SQL Project

Walmart Sales Data Analysis

Created by

Shaik Mohammad Ismael Jabivulla

Data Exploration

Open up Walmart Sale Data, explore different columns, and check for inconsistent data, if found then clean the data file with no Null Value.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	invoice id	branch	city	customer_type	gender	product_line	unit_price	quantity	tax_percentage	total	date_value	time_value	payment	cogs	gross_margin_percentage	gross_income	rating	
2	750-67-8428	A	Yangon	Member	Female	Health and beauty	74.69	7	26.1415	548.9715	05-01-2019	13:08:00	Ewallet	522.83	4.761904762	26.1415	9.1	
3	226-31-3081	C	Naypyitaw	Normal	Female	Electronic accessories	15.28	5	3.82	80.22	08-03-2019	10:29:00	Cash	76.4	4.761904762	3.82	9.6	
4	631-41-3108	A	Yangon	Normal	Male	Home and lifestyle	46.33	7	16.2155	340.5255	03-03-2019	13:23:00	Credit card	324.31	4.761904762	16.2155	7.4	
5	123-19-1176	A	Yangon	Member	Male	Health and beauty	58.22	8	23.288	489.048	27-01-2019	20:33:00	Ewallet	465.76	4.761904762	23.288	8.4	
6	373-73-7910	A	Yangon	Normal	Male	Sports and travel	86.31	7	30.2085	634.3785	08-02-2019	10:37:00	Ewallet	604.17	4.761904762	30.2085	5.3	
7	699-14-3026	C	Naypyitaw	Normal	Male	Electronic accessories	85.39	7	29.8865	627.6165	25-03-2019	18:30:00	Ewallet	597.73	4.761904762	29.8865	4.1	
8	355-53-5943	A	Yangon	Member	Female	Electronic accessories	68.84	6	20.652	433.692	25-02-2019	14:36:00	Ewallet	413.04	4.761904762	20.652	5.8	
9	315-22-5665	C	Naypyitaw	Normal	Female	Home and lifestyle	73.56	10	36.78	772.38	24-02-2019	11:38:00	Ewallet	735.6	4.761904762	36.78	8	
10	665-32-9167	A	Yangon	Member	Female	Health and beauty	36.26	2	3.626	76.146	10-01-2019	17:15:00	Credit card	72.52	4.761904762	3.626	7.2	
11	692-92-5582	B	Mandalay	Member	Female	Food and beverages	54.84	3	8.226	172.746	20-02-2019	13:27:00	Credit card	164.52	4.761904762	8.226	5.9	
12	351-62-0822	B	Mandalay	Member	Female	Fashion accessories	14.48	4	2.896	60.816	06-02-2019	18:07:00	Ewallet	57.92	4.761904762	2.896	4.5	
13	529-56-3974	B	Mandalay	Member	Male	Electronic accessories	25.51	4	5.102	107.142	09-03-2019	17:03:00	Cash	102.04	4.761904762	5.102	6.8	
14	365-64-0515	A	Yangon	Normal	Female	Electronic accessories	46.95	5	11.7375	246.4875	12-02-2019	10:25:00	Ewallet	234.75	4.761904762	11.7375	7.1	
15	252-56-2699	A	Yangon	Normal	Male	Food and beverages	43.19	10	21.595	453.495	07-02-2019	16:48:00	Ewallet	431.9	4.761904762	21.595	8.2	
16	829-34-3910	A	Yangon	Normal	Female	Health and beauty	71.38	10	35.69	749.49	29-03-2019	19:21:00	Cash	713.8	4.761904762	35.69	5.7	
17	299-46-1805	B	Mandalay	Member	Female	Sports and travel	93.72	6	28.116	590.436	15-01-2019	16:19:00	Cash	562.32	4.761904762	28.116	4.5	
18	656-95-9349	A	Yangon	Member	Female	Health and beauty	68.93	7	24.1255	506.6355	11-03-2019	11:03:00	Credit card	482.51	4.761904762	24.1255	4.6	
19	765-26-6951	A	Yangon	Normal	Male	Sports and travel	72.61	6	21.783	457.443	01-01-2019	10:39:00	Credit card	435.66	4.761904762	21.783	6.9	
20	329-62-1586	A	Yangon	Normal	Male	Food and beverages	54.67	3	8.2005	172.2105	21-01-2019	18:00:00	Credit card	164.01	4.761904762	8.2005	8.6	
21	319-50-3348	B	Mandalay	Normal	Female	Home and lifestyle	40.3	2	4.03	84.63	11-03-2019	15:30:00	Ewallet	80.6	4.761904762	4.03	4.4	
22	300-71-4605	C	Naypyitaw	Member	Male	Electronic accessories	86.04	5	21.51	451.71	25-02-2019	11:24:00	Ewallet	430.2	4.761904762	21.51	4.8	
23	371-85-5789	B	Mandalay	Normal	Male	Health and beauty	87.98	3	13.197	277.137	05-03-2019	10:40:00	Ewallet	263.94	4.761904762	13.197	5.1	
24	273-16-6619	B	Mandalay	Normal	Male	Home and lifestyle	33.2	2	3.32	69.72	15-03-2019	12:20:00	Credit card	66.4	4.761904762	3.32	4.4	
25	636-48-8204	A	Yangon	Normal	Male	Electronic accessories	34.56	5	8.64	181.44	17-02-2019	11:15:00	Ewallet	172.8	4.761904762	8.64	9.9	
26	549-59-1358	A	Yangon	Member	Male	Sports and travel	88.63	3	13.2945	279.1845	02-03-2019	17:36:00	Ewallet	265.89	4.761904762	13.2945	6	
27	227-03-5010	A	Yangon	Member	Female	Home and lifestyle	52.59	8	21.036	441.756	22-03-2019	19:20:00	Credit card	420.72	4.761904762	21.036	8.5	
28	649-29-6775	B	Mandalay	Normal	Male	Fashion accessories	33.52	1	1.676	35.196	08-02-2019	15:31:00	Cash	33.52	4.761904762	1.676	6.7	
29	189-17-4241	A	Yangon	Normal	Female	Fashion accessories	87.67	2	8.767	184.107	10-03-2019	12:17:00	Credit card	175.34	4.761904762	8.767	7.7	
30	145-94-9061	B	Mandalay	Normal	Female	Food and beverages	88.36	5	22.09	463.89	25-01-2019	19:48:00	Cash	441.8	4.761904762	22.09	9.6	

Creating Database & Table

Open up the SQL Server Database and create a new connection, after creating the Database and Data Table as per CSV data columns with datatype and constraints.

```sql
----- Create database

CREATE DATABASE WalmartSales;

----- Create table

CREATE TABLE sales(
     invoice_id VARCHAR(30) NOT NULL PRIMARY KEY,
     branch VARCHAR(5) NOT NULL,
     city VARCHAR(30) NOT NULL,
     customer_type VARCHAR(30) NOT NULL,
     gender VARCHAR(30) NOT NULL,
     product_line VARCHAR(100) NOT NULL,
     unit_price FLOAT NOT NULL,
     quantity INT NOT NULL,
     tax_percentage FLOAT NOT NULL,
     total FLOAT NOT NULL,
     date_value DATETIME NOT NULL,
     time_value TIME NOT NULL,
     payment VARCHAR(15) NOT NULL,
     cogs FLOAT NOT NULL,
     gross_margin_percentage FLOAT NOT NULL,
     gross_income FLOAT NOT NULL,
     rating FLOAT NOT NULL
);
```

Feature Engineering

- Introduce a "time_of_day" column to provide insights into sales during Morning, Afternoon, and Evening periods. This addition will aid in determining the peak sales time of the day.
- Incorporate a "day_name" column that captures the days of the week for each transaction (Mon, Tue, Wed, Thur, Fri). This inclusion facilitates an analysis of each branch's busiest days, answering questions about weekly patterns.
- Integrate a "month_name" column, extracting the months of the year for each transaction (Jan, Feb, Mar). This modification assists in identifying the months with the highest sales and profit, contributing to a comprehensive analysis of yearly trends.

Adding Column

-- Add the time_of_day column

```
----- time_of_day

SELECT time_value,
(CASE
    WHEN time_value BETWEEN '00:00:00' AND '12:00:00' THEN 'Morning'
    WHEN time_value BETWEEN '12:00:00' AND '16:00:00' THEN 'Afternoon'
ELSE 'Evening'
END) AS time_of_day
FROM sales;

ALTER TABLE sales
ADD time_of_day VARCHAR(20);
```

Adding Column

-- Add the time_of_day column

```sql
UPDATE sales
SET time_of_day = (
CASE
    WHEN time_value BETWEEN '00:00:00' AND '12:00:00' THEN 'Morning'
    WHEN time_value BETWEEN '12:00:00' AND '16:00:00' THEN 'Afternoon'
ELSE 'Evening'
END
);

SELECT invoice_id, time_of_day FROM sales;
```

Adding Column

-- Add day_name column

```
----- day_name

SELECT date_value,
DATENAME(w, date_value) AS [day_name]
FROM [sales];

ALTER TABLE sales
ADD day_name VARCHAR(10);

UPDATE sales
SET day_name = DATENAME(w, date_value);

SELECT invoice_id, day_name FROM sales;
```

Adding Column

-- Add month_name column

```
----- month_name

SELECT date_value,
DATENAME(MONTH, date_value) AS [month_name]
FROM [sales];

ALTER TABLE sales
ADD month_name VARCHAR(10);

UPDATE sales
SET month_name = DATENAME(MONTH, date_value);

SELECT invoice_id, month_name FROM sales;
```

```sql
----- How many unique cities does the data have?

SELECT DISTINCT city FROM sales;
```

100 %

Results | Messages

	city
1	Naypyitaw
2	Yangon
3	Mandalay

```sql
----- In which city is each branch?

SELECT DISTINCT city, branch FROM sales;
```

100 %

Results | Messages

	city	branch
1	Mandalay	B
2	Naypyitaw	C
3	Yangon	A

```sql
----- How many unique product lines does the data have?

SELECT DISTINCT product_line FROM sales;
```

100 %

Results | Messages

	product_line
1	Fashion accessories
2	Health and beauty
3	Electronic accessories
4	Food and beverages
5	Sports and travel
6	Home and lifestyle

```sql
----- What is the most common payment method?

SELECT payment, COUNT(payment) AS count_pay
FROM sales
GROUP BY payment
ORDER BY count_pay DESC;
```

100 %

Results | Messages

	payment	count_pay
1	Ewallet	345
2	Cash	344
3	Credit card	311

```sql
----- What is the most selling product line?

SELECT product_line, SUM(quantity) AS total_sales
FROM sales
GROUP BY product_line
ORDER BY total_sales DESC;
```

100 %

Results | Messages

	product_line	total_sales
1	Electronic accessories	971
2	Food and beverages	952
3	Sports and travel	920
4	Home and lifestyle	911
5	Fashion accessories	902
6	Health and beauty	854

```sql
------What is the total revenue by month?

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

100 %

Results | Messages

	month_name	total_revenue
1	January	116291.868041039
2	March	109455.507236481
3	February	97219.3739280701

```sql
------What month had the largest COGS?

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

100 %

Results | Messages

	month_name	total_cogs
1	January	110754.159928322
2	March	104243.339855194
3	February	92589.8799791336

```sql
------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;
```

100 %

Results | Messages

	product_line	total_revenue
1	Food and beverages	56144.8439311981
2	Sports and travel	55122.8265810013
3	Electronic accessories	54337.531457901
4	Fashion accessories	54305.8951845169
5	Home and lifestyle	53861.9131307602
6	Health and beauty	49193.7389202118

```sql
------What is the city with the largest revenue?

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

100 %

Results | Messages

	city	branch	total_revenue
1	Naypyitaw	C	110568.706618309
2	Yangon	A	106200.370399475
3	Mandalay	B	106197.672187805

```sql
-----What product line had the largest VAT?

SELECT product_line, avg(tax_percentage) AS avg_tax
FROM sales
GROUP BY product_line
ORDER BY avg_tax DESC;
```

100 %

Results | Messages

	product_line	avg_tax
1	Home and lifestyle	16.0303312379867
2	Sports and travel	15.8126295521317
3	Health and beauty	15.4115723830305
4	Food and beverages	15.3653102914492
5	Electronic accessories	15.2205970518729
6	Fashion accessories	14.5280618064859

```sql
-----Which branch sold more products than average product sold?

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

100 %

Results | Messages

	branch	(No column name)
1	A	1859
2	C	1831
3	B	1820

```sql
------What is the most common product line by gender?

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

100 %

Results | Messages

	gender	product_line	count_gender
1	Female	Fashion accessories	96
2	Female	Food and beverages	90
3	Male	Health and beauty	88
4	Female	Sports and travel	88
5	Male	Electronic accessories	86
6	Male	Food and beverages	84
7	Female	Electronic accessories	84
8	Male	Fashion accessories	82
9	Male	Home and lifestyle	81
10	Female	Home and lifestyle	79
11	Male	Sports and travel	78
12	Female	Health and beauty	64

```sql
------What is the average rating of each product line?

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

100 %

Results | Messages

	product_line	avg_rating
1	Food and beverages	7.11
2	Fashion accessories	7.03
3	Health and beauty	7
4	Electronic accessories	6.92
5	Sports and travel	6.92
6	Home and lifestyle	6.84

```sql
-- How many unique customer types does the data have?


SELECT DISTINCT customer_type
FROM sales;
```

133 %

Results | Messages

	customer_type
1	Normal
2	Member

```sql
---How many unique payment methods does the data have?


SELECT DISTINCT payment
FROM sales;
```

133 %

Results | Messages

	payment
1	Ewallet
2	Cash
3	Credit card

```sql
---What is the most common customer type?

SELECT customer_type, COUNT(customer_type) AS count_customer
FROM sales
GROUP BY customer_type
ORDER BY count_customer DESC;
```

133 %

Results | Messages

	customer_type	count_customer
1	Member	501
2	Normal	499

```sql
---Which customer type buys the most?

SELECT customer_type, COUNT(*) AS count_customer
FROM sales
GROUP BY customer_type
ORDER BY count_customer DESC;
```

133 %

Results | Messages

	customer_type	count_customer
1	Member	501
2	Normal	499

```sql
---What is the gender of most of the customers?

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

133 %

Results | Messages

	gender	count_gender
1	Female	501
2	Male	499

```sql
---What is the gender distribution per branch?

SELECT gender, branch, COUNT(*) AS count_gender
FROM sales
GROUP BY gender, branch
ORDER BY count_gender DESC;
```

133 %

Results | Messages

	gender	branch	count_gender
1	Male	A	179
2	Female	C	178
3	Male	B	170
4	Female	B	162
5	Female	A	161
6	Male	C	150

```sql
---Which time of the day do customers give most ratings?

SELECT time_of_day,ROUND(AVG(rating), 2) AS avg_rating
FROM sales
GROUP BY time_of_day
ORDER BY avg_rating DESC;
```

133 %

Results | Messages

	time_of_day	avg_rating
1	Afternoon	7.03
2	Morning	6.96
3	Evening	6.93

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

SELECT branch, time_of_day, ROUND(AVG(rating), 2) AS avg_rating
FROM sales
GROUP BY time_of_day, branch
ORDER BY avg_rating DESC;
```

133 %

Results | Messages

	branch	time_of_day	avg_rating
1	A	Afternoon	7.19
2	C	Evening	7.12
3	C	Afternoon	7.07
4	A	Morning	7.01
5	C	Morning	6.97
6	A	Evening	6.89
7	B	Morning	6.89
8	B	Afternoon	6.84
9	B	Evening	6.77

```sql
---Which day of the week has the best avg ratings?

SELECT day_name, ROUND(AVG(rating), 2) AS avg_rating
FROM sales
GROUP BY day_name
ORDER BY avg_rating DESC;
```

133 %

Results | Messages

	day_name	avg_rating
1	Monday	7.15
2	Friday	7.08
3	Sunday	7.01
4	Tuesday	7
5	Saturday	6.9
6	Thursday	6.89
7	Wednesday	6.81

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

SELECT day_name, branch, ROUND(AVG(rating), 2) AS avg_rating
FROM sales
GROUP BY day_name, branch
ORDER BY avg_rating DESC;
```

133 %

Results | Messages

	day_name	branch	avg_rating
1	Monday	B	7.34
2	Friday	A	7.31
3	Friday	C	7.28
4	Saturday	C	7.23
5	Monday	A	7.1
6	Sunday	A	7.08
7	Tuesday	A	7.06
8	Wednesday	C	7.06
9	Monday	C	7.04
10	Sunday	C	7.03

```sql
---Number of sales made in each time of the day per weekday?

SELECT time_of_day, COUNT(*) AS total_sales
FROM sales
WHERE day_name = 'Monday'
GROUP BY time_of_day
ORDER BY total_sales DESC;
```

133 %

Results | Messages

	time_of_day	total_sales
1	Evening	56
2	Afternoon	48
3	Morning	21

```sql
---Which of the customer types brings the most revenue?

SELECT customer_type, ROUND(SUM(total), 0) AS total_revenue
FROM sales
GROUP BY customer_type
ORDER BY total_revenue DESC;
```

133 %

Results | Messages

	customer_type	total_revenue
1	Member	164223
2	Normal	158743

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

SELECT city, AVG(tax_percentage) AS largest_vat
FROM sales
GROUP BY city
ORDER BY largest_vat DESC;
```

133 %

Results | Messages

	city	largest_vat
1	Naypyitaw	16.0523673185125
2	Mandalay	15.2320241358984
3	Yangon	14.8740014807266

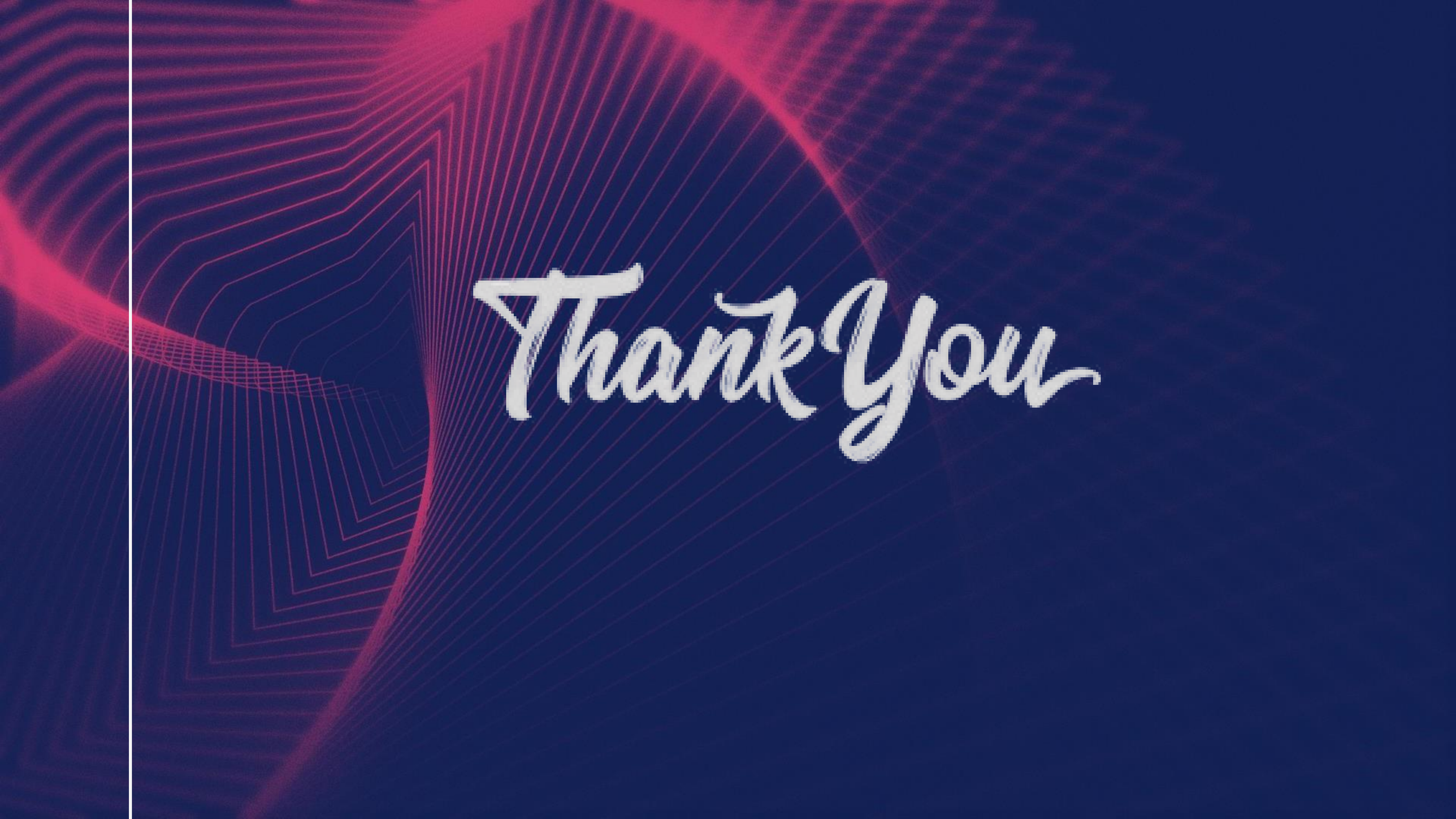
```sql
---Which customer type pays the most in VAT?

SELECT customer_type, AVG(tax_percentage) AS largest_vat
FROM sales
GROUP BY customer_type
ORDER BY largest_vat DESC;
```

133 %

Results | Messages

	customer_type	largest_vat
1	Member	15.6091097726555
2	Normal	15.1487074167074


Thank You