Report on Walmart Sales Analysis

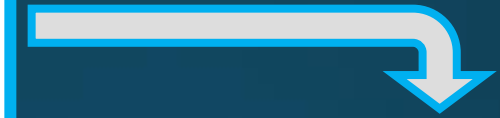
PREPARED BY
SHAKIL AHAMMED

Overview

The Walmart Sales Analysis SQL Project aims to analyze and interpret the sales data of Walmart, one of the largest retail chains in the world. The project utilizes SQL (Structured Query Language) to query and manage large datasets, providing a comprehensive view of sales patterns, customer behavior, and inventory management. The analysis is focused on identifying key trends, performance metrics, and potential areas for optimization.

1. How many unique product lines does the data have?

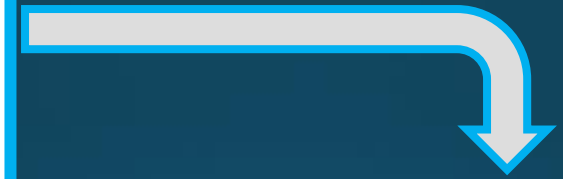
```
SELECT
    DISTINCT product_line
FROM sales;
```



	product_line character varying (100)
1	Fashion accessories
2	Electronic accessories
3	Health and beauty
4	Food and beverages
5	Sports and travel

2. What is the most selling product line?

```
SELECT
    SUM(quantity) as qty,
    product_line
FROM sales
GROUP BY product_line
ORDER BY qty DESC;
```



	qty bigint	product_line character varying (100)
1	971	Electronic accessories
2	952	Food and beverages
3	920	Sports and travel
4	911	Home and lifestyle
5	902	Fashion accessories

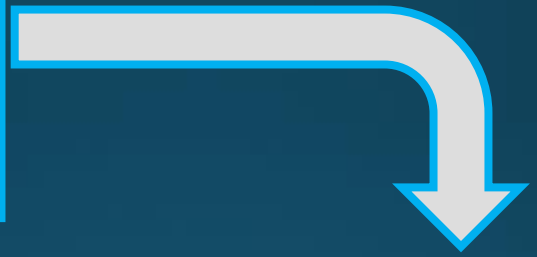
3. What is the total revenue by month?

```sql
SELECT
    month,
    SUM(total) AS total_revenue
FROM sales
GROUP BY month
ORDER BY total_revenue DESC;
```

	month character varying (10)	total_revenue numeric
1	JANUARY	116291.8680
2	MARCH	109455.5070
3	FEBRUARY	97219.3740

4. What month had the largest COGS?

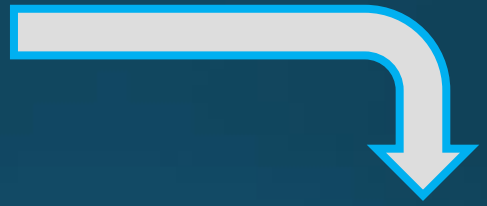
```
SELECT
 month,
    SUM(cogs) AS cogs
FROM sales
GROUP BY month
ORDER BY cogs DESC;
```



	month character varying (10)	cogs numeric
1	JANUARY	110754.16
2	MARCH	104243.34
3	FEBRUARY	92589.88

5. 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 character varying (100)	total_revenue numeric
1	Food and beverages	56144.8440
2	Sports and travel	55122.8265
3	Electronic accessories	54337.5315
4	Fashion accessories	54305.8950
5	Home and lifestyle	53861.9130

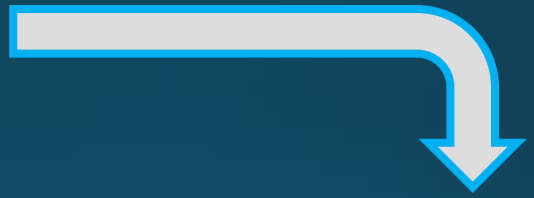
6. What is the city with the largest revenue?

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

	branch character varying (5)	city character varying (30)	total_revenue numeric
1	C	Naypyitaw	110568.7065
2	A	Yangon	106200.3705
3	B	Mandalay	106197.6720

7. What product line had the largest VAT?

```
SELECT product_line,
SUM(tax_pct) as total_tax
FROM sales
GROUP BY product_line
ORDER BY total_tax DESC;
```



	product_line character varying (100)	total_tax numeric
1	Food and beverages	2673.5640
2	Sports and travel	2624.8965
3	Electronic accessories	2587.5015
4	Fashion accessories	2585.9950
5	Home and lifestyle	2564.8530

8. Fetch each product line and add a column to those product line showing "Good", "Bad". Good if its greater than average sales?

```sql
WITH overall_avg AS (
    SELECT AVG(quantity) AS avg_qnty
    FROM sales
)
SELECT product_line,
    CASE
        WHEN AVG(quantity) > (SELECT avg_qnty FROM overall_avg) THEN 'Good
        ELSE 'Bad'
    END AS remark
FROM sales
GROUP BY product_line;
```

	product_line character varying (100)	remark text
1	Fashion accessories	Bad
2	Electronic accessories	Good
3	Health and beauty	Good
4	Food and beverages	Bad
5	Sports and travel	Good

9. Which branch sold more products than average product sold?

```sql
SELECT branch,
SUM(quantity) AS qnty
FROM sales
GROUP BY branch
HAVING SUM(quantity) > (SELECT AVG(quantity) FROM sales);
```

	branch character varying (5)	qnty bigint
1	A	1859
2	C	1831
3	B	1820

10. What is the most common product line by gender?

```sql
SELECT gender, product_line,
COUNT(gender) AS total_cnt
FROM sales
GROUP BY gender, product_line
ORDER BY total_cnt DESC;
```

	gender character varying (30)	product_line character varying (100)	total_cnt bigint
1	Female	Fashion accessories	96
2	Female	Food and beverages	90
3	Female	Sports and travel	88
4	Male	Health and beauty	88
5	Male	Electronic accessories	86

11. What is the average rating of each product line?

```sql
SELECT ROUND(AVG(rating), 2) as avg_rating,
product_line FROM sales
GROUP BY product_line
ORDER BY avg_rating DESC;
```

	avg_rating numeric	product_line character varying (100)
1	7.11	Food and beverages
2	7.03	Fashion accessories
3	7.00	Health and beauty
4	6.92	Electronic accessories
5	6.92	Sports and travel

12. What is the most common customer type?/ Which customer type buys the most?

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

	customer_type character varying (30)	count bigint
1	Normal	499
2	Member	501

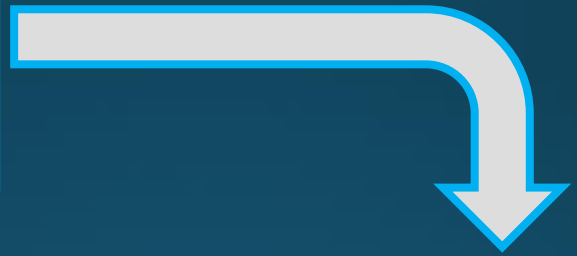
13. What is the gender of most of the customers?

```sql
SELECT gender,
COUNT(*) as gender_cnt
FROM sales
GROUP BY gender
ORDER BY gender_cnt DESC;
```

	gender character varying (30)	gender_cnt bigint
1	Female	501
2	Male	499

14. What is the gender distribution per branch?

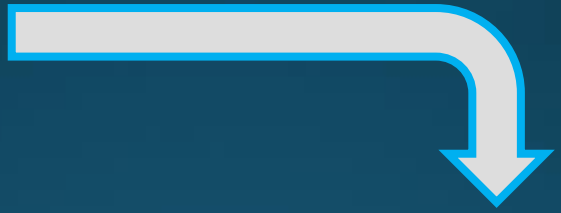
```sql
SELECT branch,
COUNT(gender) as gender_cnt
FROM sales
GROUP BY branch
ORDER BY gender_cnt DESC;
```



	branch character varying (5)	gender_cnt bigint
1	A	340
2	B	332
3	C	328

15. Gender per branch is more or less the same hence, I don't think has an effect of the sales per branch and other factors. Which time of the day do customers give most ratings?

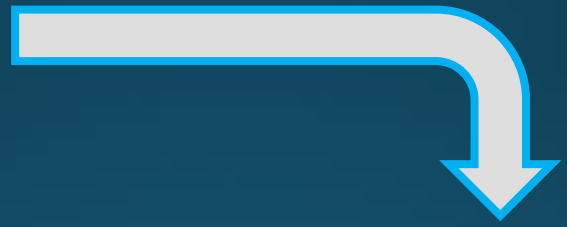
```
SELECT time_of_day,
AVG(rating) AS avg_rating
FROM sales
GROUP BY time_of_day
ORDER BY avg_rating DESC;
```



	time_of_day character varying (20)	avg_rating numeric
1	Afternoon	7.0312997347480106
2	Morning	6.9607329842931937
3	Evening	6.9268518518518519

16. Looks like time of the day does not really affect the rating, its more or less the same rating each time of the day. Which time of the day do customers give most ratings per branch?

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



	time_of_day character varying (20)	branch character varying (5)	avg_rating numeric
1	Afternoon	A	7.1888888888888889
2	Evening	C	7.1188811188811189
3	Afternoon	C	7.0666666666666667
4	Morning	A	7.0054794520547945
5	Morning	C	6.9745762711864407

17. Branch A and C are doing well in ratings, branch B needs to do a little more to get better ratings. Which day of the week has the best avg ratings?

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

	day_name character varying (10)	avg_rating numeric
1	Monday	7.1536000000000000
2	Friday	7.0762589928057554
3	Sunday	7.0112781954887218
4	Tuesday	7.0031645569620253
5	Saturday	6.9018292682926829

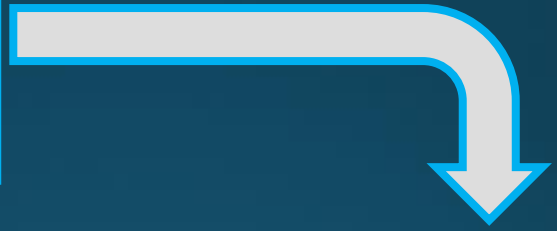
18. Monday, Friday and Sunday are the top best days for good ratings why is that the case, how many sales are made on these days? Which day of the week has the best average ratings per branch?

```sql
SELECT
day_name, branch,
COUNT(day_name), SUM(total) as total_sales
FROM sales
GROUP BY day_name, branch
ORDER BY total_sales DESC;
```

	day_name character varying (10)	branch character varying (5)	count bigint	total_sales numeric
1	Saturday	B	60	21284.4240
2	Tuesday	B	53	18859.2390
3	Saturday	C	54	18070.9725
4	Wednesday	C	50	17755.8150
5	Tuesday	C	54	17667.7935

19. 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;
```



	time_of_day character varying (20)	total_sales bigint
1	Evening	432
2	Afternoon	377
3	Morning	191

20. Evenings experience most sales, the stores are filled during the evening hours. Which of the customer types brings the most revenue?

```sql
SELECT customer_type,
SUM(total) AS total_revenue
FROM sales
GROUP BY customer_type
ORDER BY total_revenue DESC;
```

	customer_type character varying (30)	total_revenue numeric
1	Member	164223.4440
2	Normal	158743.3050

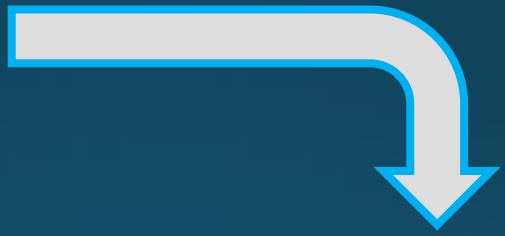
21. Which city has the largest tax/VAT percent?

```sql
SELECT city,
ROUND(AVG(tax_pct), 2) AS avg_tax_pct
FROM sales
GROUP BY city
ORDER BY avg_tax_pct DESC;
```

	city character varying (30)	avg_tax_pct numeric
1	Naypyitaw	16.05
2	Mandalay	15.23
3	Yangon	14.87

22. Which customer type pays the most in VAT?

```sql
SELECT customer_type,
AVG(tax_pct) AS total_tax
FROM sales
GROUP BY customer_type
ORDER BY total_tax DESC;
```



	customer_type character varying (30)	total_tax numeric
1	Member	15.6091097804391218
2	Normal	15.1487074148296593