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

hadoop fs -put /users/thestupidmonk/downloads/sales\_order\_data.csv /tmp/hive/assignments/

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

create table sales\_order\_data

(

order\_number int,

quantity\_ordered int,

price\_each float,

order\_line\_number int,

sales float,

status string,

qtr\_id int,

month\_id int,

year\_id int,

product\_line string,

msrp int,

product\_code string,

phone string,

city string,

state string,

postal\_code string,

country string,

territory string,

contact\_last\_name string,

contact\_first\_name string,

deal\_size 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/hive/assignments/sales\_data.csv' into tables sales\_order\_data;

5. Create an internal hive table which will store data in ORC format "sales\_order\_orc"

create table sales\_order\_data\_orc

(

order\_number int,

quantity\_ordered int,

price\_each float,

order\_line\_number int,

sales float,

status string,

qtr\_id int,

month\_id int,

year\_id int,

product\_line string,

msrp int,

product\_code string,

phone string,

city string,

state string,

postal\_code string,

country string,

territory string,

contact\_last\_name string,

contact\_first\_name string,

deal\_size string)

stored as orc;

row format delimited

fields terminated by ','

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

6. Load data from "sales\_order\_csv" into "sales\_order\_orc"

from sales\_order\_data insert overwrite table sales\_order\_data\_orc select \*;

Perform below menioned queries on "sales\_order\_orc" table :

a. Calculate total sales per year

select year\_id, Round(sum(sales), 3) as total\_sales\_per\_year from sales\_order\_data\_orc group by year\_id order by year\_id;

b. Find a product for which maximum orders were placed

select product\_code, count(order\_number) as total\_order\_per\_product\_code

from sales\_order\_data\_orc

group by product\_code

order by total\_order\_per\_product\_code desc

limit 1;

or

select MAX(x.total\_order\_per\_product\_code)

from(

select product\_code, count(order\_number) as total\_order\_per\_product\_code

from sales\_order\_data\_orc

group by product\_code

order by total\_order\_per\_product\_code desc) x;

c. Calculate the total sales for each quarter

select qtr\_id, ROUND(sum(sales), 3) as total\_sales\_per\_quarter

from sales\_order\_data\_orc

group by qtr\_id

order by qtr\_id;

d. In which quarter sales was minimum

select qtr\_id, ROUND(sum(sales), 3) as total\_sales\_per\_quarter

from sales\_order\_data\_orc

group by qtr\_id

order by qtr\_id

limit 1;

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

(select country, ROUND(SUM(sales), 3) as sales\_for\_this\_country

from sales\_order\_data\_orc

group by country

order by sales\_for\_this\_country desc limit 1)

UNION ALL

(select country, ROUND(SUM(sales), 3) as sales\_for\_this\_country

from sales\_order\_data\_orc

group by country

order by sales\_for\_this\_country limit 1);

or

select country, x.sales\_for\_countries,

row\_number() over(partition by country order by sales\_for\_countries desc)

from(select country,

SUM(sales) as sales\_for\_countries

from sales\_order\_data\_orc

group by country)x;

f. Calculate quartelry sales for each city

select city, qtr\_id, ROUND(SUM(sales), 3) as total\_sales\_per\_quarter\_for\_city

from sales\_order\_data\_orc

group by city, qtr\_id

order by city, qtr\_id;

select country, ROUND(SUM(sales), 3) as total\_sales\_per\_country

from sales\_order\_data\_orc

group by country

order by total\_sales\_per\_country;

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

with x as (select year\_id, month\_id, sum(quantity\_ordered) as quantity

from sales\_order\_data\_orc

group by year\_id, month\_id),

y as (select year\_id, month\_id, quantity,

row\_number() over(partition by year\_id order by year\_id, quantity desc) ranking

from x)

select year\_id, month\_id, quantity as quantities\_sold

from y where ranking = 1;

or

select year\_id, month\_id, x.quantity

from (select distinct month\_id, year\_id, sum(quantity\_ordered) as quantity,

row\_number() over (partition by year\_id order by quantity desc) as rn

from sales\_order\_data\_orc)x

where x.rn = 1;

or

select \* from (

select x.year\_id, x. month\_id, x.quantity,

row\_number() over(partition by x.year\_id order by x.year\_id, x.quantity desc) ranking

from

(select year\_id, month\_id, sum(quantity\_ordered) as quantity

from sales\_order\_data\_orc

group by year\_id, month\_id) x)y

where y.ranking = 1;

or

select year\_id, month\_id,quantity\_ordered, rnk from

(select year\_id, month\_id,quantity\_ordered,

dense\_rank() over(partition by year\_id order by quantity\_ordered desc) as rnk

from sales\_order\_data\_orc)a where a.rnk = 1;