

SQL PROJECT

WALMART SALES ANALYSIS

-SHAAN MANSURI



OBJECTIVE

This project aims to explore the Walmart Sales data to understand top performing branches and products, sales trend of different products, customer behaviour. The aim is to study how sales strategies can be improved and optimized. The dataset was obtained from the [Kaggle Walmart Sales Forecasting Competition](#).

The major aim of the project is to gain insight into the sales data of Walmart to understand the different factors that affect sales of the different branches.

Generic Question

SHAAN MANSURI

QUESTION - 1 How many unique cities does the data have?

```
select distinct city  
from sales;
```

The screenshot shows a database query results grid titled "Result Grid". The grid has one column labeled "city". It displays three rows of data: "Yangon", "Naypyitaw", and "Mandalay". The row for "Naypyitaw" is highlighted with a blue background.

	city
▶	Yangon
▶	Naypyitaw
▶	Mandalay

Generic Question

SHAAN MANSURI

QUESTION - 2

In which city is each branch?

- `select distinct city ,branch
from sales;`

Result Grid | Filter Rows:

	city	branch
▶	Yangon	A
	Naypyitaw	C
	Mandalay	B

QUESTION - 1

Product-Question

How many unique product lines does the data have?

```
6 • select distinct city ,branch  
7   from sales;
```

Result Grid	
	city
▶	Yangon
	Naypyitaw
	Mandalay

Product-Question

SHAAN MANSURI

QUESTION - 2

In which city is each branch?

```
• select distinct city ,branch  
from sales;
```

Result Grid | Filter Rows:

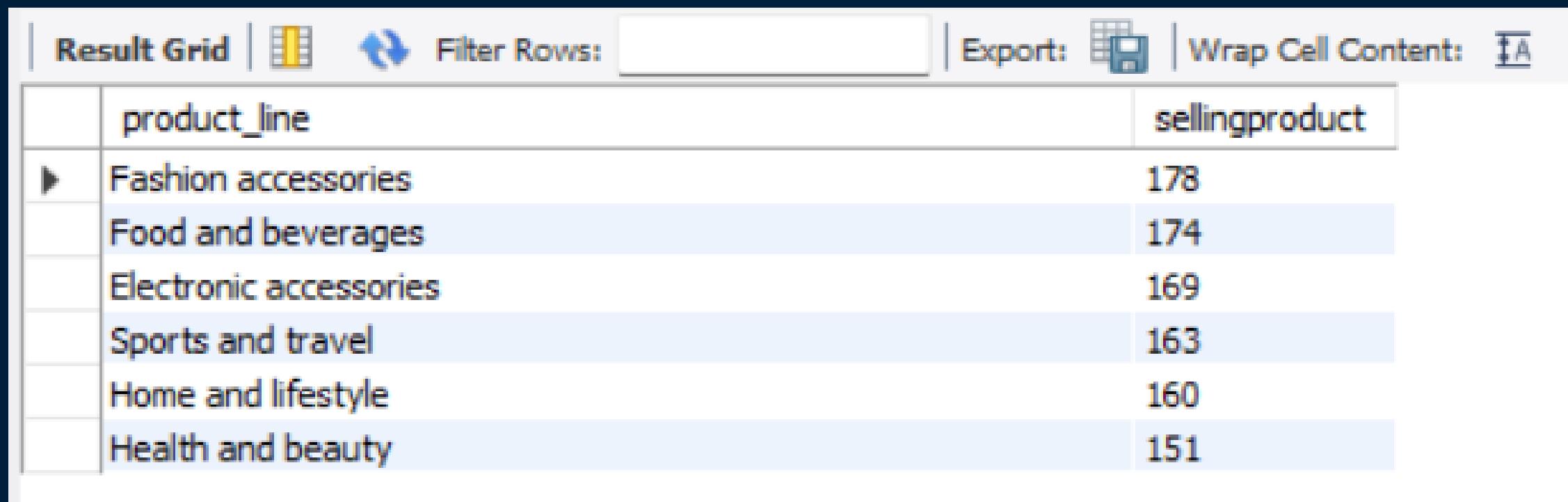
	city	branch
▶	Yangon	A
	Naypyitaw	C
	Mandalay	B

Product-Question

QUESTION - 3

What is the most selling product line?

```
7 • select product_line, count(product_line) as sellingproduct  
8   from sales  
9   group by product_line  
0   order by sellingproduct desc;
```



The screenshot shows a database query results grid. The grid has two columns: 'product_line' and 'sellingproduct'. The data is as follows:

	product_line	sellingproduct
▶	Fashion accessories	178
▶	Food and beverages	174
▶	Electronic accessories	169
▶	Sports and travel	163
▶	Home and lifestyle	160
▶	Health and beauty	151

Product-Question

SHAAN MANSURI

QUESTION - 4

What is the total revenue by month?

```
5 • select
6     month_name as month,
7     sum(total) as total_revenue
8
9 from sales
10 group by month_name
11 order by total_revenue desc;
```

	month	total_revenue
▶	January	116291.8680
▶	March	108867.1500
▶	February	95727.3765

Product-Question

SHAAN MANSURI

QUESTION - 5

What month had the largest COGS?

```
5 • SELECT
6     month_name as month,
7     sum(cogs) as cogs
8   from sales
9   group by month_name
0   order by cogs desc;
```

Result Grid | Filter Rows:

	month	cogs
▶	January	110754.16
	March	103683.00
	February	91168.93

Product-Question

QUESTION - 6

What product line had the largest revenue?

- `select`
 `product_line,`
 `sum(total) as total_revenue`
`from sales`
`group by product_line`
`order by total_revenue desc;`

product_line	total_revenue
Food and beverages	56144.8440
Fashion accessories	54305.8950
Sports and travel	53936.1270
Home and lifestyle	53861.9130
Electronic accessories	53783.2365
Health and beauty	48854.3790

Product-Question

SHAAN MANSURI

QUESTION - 7

What is the city with the largest revenue?

- **SELECT**

```
branch,  
city,  
SUM(total) AS total_revenue  
FROM sales  
GROUP BY city, branch  
ORDER BY total_revenue desc;
```

Result Grid | Filter Rows: | Export

	branch	city	total_revenue
▶	C	Naypyitaw	110490.7755
	A	Yangon	105861.0105
	B	Mandalay	104534.6085

Product-Question

SHAAN MANSURI

QUESTION - 8

What product line had the largest VAT?

- `select`
 `product_line,`
 `AVG(tax_pct) as avg_tax`
 `from sales`
 `group by product_line`
 `order by avg_tax ;`

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	product_line	avg_tax		
▶	Fashion accessories	14.52806181		
	Electronic accessories	15.15447632		
	Food and beverages	15.36531029		
	Health and beauty	15.40661591		
	Sports and travel	15.75697549		
	Home and lifestyle	16.03033124		

Product-Question

QUESTION - 10

Which branch sold more products than average product sold?

```
1 • SELECT  
2     branch, SUM(quantity) AS qty  
3 FROM  
4     sales  
5 GROUP BY branch  
6 HAVING SUM(quantity) > (SELECT  
7             AVG(quantity)  
8         FROM  
9             sales);
```

	branch	qty
▶	A	1849
	C	1828
	B	1795

Product-Question

QUESTION - 11

What is the most common product line by gender?

```
3 • SELECT
4     gender, product_line, COUNT(gender) AS total_cnt
5 FROM
6     sales
7 GROUP BY gender , product_line
8 ORDER BY total_cnt DESC;
```

	gender	product_line	total_cnt
▶	Female	Fashion accessories	96
	Female	Food and beverages	90
	Male	Health and beauty	88
	Female	Sports and travel	86
	Male	Electronic accessories	86
	Male	Food and beverages	84
	Female	Electronic accessories	83
	Male	Fashion accessories	82
	Male	Home and lifestyle	81
	Female	Home and lifestyle	79
	Male	Sports and travel	77
	Female	Health and beauty	63

Product-Question

QUESTION - 12

What is the average rating of each product line?

- **SELECT**

```
ROUND(AVG(rating), 2) AS ave_rating, product_line  
FROM  
sales  
GROUP BY product_line  
ORDER BY ave_rating DESC;
```

The screenshot shows a database query results grid with the following data:

	ave_rating	product_line
▶	7.11	Food and beverages
	7.03	Fashion accessories
	6.98	Health and beauty
	6.91	Electronic accessories
	6.86	Sports and travel
	6.84	Home and lifestyle

Sales Question

QUESTION - 1

Number of sales made in each time of the day per weekday

- ```
select
 time_of_day,
 count(*) as total_sales
 from sales
 group by time_of_day
 order by total_sales desc;
```

Result Grid | Filter Rows:

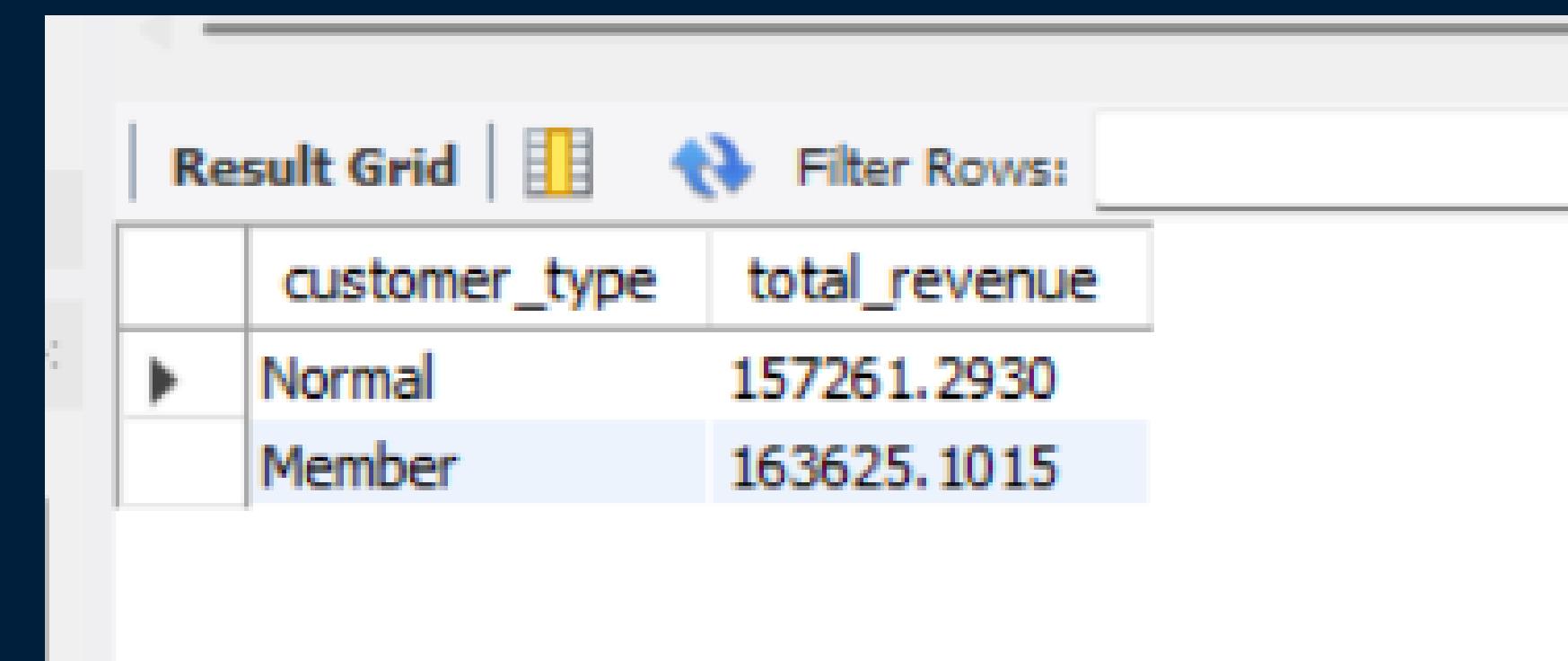
|   | time_of_day | total_sales |
|---|-------------|-------------|
| ▶ | Evening     | 429         |
|   | Afternoon   | 376         |
|   | Morning     | 190         |

# Sales Question

Which of the customer types brings the most revenue?

## QUESTION - 2

```
5 • SELECT
5 customer_type, SUM(total) AS total_revenue
6 FROM
7 sales
8 GROUP BY customer_type
9 ORDER BY total_revenue;
```



The screenshot shows a database query results grid. At the top, there are tabs for 'Result Grid' and 'Filter Rows'. The result grid displays two columns: 'customer\_type' and 'total\_revenue'. There are two rows of data: one for 'Normal' customers with a total revenue of 157261.2930, and one for 'Member' customers with a total revenue of 163625.1015.

|   | customer_type | total_revenue |
|---|---------------|---------------|
| ▶ | Normal        | 157261.2930   |
|   | Member        | 163625.1015   |

# Sales Question

## QUESTION - 3

Which city has the largest tax percent/ VAT (Value Added Tax)?

```
5 • SELECT
6 city, ROUND(AVG(tax_pct), 2) AS avg_tax_pct
7
8 FROM
9 sales
10 GROUP BY city
11 ORDER BY avg_tax_pct DESC;
```

Result Grid | Filter Rows:

|   | city      | avg_tax_pct |
|---|-----------|-------------|
| ▶ | Naypyitaw | 16.09       |
|   | Mandalay  | 15.13       |
|   | Yangon    | 14.87       |

# Sales Question

Which customer type pays the most in VAT?

## QUESTION - 4

```
3 • SELECT
4 customer_type, AVG(tax_pct) AS total_tax
5 FROM
6 sales
7 GROUP BY customer_type
8 ORDER BY total_tax;
```

Result Grid | Filter Rows:  Export

|   | customer_type | total_tax   |
|---|---------------|-------------|
| ▶ | Normal        | 15.09805040 |
| ▶ | Member        | 15.61457214 |

## Customer- Question

How many unique customer types does the data have?

### QUESTION - 1

- **SELECT DISTINCT**  
**customer\_type**  
**FROM**  
**sales;**

Result Grid | Filter

|   | customer_type |
|---|---------------|
| ▶ | Normal        |
|   | Member        |

## Customer- Question

How many unique payment methods does the data have?

### QUESTION - 2

```
SELECT DISTINCT
payment
FROM
sales;
```

| Result Grid |             |
|-------------|-------------|
|             | payment     |
| ▶           | Credit card |
|             | Ewallet     |
|             | Cash        |

# Customer- Question

SHAAN MANSURI

What is the most common customer type?

## QUESTION - 3

- **SELECT**  
`customer_type,  
count(*) as count  
FROM sales  
GROUP BY customer_type  
ORDER BY count DESC;`

Result Grid | Filter Rows:

|   | customer_type | count |
|---|---------------|-------|
| ▶ | Member        | 499   |
|   | Normal        | 496   |

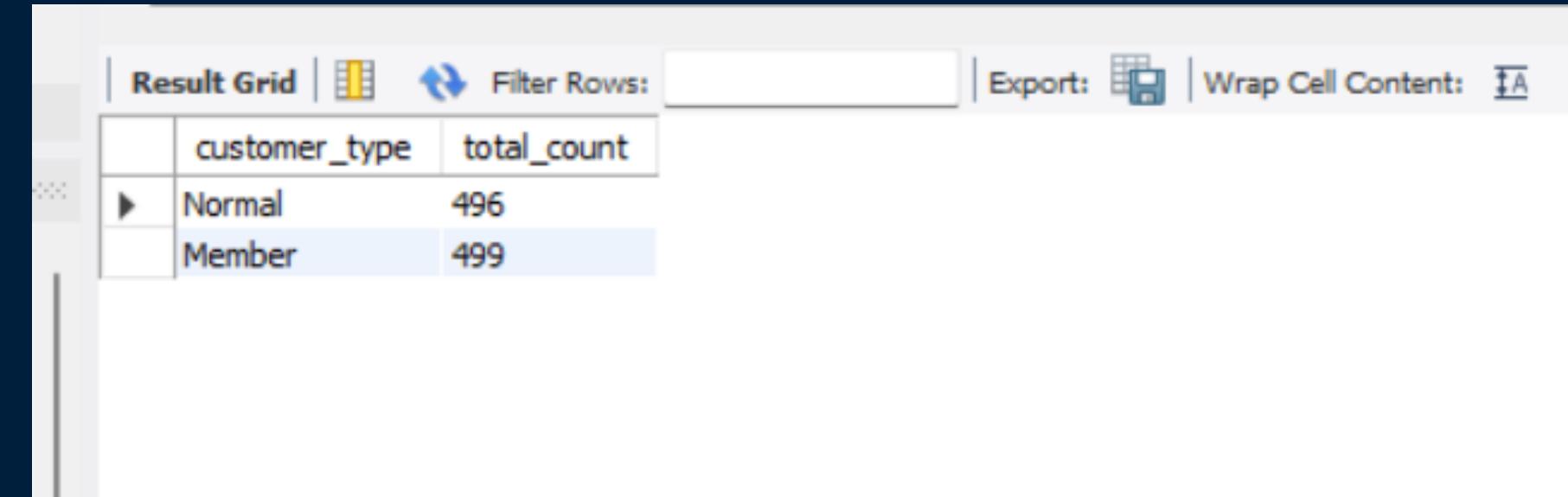
# Customer- Question

SHAAN MANSURI

Which customer type buys the most?

## QUESTION - 4

```
13 • SELECT
14 customer_type, COUNT(*) as total_count
15 FROM
16 sales
17 GROUP BY customer_type;
18
19
```



The screenshot shows the MySQL Workbench interface with a result grid. The grid has two columns: 'customer\_type' and 'total\_count'. There are two rows of data: one for 'Normal' with a count of 496, and one for 'Member' with a count of 499. The 'Result Grid' tab is selected at the top.

|   | customer_type | total_count |
|---|---------------|-------------|
| ▶ | Normal        | 496         |
|   | Member        | 499         |

## Customer- Question

SHAAN MANSURI

What is the gender of most of the customers?

### QUESTION - 5

- **SELECT**

```
gender, COUNT(*) AS gender_count
FROM
 sales
GROUP BY gender
ORDER BY gender_count DESC;
```

Result Grid | Filter Rows:

|   | gender | gender_count |
|---|--------|--------------|
| ▶ | Male   | 498          |
|   | Female | 497          |

# Customer- Question

How many unique customer types does the data have?

## QUESTION - 6

- **SELECT**

```
gender, COUNT(*) AS gender_cnt
FROM
sales
WHERE
branch = 'B'
GROUP BY gender
ORDER BY gender_cnt DESC;
```

The screenshot shows the MySQL Workbench interface with the 'Result Grid' tab selected. The results of the query are displayed in a table:

|   | gender | gender_cnt |
|---|--------|------------|
| ▶ | Male   | 169        |
|   | Female | 160        |

## Customer- Question

Which time of the day do customers give most ratings?

### QUESTION - 7

- `SELECT  
 time_of_day, AVG(rating) AS avg_rating  
 FROM  
 sales  
 GROUP BY time_of_day  
 ORDER BY avg_rating DESC;`

Result Grid | Filter Rows:

|   | time_of_day | avg_rating |
|---|-------------|------------|
| ▶ | Afternoon   | 7.02340    |
|   | Morning     | 6.94474    |
|   | Evening     | 6.90536    |

## Customer- Question

Which time of the day do customers give most ratings per branch?

### QUESTION - 8

```
SELECT
 time_of_day, AVG(rating) AS avg_rating
FROM
 sales
WHERE
 branch = 'A'
GROUP BY time_of_day
ORDER BY avg_rating DESC;
```

Result Grid | Filter Rows:

|   | time_of_day | avg_rating |
|---|-------------|------------|
| ▶ | Afternoon   | 7.02340    |
|   | Morning     | 6.94474    |
|   | Evening     | 6.90536    |

# Customer- Question

Which day fo the week has the best avg ratings?

## QUESTION - 9

- **SELECT**

```
 day_name, AVG(rating) AS avg_rating
```

```
FROM
```

```
sales
```

```
GROUP BY day_name
```

```
ORDER BY avg_rating DESC;
```

|   | day_name  | avg_rating |
|---|-----------|------------|
| ▶ | Monday    | 7.13065    |
|   | Friday    | 7.05507    |
|   | Tuesday   | 7.00316    |
|   | Sunday    | 6.98864    |
|   | Saturday  | 6.90183    |
|   | Thursday  | 6.88986    |
|   | Wednesday | 6.76028    |

## Customer- Question

Which day of the week has the best average ratings per branch?

### QUESTION - 10

```
39 • SELECT
40 day_name, COUNT(day_name) total_sales
41 FROM
42 sales
43 WHERE
44 branch = 'C'
45 GROUP BY day_name
46 ORDER BY total_sales DESC;
47
```

|   | day_name  | total_sales |
|---|-----------|-------------|
| ▶ | Tuesday   | 54          |
|   | Saturday  | 54          |
|   | Wednesday | 50          |
|   | Thursday  | 48          |
|   | Sunday    | 46          |
|   | Monday    | 38          |
|   | Friday    | 37          |

**THANK  
YOU!!**

SHAAN MANSURI