```
-- Create table
CREATE TABLE IF NOT EXISTS 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 DECIMAL(10,2) NOT NULL,
    quantity INT NOT NULL,
    tax_pct FLOAT(6,4) NOT NULL,
    total DECIMAL(12, 4) NOT NULL,
    date DATETIME NOT NULL,
    time TIME NOT NULL,
    payment VARCHAR(15) NOT NULL,
    cogs DECIMAL(10,2) NOT NULL,
    gross_margin_pct FLOAT(11,9),
    gross_income DECIMAL(12, 4),
    rating FLOAT(2, 1)
);
---------------------------------------------------------------------------------------------------
------------------------------------ feature engineering ----------------------------------------




select * from sales;


---------------------------- adding new columns  -------------------------


----- adding new time column labelled as morning,afternoon , evening ----------
select time,time_of_day from sales;
```

```
alter table sales
add column time_of_day varchar(10);

update sales
set time_of_day = case
                                                    when hour(time_sold)>=00 and hour(time_sold)<12
then   'Morning '
                when hour(time_sold) >=12 and hour(time_sold) <16 then   'Afternoon'
                when hour(time_sold) >=16  then   'Evening'
                                        end;

select time_sold , time_of_day from sales;
select * from sales;

---- dayname column ----------------------
alter table sales
add column name_of_the_day varchar(10);



UPDATE sales
SET name_of_the_day = DAYNAME(date);

select date , name_of_the_day from sales;

--------------------------   month name column ----------------
alter table sales
add column name_of_the_month varchar(10);

UPDATE sales
SET name_of_the_month = monthname(date);
```

```
select date , name_of_the_month from sales;

select * from sales;
```

Business Questions To Answer

Generic Question

```
---------- 1. How many unique cities does the data have? ----------

select distinct(city) from sales;

select count(distinct(city)) from sales;

----------  2. In which city is each branch? ----------

select city , branch
from sales
group by city,branch
order by city;

----- ### Product related questions ----------

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

select distinct(product_line) from sales;
```

```
select count(distinct(product_line)) from sales;

-- 2. What is the most common payment method? --

select * from
(
select payment , count(payment) as num_of_times_trans ,
dense_rank () over( order by count(payment) desc) as ranking
from sales
group by payment
) as s
where ranking=1;

-- 3. What is the most selling product line?---

select product_line,counting from
(
select product_line , count(product_line) as counting ,
dense_rank () over( order by count(product_line) desc) as ranking
from sales
group by product_line
) as s
where ranking=1;

-- 4. What is the total revenue by month? --

select unit_price,quantity,cogs,tax_pct,total,gross_income,gross_margin_pct from sales;

select date from sales;

select name_of_the_month,sum(total)
```

```
from sales
group by name_of_the_month
order by name_of_the_month;

-- 5. What month had the largest COGS? ------

select name_of_the_month , sum(cogs) as sum_of_cogs
from sales
group by name_of_the_month
order by sum_of_cogs desc;

-- 6. What product line had the largest revenue?

select product_line,sum(total) as tot_rev
from sales
group by product_line
order by tot_rev desc;

-- 7. What is the city with the largest revenue? --

select city,sum(total) as tot_rev
from sales
group by city
order by tot_rev desc;

-- 8. What product line had the largest VAT collection? ----

select product_line, sum(tax_pct) as tot_tax
from sales
group by product_line
order by tot_tax desc;
```

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

select product_line, (select avg(total) from sales) as tot_avg, avg(total) as avg_of_pro_line,
case
when avg(total) >( select avg(total) from sales) then 'GOOD'
     else 'BAD'
end as good_bad_category
from sales
group by product_line;


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

select branch , sum(quantity) as qt_sold_in_this_branch,
(select sum(quantity)/count(distinct branch) from sales) as avg_qt_sold_at_all_branches
from sales
group by branch
having sum(quantity) >(select sum(quantity)/count(distinct branch) from sales);

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

select gender , product_line , product_purchased from
(
select gender , product_line , sum(quantity) as product_purchased,
dense_rank() over( partition by gender order by sum(quantity) desc) as rnk
from sales
group by gender , product_line
) as s
```

```
where rnk=1;
```

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

```
select product_line , avg(rating) as avg_rating
from sales
group by product_line
order by avg_rating desc;
```

Sales

---------- 1. Number of qt. and amount of sales made in each time of the day per weekday ------

```
select name_of_the_day,time_of_day,sum(quantity),sum(total) ,
dense_rank()  over(partition by name_of_the_day order by time_of_day desc)
from sales
group by name_of_the_day,time_of_day;
```

---------- 2. Which of the customers spend the most amount of money and

---------- which customers purchase the most no. of goods quantity wise ? --------

```
select distinct(customer_type)
from sales;
```

```
select customer_type , sum(total) as tot_amt
from sales
group by customer_type
order by tot_amt desc;
```

```
select customer_type ,product_line, sum(total) as tot_amt
from sales
group by customer_type,product_line
order by customer_type,tot_amt ;


select customer_type , sum(quantity) as tot_cnt
from sales
group by customer_type
order by tot_cnt desc;

select customer_type ,product_line, sum(quantity) as tot_cnt
from sales
group by customer_type,product_line
order by customer_type,tot_cnt ;


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

select city , sum(tax_pct) as vat
from sales
group by city
order by vat desc;

-------- 4. Which customer type pays the most in VAT? -------

select customer_type , sum(tax_pct) as vat
from sales
group by customer_type
order by vat desc;
```

Customer

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

```
select distinct(customer_type)
from sales;

select count(distinct(customer_type))
from sales;
```

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

```
select distinct(payment)
from sales;

select count(distinct(payment))
from sales;
```

---------- 3. What is the most common customer type? -----

```
select customer_type , count(customer_type) cnt
from sales
group by customer_type
order by cnt desc ;
```

---------- 4. Which customer type buys the most moneywise,quantitywise? -------

```
select customer_type , sum(total) amt
from sales
```

```
group by customer_type
order by amt desc;

select customer_type , sum(quantity) qt
from sales
group by customer_type
order by qt desc;
```

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

```
select customer_type , gender , count(gender) cnt
from sales
group by customer_type,gender
order by customer_type asc ,gender desc;
```

---------- 6. What is the gender distribution per branch? ----------

```
select branch ,gender , count(gender) as count_of_individuals
from sales
group by branch , gender
order by branch , gender desc ;
```

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

```
select time_of_day , count(rating) as cnt_of_ratings
from sales
group by time_of_day
order by cnt_of_ratings desc ;
```

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

```
select branch , time_of_day , count(rating) as cnt_of_ratings
from sales
group by branch , time_of_day
order by branch , cnt_of_ratings desc ;
```

---------- 9. Which day of the week has the best avg ratings? ------

```
select name_of_the_day , avg(rating) as avg_ratings
from sales
group by name_of_the_day
order by avg(rating) desc limit 1;
```

```
select name_of_the_day,avg_ratings from
(select name_of_the_day , avg(rating) as avg_ratings ,
dense_rank() over(order by avg(rating) desc) rnk
from sales
group by name_of_the_day) s
where rnk=1;
```

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

```
select branch,name_of_the_day,avg_ratings from
(select branch,name_of_the_day , avg(rating) as avg_ratings ,
dense_rank() over(partition by branch order by avg(rating) desc) rnk
from sales
group by name_of_the_day,branch) s
where rnk=1;
```