CUSTOMER SEGMENTATION ANALYSIS

PTID-CDA-NOV-24-241

Team members:

Ayyar Swetha Ponugupati - ayyarswetha123@gmail.com
Shaik Shaistha Naaz - shaikshaisthanaaz@gmail.com
Sruti Rekha Pradhan - srutirekhapradhan785@gmail.com

Dataset Overview:

- The dataset contains customer shopping information from 10 different shopping malls in Istanbul, spanning the years 2021 to 2023.
- It includes various attributes that provide insights into customer behaviour, purchasing patterns, and payment preferences.
- The key columns in the dataset are:
  - *invoice_no*: A unique identifier for each transaction.
  - *customer_id*: A unique identifier for each customer.
  - *gender*: The gender of the customer (Male/Female).
  - *age*: The customer's age.
  - *category*: The category of the purchased product (e.g., electronics, clothing, etc.).
  - *quantity*: The number of items purchased in each transaction.
  - *price*: The price of a single unit of the product.
  - *payment_method*: The method of payment used (cash, credit card, or debit card).
  - *invoice_date*: The date when the transaction occurred.
  - *shopping_mall*: The name of the shopping mall where the transaction took place

Task 1: How is the shopping distribution according to gender?

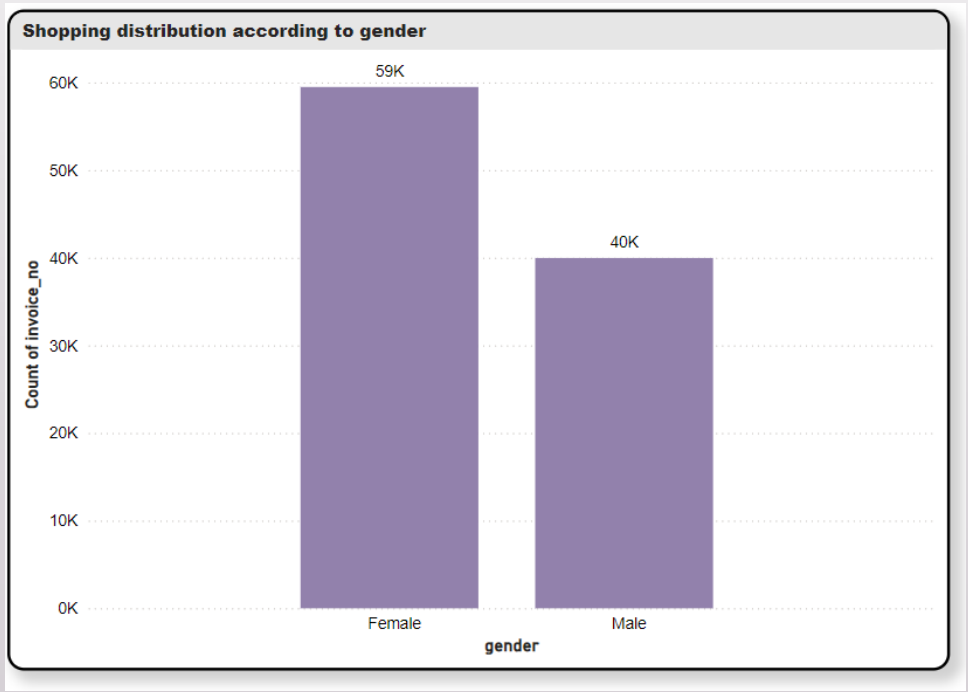
project* | project

Limit to 1000 rows

```
11
12      -- 1. Shopping distribution according to gender
13
14      SELECT gender, COUNT(invoice_no) AS distribution
15      from customer GROUP BY gender;
16
```

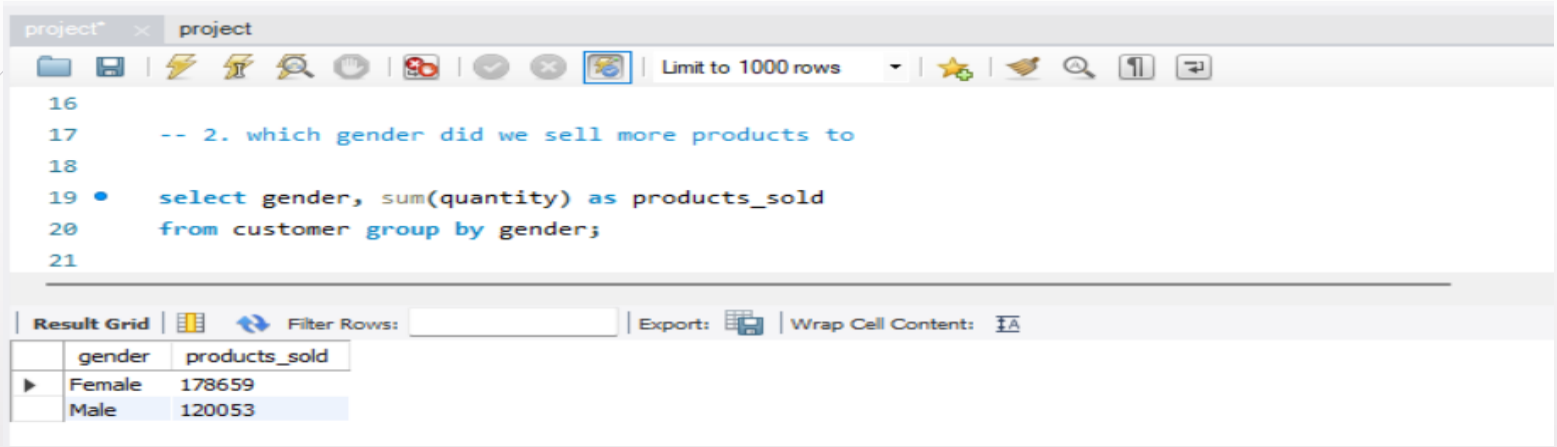
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

gender	distribution
Female	59482
Male	39975



- The analysis of shopping distribution by gender shows that females have the highest distribution.

Task 2: Which gender did we sell more products to?



Gender based product sales comparison

gender	Sum of quantity
Female	178659
Male	120053
Total	298712

- The analysis reveals that females purchased more products

Task 3: Which gender generated more revenue?

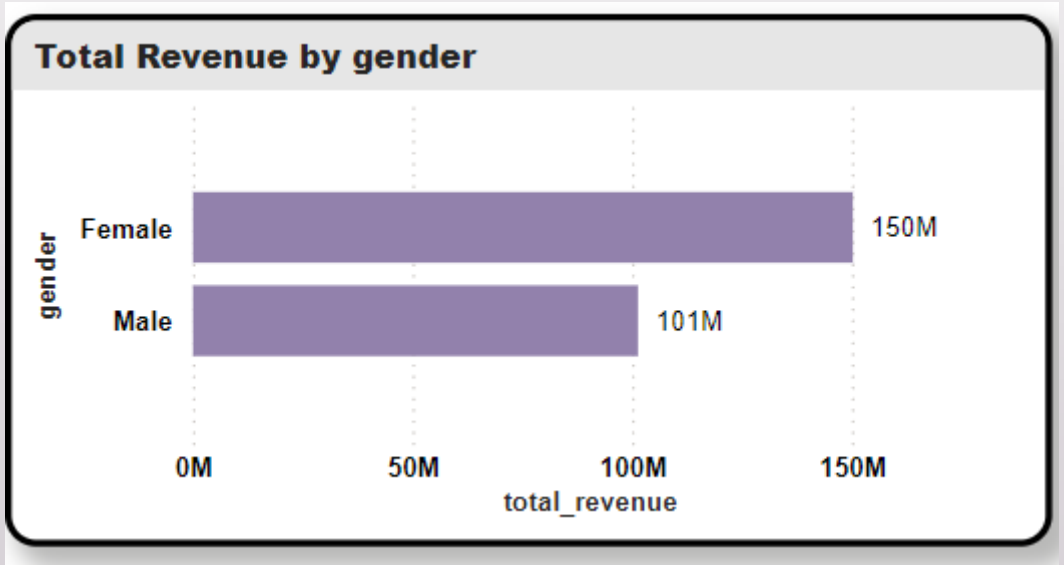
project* | project

Limit to 1000 rows

```
21
22      -- 3. Which gender generated more revenue
23
24 •    select gender, sum(quantity*price) as total_revenue
25      from customer group by gender;
26
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

gender	total_revenue
Female	150207136.02
Male	101298658.23



- The analysis indicates that females generated more revenue.

Task 4: Distribution of purchase categories relative to other columns?

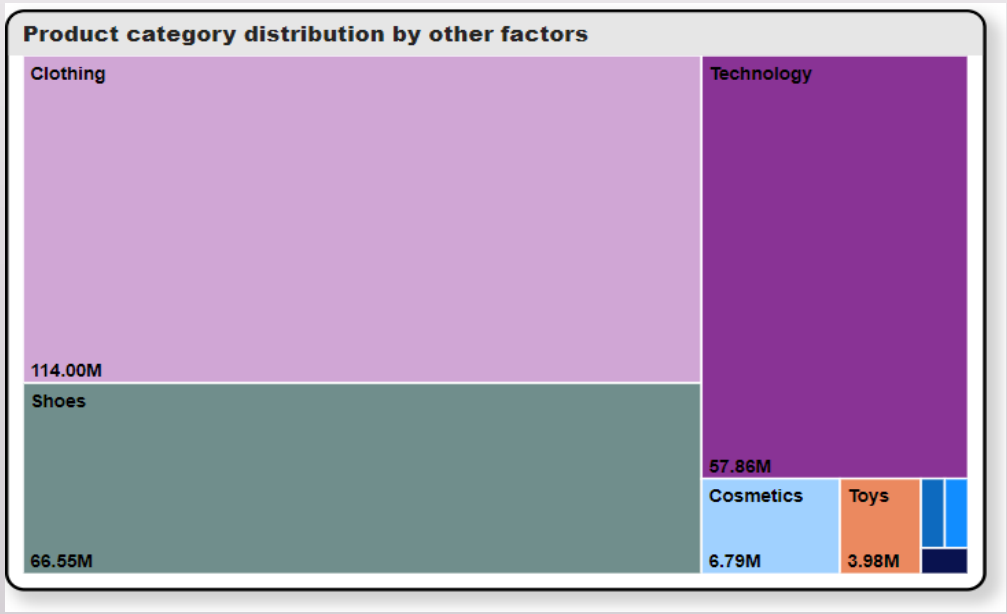
project* | project

Limit to 1000 rows

```
26
27      -- 4. Distribution of purchase category relative to other column
28
29  •   select category, sum(quantity*price) as total_revenue
30      from customer group by category order by total_revenue desc;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

category	total_revenue
Clothing	113996791.04
Shoes	66553451.47
Technology	57862350.00
Cosmetics	6792862.90
Toys	3980426.24
Food & Beverage	849535.05
Books	834552.90
Souvenir	635824.65



- The analysis shows the distribution of purchase categories by revenue, with **clothing leading at 114 million**, followed by shoes (66.55M) and technology (57.86M). Other categories include cosmetics (6.79M), toys (3.98M), food & beverage (849k), books (834k), and souvenirs (635k).

Task 5: How is the shopping distribution according to age?

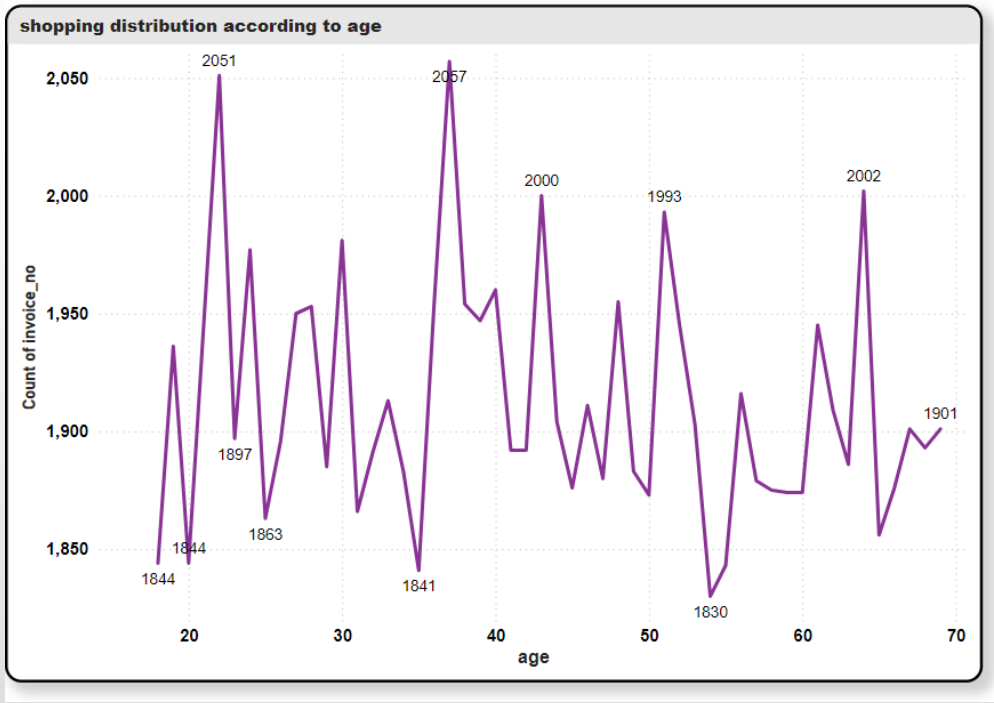
```sql
-- 5. Shopping distribution according to age

select age, count(invoice_no) as distribution
from customer group by age order by distribution desc;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content

age	distribution
37	2057
22	2051
64	2002
43	2000
51	1993
30	1981
24	1977
40	1960
48	1955
38	1954
36	1954
28	1953
27	1950
39	1947
21	1947
52	1945
61	1945
19	1936
56	1916
33	1913
46	1911
62	1909
44	1904
53	1903

Result 18



- The analysis of shopping distribution by age shows that the age 37 represent the highest distribution, while the age 54 have the lowest distribution

Task 6: Which age cat did we sell more products to?

project* | project

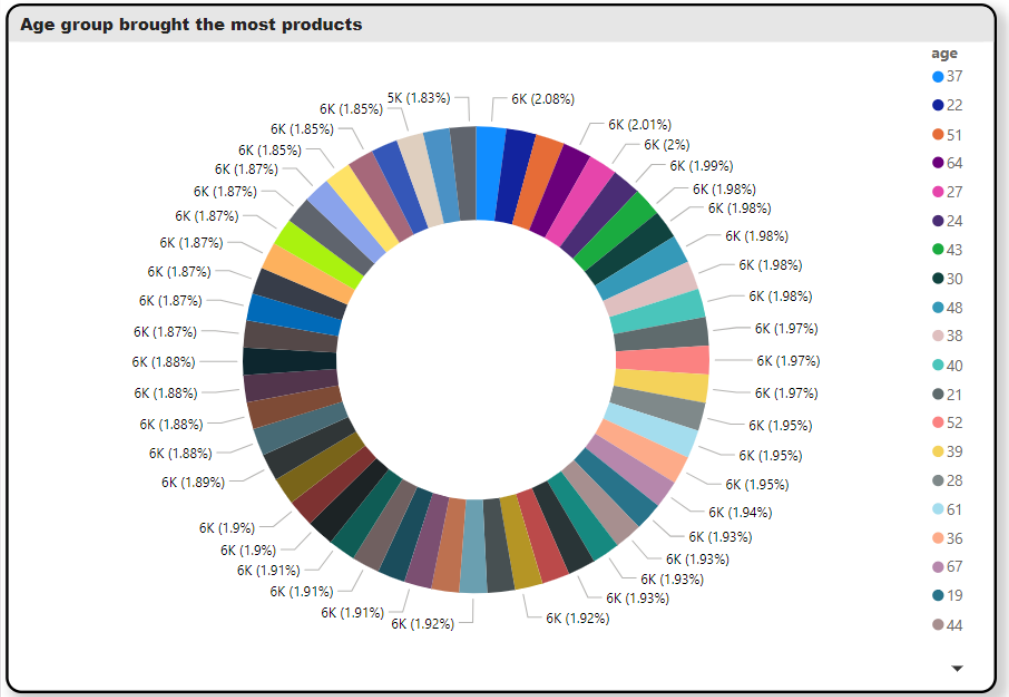
Limit to 1000 rows

```
37      -- 6. Which age group bought more products
38
39  •   select age, sum(quantity) as total_products
40      from customer group by age order by total_products desc;
41
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

age	total_products
37	6217
22	6148
51	6014
64	5991
27	5969
24	5957
43	5928
30	5927
48	5918
38	5910
40	5902
21	5894
52	5892
39	5874
28	5832
61	5829
36	5826
67	5788
19	5778
44	5777
33	5756
46	5751
29	5744
68	5737

Result 19



- The analysis shows that the highest number of products were sold to customers aged 37, while the least were sold to those aged 54.

Task 7: Which age cat generated more revenue?

project* | project

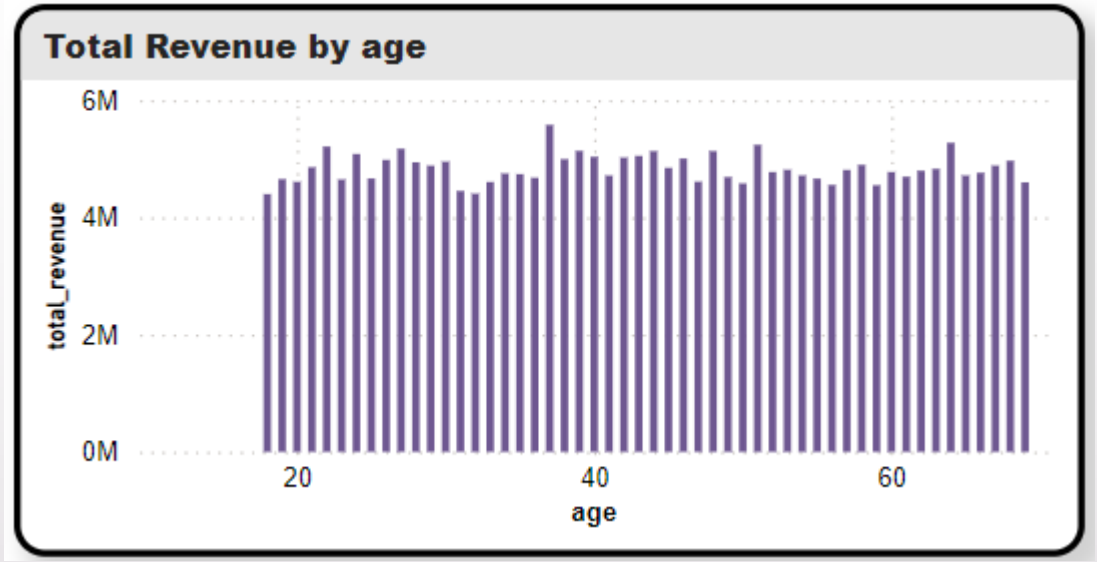
Limit to 1000 rows

```
42    -- 7. Which age group generated more revenue
43
44 •  select age, sum(quantity*price) as total_revenue
45    from customer group by age order by total_revenue desc;
46
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

age	total_revenue
37	5578539.57
64	5272475.10
51	5238724.74
22	5208841.41
27	5171859.06
39	5135672.77
48	5131747.86
44	5131686.80
24	5082409.90
43	5050323.72
40	5034207.14
42	5021505.35
46	5004837.20
38	4994225.53
26	4981414.82
68	4967120.72
30	4951952.64
28	4940135.42
58	4896507.56
67	4884731.33
29	4883768.82
21	4856158.41
45	4845927.03
63	4829099.00

Result 20



- The analysis shows that the highest revenue is generated by individuals aged 37, 64, and 51, while the lowest revenue comes from individuals aged 18, 19, and 20.

Task 8: Distribution of purchase categories relative to other columns?

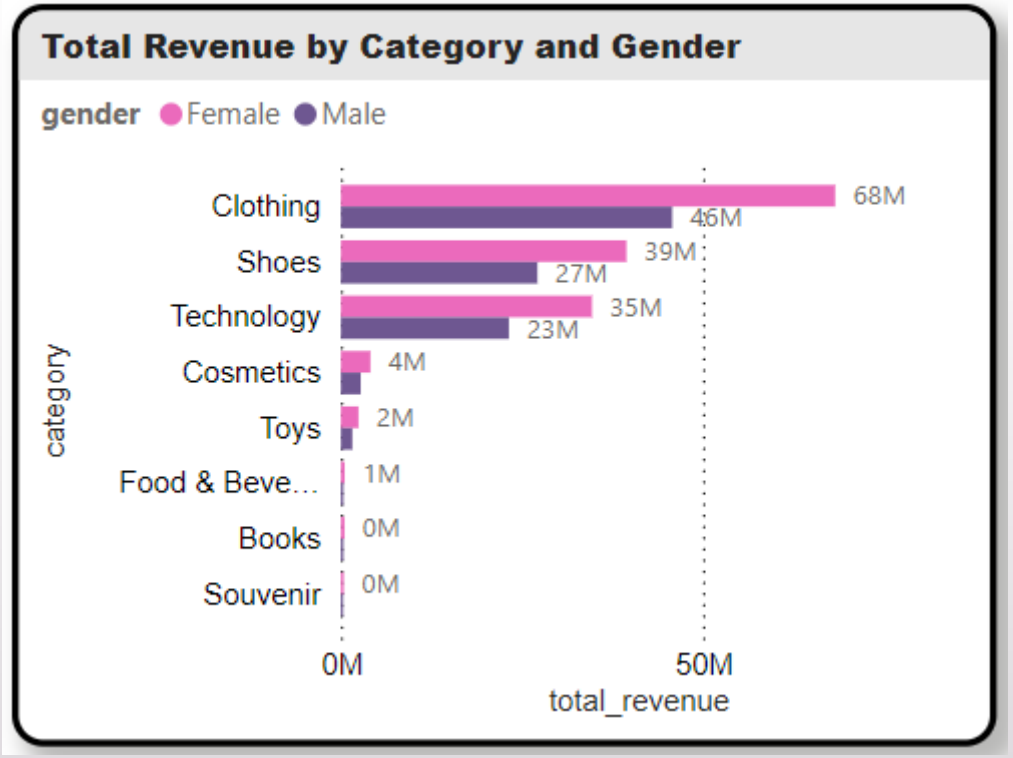
project* | project

Limit to 1000 rows

```
47  -- 8. Does the payment method related to other column (e.g., gender,age,category)
48
49  select category,gender,sum(quantity*price) as total_revenue
50  from customer group by category, gender order by total_revenue desc;
51
```

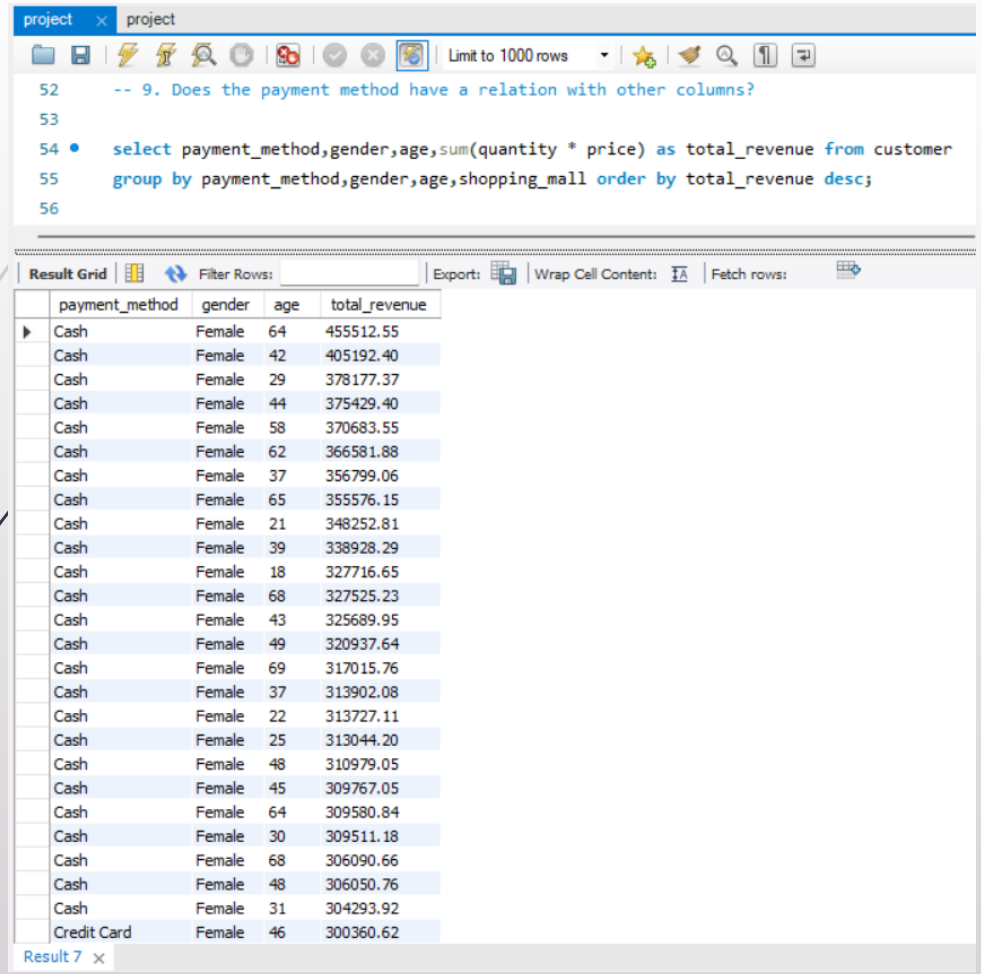
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

category	gender	total_revenue
Clothing	Female	68251695.60
Clothing	Male	45745095.44
Shoes	Female	39425167.30
Technology	Female	34669950.00
Shoes	Male	27128284.17
Technology	Male	23192400.00
Cosmetics	Female	4066772.54
Cosmetics	Male	2726090.36
Toys	Female	2416046.08
Toys	Male	1564380.16
Food & Be...	Female	505322.60
Books	Female	489314.70
Souvenir	Female	382867.20
Books	Male	345238.20
Food & Be...	Male	344212.45
Souvenir	Male	252957.45



- The Analysis shows that in all the categories the female has more contribution compared to male in total revenue.

Task 9: Does the payment method have a relation with other columns?



project | project

Limit to 1000 rows

```
52    -- 9. Does the payment method have a relation with other columns?
53
54 •  select payment_method,gender,age,sum(quantity * price) as total_revenue from customer
55    group by payment_method,gender,age,shopping_mall order by total_revenue desc;
56
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

payment_method	gender	age	total_revenue
Cash	Female	64	455512.55
Cash	Female	42	405192.40
Cash	Female	29	378177.37
Cash	Female	44	375429.40
Cash	Female	58	370683.55
Cash	Female	62	366581.88
Cash	Female	37	356799.06
Cash	Female	65	355576.15
Cash	Female	21	348252.81
Cash	Female	39	338928.29
Cash	Female	18	327716.65
Cash	Female	68	327525.23
Cash	Female	43	325689.95
Cash	Female	49	320937.64
Cash	Female	69	317015.76
Cash	Female	37	313902.08
Cash	Female	22	313727.11
Cash	Female	25	313044.20
Cash	Female	48	310979.05
Cash	Female	45	309767.05
Cash	Female	64	309580.84
Cash	Female	30	309511.18
Cash	Female	68	306090.66
Cash	Female	48	306050.76
Cash	Female	31	304293.92
Credit Card	Female	46	300360.62

Result 7

Payment Method based on other Columns

payment_method	gender	total_revenue
⊟ Cash	⊞ Female	6,75,73,638.45
	⊞ Male	4,52,58,604.57
	Total	11,28,32,243.02
⊟ Credit Card	⊞ Female	5,28,75,809.90
	⊞ Male	3,52,01,313.87
	Total	8,80,77,123.77
⊟ Debit Card	⊞ Female	2,97,57,687.67
	⊞ Male	2,08,38,739.79
	Total	5,05,96,427.46
Total		25,15,05,794.25

- The Analysis shows that the payment method preferences vary across age and gender which in turn effects the Total Revenue.

Task 10: How is the distribution of the payment method?

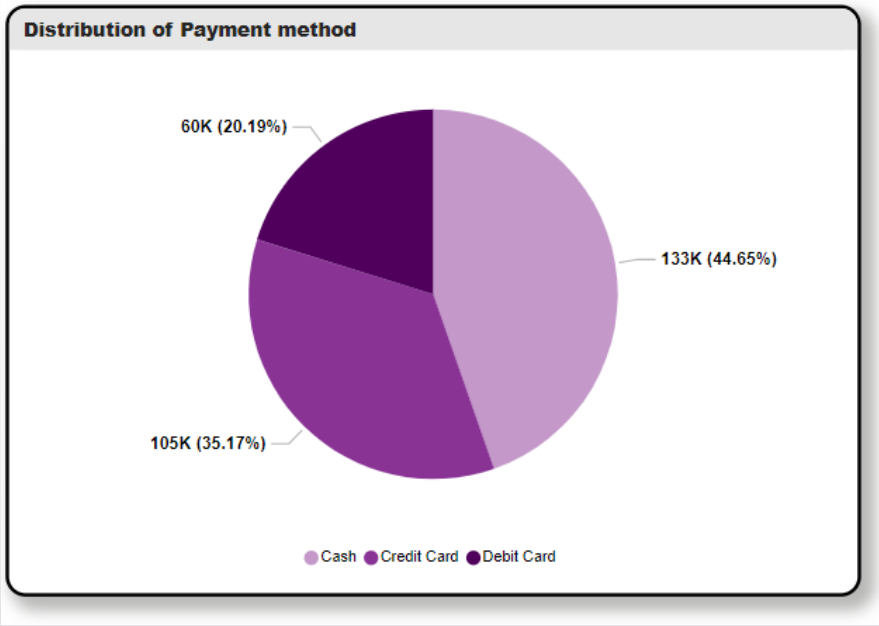
project* | project

Limit to 1000 rows

```
57    -- 10.How is the distribution of the payment method?
58
59    select payment_method,sum(quantity) as distribution from customer
60    group by payment_method;
61
```

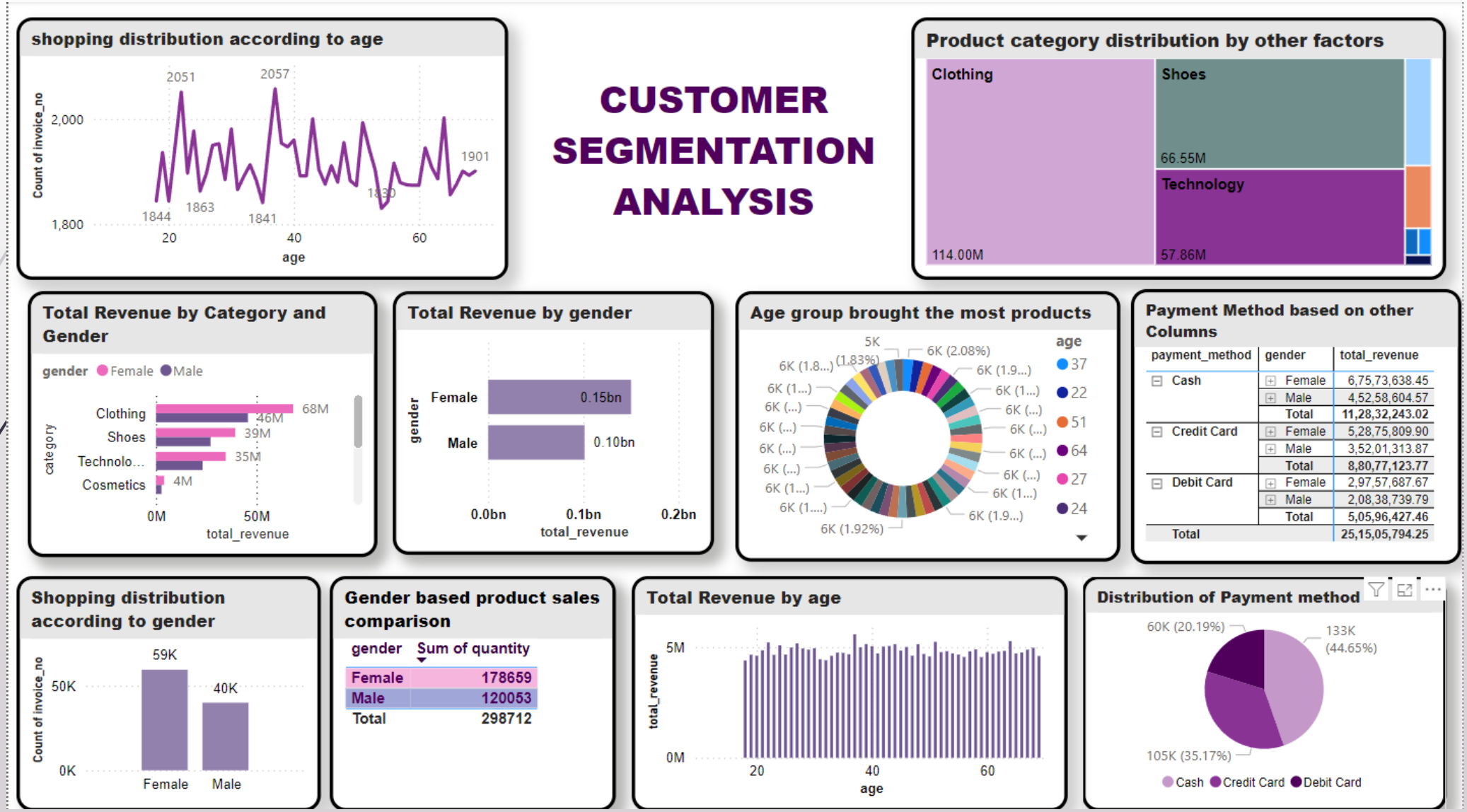
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

payment_method	distribution
Cash	133370
Credit Card	105045
Debit Card	60297



- The distribution of payment methods reveals customer preferences for how they pay for their purchases. Some customers prefer paying with **credit cards**, while others use **debit cards** or **cash**.

DASHBOARD


shopping distribution according to age
Count of invoice_no
2,000
1,800
2051
2057
1901
1844
1863
1841
1830
20
40
60
age
CUSTOMER SEGMENTATION ANALYSIS
Product category distribution by other factors
Clothing
114.00M
Shoes
66.55M
Technology
57.86M
Total Revenue by Category and Gender
gender Female Male
category
Clothing
Shoes
Technolo...
Cosmetics
68M
46M
39M
35M
4M
0M
50M
total_revenue
Total Revenue by gender
gender
Female
Male
0.15bn
0.10bn
0.0bn
0.1bn
0.2bn
total_revenue
Age group brought the most products
age
37
22
51
64
27
24
5K
6K (2.08%)
6K (1.83%)
6K (1.92%)
Payment Method based on other Columns
payment_method | gender | total_revenue
Cash | Female | 6,75,73,638.45
Male | 4,52,58,604.57
Total | 11,28,32,243.02
Credit Card | Female | 5,28,75,809.90
Male | 3,52,01,313.87
Total | 8,80,77,123.77
Debit Card | Female | 2,97,57,687.67
Male | 2,08,38,739.79
Total | 5,05,96,427.46
Total | 25,15,05,794.25
Shopping distribution according to gender
Count of invoice_no
50K
0K
59K
40K
Female
Male
Gender based product sales comparison
gender | Sum of quantity
Female | 178659
Male | 120053
Total | 298712
Total Revenue by age
total_revenue
5M
0M
20
40
60
age
Distribution of Payment method
60K (20.19%)
133K (44.65%)
105K (35.17%)
Cash
Credit Card
Debit Card

CONCLUSION

- The shopping habits differ across age groups and genders, influencing both the quantity of products purchased and the revenue generated.
- We found that certain genders tend to purchase specific product categories and generate higher revenue, which can help tailor marketing efforts.
- The payment method preferences vary across age and gender, and understanding these patterns can help businesses adapt their payment options and promotions.
- By examining product categories, we identified which categories are popular with certain age groups and genders, helping businesses make better stocking and promotional decisions.
- Based on the insights, businesses should target their marketing at the age groups and genders that spend the most, encourage the use of popular payment methods, and focus on promoting the most popular product categories.