Name : Omkar Gharat
Email : gharatomkar10@gmail.com

Mobile No: 8087087938

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

=> copyFromLocal sales_order_data.csv/

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_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"

=>

```
hive> load data local inpath 'sales_order_data.csv' into table sales_order_csv;
Loading data to table assi_3.sales_order_csv
Table assi_3.sales_order_csv stats: [numFiles=1, totalSize=357411]
OK
Time taken: 1.71 seconds
```

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

=>

```
create table sales_order__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;
OK
Time taken: 0.243 seconds
```

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

=>

```
hive> insert overwrite table sales_order__orc select * from sales_order_csv;
```

Perform below menioned queries on "sales_order_orc" table :

a. Calculatye total sales per year

=>select year_id , sum(sales) from sales_order__orc group by year_id;

output :

2003 3516979.547241211

2004 4724162.593383789

2005 1791486.7086791992

Time taken: 66.56 seconds, Fetched: 3 row(s)

b. Find a product for which maximum orders were placed

=>select PRODUCTLINE, count(quantityordered) as maxordered from sales_order__orc group by productline order by maxordered desc limit 1;

output:

Classic Cars 967

Time taken: 120.017 seconds, Fetched: 1 row(s)

c. Calculate the total sales for each quarter

=> select qtr_id , count(sales) from sales_order__orc group by qtr_id;

output:

1 665

2 561

3 503

4 1094

Time taken: 57.364 seconds, Fetched: 4 row(s)

d. In which quarter sales was minimum

=> select qtr_id, count(sales) from sales_order__orc group by qtr_id order by qtr_id asc limit 1;

output :

1 665

Time taken: 117.683 seconds, Fetched: 1 row(s)

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

=>

country sales was maximum

select country , count(sales) as tot_sales from sales_order__orc group by country order by tot_sales desc limit 1;

output:

USA 1004

Time taken: 125.78 seconds, Fetched: 1 row(s)

country sales was minimum

select country , count(sales) as tot_sales from sales_order__orc group by country order by tot_sales asc limit 1;

output :

Ireland 16

Time taken: 128.562 seconds, Fetched: 1 row(s)

f. Calculate quartelry sales for each city

=>

```
select qtr_id,city , count(sales) from sales_order__orc group by  qtr_id, city ;
```

output :

1	Bergamo	18
1	Boras	8
1	Brickhaven	8
1	Brisbane	4
1	Bruxelles	5
1	Burbank	10
1	Burlingame	4
1	Cambridge	6
1	Charleroi	3
1	Cowes	8
1	Dublin	9
1	Espoo	15
1	Frankfurt	12
1	Gensve	14
1	Glendale	1
1	Graz	3
1	Helsinki	6
1	Kobenhavn	15
1	Lille	6
1	London	3
1	Los Angeles	8
1	Lule	2
1	Lyon	30

1	Madrid	104
1	Makati City	16
1	Manchester	16
1	Marseille	1
1	Melbourne	15
1	Minato-ku	9
1	NYC	10
1	Nantes	15
1	Nashua	4
1	New Bedford	11
1	Newark	3
1	North Sydney	19
1	Osaka	17
1	Oulu	17
1	Paris	19
1	Pasadena	13
1	Philadelphia	6
1	Reims	18
1	San Diego	25
1	San Francisco	19
1	San Rafael	74
1	Singapore	8
1	South Brisbane	6
1	Stavern	16
1	Toulouse	5
1	Versailles	1
2	Allentown	2
2	Barcelona	1
2	Boston	23

2	Brickhaven	2
2	Bridgewater	17
2	Bruxelles	5
2	Cambridge	5
2	Charleroi	1
2	Chatswood	13
2	Espoo	6
2	Glen Waverly	5
2	Glendale	8
2	Kobenhavn	13
2	Las Vegas	13
2	Liverpool	23
2	London	9
2	Madrid	88
2	Marseille	17
2	Melbourne	17
2	Minato-ku	7
2	Montreal	18
2	NYC	44
2	Nantes	17
2	New Haven	9
2	Newark	18
2	Osaka	3
2	Oulu	5
2	Paris	22
2	Philadelphia	3
2	Reggio Emilia	14
2	Reims	5
2	Salzburg	28

2	San Jose	40
2	San Rafael	2
2	Singapore	27
2	Strasbourg	19
2	Tsawassen	12
3	Allentown	20
3	Bergen	4
3	Boras	14
3	Boston	5
3	Brickhaven	33
3	Brisbane	11
3	Bruxelles	15
3	Burlingame	11
3	Cambridge	13
3	Charleroi	1
3	Chatswood	19
3	Dublin	7
3	Espoo	9
3	Gensve	17
3	Glen Waverly	2
3	Glendale	2
3	Helsinki	11
3	Las Vegas	10
3	Madrid	19
3	Munich	14
3	NYC	18
3	Nantes	20
3	New Bedford	16
3	North Sydney	14

3	Oslo	13
3	Oulu	10
3	Paris	8
3	Pasadena	16
3	Reggio Emilia	15
3	Reims	4
3	Salzburg	2
3	San Rafael	57
3	Singapore	26
3	South Brisbane	4
3	Torino	26
3	Toulouse	3
3	Tsawassen	14
4	Aaarhus	27
4	Allentown	9
4	Barcelona	22
4	Bergamo	30
4	Bergen	25
4	Boras	16
4	Boston	16
4	Brickhaven	4
4	Bridgewater	8
4	Burbank	3
4	Burlingame	19
4	Cambridge	14
4	Charleroi	3
4	Chatswood	14
4	Cowes	18
4	Frankfurt	10

4	Glen Waverly	16
4	Glendale	11
4	Graz	12
4	Helsinki	13
4	Kobenhavn	8
4	Koln	26
4	Las Vegas	6
4	Lille	14
4	Liverpool	6
4	London	26
4	Los Angeles	6
4	Lule	17
4	Lyon	11
4	Madrid	93
4	Makati City	10
4	Manchester	35
4	Marseille	7
4	Melbourne	23
4	Minato-ku	16
4	Montreal	4
4	NYC	80
4	Nantes	8
4	Nashua	30
4	New Bedford	34
4	New Haven	8
4	North Sydney	13
4	Oslo	11
4	Paris	21
4	Pasadena	1

4 Philadelphia 35

4 Reggio Emilia 10

4 Reims 14

4 Salzburg 10

4 San Francisco 43

4 San Rafael 47

4 Sevilla 15

4 Singapore 18

4 South Brisbane 5

4 Stavern 16

4 Toulouse 12

4 Vancouver 22

4 Versailles 17

4 White Plains 26

Time taken: 60.828 seconds, Fetched: 182 row(s)

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

for 2003

=> select month_id, year_id, count(quantityordered) as qty from sales_order__orc where year_id = 2003 group by year_id, month_id order by qty desc limit 1;

output:

11 2003 296

Time taken: 126.788 seconds, Fetched: 1 row(s)

for 2004

=> select month_id, year_id, count(quantityordered) as qty from sales_order__orc where year_id = 2004 group by year_id, month_id order by qty desc limit 1;

output:

11 2004 301

Time taken: 127.062 seconds, Fetched: 1 row(s)

for 2005

=> select month_id, year_id, count(quantityordered) as qty from sales_order__orc where year_id = 2005 group by year_id, month_id order by qty desc limit 1;

output :

5 2005 120

Time taken: 112.273 seconds, Fetched: 1 row(s)