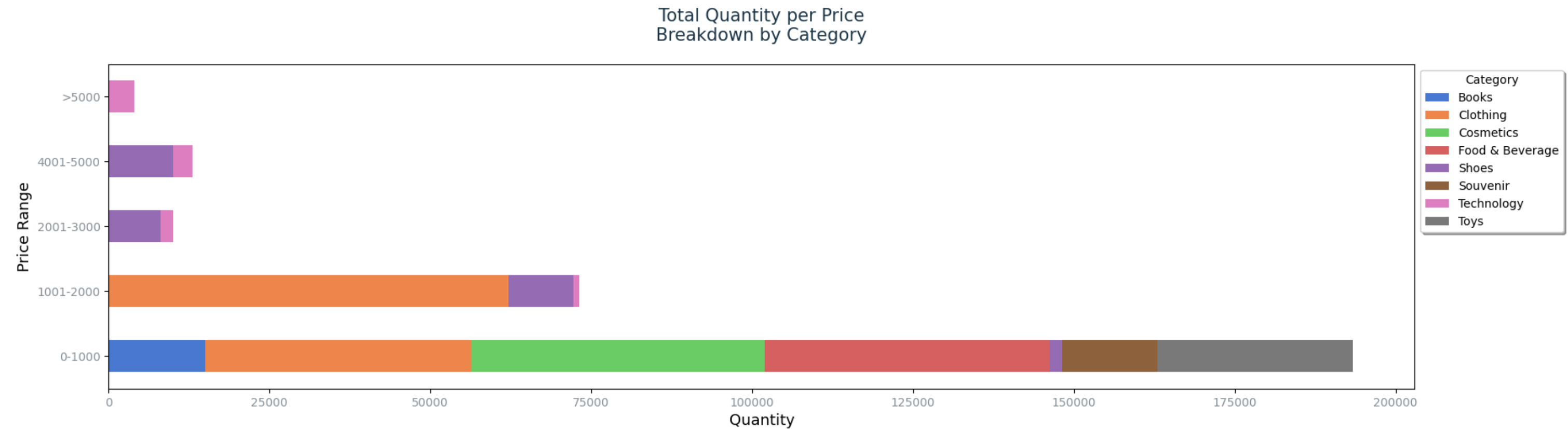
Most visited by customer is Mall of Istanbul and the least visited is Emaar Square Mall.



Customer shopping is dominated by customers aged 21-30 years or teenagers. Customers aged 31-40 are in second place after teenagers which is only about 200 fewer.

The category most purchased by every age range is clothing

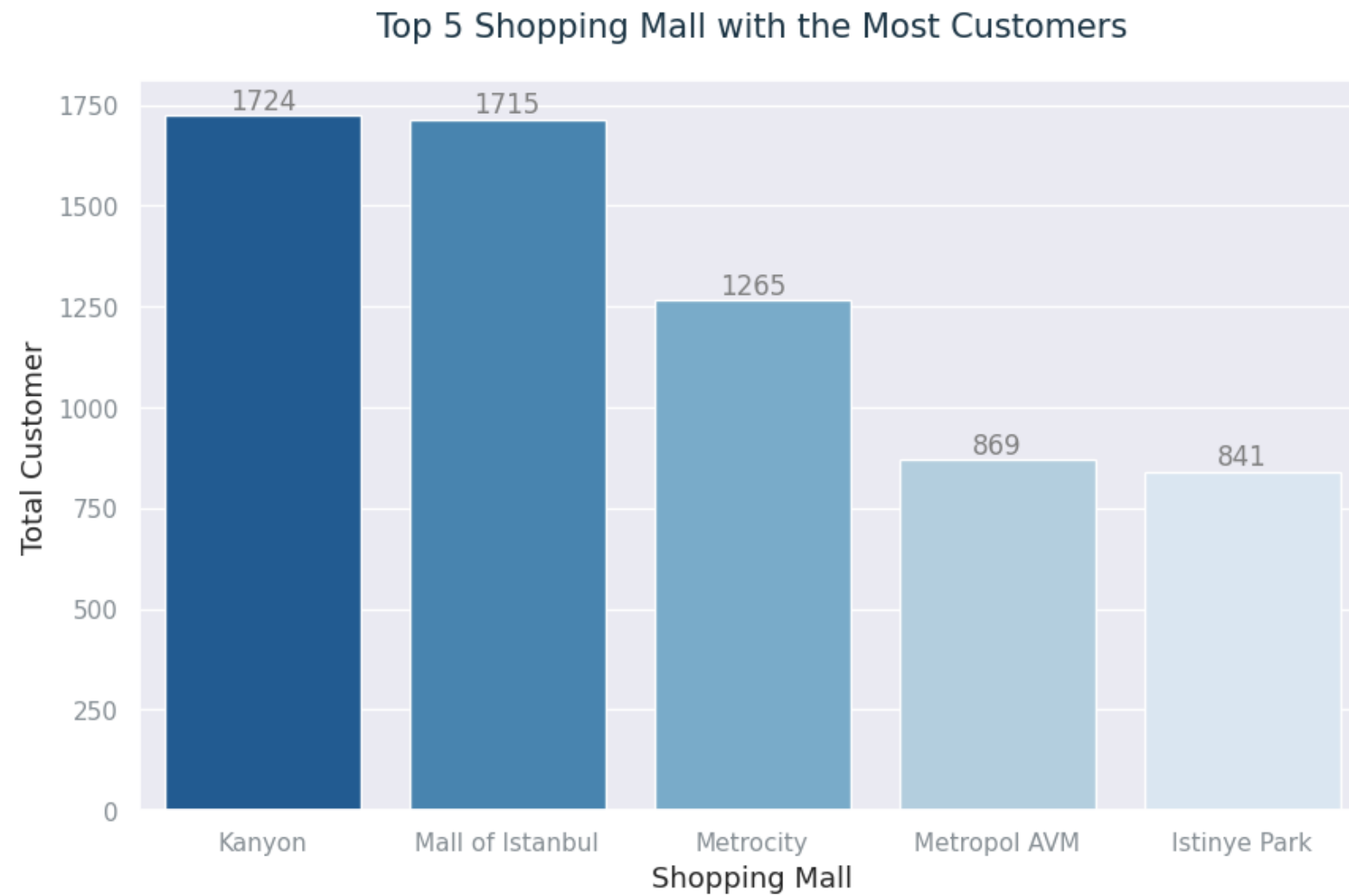




Customers buy more categories with a price range of 0-1000 TL and the most purchased category is clothing

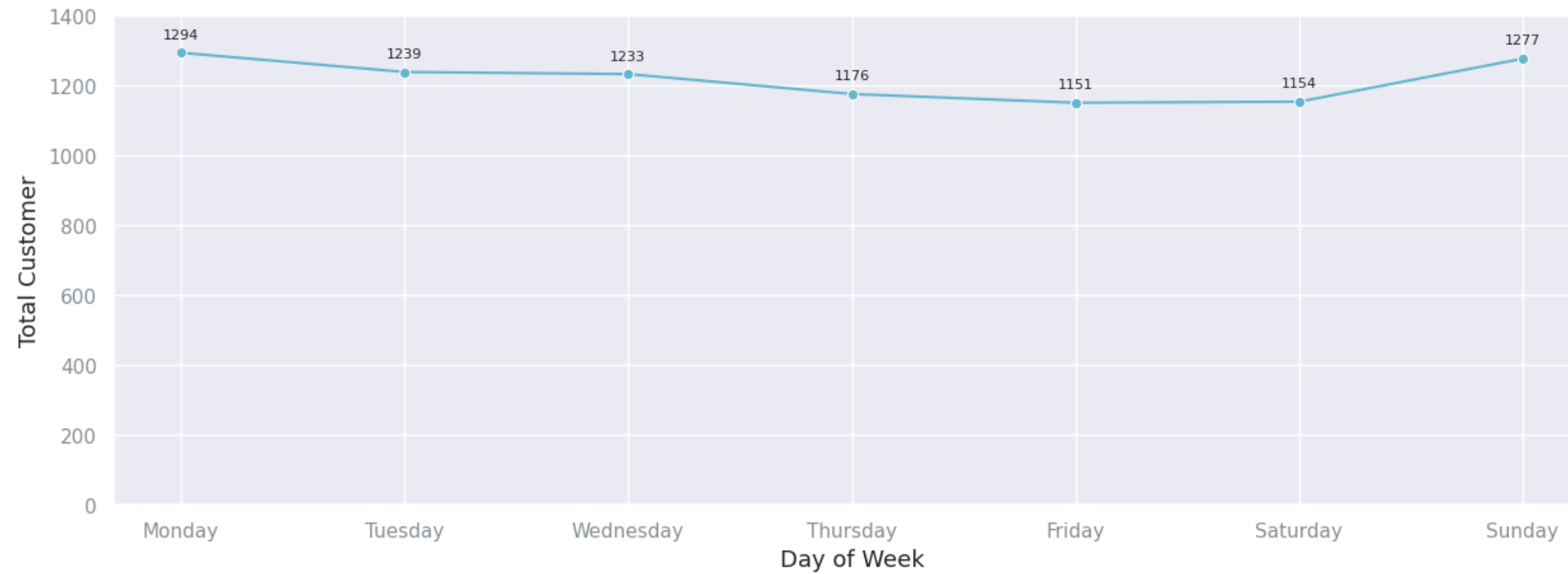
Customer Shopping Analysis in 2023





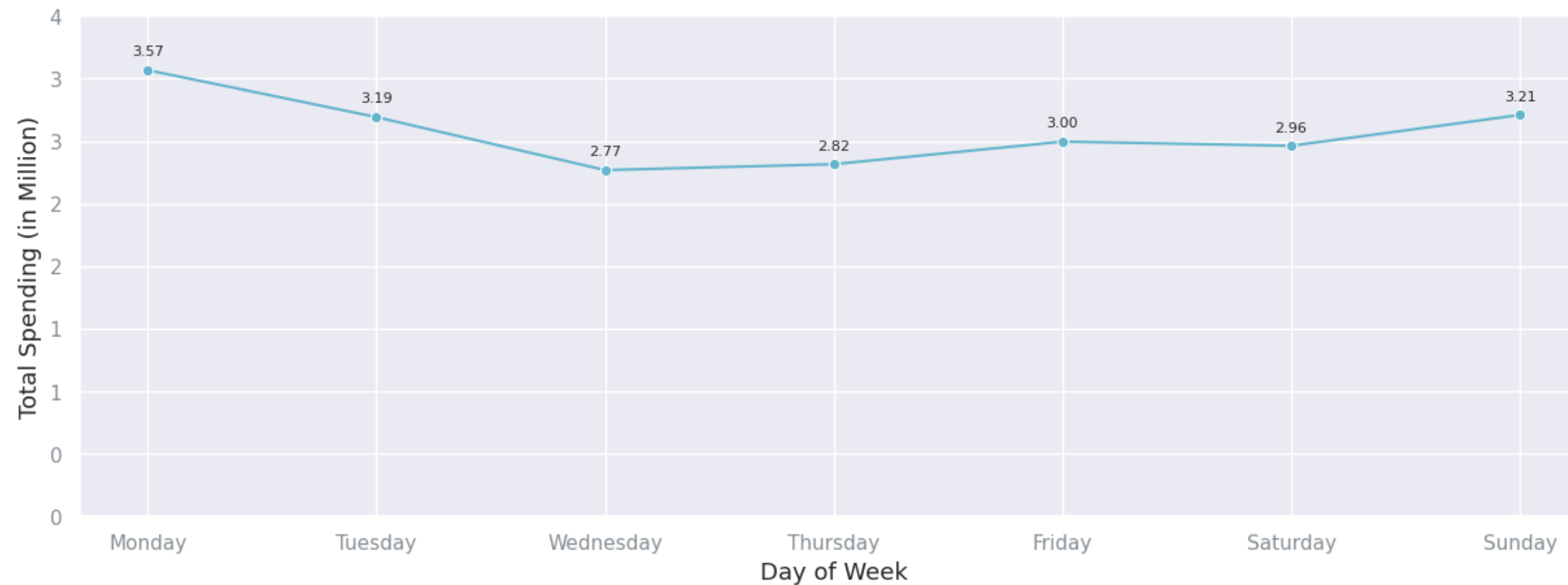
Charts show 5 shopping malls with the most of customers and the highest total revenue. In 2023, Kanyon is the mall with the most number of customers with a total revenue of more than 4 million TL.

Total Customer by Weekday - 2023



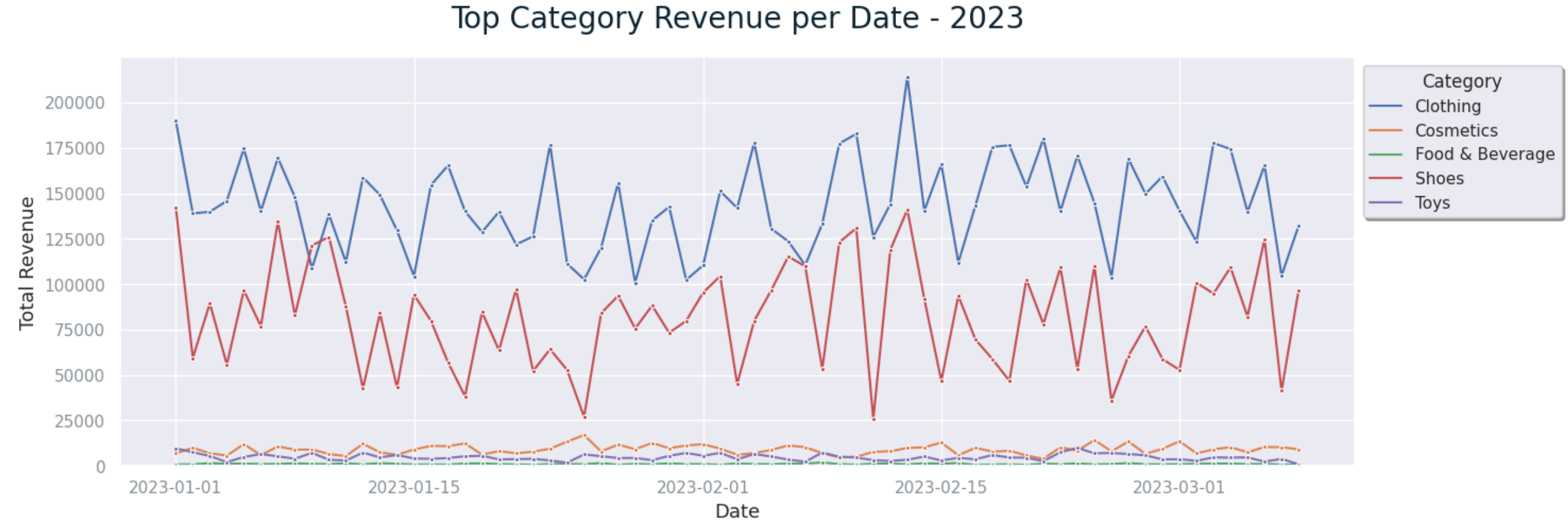
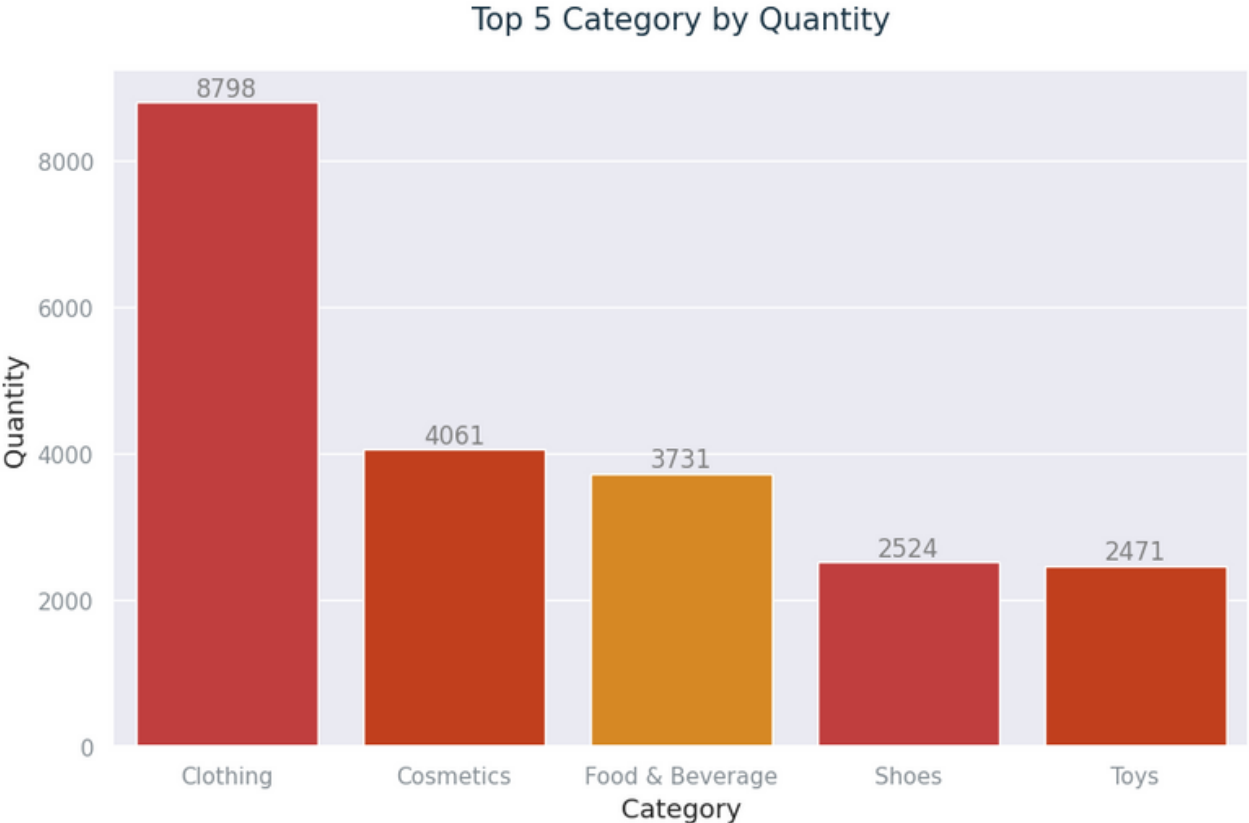
In 2023, Customers do more transactions and spend money to shop on **Monday**.

Total Spending by Weekday - 2023



On Monday, there were 1294 transactions by customers with a total spending of 3.57 million TL.

Category with the most total purchased and the highest revenue in 2023.



Recommendation for Business

Based on the visualization that has been done, there are several business strategy recommendations that can be used by companies/malls to increase sales.

Customers are dominated by female customers with an age range of 21-30 years old. Companies can attract the attention of these customers by adding the types of products that they usually buy.

Category with the highest price that is bought the most is technology. Companies can provide discounts or guarantees for every purchase of technology products to increase sales and maintain customer trust.

Thank you

Let's Work Together