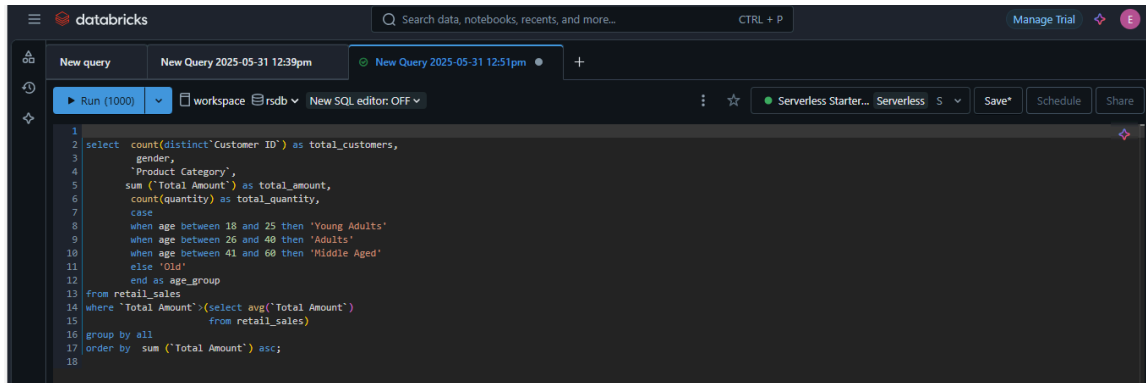


## Data Analysis of the retail sales dataset on databricks

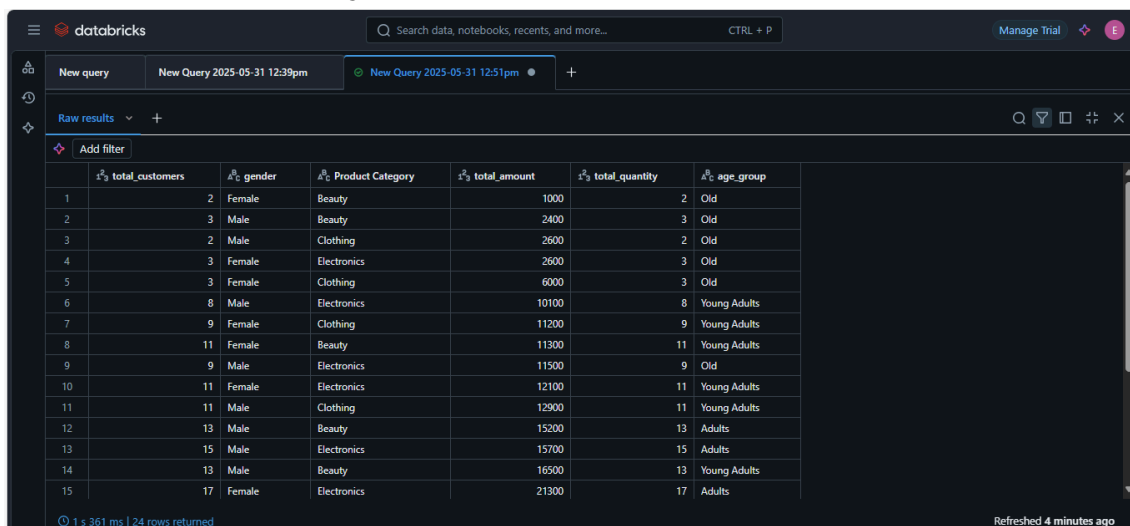
The objective was to analyse and make sense of the dataset. Looking at the distribution of sales amongst various customers.

### 1. The syntax for analysing data



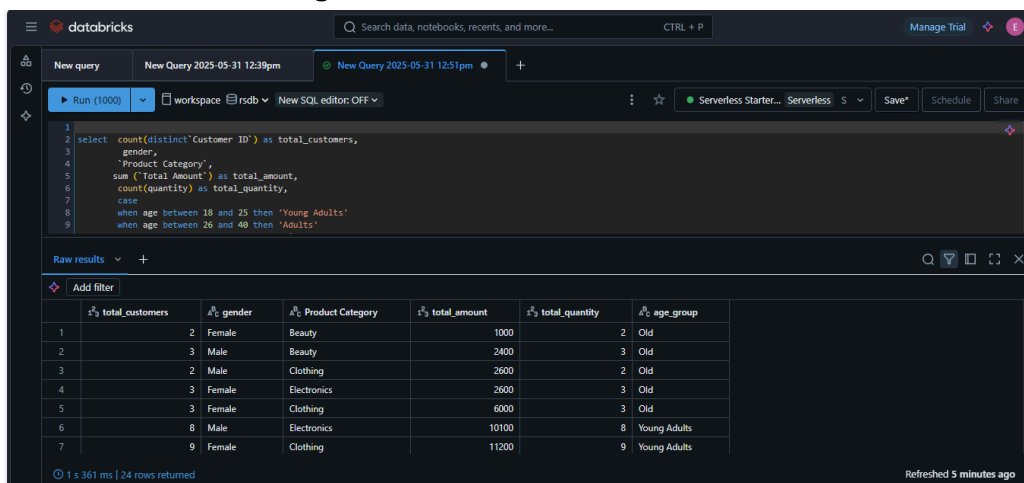
```
1 select count(distinct Customer ID) as total_customers,
2 gender,
3 Product Category,
4 sum (Total Amount) as total_amount,
5 count(quantity) as total_quantity,
6 case
7   when age between 18 and 25 then 'Young Adults'
8   when age between 26 and 40 then 'Adults'
9   when age between 41 and 60 then 'Middle Aged'
10  else 'Old'
11 end as age_group
12 from retail_sales
13 where Total Amount > (select avg(Total Amount)
14   from retail_sales)
15 group by all
16 order by sum (Total Amount) asc;
```

### 2. The result of the syntax as a table



	total_customers	gender	Product Category	total_amount	total_quantity	age_group
1	2	Female	Beauty	1000	2	Old
2	3	Male	Beauty	2400	3	Old
3	2	Male	Clothing	2600	2	Old
4	3	Female	Electronics	2600	3	Old
5	3	Female	Clothing	6000	3	Old
6	8	Male	Electronics	10100	8	Young Adults
7	9	Female	Clothing	11200	9	Young Adults
8	11	Female	Beauty	11300	11	Young Adults
9	9	Male	Electronics	11500	9	Old
10	11	Female	Electronics	12100	11	Young Adults
11	11	Male	Clothing	12900	11	Young Adults
12	13	Male	Beauty	15200	13	Adults
13	15	Male	Electronics	15700	15	Adults
14	13	Male	Beauty	16500	13	Young Adults
15	17	Female	Electronics	21300	17	Adults

### 3. The overall thing



```
1 select count(distinct Customer ID) as total_customers,
2 gender,
3 Product Category,
4 sum (Total Amount) as total_amount,
5 count(quantity) as total_quantity,
6 case
7   when age between 18 and 25 then 'Young Adults'
8   when age between 26 and 40 then 'Adults'
```

	total_customers	gender	Product Category	total_amount	total_quantity	age_group
1	2	Female	Beauty	1000	2	Old
2	3	Male	Beauty	2400	3	Old
3	2	Male	Clothing	2600	2	Old
4	3	Female	Electronics	2600	3	Old
5	3	Female	Clothing	6000	3	Old
6	8	Male	Electronics	10100	8	Young Adults
7	9	Female	Clothing	11200	9	Young Adults