

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

2. Store raw data into hdfs location

Ans : `hadoop fs -put /home/dipanwita_data/sales_order_data.csv /home/dipanwita_data`

`hadoop fs -ls /home/dipanwita_data`

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_agm
(
  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 '/home/dipanwita_data' into table 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_agm
(
  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 ','
stored as orc;

```

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

```

from sales_order_data_agm insert overwrite table
sales_order_data_orc_agm select *;

```

Perform below mentioned queries on "sales_order_orc" table :

a. Calculate total sales per year

```

select year_id, sum(sales) as total_sales from sales_order_data
group by year_id;

```

Ans:

2003	3516979.547241211
2004	4724162.593383789
2005	1791486.7086791992

b. Find a product for which maximum orders were placed

```

select productline, sum(quantityordered) as total_order from
sales_order_data group by productline order by total_order desc
limit 1

```

Ans Classic Cars 33992

c. Calculate the total sales for each quarter

```

select qtr_id, sum(sales) as total_sales from sales_order_data group
by qtr_id;

```

Ans:

```
1 2350817.726501465
2 2048120.3029174805
3 1758910.808959961
4 3874780.010925293
```

d. In which quarter sales was minimum

Ans 3 1758910.808959961

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

```
select country, sum(sales) as total_sales from sales_order_data
group by country order by total_sales;
```

Maximum: Ireland 57756.43029785156

Minimum: USA 3627982.825744629

f. Calculate quartelry sales for each city

```
select city, year_id, qtr_id, sum(sales) as total_sales from
sales_order_data group by city, year_id, qtr_id order by
total_sales;
```

```
Minato-ku 20052 553.9500122070312
Stavern 20034 1474.6600341796875
Charleroi 20033 1637.199951171875
Charleroi 20032 1711.260009765625
Paris 20043 2173.60009765625
Marseille 20051 2317.43994140625
Singapore 20052 2587.5799560546875
Burlingame 20044 2916.199951171875
Sevilla 20044 3220.8800659179688
Glendale 20051 3987.199951171875
Singapore 20051 4175.60009765625
Barcelona 20032 4219.2001953125
Pasadena 20034 4512.47998046875
Glendale 20034 5142.14990234375
Versailles 20041 5759.419921875
Allentown 20052 6166.7998046875
Salzburg 20043 6693.2802734375
San Rafael 20052 7261.75
Brickhaven 20032 7277.35009765625
Philadelphia 20042 7287.240234375
Glendale 20033 7600.1201171875
Burbank 20034 8234.559936523438
Bruxelles 20052 8411.949829101562
London 20041 8477.219970703125
Newark 20041 8722.1201171875
Graz 20051 8775.159912109375
Frankfurt 20044 9214.969970703125
Lule 20031 9748.999755859375
South Brisbane 20033 10640.290161132812
```

Frankfurt	20031	11432.33984375
Brickhaven	20044	11528.52978515625
Nashua	20031	12133.25
Glen Waverly	20043	12334.819580078125
San Rafael	20031	12398.56005859375
Charleroi	20044	13463.480224609375
Burlingame	20051	13529.570190429688
London	20044	13739.900024414062
Paris	20052	14066.7998046875
Glen Waverly	20052	14378.089965820312
Cambridge	20042	14380.920043945312
Las Vegas	20044	14449.609741210938
Toulouse	20051	15139.1201171875
Reims	20033	15146.31982421875
Boston	20043	15344.640014648438
Makati City	20044	15928.750244140625
Montreal	20034	15947.290405273438
Brisbane	20041	16118.479858398438
Bergen	20043	16363.099975585938
Nantes	20032	16560.300048828125
Charleroi	20051	16628.16015625
Osaka	20042	17114.43017578125
Toulouse	20033	17251.08056640625
Oulu	20042	17813.40008544922
San Francisco	20031	18695.579833984375
Bruxelles	20041	18800.089721679688
Reims	20032	18971.959716796875
Dublin	20043	18971.959838867188
Marseille	20044	20136.859985351562
Lille	20041	20178.1298828125
Paris	20051	20321.52978515625
Glendale	20032	20350.949768066406
South Brisbane	20051	21730.029907226562
Cambridge	20041	21782.699951171875
Madrid	20043	21986.269897460938
Makati City	20034	22841.960083007812
Nantes	20044	23031.589599609375
Los Angeles	20041	23889.320068359375
Kobenhavn	20044	24078.610107421875
Los Angeles	20034	24159.14013671875
Singapore	20041	24219.58984375
Montreal	20042	24564.530029296875
Paris	20033	25624.880004882812
Minato-ku	20042	25928.750244140625
Kobenhavn	20052	26012.87060546875
Bridgewater	20034	26115.800537109375
Helsinki	20051	26422.819458007812
Boston	20042	26677.350219726562
Liverpool	20034	26797.210083007812
Cowes	20041	26906.68017578125
South Brisbane	20034	27098.800048828125
Frankfurt	20034	27257.790283203125
Philadelphia	20031	27398.820434570312
Paris	20042	27931.210083007812

Newark 20042 27987.069580078125
 Chatswood 20033 28397.259521484375
 Glendale 20044 29343.349975585938
 Barcelona 20044 30183.35009765625
 San Francisco 20044 30348.720336914062
 Espoo 20042 31018.230102539062
 Tsawassen 20042 31302.500244140625
 Koln 20034 31363.1796875
 Brickhaven 20051 31474.7802734375
 Espoo 20033 31569.430053710938
 Boras 20051 31606.72021484375
 Bridgewater 20052 31648.469970703125
 London 20032 32376.29052734375
 NYC 20031 32647.809814453125
 Montreal 20052 33692.97009277344
 Las Vegas 20032 33847.61975097656
 Brisbane 20033 34100.030029296875
 Oslo 20043 34145.47021484375
 Las Vegas 20043 34453.84973144531
 Brickhaven 20033 34992.39978027344
 Munich 20043 34993.92004394531
 Strasbourg 20052 35680.350341796875
 Kobenhavn 20042 36079.010009765625
 Vancouver 20044 36576.70983886719
 New Haven 20042 36973.309814453125
 Frankfurt 20041 37266.48937988281
 Oulu 20033 37501.580322265625
 Burbank 20041 37850.07958984375
 Glen Waverly 20034 37878.54992675781
 Chatswood 20044 37905.14990234375
 Toulouse 20034 38098.240234375
 Minato-ku 20051 38191.38977050781
 Paris 20032 38217.41046142578
 Salzburg 20032 38629.14001464844
 Vancouver 20034 38662.209716796875
 White Plains 20034 38682.949462890625
 Dublin 20041 38784.470458984375
 Philadelphia 20034 40061.66003417969
 Bergamo 20034 40077.71026611328
 Aarhus 20034 40321.60998535156
 Liverpool 20052 40802.8095703125
 Chatswood 20043 41297.14050292969
 Reggio Emilia 20052 41509.94006347656
 Lyon 20034 41535.11022949219
 Bergamo 20044 41696.68981933594
 North Sydney 20034 41791.949462890625
 Burlingame 20033 42031.83020019531
 Helsinki 20034 42083.499755859375
 New Haven 20034 42498.760498046875
 Helsinki 20043 42744.0595703125
 Tsawassen 20043 43332.349609375
 Graz 20034 43488.740234375
 Singapore 20032 43657.47009277344
 Nantes 20052 43784.69012451172

Chatswood	20052	43971.429931640625
Barcelona	20034	44009.30993652344
Allentown	20044	44040.729736328125
Bridgewater	20042	44130.52062988281
Singapore	20033	44219.36022949219
Pasadena	20051	44273.359436035156
Madrid	20031	44621.96008300781
Reggio Emilia	20034	44669.740478515625
Strasbourg	20042	44758.12951660156
Salzburg	20044	45001.10986328125
Oslo	20034	45078.759765625
New Bedford	20033	45738.38952636719
Singapore	20042	45788.72009277344
Singapore	20043	46030.7197265625
Newark	20052	46518.99951171875
White Plains	20044	46873.04016113281
North Sydney	20033	47191.76013183594
Madrid	20033	47727.82019042969
Bruxelles	20043	47760.479736328125
Boston	20052	48316.89001464844
New Bedford	20051	48578.95935058594
Boras	20034	48710.92053222656
Cambridge	20043	48828.71942138672
Lille	20034	48874.28088378906
Reims	20044	48895.59014892578
San Rafael	20041	48922.76989746094
Oulu	20051	49055.40026855469
New Bedford	20044	49504.379150390625
Melbourne	20041	49637.57067871094
San Rafael	20034	50360.88977050781
Liverpool	20042	50408.25
Gensve	20041	50432.549560546875
Osaka	20041	50490.64013671875
Manchester	20031	51017.919860839844
Paris	20041	51172.649658203125
Cowes	20044	51334.15966796875
Espoo	20051	51373.49072265625
Sevilla	20034	51502.74108886719
Reims	20051	52029.07043457031
Marseille	20032	52481.840087890625
Boras	20043	53941.68981933594
San Francisco	20051	54203.6201171875
Cambridge	20044	54251.659912109375
Stavern	20031	54701.999755859375
Makati City	20031	55245.02014160156
Nashua	20044	55570.59973144531
Pasadena	20033	55776.119873046875
Minato-ku	20044	55888.65026855469
Bergamo	20031	56181.320068359375
Reggio Emilia	20043	56421.650390625
Kobenhavn	20031	58871.110107421875
Versailles	20044	59074.90026855469
Salzburg	20052	59475.100036621094
Nantes	20051	59617.39978027344

Melbourne	20032	60135.84033203125
Aarhus	20044	60273.93981933594
Stavern	20044	60422.530029296875
Nantes	20043	61310.880126953125
Burlingame	20034	62305.47009277344
NYC	20043	63027.92004394531
Boston	20034	63730.7802734375
Nashua	20034	63981.449768066406
New Bedford	20034	64053.130615234375
San Jose	20042	64600.33984375
North Sydney	20051	65012.41955566406
Lule	20044	66005.8798828125
Gensve	20043	67281.00903320312
Koln	20044	68943.40051269531
London	20034	70230.12927246094
NYC	20042	71860.779296875
Allentown	20043	71930.61041259766
Philadelphia	20044	76441.41040039062
Singapore	20034	77809.37023925781
Brickhaven	20043	79982.13989257812
San Diego	20041	87489.23010253906
Paris	20044	89436.60034179688
NYC	20034	89600.84008789062
Melbourne	20044	91221.99914550781
NYC	20032	93239.56018066406
San Rafael	20043	93928.72961425781
Torino	20043	94117.25988769531
Bergen	20034	95277.17993164062
San Jose	20052	95409.93041992188
Madrid	20032	100689.03051757812
Lyon	20041	101339.13977050781
Madrid	20041	105491.33990478516
Manchester	20044	106789.88977050781
Madrid	20034	112573.33026123047
San Rafael	20044	113622.75903320312
Madrid	20052	119242.98052978516
Madrid	20042	119656.04028320312
San Francisco	20034	121110.76025390625
San Rafael	20033	122368.67102050781
Madrid	20044	203007.4793701172
San Rafael	20051	205993.9287109375
Madrid	20051	207555.18994140625
NYC	20044	210410.85986328125

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

```

Select * from
  (Select *,
    rank() over (partition by year_id order by quantity desc) as rk
  from (select year_id, month_id , sum(quantityordered) as quantity
  from sales_order_data group by year_id, month_id) as sales ) as
query where rk=1;

```

Ans 2003	11	10179	1
2004	11	10678	1
2005	5	43571	