



SQL Project

Data Exploration &  
Visualization

Turkiye Ecommerce Sales Dashboard

By

Agif Ridho Rizkiawan

# Tools

- MySQL
- Visual Studio Code
- Tableau
- Microsoft Excel

# Dataset

- Kaggle

# Skill used

- CTE
- Windows Function
- Aggregate Function
- Date Function
- Case
- String Function





# Data Cleaning

- Standardize Date Format

## Problem Statement

1. Create a query to get the total transaction per month. Please use time frame from 01 January 2021 until 31 Desember 2021.
2. Create a query to get total revenue grouped by shopping mall in 2022.
3. Create a query to get which payment method most used by customer and total revenue of each payment method 3 month before 01 January 2022.
4. Get the total\_revenue of each category\_variant. 1 month after 30 April 2022.
5. Find monthly growth of total\_revenue in percentage breakdown by Shopping Mall, ordered by time descendingly.
6. Find from which age group the most transaction, and ratio percentage.
7. Query to get revenue ratio percentage from total revenue in each year revenue.



## Problem Statement 1

Create a query to get the total transaction per month. Please use time frame from 01 January 2021 until 31 Desember 2021.

### Syntax



```
1 select
2     date_format(invoice_date,'%m-%Y') month
3     , count(invoice_date) total_transactions
4 from customer_transactions
5 where year(invoice_date) = 2021
6 group by 1
7 order by 1
```

month	total_transactions
abc Filter...	abc Filter...
01-2021	3835
02-2021	3407
03-2021	3813
04-2021	3724
05-2021	3848
06-2021	3783
07-2021	3984
08-2021	3723
09-2021	3670
10-2021	3916
11-2021	3798
12-2021	3881

## Problem Statement 2

Create a query to get total revenue grouped by shopping mall in 2022.

### Syntax



```
1 select
2     shopping_mall mall
3     , round (sum(quantity*price),2) total_revenue
4 from customer_transactions
5 where year(invoice_date) = 2022
6 group by 1
7 order by 2 desc
```

mall	total_revenue
abc Filter...	abc Filter...
Mall of Istanbul	23410362.3
Kanyon	22922200.33
Metrocity	17244618.86
Istinye Park	11469526.74
Metropol AVM	11375841
Zorlu Center	6037789.44
Viaport Outlet	5920500.33
Cevahir AVM	5831795.35
Forum Istanbul	5800267.22
Emaar Square Mall	5423912.51

## Problem Statement 3

Create a query to get which payment method most used by customer and total revenue of each payment method 3 month before 01 January 2022.

### Syntax

```
1  select
2    payment_method payment_by
3    , invoice_date
4    , count(payment_method) count_paymethod
5    , round(sum(quantity*price),2) total_revenue
6  from customer_transactions
7  where invoice_date < date_sub('2022-01-01', interval 3 month)
8  group by 1,2
9  order by 3 desc
```

payment by	invoice_date	count_paymethod	total_revenue
Cash	2021-01-28	77	180667.02
Cash	2021-06-19	74	209764.22
Cash	2021-01-06	73	218741.57
Cash	2021-07-11	72	143769.96
Cash	2021-06-23	69	177191.74
Cash	2021-07-19	69	170026.49
Cash	2021-06-04	69	88727.37
Cash	2021-07-18	69	202050.21
Cash	2021-09-19	69	140786.1
Cash	2021-05-17	68	151541.73
Cash	2021-05-22	68	130887.51
Cash	2021-09-03	68	251034.97
Cash	2021-07-14	68	167213.62
Cash	2021-08-09	68	199618.67
Cash	2021-09-24	68	160336.21
Cash	2021-04-10	67	175163.57
Cash	2021-08-21	66	138473.73
Cash	2021-03-04	66	170023.94
Cash	2021-06-26	66	212777.85

## Problem Statement 4

Get the total\_revenue of each category\_variant. 1 month after 30 April 2022.

### Syntax

```
1 with category_bef_variant as
2 (
3   select
4     *
5   from customer_transactions
6 ),
7 variant_make as
8 (
9   select
10    *
11    , dense_rank () over (partition by category order by price
12 desc) rank_item
13   from category_bef_variant
14 ),
15 category_variant_made as
16 (
17   select
18     invoice_no
19     , customer_id
20     , gender
21     , age
22     , category
23     , quantity
24     , price
25     , payment_method
26     , invoice_date
27     , shopping_mall
28     , concat (category,"-",rank_item) category_variant
29   from variant_make
30 )
31 select
32   category
33   , price
34   , category_variant
35   , round(sum(quantity*price),2) total_revenue
36 from category_variant_made
37 where invoice_date > date_add('2022-04-30', interval 1 month)
38 group by 1,2,3
```

category	price	category_variant	total_revenue
Books	75.75	Books-1	156802.5
Books	60.6	Books-2	76113.6
Books	45.45	Books-3	50449.5
Books	30.3	Books-4	23452.2
Books	15.15	Books-5	5423.7
Clothing	900.24	Clothing-1	6573552.48
Clothing	600.16	Clothing-2	2964790.4
Clothing	300.08	Clothing-3	726193.6
Clothing	1500.4	Clothing-4	18379900
Clothing	1200.32	Clothing-5	11302213.12
Cosmetics	81.32	Cosmetics-1	171422.56
Cosmetics	40.66	Cosmetics-2	44685.34
Cosmetics	203.3	Cosmetics-3	1148645
Cosmetics	162.64	Cosmetics-4	713664.32
Cosmetics	121.98	Cosmetics-5	402534
Food & Beverage	5.23	Food & Beverage-1	5596.1
Food & Beverage	26.15	Food & Beverage-2	137026
Food & Beverage	20.92	Food & Beverage-3	85604.64
Food & Beverage	15.69	Food & Beverage-4	49141.08
Food & Beverage	10.46	Food & Beverage-5	21422.08



## Problem Statement 5

Find monthly growth of total\_revenue in percentage breakdown by Shopping Mall, ordered by time descendingly.

### Syntax

```
1 with total_revenue as
2 (
3     select
4         date_format(invoice_date,'%Y-%m') month
5         , shopping_mall mall
6         , round(sum(quantity*price),2) total_rev
7     from customer_transactions
8     group by 1,2
9     order by 1
10 ),
11 prev_total_revenue as
12 (
13     select
14         *
15         , lag(total_rev) over (partition by mall order by month) prevmonth_total_rev
16     from total_revenue
17 )
18 select
19     *
20     , round((total_rev - prevmonth_total_rev)/prevmonth_total_rev*100,2) growth_prctng
21 from prev_total_revenue
```

month	mall	total_rev	prevmonth_tota...	growth_prctng
2021-01	Cevahir AVM	483830.06	NULL	NULL
2021-02	Cevahir AVM	329264.45	483830.06	-31.95
2021-03	Cevahir AVM	512891.12	329264.45	55.77
2021-04	Cevahir AVM	432312.84	512891.12	-15.71
2021-05	Cevahir AVM	489344.99	432312.84	13.19
2021-06	Cevahir AVM	400583.19	489344.99	-18.14
2021-07	Cevahir AVM	547457.71	400583.19	36.67
2021-08	Cevahir AVM	423705.15	547457.71	-22.6
2021-09	Cevahir AVM	488999.18	423705.15	15.41
2021-10	Cevahir AVM	537142.92	488999.18	9.85
2021-11	Cevahir AVM	586604.8	537142.92	9.21
2021-12	Cevahir AVM	526692.14	586604.8	-10.21
2022-01	Cevahir AVM	504034.8	526692.14	-4.3
2022-02	Cevahir AVM	506621.96	504034.8	0.51
2022-03	Cevahir AVM	382896.96	506621.96	-24.42
2022-04	Cevahir AVM	406154.18	382896.96	6.07
2022-05	Cevahir AVM	562474.8	406154.18	38.49
2022-06	Cevahir AVM	533126.41	562474.8	-5.22
2022-07	Cevahir AVM	494951.64	533126.41	-7.16
2022-08	Cevahir AVM	521878.16	494951.64	5.44

## Problem Statement 6

Find from which age group the most transaction, and ratio percentage.

### Syntax

```
1 with age_gr as
2 (
3   select
4     -- age
5     case
6       when age >= 50 then 'Elderly Customer'
7       when age < 50 and age >= 30 then 'Middle Age Customer'
8       when age < 30 then 'Adolescent Customer'
9     end age_group
10    , count(age) total_customer
11 from customer_transactions
12 where year(invoice_date) in (2021,2022)
13 group by 1
14 )
15 select
16   *
17   , round((total_customer/sum(total_customer) over ())*100,2) cnt_ratio_prcgt
18 from age_gr
```

age_group	total_customer	cnt_ratio_prcgt
abc Filter...	abc Filter...	abc Filter...
Adolescent Customer	21044	23.14
Elderly Customer	34749	38.21
Middle Age Customer	35140	38.64

## Problem Statement 7

Query to get revenue ratio percentage from total revenue in each year revenue.

### Syntax

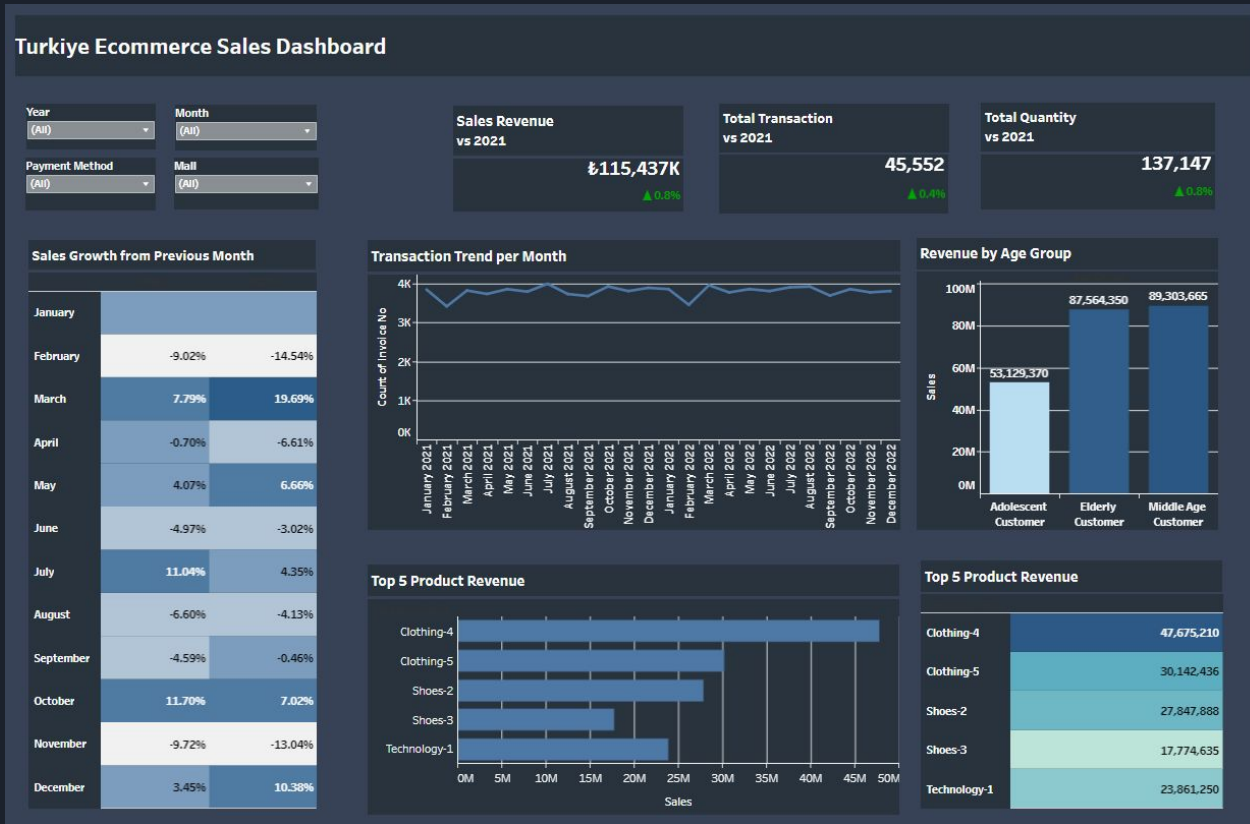
```
1 with revenue_per_year as
2 (
3   select
4     year(invoice_date) year
5     , round(sum(price*quantity),2) total_revenue
6   from customer_transactions
7   group by 1
8 )
9 select
10  *
11  , sum(total_revenue) over () allyears_total_revenue
12  , round(total_revenue/sum(total_revenue) over (),2) ratio_prcntg
13 from revenue_per_year
```

year	total_revenue	allyears_total_re...	ratio_prcntg
Filter...	Filter...	Filter...	Filter...
2022	115436814.08	251505794.25	0.46
2021	114560570.59	251505794.25	0.46
2023	21508409.58	251505794.25	0.09

# Query for Visualization

```
1  -- Data for Visualization
2  with category_bef_variant as
3  (
4      select
5          *
6      from customer_transactions
7  ),
8  variant_make as
9  (
10     select
11         *
12         , dense_rank () over (partition by category order by price desc) rank_item
13     from category_bef_variant
14 )
15 select
16     invoice_no
17     , customer_id
18     , gender
19     , age
20     , category
21     , quantity
22     , price
23     , payment_method
24     , invoice_date
25     , shopping_mall
26     , concat (category,"-",rank_item) category_variant
27     , case
28         when age >= 50 then 'Elderly Customer'
29         when age < 50 and age >= 30 then 'Middle Age Customer'
30         when age < 30 then 'Adolescent Customer'
31     end age_group
32 from variant_make
33 where year(invoice_date) in (2021,2022)
```

## Turkiye Ecommerce Sales Dashboard



# Let's Connect



+62 812-8592-0945



agif.rizkiawan@gmail.com



<https://www.linkedin.com/in/agif0694/>



<https://www.mavenanalytics.io/profile/Agif-Ridho-Rizkiawan/159809712>

THANK YOU

