CREATE DATABASE WALMARTSALES;

CREATE TABLE SALES\_DATA(

INVOICE\_ID VARCHAR(30) NOT NULL PRIMARY KEY,

BRANCH VARCHAR(50) 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 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,

GROSS\_INCOME DECIMAL(12,4) NOT NULL,

RATING FLOAT

);

-- ---- ------- FEATURE ENGINEERING------

#-----time\_of\_day---------

SELECT TIME,

(CASE

WHEN TIME\_FORMAT(TIME, '%H:%i:%s') BETWEEN '00:00:00' AND '12:00:00' THEN 'MORNING'

WHEN TIME\_FORMAT(TIME, '%H:%i:%s') BETWEEN '12:01:00' AND '16:00:00' THEN 'AFTERNOON'

ELSE 'EVENING'

END) AS TIME\_OF\_DAY

FROM SALES\_DATA;

ALTER TABLE SALES\_DATA ADD COLUMN TIME\_OF\_DAY VARCHAR(20);

SELECT \* FROM SALES\_DATA;

UPDATE SALES\_DATA

SET TIME\_OF\_DAY=(

CASE

WHEN TIME\_FORMAT(TIME, '%H:%i:%s') BETWEEN '00:00:00' AND '12:00:00' THEN 'MORNING'

WHEN TIME\_FORMAT(TIME, '%H:%i:%s') BETWEEN '12:01:00' AND '16:00:00' THEN 'AFTERNOON'

ELSE 'EVENING'

END

) ;

#DAY\_NAME

SELECT DATE,DAYNAME(DATE) AS DAY\_NAME FROM SALES\_DATA;

ALTER TABLE SALES\_DATA ADD COLUMN DAY\_NAME VARCHAR(10);

UPDATE SALES\_DATA

SET DAY\_NAME = DAYNAME(DATE);

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

SELECT DATE,MONTHNAME(DATE) AS MONTH\_NAME FROM SALES\_DATA;

ALTER TABLE SALES\_DATA

ADD COLUMN MONTH\_NAME VARCHAR(10);

UPDATE SALES\_DATA

SET MONTH\_NAME=MONTHNAME(DATE);

-- HOW MANY UNIQUE CITIES DOES DATA HAS?

SELECT DISTINCT CITY FROM SALES\_DATA;

-- IN WHICH CITY IS EACH BRANCH ?

SELECT DISTINCT CITY,BRANCH FROM SALES\_DATA;

-- ------------------------------------------------------------------

-- -----------------PRODUCT-----------------------------------------

-- HOW MANY UNIQUE PRODUCT LINES DOES THE DATA HAVE?

SELECT \* FROM SALES\_DATA;

SELECT COUNT(DISTINCT PRODUCT\_LINE) FROM SALES\_DATA;

-- WHAT IS THE MOST COMMON PYAMENT METHOD?

SELECT PAYMENT\_METHOD,COUNT(PAYMENT\_METHOD) AS CNT FROM SALES\_DATA GROUP BY PAYMENT\_METHOD

ORDER BY CNT DESC;

-- WHAT IS THE MOST SELLING PRODUCT\_LINE

SELECT PRODUCT\_LINE,COUNT(PRODUCT\_LINE) AS CNT FROM SALES\_DATA GROUP BY PRODUCT\_LINE

ORDER BY CNT DESC;

-- WHAT IS THE TOTAL REVENUE BY MONTH

SELECT \* FROM SALES\_DATA;

SELECT MONTH\_NAME AS MONTH,CAST(SUM(TOTAL) AS DECIMAL(10,2))AS TOTAL\_REVENUE

FROM SALES\_DATA GROUP BY MONTH\_NAME ORDER BY TOTAL\_REVENUE DESC ;

-- WHAT MONTH HAS LAREST COGS?

SELECT MONTH\_NAME AS MONTH ,SUM(COGS ) AS LARGEST\_COGS

FROM SALES\_DATA GROUP BY MONTH ORDER BY LARGEST\_COGS DESC;

-- WHAT PRODUCT LINE HAS THE LARGEST REVENUE?

SELECT PRODUCT\_LINE ,CAST(SUM(TOTAL) AS DECIMAL(10,2)) AS TOTAL\_REVENUE

FROM SALES\_DATA GROUP BY PRODUCT\_LINE ORDER BY TOTAL\_REVENUE DESC ;

-- WHAT IS THE CITY WITH THE LARGEST REVENUE?

SELECT BRANCH,CITY ,CAST(SUM(TOTAL) AS DECIMAL(10,2)) AS TOTAL\_REVENUE

FROM SALES\_DATA GROUP BY CITY,BRANCH ORDER BY TOTAL\_REVENUE DESC ;

-- WHAT PRODUCT LINE HAD THE LARGEST VAT?

SELECT \* FROM SALES\_DATA;

SELECT PRODUCT\_LINE ,CAST(AVG(VAT) AS DECIMAL(10,2)) AS AVG\_VAT

FROM SALES\_DATA GROUP BY PRODUCT\_LINE ORDER BY AVG\_VAT DESC ;

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

SELECT BRANCH,SUM(QUANTITY) FROM SALES\_DATA GROUP BY BRANCH HAVING

SUM(QUANTITY) > (SELECT AVG(QUANTITY) FROM SALES\_DATA) ;

-- 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;

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

SELECT PRODUCT\_LINE ,GENDER,COUNT(GENDER) AS TOTAL FROM SALES\_DATA

GROUP BY PRODUCT\_LINE,GENDER ORDER BY TOTAL DESC;

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

SELECT PRODUCT\_LINE,ROUND(AVG(RATING),2) AS AVG\_RATING FROM SALES\_DATA

GROUP BY PRODUCT\_LINE ORDER BY AVG\_RATING DESC;

-- ---------------------------------------------------------------------------------------------

-- --------------------------------------- SALES -------------------------------------------------

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

SELECT TIME\_OF\_DAY,COUNT(\*) AS TOTAL\_SALES FROM SALES\_DATA

WHERE DAY\_NAME="MONDAY" GROUP BY TIME\_OF\_DAY ORDER BY TOTAL\_SALES DESC;

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

select \*from sales\_data;

SELECT CUSTOMER\_TYPE,ROUND(SUM(TOTAL),2) AS HIGH\_REVENUE FROM SALES\_DATA

GROUP BY CUSTOMER\_TYPE ORDER BY HIGH\_REVENUE DESC;

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

SELECT CITY,AVG(VAT) AS VAT FROM SALES\_DATA GROUP BY CITY ORDER BY VAT DESC;

-- Which customer type pays the most in VAT?

SELECT CUSTOMER\_TYPE,AVG(VAT) AS VAT FROM SALES\_DATA GROUP BY CUSTOMER\_TYPE ORDER BY VAT DESC;

-- ------------------------------------------------------------------------------------------------

-- ------------------------CUSTOMER---------------------------------------------------------------

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

SELECT DISTINCT CUSTOMER\_TYPE FROM SALES\_DATA;

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

SELECT DISTINCT PAYMENT\_METHOD FROM SALES\_DATA;

-- What is the most common customer type?

SELECT customer\_type, COUNT(\*) as count

FROM SALES\_DATA

GROUP BY customer\_type

ORDER BY count DESC

LIMIT 1;

-- Which customer type buys the most?

SELECT CUSTOMER\_TYPE ,COUNT(\*) AS CUSTOMER\_COUNT FROM SALES\_DATA

GROUP BY CUSTOMER\_TYPE ;

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

SELECT GENDER,COUNT(\*) AS GENDER\_CNT FROM SALES\_DATA GROUP BY GENDER ORDER BY GENDER\_CNT DESC;

-- What is the gender distribution per branch?

SELECT BRANCH, GENDER, COUNT(\*) AS GENDER\_CNT

FROM SALES\_DATA

GROUP BY BRANCH, GENDER

ORDER BY BRANCH, GENDER\_CNT DESC;

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

SELECT

time\_of\_day,

AVG(rating) AS avg\_rating

FROM sales\_data

GROUP BY time\_of\_day

ORDER BY avg\_rating DESC;

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

SELECT

time\_of\_day,

round(AVG(rating),2) AS avg\_rating

FROM sales\_data

WHERE branch = "A"

GROUP BY time\_of\_day

ORDER BY avg\_rating DESC;

-- Which day fo the week has the best avg ratings?

SELECT

day\_name,

round(AVG(rating),2) AS avg\_rating

FROM sales\_data

GROUP BY day\_name

ORDER BY avg\_rating DESC;

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

SELECT

day\_name,

COUNT(day\_name) total\_sales

FROM sales\_data

WHERE branch = "C"

GROUP BY day\_name

ORDER BY total\_sales DESC;