

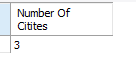
**Walmart Sales Data Analysis**

**Calculating some general information :**

How many distinct cities are represented in the dataset where Walmart stores are located?

SELECT COUNT( DISTINCT city) AS 'Number Of Citites'

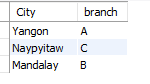
FROM sales;



In which city is each Walmart branch located?

SELECT DISTINCT city AS 'City', branch

FROM sales;

****

**Calculate some product related information:**

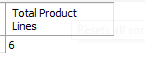
How many product lines the data have ?

SELECT

COUNT(DISTINCT product\_line) AS 'Total Product Lines'

FROM

sales;



Calculate the most used Payment Method.

SELECT

payment\_method, COUNT(payment\_method) AS 'Used\_in\_payments'

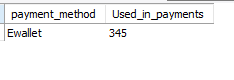
FROM

sales

GROUP BY payment\_method

ORDER BY Used\_in\_payments DESC

LIMIT 1;



Calculate the most common customer type.

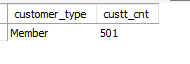
SELECT customer\_type, COUNT(customer\_type) AS custt\_cnt

FROM sales

GROUP BY customer\_type

ORDER BY custt\_cnt

DESC LIMIT 1;



Which customer type buys the most?

SELECT customer\_type

FROM (SELECT COUNT(invoice\_id), customer\_type FROM sales

GROUP BY customer\_type ORDER BY COUNT(invoice\_id) DESC

LIMIT 1) AS mosyByCustomer;

****

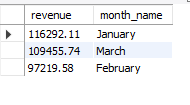
What is the total revenue by month?

SELECT SUM(total) AS revenue, month\_name

FROM sales

GROUP BY month\_name

ORDER BY revenue DESC;



What month had the largest COGS ?

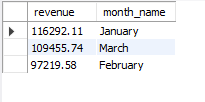
SELECT SUM(cogs) AS total\_cogs, month\_name

FROM sales

GROUP BY month\_name

ORDER BY total\_cogs DESC

LIMIT 1;

****

What product line had the largest revenue?

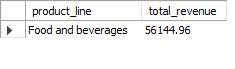
SELECT product\_line, SUM(total) AS total\_revenue

FROM sales

GROUP BY product\_line

ORDER BY total\_revenue DESC

LIMIT 1;

****

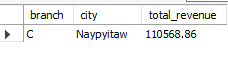
City with largest revenue including specific branch in the city.

SELECT branch,city, SUM(total) AS total\_revenue

FROM sales

GROUP BY branch,city

ORDER BY total\_revenue DESC LIMIT 1;

****

Which product line had the largect VAT ?

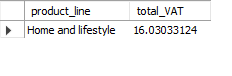
SELECT product\_line, AVG(VAT) AS total\_VAT

FROM sales

GROUP BY product\_line

ORDER BY TOTAL\_VAT DESC

LIMIT 1;

****

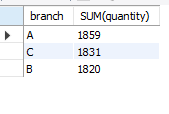
which branch sold more products than the average product sold ?

SELECT branch, SUM(quantity)

FROM sales

GROUP BY branch

HAVING SUM(quantity) > (SELECT AVG(quantity) FROM sales);

****

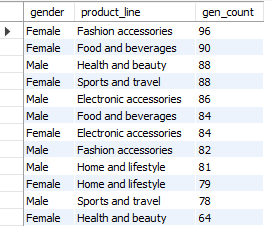
which is the most common product line by gender ?

SELECT gender, product\_line, COUNT(gender) AS gen\_count

FROM sales

GROUP BY gender, product\_line

ORDER BY gen\_count DESC;



Calculate average rating for each product line

SELECT product\_line, ROUND(AVG(rating),2) AS avg\_rating

FROM sales

GROUP BY product\_line

ORDER BY avg\_rating DESC;



Fetch each product line and add a column to those product

-- line showing "Good", "Bad". Good if its greater than average sales

SELECT AVG(quantity) AS avg\_qnty

FROM sales;

SELECT product\_line,

CASE

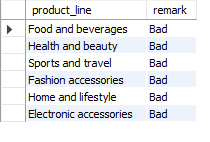
WHEN AVG(quantity) > 6 THEN "Good"

ELSE "Bad"

END AS remark

FROM sales

GROUP BY product\_line;

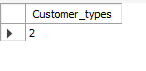


**Calculate some customer related information:**

How many unique customer types does the Data have?

SELECT COUNT(DISTINCT customer\_type) AS 'Customer\_types'

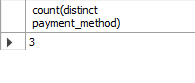
FROM sales;



How many unique payment method does the Data have?

Select count(distinct payment\_method)

From sales;



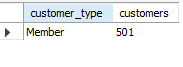
What is the most common customer type?

Select customer\_type, count(customer\_type) as customers

from sales

group by customer\_type

order by customers desc limit 1;



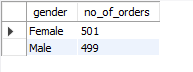
What is the gender of most of the customers?

SELECTbgender, COUNT(\*) AS 'no\_of\_orders'

FROM sales

GROUP BY gender

ORDER BY count(\*) DESC LIMIT 2;

****

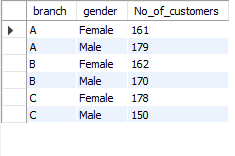
What is the gender distribution per branch?

SELECT branch, gender,COUNT(gender) AS 'No\_of\_customers'

FROM sales

GROUP BY branch,gender

order by branch ;



Which time of the day do the customers gives most ratings?

Select time\_of\_day, count(rating) as rating\_cnt

From sales

group by time\_of\_day

order by rating\_cnt desc limit 1;

****

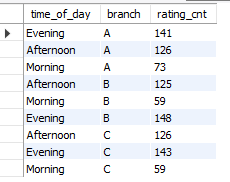
Which time of the day do the customers gives most ratings per branch?

Select time\_of\_day, branch, count(rating) as rating\_cnt

From sales

group by time\_of\_day,branch

order by branch;

****

Which day of the week has the best average ratings?

Select day\_name,avg(rating) as avg\_rating

From sales

group by day\_name

order by avg\_rating desc limit 1;

****

**Calculate some sales related information:**

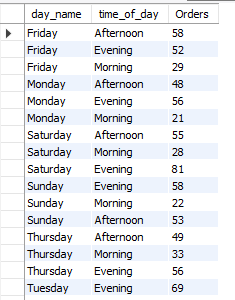
The number of sales made in each time of day per weekday?

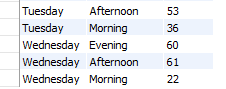
SELECT day\_name,time\_of\_day,COUNT(\*) AS 'Orders'

FROM sales

GROUP BY time\_of\_day,day\_name

ORDER BY day\_name;

****

****

Which of the customer types bring the most revenue?

SELECT customer\_type, SUM(total) AS revenue

FROM sales

GROUP BY customer\_type

ORDER BY revenue DESC LIMIT 1;



Which city has the largest tax percent / VAT(value add tax)?

SELECT city, AVG(VAT) AS vat

FROM sales

GROUP BY city

ORDER BY vat DESC LIMIT 1;

****

Which customer type pays the most in vat?

SELECT customer\_type, AVG(VAT) AS vat

FROM sales

GROUP BY customer\_type

ORDER BY vat DESC LIMIT 1;

****

**Creating Table in MySQL Workbench in which we insert our csv file.**

CREATE TABLE IF NOT EXISTS SALES(

invoice\_id varchar(30) not null primary key,

branch varchar(10) 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(10,2),

Date datetime not null ,

time time not null,

payment\_method varchar(15) not null,

cogs decimal(10,2) not null,

cogs\_margin\_pct float(11,9) not null,

gross\_income decimal(12,4) not null,

rating float(4,2) not null

);

**Data Cleaning And Transformations:**

-- -----------------------------FEATURE ENGINEERING -------------------------------------------- time\_of\_day

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 sales;

**Create Table which show’s the time of the day**

alter table sales add column time\_of\_day varchar(30);

SET SQL\_SAFE\_UPDATES = 0;

update sales

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

);

-- ------------DAY\_NAME--------------------------

select date, dayname(date) as day\_name

from sales;

alter table sales add column day\_name varchar(10);

update sales

set day\_name = dayname(date);

-- ------------MONTH\_NAME--------------------------

select date, monthname(date) as month\_name

from sales;

ALTER table sales add column month\_name varchar(10);

update sales

set month\_name = monthname(date);