1. Download vechile sales data -> https://github.com/shashank-mishra219/Hive-Class/blob/main/sales_order_data.csv

2. Store raw data into hdfs location
3. Create a internal hive table "sales_order_csv" which will store csv data sales_order_csv .. make sure to skip header row while creating table
4. Load data from hdfs path into "sales_order_csv"
5. Create an internal hive table which will store data in ORC format "sales_order_orc"
6. Load data from "sales_order_csv" into "sales_order_orc"
Perform below menioned queries on "sales_order_orc" table :
a. Calculatye total sales per year
b. Find a product for which maximum orders were placed
c. Calculate the total sales for each quarter
d. In which quarter sales was minimum
e. In which country sales was maximum and in which country sales was minimum
f. Calculate quartelry sales for each city
h. Find a month for each year in which maximum number of quantities were sold

2. Store raw data into hdfs location

hadoop fs -put sales_order_data.csv /tmp

3. Create a internal hive table “sales_order_csv” which will stored csv data sales_order_csv

create table sales_data_csv

> (

> ORDERNUMBER int, QUANTITYORDERED int, PRICEEACH float, ORDERLINENUMBER int, SALES float, STATUS string, QTR_ID int, MONTH_ID int, YEAR_ID int, PRODUCTLINE string, MSRP int, PRODUCTCODE string, PHONE string, CITY string, STATE string, POSTALCODE string, COUNTRY string, TERRITORY string, CONTACTLASTNAME string, CONTACTFIRSTNAME string, DEALSIZE string

>)

> row format delimited

> fields terminated by ','

> tblproperties("skip.header.line.count"="1");

4. Load data from hdfs path into "sales_order_csv"
load data inpath '/tmp/sales_order_data.csv' into table sales_data_csv;

5. Create an internal hive table which will store data in ORC format "sales_order_orc"

create table sales_data_orc

> (

> ORDERNUMBER int, QUANTITYORDERED int, PRICEEACH float, ORDERLINENUMBER int, SALES float, STATUS string, QTR_ID int, MONTH_ID int, YEAR_ID int, PRODUCTLINE string, MSRP int, PRODUCTCODE string, PHONE string, CITY string, STATE string, POSTALCODE string, COUNTRY string, TERRITORY string, CONTACTLASTNAME string, CONTACTFIRSTNAME string, DEALSIZE string)

> stored as orc;

6. Load data from "sales_order_csv" into "sales_order_orc"

from sales_data_csv insert overwrite table sales_data_orc select *;

Perform below menioned queries on "sales_order_orc" table :

a. Calculatye total sales per year

select year_id as year,sum(sales) as total_sales

> from sales_data_orc

> group by year_id;

b. Find a product for which maximum orders were placed

```
select productline, sum(quantityordered) as total_orders
    > from sales_data_orc
    > group by productline
    > order by total_orders desc limit 1;
```

c. Calculate the total sales for each quarter

```
select qtr_id,sum(sales) as total_sales
    > from sales_data_orc
    > group by qtr_id;
```

d. In which quarter sales was minimum In which quarter sales was minimum

```
select qtr_id,sum(sales)
    > from sales_data_orc
    > group by qtr_id
    > order by sum(sales) asc limit 1;
```

e. In which country sales was maximum and in which country sales was minimum

```
select country,sum(sales) as max_sale
    > from sales_data_orc
    > group by country
    > order by max_sale desc limit 1;

select country,sum(sales) as min_sale
    > from sales_data_orc
    > group by country
    > order by max_sale asc limit 1;
```

f. Calculate quartelry sales for each city

```
select city,qtr_id,sum(sales) as total_sales
    > from sales_data_orc
    > group by city,qtr_id;
```

h. Find a month for each year in which maximum number of quantities were sold

```
select year_id,month_id,sum(quantityordered) as total_quantity
    > from sales_data_orc
    > group by year_id,month_id
    > order by year_id,total_quantity desc limit 1;
```