Walmart Sales Analysis SQL Project

9 min read · Sep 8, 2023

The goal of this project is to examine Walmart Sales data in order to comprehend top-performing branches and items, sales trends for various products, and consumer behavior. The goal is to investigate how sales methods might be made better and more effective. The [Kaggle Walmart Sales Forecasting Competition](https://www.kaggle.com/c/walmart-recruiting-store-sales-forecasting) is where the dataset was collected.

“In this hiring contest, job hopefuls are given historical sales information for 45 Walmart locations spread across several geographies. Participants must forecast the sales for each department in each store, which has multiple departments. Selected holiday markdown events are included in the dataset to increase the difficulty. Although it is recognized that these markdowns have an influence on sales, it can be difficult to forecast which departments will be impacted and how much.

Purpose

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

About Data

This dataset contains sales transactions from a three different branches of Walmart, respectively located in Mandalay, Yangon and Naypyitaw. The data contains 17 columns and 1000 rows:

Column	Description	Data Type
invoice_id	Invoice of the sales made	VARCHAR(30)
branch	Branch at which sales were made	VARCHAR(5)
city	The location of the branch	VARCHAR(30)
customer_type	The type of the customer	VARCHAR(30)
gender	Gender of the customer making purchase	VARCHAR(10)
product_line	Product line of the product solf	VARCHAR(100)
unit_price	The price of each product	DECIMAL(10, 2)
quantity	The amount of the product sold	INT
VAT	The amount of tax on the purchase	FLOAT(6, 4)
total	The total cost of the purchase	DECIMAL(10, 2)
date	The date on which the purchase was made	DATE
time	The time at which the purchase was made	TIMESTAMP
payment_method	The total amount paid	DECIMAL(10, 2)
cogs	Cost Of Goods sold	DECIMAL(10, 2)
gross_margin_percentage	Gross margin percentage	FLOAT(11, 9)
gross_income	Gross Income	DECIMAL(10, 2)
rating	Rating	FLOAT(2, 1)

Data Wrangling

This is the first step where inspection of data is done to make sure **NULL** values and missing values are detected and data replacement methods are used to replace, missing or **NULL** values.

1. *Build a database*

```
create database  if not exists  SalesDataWalmart;
```

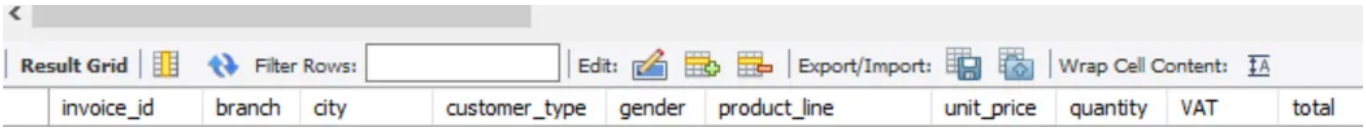
2. Create table and insert the data.

3. Select columns with null values in them. There are no null values in our database as in creating the tables, we set **NOT NULL** for each field, hence null values are filtered out.

```
Create table if not exists WMsales (
 invoice_id  varchar(30) Not null primary Key,
```

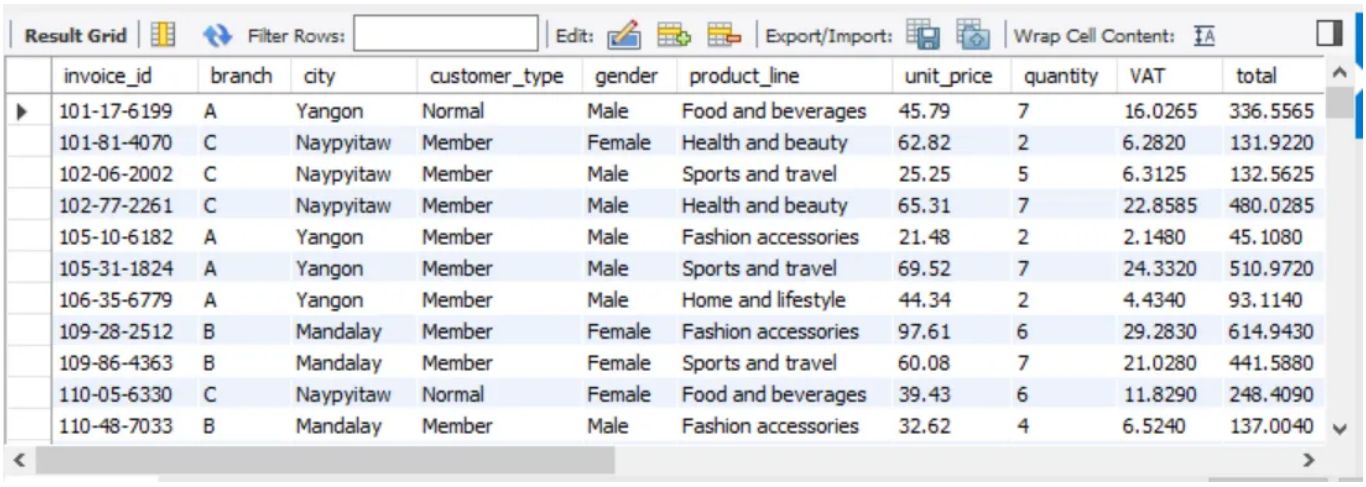
```
 branch  varchar (5) not null,
    city varchar (30) not null,
    customer_type varchar (30) not null,
    gender  varchar(10) not null,
    product_line varchar (100) not null,
    unit_price  decimal(10,2) not null,
    quantity int not null,
    VAT  float(6,4) not null,
    total decimal (12,4) not null,
    date datetime not null,
    time TIME not null,
    payment_method varchar (15)  not null,
    cogs  decimal (10, 2) not null,
    gross_margin_pct float (11,9),
    gross_income decimal(12, 4) not null,
    rating float (2, 1)
);
```

Import Data from CSV



After creating a database, it's time to import data from a CSV file. For this select import and a new window will open Select the path from the local machine.

A snippet of the table view



invoice_id	branch	city	customer_type	gender	product_line	unit_price	quantity	VAT	total
101-17-6199	A	Yangon	Normal	Male	Food and beverages	45.79	7	16.0265	336.5565
101-81-4070	C	Naypyitaw	Member	Female	Health and beauty	62.82	2	6.2820	131.9220
102-06-2002	C	Naypyitaw	Member	Male	Sports and travel	25.25	5	6.3125	132.5625
102-77-2261	C	Naypyitaw	Member	Male	Health and beauty	65.31	7	22.8585	480.0285
105-10-6182	A	Yangon	Member	Male	Fashion accessories	21.48	2	2.1480	45.1080
105-31-1824	A	Yangon	Member	Male	Sports and travel	69.52	7	24.3320	510.9720
106-35-6779	A	Yangon	Member	Male	Home and lifestyle	44.34	2	4.4340	93.1140
109-28-2512	B	Mandalay	Member	Female	Fashion accessories	97.61	6	29.2830	614.9430
109-86-4363	B	Mandalay	Member	Female	Sports and travel	60.08	7	21.0280	441.5880
110-05-6330	C	Naypyitaw	Normal	Female	Food and beverages	39.43	6	11.8290	248.4090
110-48-7033	B	Mandalay	Member	Male	Fashion accessories	32.62	4	6.5240	137.0040

Feature Engineering

1. Add a new column named `**time_of_day**` to give insight of sales in the Morning, Afternoon and Evening. This will help answer the question on which part of the day most sales are made.

```
select
 time,
    (case
  when `time` between "00:00:00" and "12:00:00" then "Morning"
        when `time` between "12:01:00" and "16:00:00" then "Afternoon"
        else  "Evening"
    end
    ) As time_of_day
from wmsales;
alter table wmsales add column time_of_day  varchar(30);
update wmsales
set time_of_day= (
 case
  when `time` between "00:00:00" and "12:00:00" then "Morning"
        when `time` between "12:01:00" and "16:00:00" then "Afternoon"
        else  "Evening"
    end
);
```

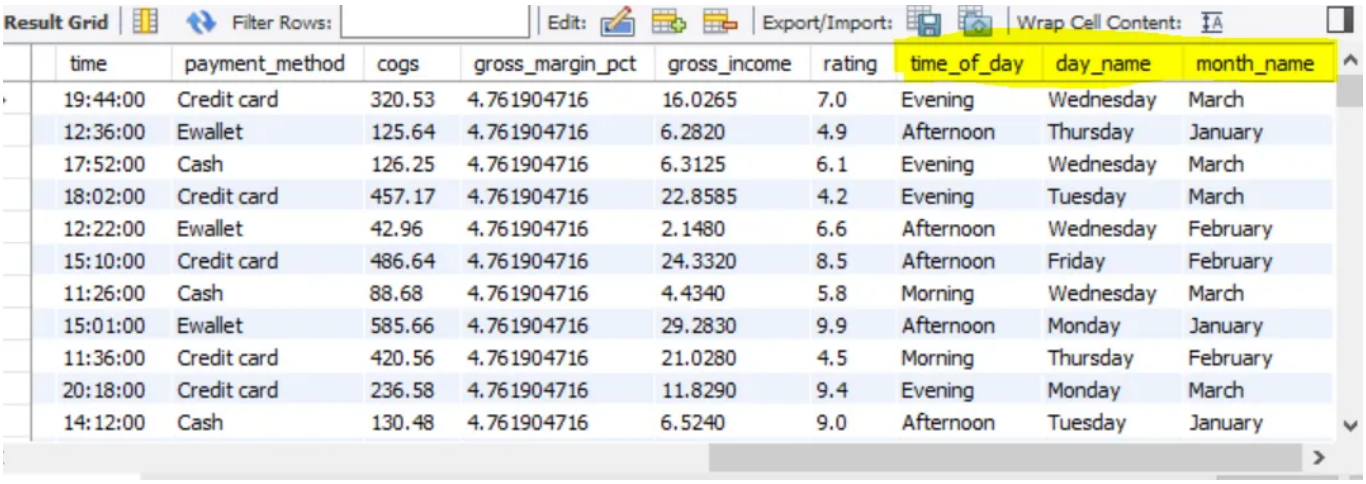
2. Add a new column named `**day_name**` that contains the extracted days of the week on which the given transaction took place (Mon, Tue, Wed, Thur, Fri). This will help answer the question on which week of the day each branch is busiest.

```
--- -day name
select
 date,
    dayname(date)
from wmsales;
alter table wmsales add column day_name varchar(10);
update  wmsales
set day_name= dayname(date);
```

3. Add a new column named `**month_name**` that contains the extracted months of the year on which the given transaction took place (Jan, Feb, Mar). Help determine which month of the year has the most sales and profit.

```
-----------  Month_name
select
 date,
    monthname(date)
from wmsales;
alter table wmsales add column month_name varchar(15);
update  wmsales
set month_name= monthname(date);
```

An updated data set with day_name, month_name and time_of_day



Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

time	payment_method	cogs	gross_margin_pct	gross_income	rating	time_of_day	day_name	month_name
19:44:00	Credit card	320.53	4.761904716	16.0265	7.0	Evening	Wednesday	March
12:36:00	Ewallet	125.64	4.761904716	6.2820	4.9	Afternoon	Thursday	January
17:52:00	Cash	126.25	4.761904716	6.3125	6.1	Evening	Wednesday	March
18:02:00	Credit card	457.17	4.761904716	22.8585	4.2	Evening	Tuesday	March
12:22:00	Ewallet	42.96	4.761904716	2.1480	6.6	Afternoon	Wednesday	February
15:10:00	Credit card	486.64	4.761904716	24.3320	8.5	Afternoon	Friday	February
11:26:00	Cash	88.68	4.761904716	4.4340	5.8	Morning	Wednesday	March
15:01:00	Ewallet	585.66	4.761904716	29.2830	9.9	Afternoon	Monday	January
11:36:00	Credit card	420.56	4.761904716	21.0280	4.5	Morning	Thursday	February
20:18:00	Credit card	236.58	4.761904716	11.8290	9.4	Evening	Monday	March
14:12:00	Cash	130.48	4.761904716	6.5240	9.0	Afternoon	Tuesday	January

Exploratory Data Analysis (EDA)

Exploratory data analysis is done to answer the listed questions and aims of this project.

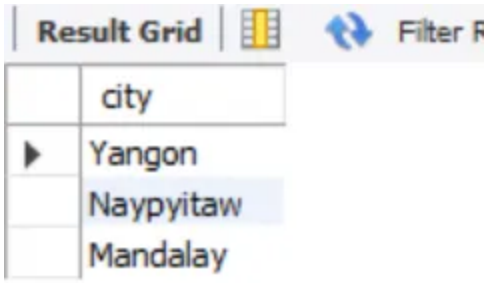
Business Questions To Answer

Generic Question

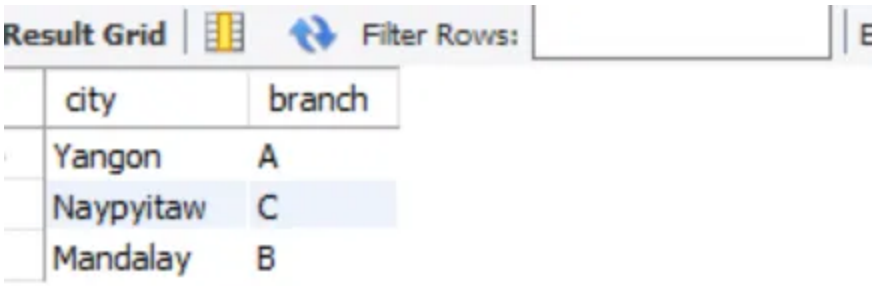
1. *How many unique cities does the data have?*

2. *In which city is each branch?*

```
select
 distinct city
from wmsales;
-- 2. In which city is each branch?
select
 distinct city,
    branch
from wmsales;
```



Distinct City

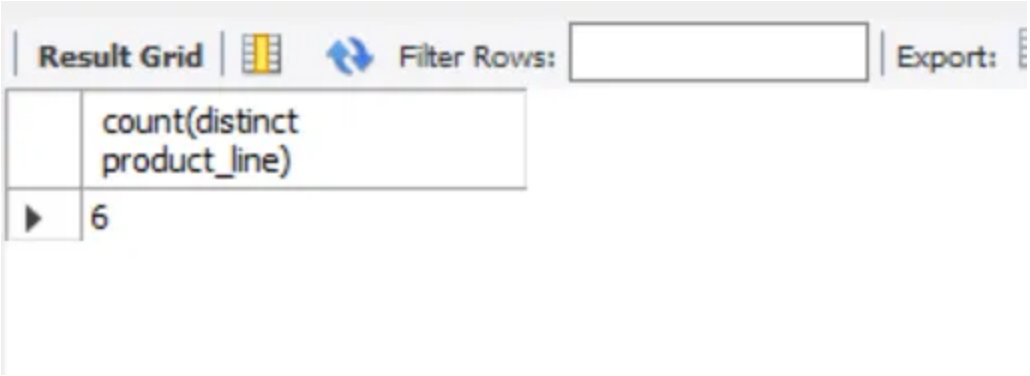


Distinct Branches in each city

Product Analysis

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

```
-------------- 1. How many unique product lines does the data have?----------
select
 count(distinct product_line)
from wmsales;
```



No of Product lines

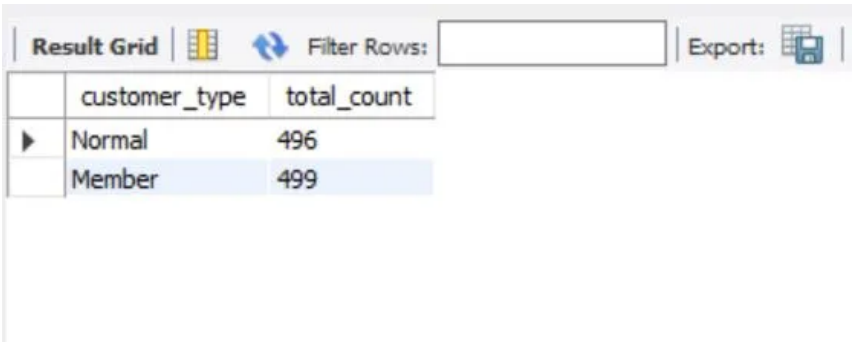
2. What is the most common payment method?

3. What is the most selling product line?

4. What is the total revenue by month?

```
------------------------- 2. What is the most common payment method?----------
select
 payment_method,
 count( payment_method) as cnt
from wmsales
group by payment_method
order by  cnt desc;
----------- 3. What is the most selling product line?---------------
select
 product_line,
 count( product_line) as cnt
from wmsales
group by product_line
order by  cnt desc;
--------------- 4. What is the total revenue by month? ------------
Select
 month_name as Month,
```

```
    sum(total) as Total_Revenue
from wmsales
group by month_name
order by Total_Revenue Desc ;
```



5. What month had the largest COGS?

6. What product line had the largest revenue?

7. What is the city with the largest revenue?

8. What product line had the largest VAT?

```
---------------------- 5. What month had the largest COGS?----------
select
 month_name as month,
    sum(cogs) as cogs
from wmsales
group by month_name
order by cogs desc;
---------------- 6. What product line had the largest revenue? ---------------
select
 product_line,
    sum(total) as Total_revenue
from wmsales
group by product_line
order by Total_revenue  desc;
---------------------- 7. What is the city with the largest revenue?-----------
select
```

```
 city,
    sum(total) as Total_revenue
from wmsales
group by city
order by Total_revenue  desc;
-------------- -8. What product line had the largest VAT?
select
 product_line,
    sum(VAT) as Valueable_Tax
from wmsales
group by product_line
order by Valueable_Tax  desc;
```

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

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

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

```
---- 10. Which branch sold more products than average product sold?-----
select
 branch ,
    sum(quantity) as qty
from wmsales
group by branch
having sum(quantity) >
 (select  avg(quantity) from wmsales);
------------ 11. What is the most common product line by gender?-------
select
 gender,
    product_line,
    count(gender) as total_count
from wmsales
group by gender, product_line
order by total_count desc;

------- 12. What is the average rating of each product line?-----------
select
 round(avg (rating) , 2) as avg_rating,
    product_line
from wmsales
```

```
group by product_line
order by avg_rating desc;
```

Sales Analysis

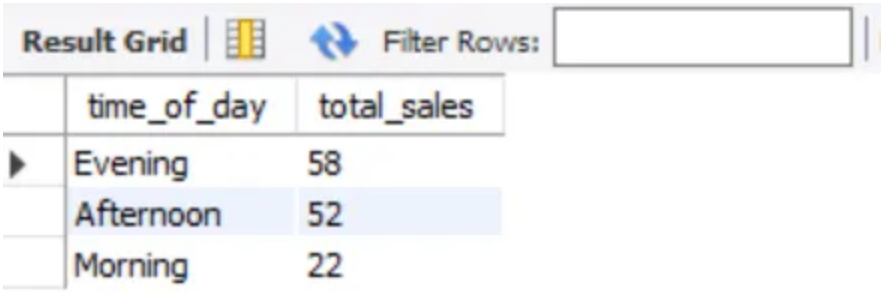
1. Number of sales made in each time of the day per weekday

2. Which of the customer types brings the most revenue?

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

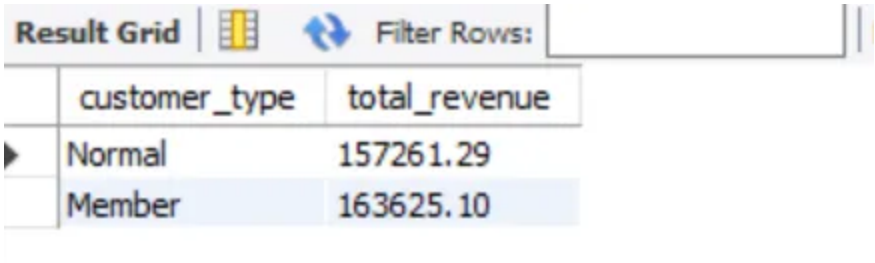
4. Which customer type pays the most in VAT?

```
-------- -----1. Number of sales made in each time of the day per weekday-------
select
 time_of_day,
    count(*) as total_sales
from wmsales
where day_name = "Sunday"
group by time_of_day
order by total_sales desc;
```



Number of sales made each day

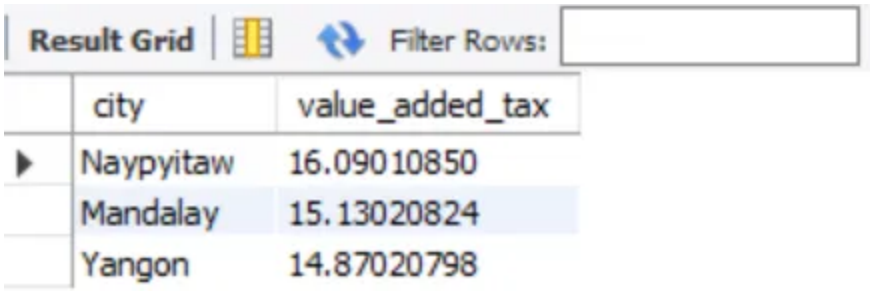
```sql
---------- 2. Which of the customer types brings the most revenue?---
select
 customer_type,
    round (sum(total), 2) as total_revenue
from wmsales
group by customer_type
order by total_revenue;
```



customer_type	total_revenue
Normal	157261.29
Member	163625.10

The customer brings the most revenue

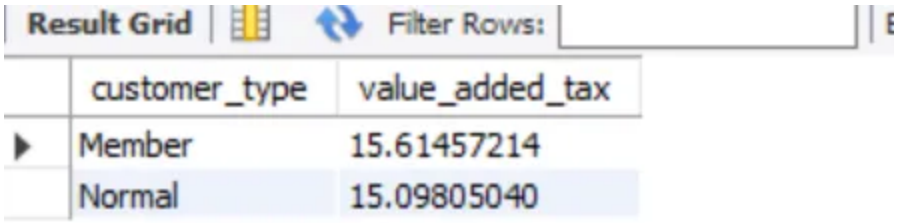
```sql
--------------------------- 3. Which city has the largest tax percent/ VAT (**Va
select
 city,
    avg(VAT) as value_added_tax
from wmsales
group by city
order by value_added_tax desc;
```



city	value_added_tax
Naypyitaw	16.09010850
Mandalay	15.13020824
Yangon	14.87020798

City Contribution in VAT

```
---------------------- 4. Which customer type pays the most in VAT?-----------
select
 customer_type,
    avg(VAT) as value_added_tax
from wmsales
group by customer_type
order by value_added_tax desc;
```



Result Grid | Filter Rows:

customer_type	value_added_tax
Member	15.61457214
Normal	15.09805040

Customers contributing in VAT

Customer Analysis

1. *How many unique customer types does the data have?*

```
------------- 1. How many unique customer types does the data have?----------
select
 distinct (customer_type)
from wmsales;
```

2. How many unique payment methods does the data have?

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

```
 distinct (payment_method)
from wmsales;
```

3. What is the most common customer type?

```
-------------------------- 3. What is the most common customer type?--
select
 customer_type,
    count(*) as total_count
from wmsales
group by customer_type
order by total_count;
```

4. Which customer type buys the most?

```
------------ 4. Which customer type buys the most?----------
select
 customer_type,
    count(*) as total_count
from wmsales
group by customer_type
order by total_count;
```

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

```
-------------------------- 5. What is the gender of most of the customers?------
select
 gender,
    count(*) as gender_count
from wmsales
group by gender
order by gender_count desc;
```

6. What is the gender distribution per branch?

```
------------------ 6. What is the gender distribution per branch?---------
select
 gender,
    count(*) as gender_count
from wmsales
where branch = "C"
group by gender
order by gender_count desc;
```

7. Which time of the day do customers give most ratings?

```
-------------- 7. Which time of the day do customers give most ratings?-----
select
 time_of_day,
    avg(rating) as avg_rating
from wmsales
group by time_of_day
order by avg_rating desc;
```

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

```
------------ 8. Which time of the day do customers give most ratings per branch?
select
 time_of_day,
    branch,
    avg(rating) as avg_rating
from wmsales
group by time_of_day, branch
order by avg_rating ;


select
 time_of_day,
    avg(rating) as avg_rating
```

```
from wmsales
where branch = "C"
group by time_of_day
order by avg_rating ;
```

9. Which day fo the week has the best avg ratings?

```
---------------------- 9. Which day fo the week has the best avg ratings?-------
select
 day_name,
    avg(rating) as avg_rating
from wmsales
group by day_name
order by avg_rating desc;
```

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

```
----- 10. Which day of the week has the best average ratings per branch?--
select
 day_name,
    avg(rating) as avg_rating
from wmsales
where branch= "A"
group by day_name
order by avg_rating desc;
```

Revenue And Profit Calculations

```
------ $ total(gross_sales)
select
 sum(VAT+cogs) as total_grass_sales
from  wmsales;
----- gross profit-----------
```

```
SELECT
 (SUM(VAT + COGS) - COGS)
FROM wmsales;

SELECT (SUM(ROUND(VAT, 2) + COGS) - COGS) FROM wmsales;
```

Conclusion

In this data analysis project, we embarked on a journey to gain valuable insights from Walmart's sales data. We began by preparing and exploring the dataset, cleaning any missing or null values, and engineering new features to help us uncover meaningful patterns.

We addressed a variety of questions, ranging from understanding product performance and sales trends to diving into customer behavior. These insights are crucial for Walmart's sales strategies and can guide future optimizations.

Throughout our analysis, we made use of SQL queries to extract relevant information from the dataset. We also computed important metrics such as COGS (Cost of Goods Sold), VAT (Value Added Tax), total revenue, and gross profit to better understand the financial aspects of Walmart's operations.

Some key takeaways from our analysis include:

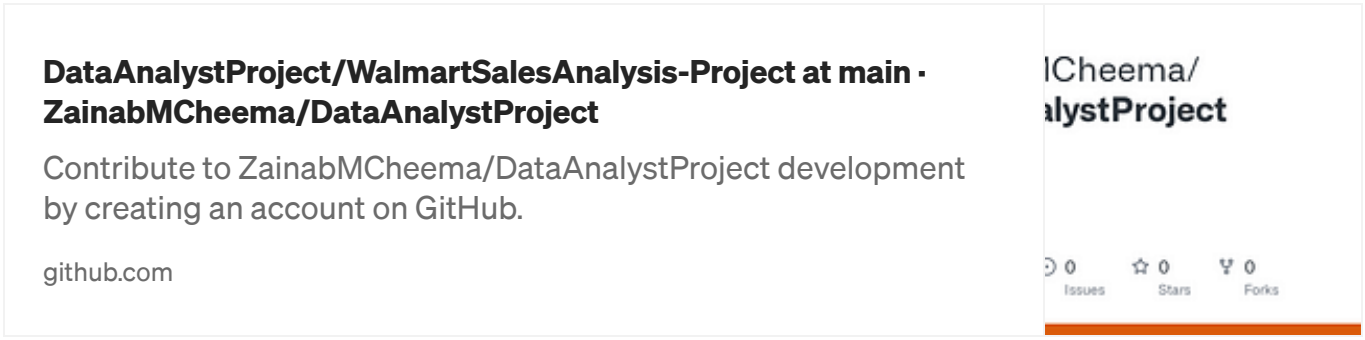
- Identification of top-performing product lines and branches.
- Analysis of sales trends, which can inform sales strategies and modifications.
- Profiling of customer segments and their profitability.

The data-driven decisions made possible through this analysis can contribute to improved sales strategies, inventory management, and overall business performance.

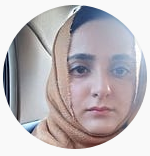
While this project has provided valuable insights, it's important to note that data analysis is an ongoing process. As Walmart continues to collect sales data, the potential for further analysis and refinement of strategies remains. This project serves as a foundation for future endeavors in enhancing Walmart's sales forecasting and optimizing its operations.

Feel free to include any specific findings or results from your analysis in this section to provide a more detailed conclusion. Additionally, you can summarize the impact of your analysis on Walmart's business strategies and operations.

Github Link for source code file



Sql | Walmart | Data Science | Data Analysis | Data Analytics



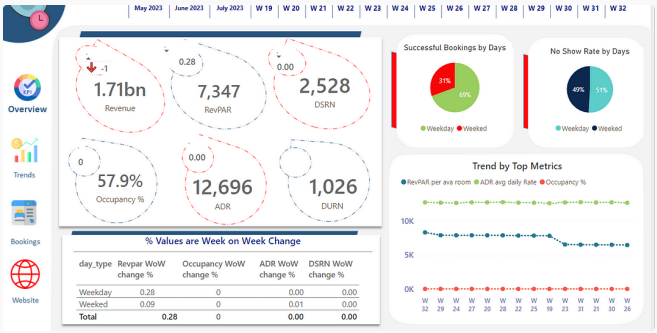
Written by Zainab Mehmood

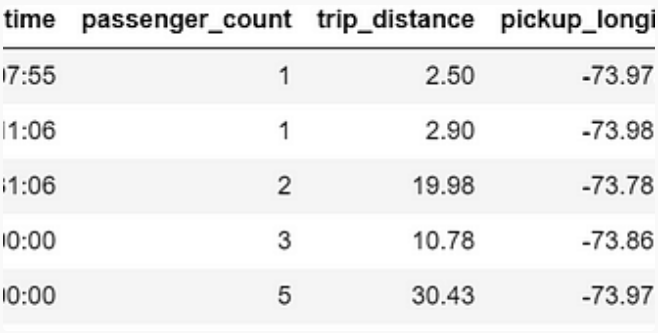
Follow

70 Followers

Data Analyst | Power BI | Python, Excel Data-driven Decision Making

More from Zainab Mehmood





time	passenger_count	trip_distance	pickup_longi
7:55	1	2.50	-73.97
1:06	1	2.90	-73.98
1:06	2	19.98	-73.78
0:00	3	10.78	-73.86
0:00	5	30.43	-73.97

Zainab Mehmood

PowerBI Dashboard: HospitalityManagement System...

Imagine a canvas where data comes alive—a place where insights flourish, and patterns...

3 min read · Aug 24, 2023

6 1

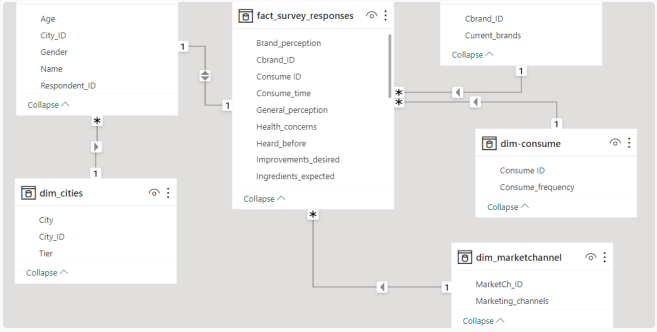
Zainab Mehmood

Data Analytics Project—Exploring Uber Data Set in New York City

Uber has revolutionized the way people travel, and its vast data stores have become a...

5 min read · Aug 26, 2023

23



Zainab Mehmood

Unleashing Market Insights: A Data Analysis Project in the Food &...

In today's data-driven world, businesses are constantly seeking ways to gain a competitiv...

Zainab Mehmood

Data Exploration With SQL

In the realm of data exploration, SQL emerges as a guiding compass, steering us through...

8 min read · Sep 6, 2023

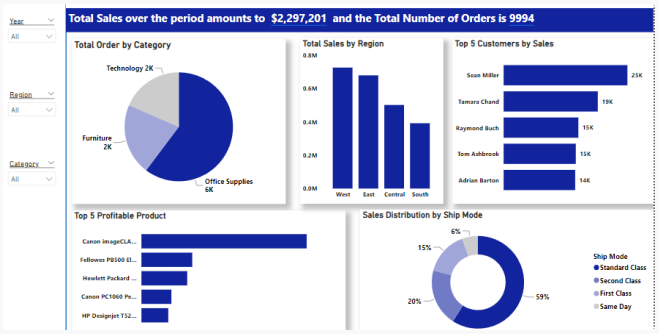
1

3 min read · Aug 26, 2023

12

See all from Zainab Mehmood

Recommended from Medium



Sefiat Kemi Adedeji

Superstore Sales Analysis using SQL

Optimizing efficiency and profitability by analyzing patterns across our products,...

11 min read · Jan 15

3

Faisal Afif

Sales Strategy Analysis on New Product Launch

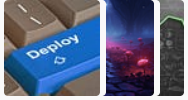
Providing insights and recommendations for sales strategy analysis

14 min read · Oct 14, 2023

7

1

Lists

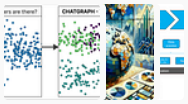


Predictive Modeling w/ Python

20 stories · 849 saves

Practical Guides to Machine Learning

10 stories · 992 saves



ChatGPT prompts

36 stories · 1055 saves

Coding & Development

11 stories · 413 saves

Promise Chinonso

Recommeded Resources for Learning Data Analytics in 2024

A list of courses for every beginner Data Analyst

2 min read · Jan 18

56

Thales Vignoli

Telecom Customer Churn Analysis using SQL and Power BI

Understading why customers are leaving a telecom company.

18 min read · Nov 23, 2023

4

Muhammad Arief Rachman

Credit Scoring Prediction

Scorecard modeling

15 min read · Nov 15, 2023

Jacob A Odey

EXPLORING EMPLOYEE DATA: A COMPREHENSIVE ANALYSIS...

INTRODUCTION

11 min read · Sep 23, 2023