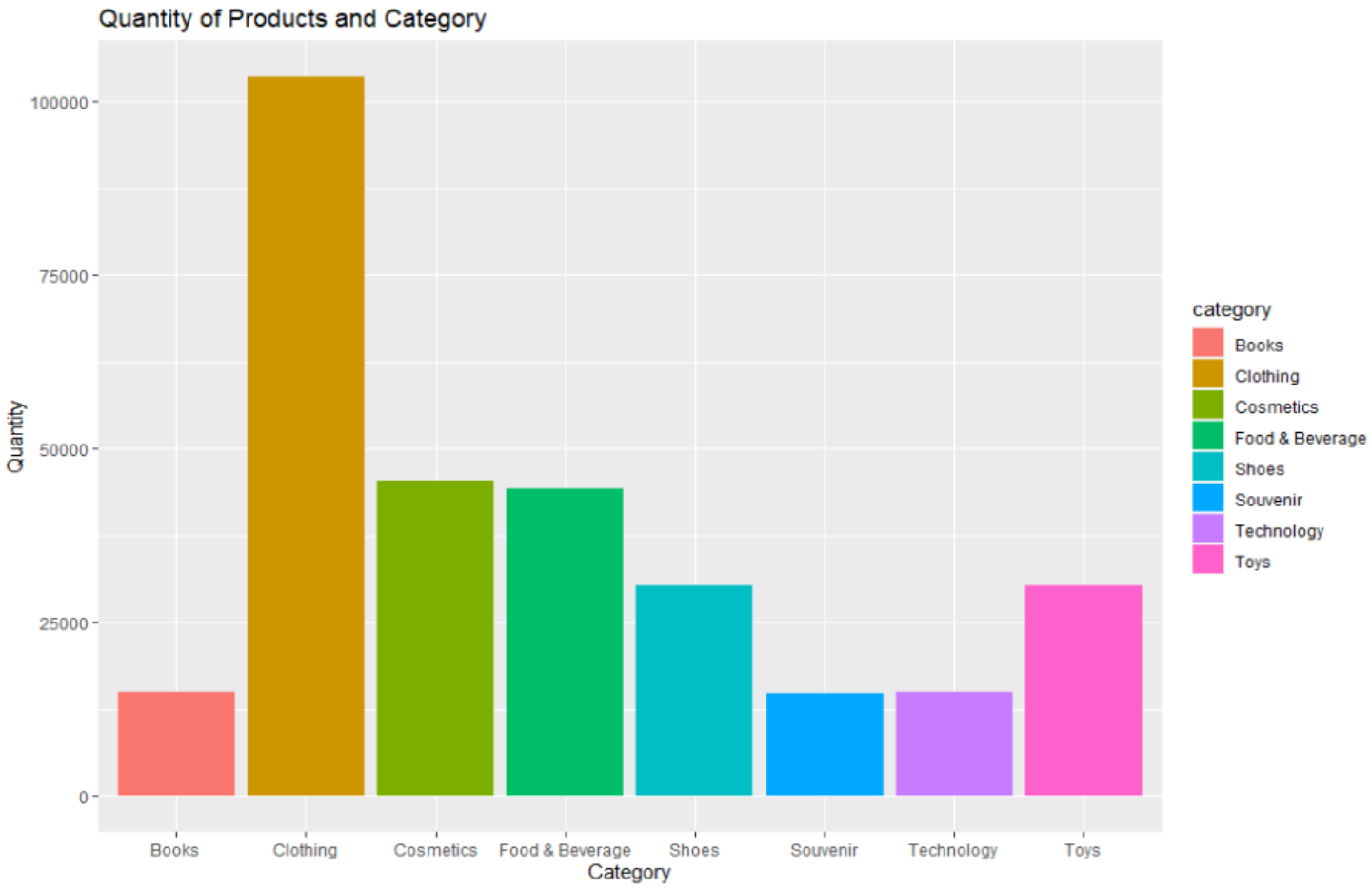
# CUSTOMER SHOPPING HABIT ANALYSIS

## 1. Which product categories are most commonly purchased?

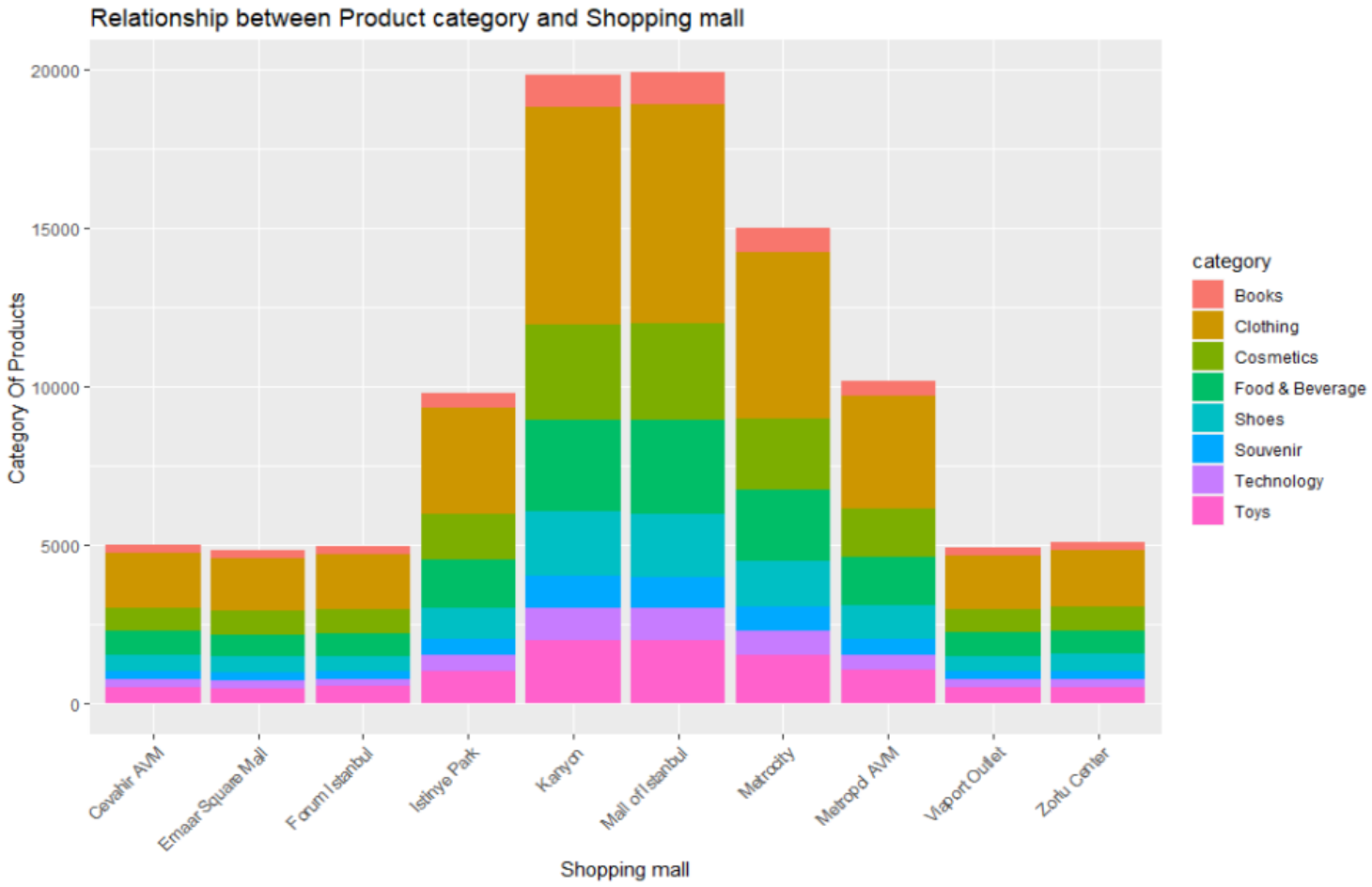
```
> library(ggplot2)
> ggplot(data=data,aes(x=category,y=quantity,fill=category))+
+ geom_bar(stat="identity")+
+ labs(title="Quantity of Products and Category",x="Category",y="Quantity")
```




From the above barplot we can conclude that the Customers shopping habits includes a large extent of customers spending on “Clothing” products bagging the top spot with the highest-level of more than a lakh quantities of clothes being purchased as compared to other categories of the products in the dataset. Followed by a good number of people making a purchase of “Cosmetics” and relatively same amount (approximately 22000-23000) of customers are observed to invest in “Food and Beverages”. It has been recored that almost 30000 quanities of “Toys” and “Shoes” are sold. “Books”, “Technology”, “Souvenir” (usually inexpensive article purchased as a reminder of a place visited) are among the least products brought by the customers,with the rough numbers of quantities of 12000 each being sold.

2. Which product categories are most commonly purchased from a particular shopping m all?

```
>ggplot(data=data,aes(x=shopping_mall,fill=category))+
+ geom_bar(position="stack")+
+ labs(title="Relationship between Product category and Shopping mall",x="Shopping mal
l",y="Category Of Products")+
+ theme(axis.text.x = element_text(angle = 45, hjust = 1))   #to rotate x-axis labels
for readability
```




- From the above given stacked barplot, we can say that most number of products are purchased from the "Mall of Istanbul" and then in close approximity with it is another shopping mall - "Kanyon". The products which are **largely purchased** from these malls are **"Clothes".** Also all other product categories are extensively purchased from these malls as compared to others.

- It is also noticible that the **books, souvenirs and shoes** purchased are in good numbers from the malls of Instanbul, Kanyon and moderate numbers are observe d from the malls such as Metrocity, Metropd AVM and Istinye park and to be men tioned that there are hardly any numbers of books brought from other malls such as Cevahir AVM, Emaar Square Mall, Forum Istanbul, Viaport Outlet, Zorlu center.

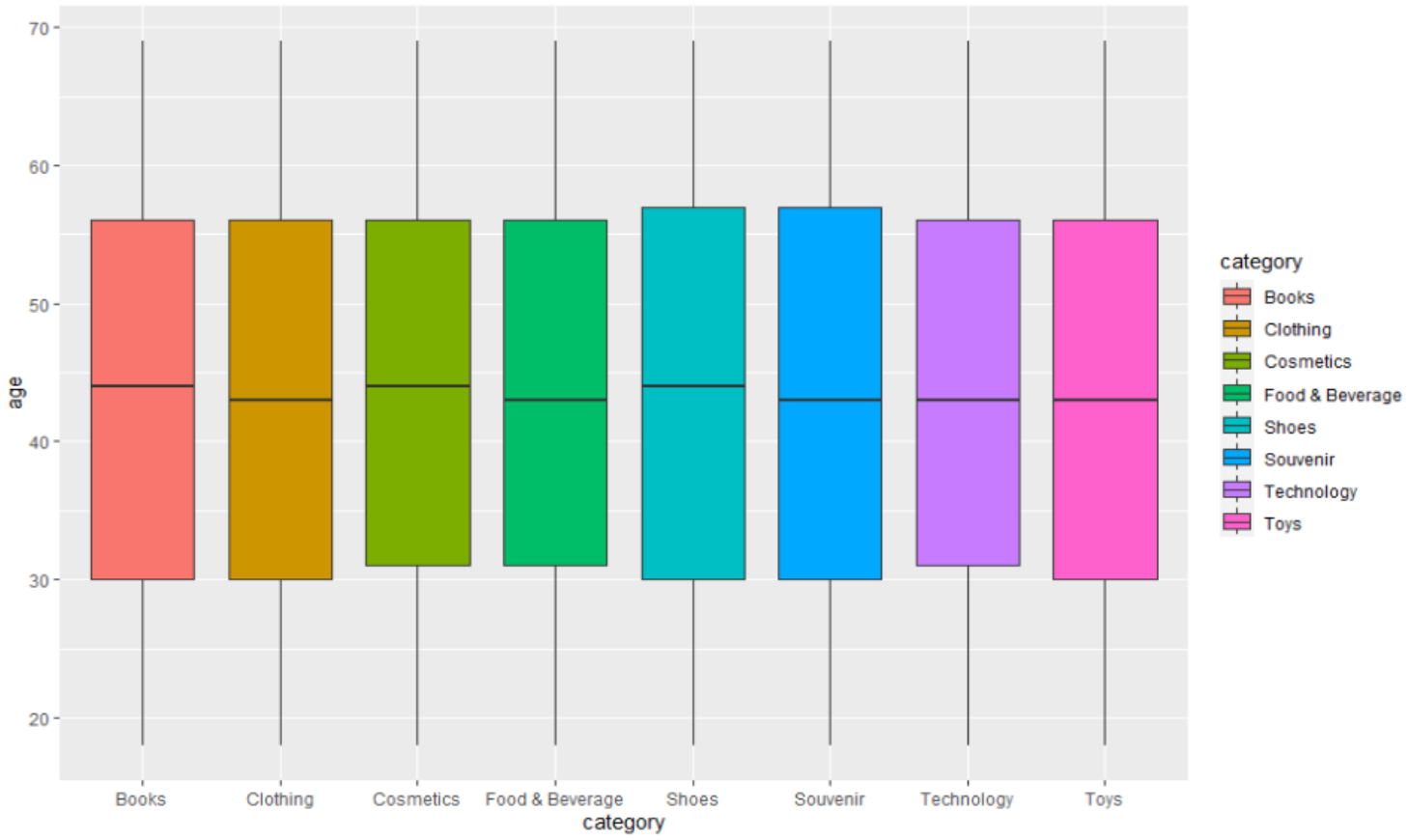
- It is also observed that product category such as **Cosmetics, food and beverage, toys,Technology** follows a very similar trend wherein there are highest purchas es equally purchased from the mall of istanbul and kanyon, moderate purchases

equally purchased from malls such as– Metrocity, Metropd AVM, Istinye Park, and low numbers of pruchases from other malls.

- Conclusion: 'Mall of Istanbul' and 'Kanyon' shopping malls seems to be the most c ustomers top choice for making their purchases for any product category, follow ed by in the list of top malls for maximum purchases of products is 'Metrocity' an d almost with the same amount of product purchase are 'Metropd AVM' and 'Isti nye Park'.

## 3. Which product categories are most commonly purchased among an age group?

```
> ggplot(data = data, aes(x = category, y = age,fill = category)) +
+ geom_boxplot()
```



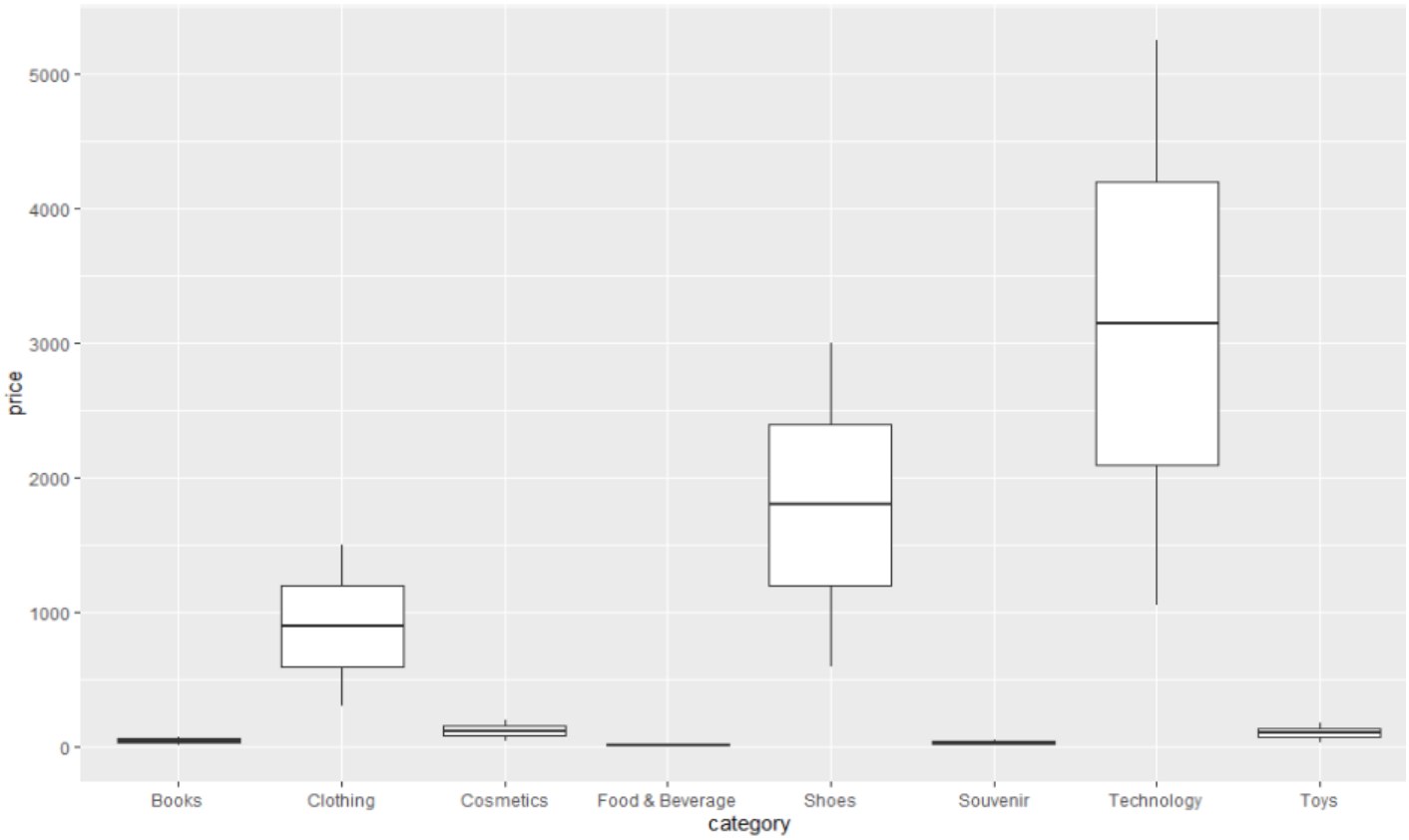

- The above boxplot shows that the median age of customers making a purchase of "Books", "Cosmetics" and "Shoes" is higher as compared to others and hence we can conclude that these customers are older as compared to others

.

- It is also observed that the median age of customers who purchased products such as – Clothing, Food & Beverages, Souvenir, Technology, and Toys are nearly same and thereafter we can conclude that these customers belong to a younger lot of customers.

4. Is there any relationship between Price of a product and the product category? Also compare the number of customers making a purchase of those products.

```
> ggplot(data = data, aes(x = category, y = price)) +
+ geom_boxplot()
```




- It is essential to note here that eventhough there is larger chunk of customers making a purchase of clothes, it is primarily because the prices of clothes are less. The average prices of clothes is nearly 900.

- It is important for us to note that the average prices of shoes is more than clothing products. Eventhough the shoes are not among the top demand of purchases; they are neither the least and hence we can conclude that there could be enough revenue generated from shoe sales.

- The average prices of technology grabs the top spot among all the categories of the product. If we compare the average price of clothes with that of technology there is a remarkable difference between the two. The average prices of tech products are more than 3 times of clothing products.

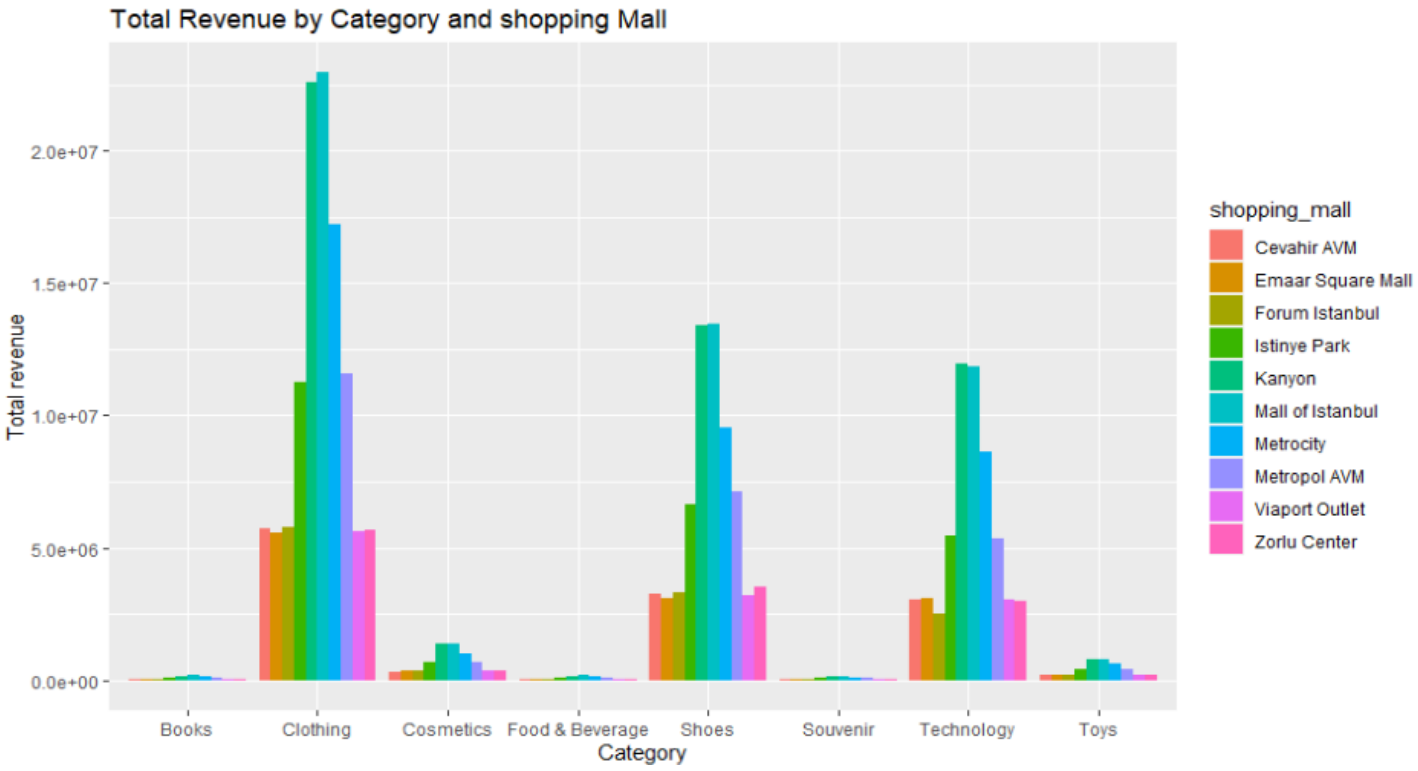
- It is also of atmost importance to note here that the average prices of other Categories of products are nearly negligible when compared to the rest.

5. Find which shopping mall has the highest revenue generated & from which product?

```
> total_revenue=data %>%
```

```
+ group_by(shopping_mall,category)%>%
+ summarize(total=sum(price*quantity))

> ggplot(total_revenue, aes(x=category, y=total, fill=shopping_mall))+
+ geom_bar(stat="identity", position = "dodge")+
+ labs(title="Total Revenue by Category and shopping Mall", x="Category",y= "Total
  revenue")
```




- From the above barplot , we can conclude that highest revenue generated is from sales of clothes followed by shoes and technology.

- At the peak is the Mall of Istanbul who's total revenue is the highest from the Clothing purchases. With more or less numbers is the Mall- Kanyon and then followed by malls such as – Metrocity, Metropd AVMand Istinye Park

- If we consider the sales for shoes the total revenue generated by Mall of istanbul and kanyon is nearly same followed by Metrocity, Metropd AVM and Istinye Park.

- If we consider the sales of technology the total revenue generated by kanyon mall is slightly more than mall of istanbul followed by the Metrocity and nearly equal Revenue by Metropd AVM and Istinye park.

- Rest all malls follow almost similar trend in terms of clothing, shoes and technolo gy respectively.