EXPLORING SHOPPING HABITS FROM 10 MALLS IN ISTANBUL

INTRODUCTION

In today's world, technology has drastically changed the way customers shop, making it imperative for businesses to adapt to the changing trends. The purpose of this analysis is to shed light on the

shopping behavior and habits of customers in Istanbul, which can help retail businesses thrive in the region.

This article will delve into the data collection process, analysis methods, and project findings. So, let's explore the world of data analysis and gain insights into the shopping habits of Istanbul's customers.

THE DATASET

This report is based on a vast dataset obtained from ten shopping malls in Istanbul, collected between 2021 and 2023. The dataset includes over 99,000 transactions, which offer valuable insights into consumer behavior in this bustling city. Each transaction is detailed with important information such as invoice numbers, customer IDs, age, gender, payment methods, product categories, quantity, price, order dates, and shopping mall locations. These details enable a thorough analysis of consumer preferences and behaviors.

The Data Collection

This dataset was carefully curated from Kaggle, providing valuable information for analysis. It can be assessed *HERE*!

Problem Statement

In order for businesses to remain competitive in the constantly evolving retail industry, they must stay ahead of the curve. Understanding customer behavior and spending habits is crucial for determining a business's success. Without this knowledge, they may struggle to optimize marketing strategies, product offerings, and store layouts, which can hinder their growth and profitability.

To address this issue, this analysis aims to explore customer demographics, popular product categories, payment methods, shopping mall preferences, seasonal spending trends, and profitability by product/category. By doing so, retailers can identify opportunities for growth and improvement.

The insights gained from this analysis can help retailers identify the most effective marketing strategies, product offerings, and store layouts to increase customer engagement and maximize profitability. This can lead to sustainable growth in the long run.

Attribute Information:

- **Invoice_no:** Invoice number. Nominal. A combination of the letter ‘I’ and a 6-digit integer is uniquely assigned to each operation.
- **Customer_id:** Customer number. Nominal. A combination of the letter ‘C’ and a 6-digit integer is uniquely assigned to each operation.
- **Gender:** String variable of the customer’s gender.
- **Age:** Positive Integer variable of the customers' age.

- **Category:** String variable of the category of the purchased product.
- **Quantity:** The quantities of each product (item) per transaction. Numeric.
- **Price:** Unit price. Numeric. Product price per unit in Turkish Liras (TL).
- **Payment_method:** String variable of the payment method (cash, credit card, or debit card) used for the transaction.
- **Invoice_date:** Invoice date. The day when a transaction was generated.
- **Shopping_mall:** String variable of the name of the shopping mall where the transaction was made.

DATA CLEANING

Prior to analyzing the dataset, I took the necessary step of thoroughly cleaning the data to ensure its accuracy and reliability. This process involved meticulously inspecting each data column for any inconsistencies, errors, or outliers that may have skewed the results.

To begin, I carefully combed through the dataset, paying close attention to any missing or null values that needed to be filled in or removed. Next, I looked for duplicate entries so I could eliminate them.

However, upon inspection, I discovered that the dataset had already been cleaned. As a result, I was confident that the dataset was ready for thorough analysis, as it was free from any errors that could have otherwise impacted the results.

Spending time to carefully clean and prepare data ensures that the insights gleaned from the analysis are accurate, reliable, and meaningful.

Exploratory Data Analysis (EDA)

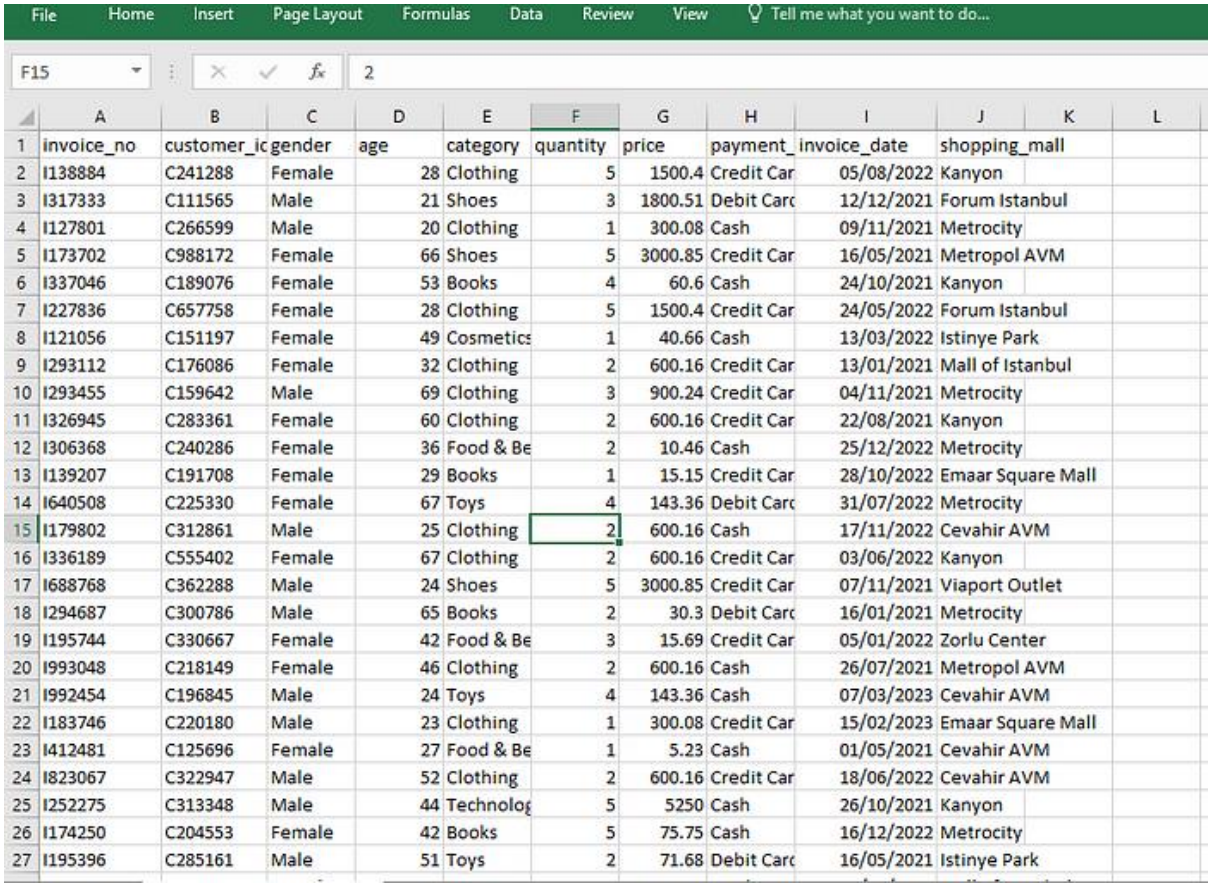
The term "Exploratory Data Analysis" (EDA) may be unfamiliar to those who are new to data analysis. Allow me to explain briefly what it means. Essentially, EDA is a crucial step in any data analysis project. It involves examining and comprehending the structure and patterns of the data in order to uncover insights. By doing so, one can gain a better understanding of the data before developing models or making business decisions based on it. In this particular case, what were the results of the EDA performed on the Istanbul shopping dataset? Keep reading to find out!

Data Exploration Tool

I used ***Microsoft Excel*** for my data exploration and analysis.

METHODOLOGY

To begin, I imported the dataset into Microsoft Excel. I opened a new Excel sheet and clicked on the “Data” tab from the top menu. I chose “From Text/CSV” and then selected the dataset file. After that, I clicked on “Import” to import the data into Excel.



	A	B	C	D	E	F	G	H	I	J	K	L
1	invoice_no	customer_ic	gender	age	category	quantity	price	payment_	invoice_date	shopping_mall		
2	I138884	C241288	Female	28	Clothing	5	1500.4	Credit Car	05/08/2022	Kanyon		
3	I317333	C111565	Male	21	Shoes	3	1800.51	Debit Carc	12/12/2021	Forum Istanbul		
4	I127801	C266599	Male	20	Clothing	1	300.08	Cash	09/11/2021	Metrocity		
5	I173702	C988172	Female	66	Shoes	5	3000.85	Credit Car	16/05/2021	Metropol AVM		
6	I337046	C189076	Female	53	Books	4	60.6	Cash	24/10/2021	Kanyon		
7	I227836	C657758	Female	28	Clothing	5	1500.4	Credit Car	24/05/2022	Forum Istanbul		
8	I121056	C151197	Female	49	Cosmetics	1	40.66	Cash	13/03/2022	Istinye Park		
9	I293112	C176086	Female	32	Clothing	2	600.16	Credit Car	13/01/2021	Mall of Istanbul		
10	I293455	C159642	Male	69	Clothing	3	900.24	Credit Car	04/11/2021	Metrocity		
11	I326945	C283361	Female	60	Clothing	2	600.16	Credit Car	22/08/2021	Kanyon		
12	I306368	C240286	Female	36	Food & Be	2	10.46	Cash	25/12/2022	Metrocity		
13	I139207	C191708	Female	29	Books	1	15.15	Credit Car	28/10/2022	Emaar Square Mall		
14	I640508	C225330	Female	67	Toys	4	143.36	Debit Carc	31/07/2022	Metrocity		
15	I179802	C312861	Male	25	Clothing	2	600.16	Cash	17/11/2022	Cevahir AVM		
16	I336189	C555402	Female	67	Clothing	2	600.16	Credit Car	03/06/2022	Kanyon		
17	I688768	C362288	Male	24	Shoes	5	3000.85	Credit Car	07/11/2021	Viaport Outlet		
18	I294687	C300786	Male	65	Books	2	30.3	Debit Carc	16/01/2021	Metrocity		
19	I195744	C330667	Female	42	Food & Be	3	15.69	Credit Car	05/01/2022	Zorlu Center		
20	I993048	C218149	Female	46	Clothing	2	600.16	Cash	26/07/2021	Metropol AVM		
21	I992454	C196845	Male	24	Toys	4	143.36	Cash	07/03/2023	Cevahir AVM		
22	I183746	C220180	Male	23	Clothing	1	300.08	Credit Car	15/02/2023	Emaar Square Mall		
23	I412481	C125696	Female	27	Food & Be	1	5.23	Cash	01/05/2021	Cevahir AVM		
24	I823067	C322947	Male	52	Clothing	2	600.16	Credit Car	18/06/2022	Cevahir AVM		
25	I252275	C313348	Male	44	Technolog	5	5250	Cash	26/10/2021	Kanyon		
26	I174250	C204553	Female	42	Books	5	75.75	Cash	16/12/2022	Metrocity		
27	I195396	C285161	Male	51	Toys	2	71.68	Debit Carc	16/05/2021	Istinye Park		

The imported data

After collecting the necessary data, I converted it into a table format to present it in an organized and structured manner. By using rows and columns, each row represents a single record or observation, while each column represents a variable or attribute of that record. This makes it easier to analyze and interpret the data.

A	B	C	D	E	F	G	H	I	J	K
Invoice_No	Customer_Id	Gender	Age	Category	Quantit	Price	Payment_Method	Invoice_Date	Shopping_Mall	Column
I138884	C241288	Female	28	Clothing	5	1500.4	Credit Card	5/8/2022	Kanyon	
I317333	C111565	Male	21	Shoes	3	1800.51	Debit Card	12/12/2021	Forum Istanbul	
I127801	C266599	Male	20	Clothing	1	300.08	Cash	9/11/2021	Metrocity	
I173702	C988172	Female	66	Shoes	5	3000.85	Credit Card	16/05/2021	Metropol AVM	
I337046	C189076	Female	53	Books	4	60.6	Cash	24/10/2021	Kanyon	
I227836	C657758	Female	28	Clothing	5	1500.4	Credit Card	24/05/2022	Forum Istanbul	
I121056	C151197	Female	49	Cosmetics	1	40.66	Cash	13/03/2022	Istinye Park	
I293112	C176086	Female	32	Clothing	2	600.16	Credit Card	13/01/2021	Mall of Istanbul	
I293455	C159642	Male	69	Clothing	3	900.24	Credit Card	4/11/2021	Metrocity	
I326945	C283361	Female	60	Clothing	2	600.16	Credit Card	22/08/2021	Kanyon	
I306368	C240286	Female	36	Food & Beverage	2	10.46	Cash	25/12/2022	Metrocity	
I139207	C191708	Female	29	Books	1	15.15	Credit Card	28/10/2022	Emaar Square Mall	
I640508	C225330	Female	67	Toys	4	143.36	Debit Card	31/07/2022	Metrocity	
I179802	C312861	Male	25	Clothing	2	600.16	Cash	17/11/2022	Cevahir AVM	
I336189	C555402	Female	67	Clothing	2	600.16	Credit Card	3/6/2022	Kanyon	
I688768	C362288	Male	24	Shoes	5	3000.85	Credit Card	7/11/2021	Viaport Outlet	
I294687	C300786	Male	65	Books	2	30.3	Debit Card	16/01/2021	Metrocity	
I195744	C330667	Female	42	Food & Beverage	3	15.69	Credit Card	5/1/2022	Zorlu Center	
I993048	C218149	Female	46	Clothing	2	600.16	Cash	26/07/2021	Metropol AVM	
I992454	C196845	Male	24	Toys	4	143.36	Cash	7/3/2023	Cevahir AVM	
I183746	C220180	Male	23	Clothing	1	300.08	Credit Card	15/02/2023	Emaar Square Mall	

Format Data into Tables

Tables are useful for presenting large amounts of data in a compact and easy-to-read format, and they also allow for easy exploration and analysis of the data.

Summary Statistics

In order to calculate summary statistics for the numerical variables in the dataset, I utilized Excel's in-built functions such as AVERAGE, MAX, MIN, MEAN, MEDIAN, and STANDARD DEVIATION. Additionally, I used the percentile function = Percentile (range of cell,25) to determine the 25th percentile of the range of cells. Finally, I rounded up the age distribution to whole numbers.

A	B	C	D
	Age	Quantity	Price
Mean	43	3.00	689.26
Standard Deviation	15	1.41	941.18
Maximum	69	5	5250
Min	18	1	5.23
25th percentile	30	2	45.45
50th	43	3	203.3
75th	56	4	1200.32

Summary Statistics

Age

The analysis of the age distribution of customers showed that the average age was 43 years, ranging from 18 to 69 years. The moderate standard deviation of 15 indicated that there was considerable variability in the ages of the customers. This means that the ages were not uniformly distributed around the mean age of 43 years, but rather were spread out over a moderate range of values, with some customers being much younger or older than the average age.

Furthermore, the 25th percentile of 30 indicates that 25% of the customers were 30 years old or younger, while 75% were older than 30. Similarly, the 50th percentile, also known as the median, of 43 shows that half of the customers were younger than 43 years, and half were older than 43.

Finally, the 75th percentile value of 56 implies that 75% of the customers were 56 years old or younger, while 25% of the customers were older than 56 years. These percentiles provide information on how the dataset is distributed by dividing it into equal parts and identifying the values that correspond to each part.

Quantity

After examining the quantity variable, I found out that the average number of items bought per transaction was 3.00, with a minimum and maximum of 1 and 5 respectively. The standard deviation of 1.41 indicates some variability in the quantity purchased across the transactions. The 25th percentile, which was 2, shows that 25% of the transactions involved the purchase of 2 or fewer items.

Furthermore, the median or 50th percentile of 3 implies that half of the transactions involved the purchase of 3 or fewer items while half of them involved the purchase of 4 or more items. Finally, the 75th percentile value of 4 suggests that 75% of the transactions involved the purchase of 4 or fewer items.

Price

After analyzing the price variable, I noticed that the average price per item sold was 689.26 Turkish Lira. The prices ranged from a minimum of 5.23 TL to a maximum of 5250 TL. The standard deviation value of 941.18 TL indicated a wide range of prices for the products sold.

The 25th percentile value showed that 25% of the products sold were priced at 45.45 TL or lower. The median, or 50th percentile value, indicated that half of the products sold were priced at 203.3 TL or lower, while the other half were priced higher. Finally, the 75th percentile value revealed that 75% of the products sold were priced at 1200.32 TL or lower.

Pivot Table

In this case, data related to product categories, shopping mall locations, payment methods, and gender were analyzed using pivot tables.

Row Labels	Average of Price	Count of Invoice_No
Books	45.57	5.01%
Clothing	901.08	34.68%
Cosmetics	122.45	15.18%
Food & Beverage	15.67	14.86%
Shoes	1807.39	10.09%
Souvenir	34.89	5.03%
Technology	3156.94	5.02%
Toys	107.73	10.14%
Grand Total	**689.26**	**100.00%**

Pivot Table for Product Category

From the image above, the following can be deduced:

- Technology is the most expensive product category with an average price of 3156.94 TL and an invoice count of 5.02%. This high average price is likely due to the advanced features or higher quality of the products in this category.
- The second most expensive category is Shoes, with an average price of 1807.39 TL and an invoice count of 10.9%. This suggests that the products in this category are relatively expensive, possibly due to the use of high-quality materials or advanced features.
- Clothing is the third most expensive category with an average price of 901.08 TL and a sales count of 34.68%. Based on my analysis, clothing has the highest number of invoice counts.
- Cosmetics has an average price of 122.45 TL and accounts for 15.18% of total invoices. This indicates that it is a medium-range product.
- Toys have an average price of 107.73 TL and account for 10.14% of sales, indicating a moderate number of transactions. This suggests that it is a relatively popular product category.
- Books have the lowest average price of 45.57 TL and an invoice count of 5.01%.
- Souvenirs have an average price of 34.89 TL and an invoice count of 5.03%. This indicates that it is not an expensive product, and many customers buy it.
- Finally, the food and beverage category has the lowest average price of 15.67 TL and accounts for 14.86% of total sales.

The next pivot table is based on the payment method customers use in carrying out the above transactions.

Row Labels	Count of Invoice_No	Average of Price
Cash	44.69%	690.82
Credit Card	35.12%	688.54
Debit Card	20.19%	687.03
Grand Total	**100.00%**	**689.26**

Pivot Table based on payment methods.

The data provided comprises of three different payment methods, namely cash, credit card, and debit card. The total average price for all three payment methods combined was $689.26.

Amongst the three payment methods, cash was the most popular with a majority of 44.69% of total sales. Credit cards came second with 35.12%, while debit cards were used the least, accounting for only 20.19% of total transactions.

The popularity of cash as a payment method may indicate that customers prefer the use of physical currency or they may have a preference for the privacy and security offered by using cash for transactions. Credit cards, on the other hand, may be popular due to the rewards and benefits that

customers can receive from them, as well as the ease of tracking expenses and making larger purchases. Lastly, the low popularity of debit cards may be due to a lack of incentives or benefits for customers to use them, or a preference for other payment methods.

The transactions in different Istanbul malls are summarized in the pivot table below.

Row Labels	Sum of Price
Cevahir AVM	5.01%
Emaar Square Mall	4.95%
Forum Istanbul	4.87%
Istinye Park	9.80%
Kanyon	20.00%
Mall of Istanbul	20.21%
Metrocity	14.95%
Metropol AVM	10.12%
Viaport Outlet	4.98%
Zorlu Center	5.12%
Grand Total	**100.00%**

Pivot table showing the shopping malls in Istanbul

After analyzing the data, we found that Kanyon and Mall of Istanbul were the highest sales-generating malls, contributing to 20% of the total sales each. Istinye Park and Metrocity were the second in line, contributing approximately 10% each.

On the other hand, Emaar Square Mall, Forum Istanbul, and Viaport Outlet had the lowest sales amounts, contributing less than 5% of the total sales.

It is worth noting that Kanyon and Mall of Istanbul had the highest average price per item among all the malls. This indicates that they might be targeting a more affluent market segment who are willing to pay a premium price for high-end products.

In contrast, Istinye Park and Metrocity had relatively lower average prices per item but still managed to generate significant sales. This suggests that they cater to a wider customer base with a diverse range of products and price points.

The next pivot table shows the purchase of different products made by each gender.

Sum of Price	Column Labels								
Row Labels	Books	Clothing	Cosmetics	Food & Beverage	Shoes	Souvenir	Technology	Toys	Grand Total
Female	0.19%	27.16%	1.62%	0.20%	15.68%	0.15%	13.75%	0.96%	59.71%
Male	0.14%	18.17%	1.08%	0.14%	10.78%	0.10%	9.26%	0.63%	40.29%
Grand Total	**0.33%**	**45.33%**	**2.70%**	**0.34%**	**26.46%**	**0.25%**	**23.01%**	**1.59%**	**100.00%**

Pivot Table showing the purchases of products by gender.

According to the table above, clothing was the most purchased product by females, accounting for 27.16% of their total purchases. On the other hand, souvenirs were the least purchased product by females, accounting for only 0.15% of their total purchases. Similarly, males also purchased clothing the most, accounting for 18.17% of their total purchases, while souvenirs were the least purchased product, accounting for only 0.10% of their total purchases.

The next most purchased product by both genders was shoes, with females having a higher purchase rate of 15.68% compared to males with a purchase rate of 10.78%.

In contrast, both genders had a lower purchase rate for food & beverage and cosmetics. Females had a purchase rate of 1.62% for cosmetics and 0.20% for food & beverage. Males had a purchase rate of 1.08% for cosmetics and 0.14% for food & beverage.

Moreover, technology products had a significant sales rate for both genders, with females accounting for 13.75% and males accounting for 9.26%. This was second only to clothing in terms of sales.

In summary, the sales distribution between genders was fairly balanced, with females making up 59.71% of total sales and males accounting for 40.29%.

The next pivot table shows the age group of customers.

Row Labels	Sum of Price
18-27	19.19%
28-37	19.22%
38-47	19.65%
48-57	19.04%
58-67	19.12%
68-77	3.78%
Grand Total	**100.00%**

Age group and amount spent by them

The sales distribution across various age brackets is shown in the table above. The highest proportion of sales, which amounts to 19.65%, is contributed by the age group "38-47". The age group "28-37" accounts for 19.22%, while the age group "18-27" accounts for 19.19%. Similarly, the age group "58-67" accounts for 19.12%, and the group "48-57" contributes 19.04% to the total sales. However, the age category "68-77" had a comparatively lower contribution to the total sales, at 3.78%.

Dashboard

CONCLUSION

- *Shopping Mall Sales Performance: The analysis indicates that Emaar Square Mall had the highest sales of 704.72%, while Viaport Outlet had the lowest.*
- *Product Categories:* The most popular products in these malls are clothing, shoes, and technology, while books, cosmetics, and toys are less in demand.
- *Method of Payment:* Customers primarily use cash, followed by credit and debit cards.