Insights from online retail industry Dataset

Presented By : -
Mr.Aman Mulla

#Questions

1. In How many countries working in. List down all countries name?
2. What is date range of given date. OR Data belongs to how many years or Months?
3. What is maximum and minimum quantity of description?
4. Which are 5 top countries in which maximum quantity of product sold?
5. Which are 5 bottom countries in which maximum quantity of product sold?
6. List down which are top 5 description sold in UK and Other than UK.
7. Plot Retention Chart.
8. What is Total sum of quantity sold per year and every month of 2011?
9. Which one is most demanded description along with stock code in UK?
10. Which one is most demanded description along with stock code other than UK.
11. What is Total revenue?
12. What is Yearly/Monthly/Quarterly revenue?
13. Calculate growth rate for Year on year and quarter on quarter on quarter?

#About Dataset

Column Name with respective data-type : -

column_name	data_type
invoicedate	timestamp without time zone
unitprice	double precision
customerid	integer
quantity	integer
country	character varying
stockcode	character varying
description	character varying
invoiceno	character varying

#About Dataset

- Taken Dataset belongs to online retail industry in which having data from **12-2010** to **12-2012.**

- Column Name in database as listed : - 1 . InvoiceNo
  2. StockCode
  3. Description
  4. Quantity
  5. InvoiceDate
  6. UnitPrice
  7. Customer Id
  8. Country

#SQL queries to copy/load dataset in Postgrey

```sql
drop table if exists retail_store;

create table retail_store(InvoiceNo varchar(25),StockCode varchar(25),Description varchar(350),Quantity int,
                        InvoiceDate timestamp,UnitPrice float,CustomerID int,Country varchar(50));

copy retail_store from 'G:\MySQL\SelfPractice Dataset\Tata\Online_Retail.csv' with delimiter ',' csv header encoding 'windows-1251';

select * from retail_store;
```

Kaggle Link for dataset

https://www.kaggle.com/datasets/vijayuv/onlineretail

#Insights

- Company Operating in total *37* countries.(One country Name is Unspecified so total count of country taken as 37)

```
8  select * from retail_store;
9  select count(distinct(country)) from retail_store;
```

- List of operating countries -

```
8  select * from retail_store;
9  select distinct(country) from retail_store;
```

#Insights

- Data Period

```
16  select distinct(InvoiceDate) from retail_store;
17  select min(InvoiceDate) from retail_store;
18  select max(InvoiceDate) from retail_store;
```

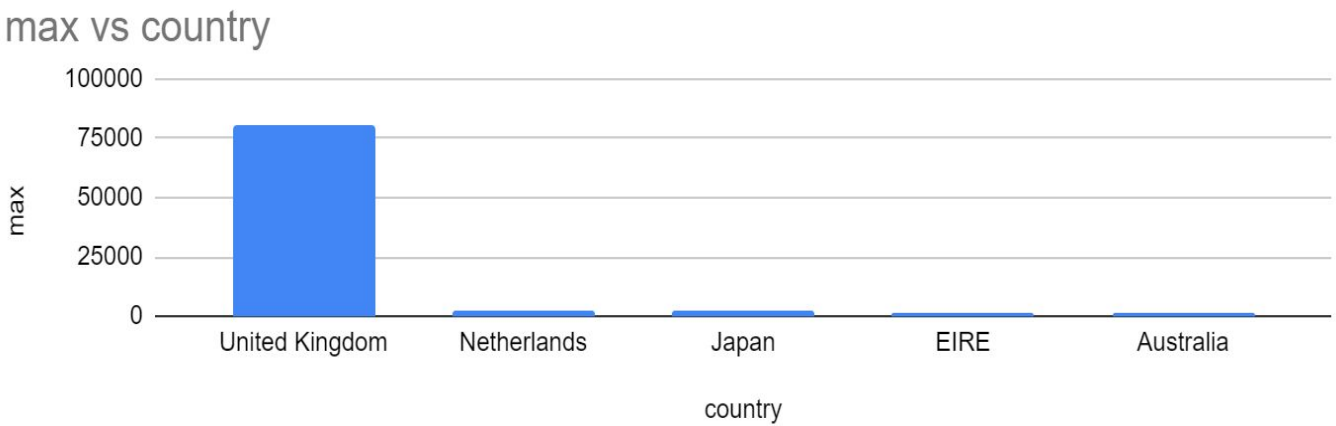
- Max/Min Quantity order

```
20  #Maximum Quantity order
21  select min(Quantity) from retail_store;
22  select max(Quantity) from retail_store;
```

#Insights

- Countries which having highest quantity (Descending and ascending order)

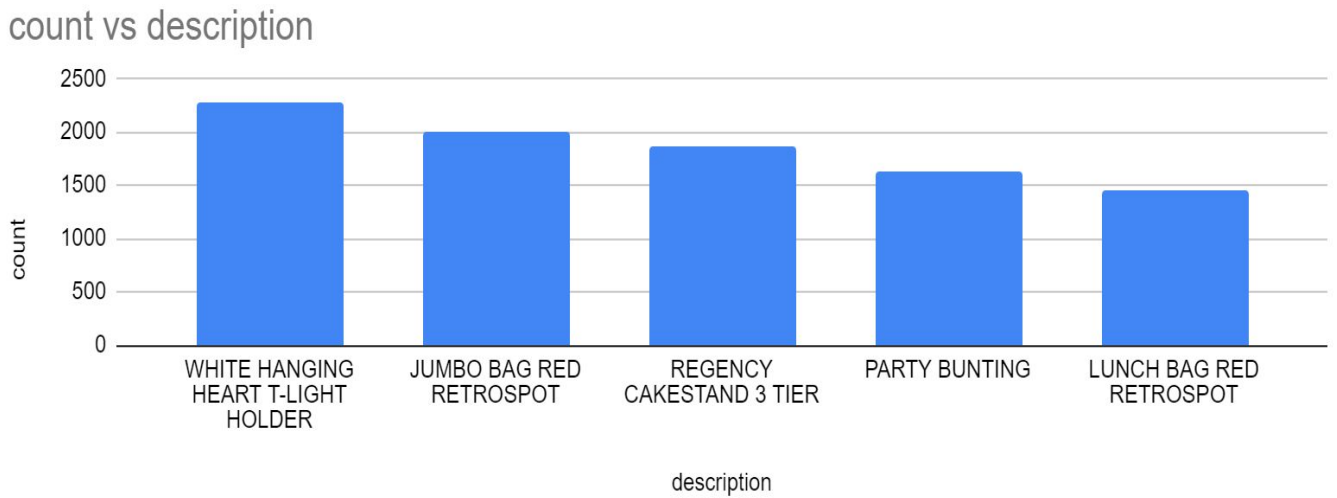
```
25  select max(Quantity),Country from retail_store group by Country order by max(Quantity) desc limit 5;
26  select max(Quantity),Country from retail_store group by Country order by max(Quantity) limit 5;
```



#Insights

- Highest Description count with Description in UK.

```
40  select distinct count(Description),Description, country from retail_store group by country,Description
41  having country ='United Kingdom' order by count(Description) desc limit 5
```

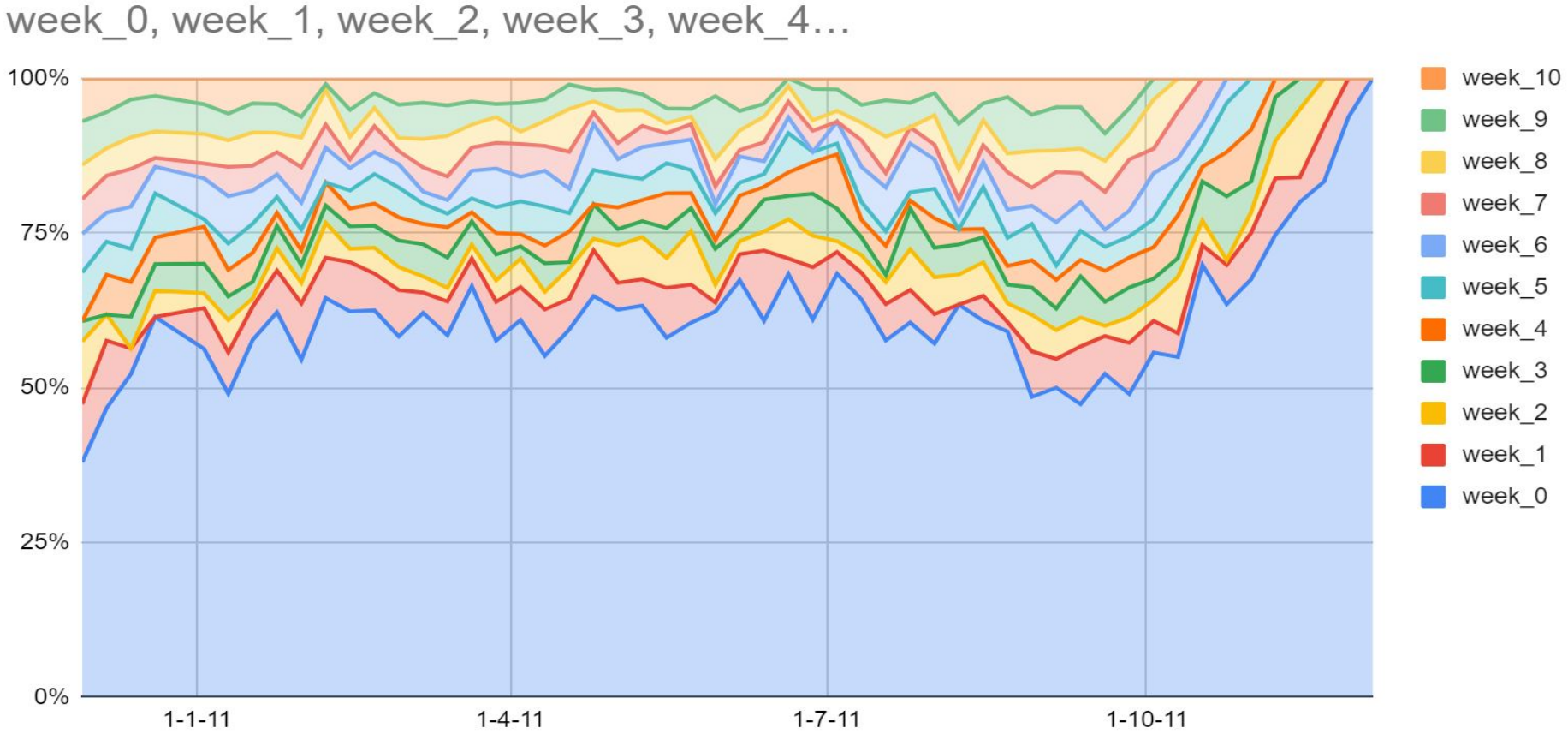


#Insights

- Retention Table - Retention table shows that how many customer join on 0th week and how many of them are come back in next weeks.

```
#Retention Table
select * from retail_store;
drop table if exists retention_table;
create temporary table retention_table as select *,date(date_trunc('week',invoicedate)) as first_day_of_week from retail_store;
select * from retention_table;
select * from retention_table order by customerid;
drop table if exists first_invoice_date;
create temporary table first_invoice_date as select customerid,min(first_day_of_week) as weekstartday
from retention_table group by 1 order by 1;
select * from first_invoice_date;
drop table if exists week_diff;
create temporary table week_diff as select a.customerid, a.first_day_of_week,b.weekstartday,(a.first_day_of_week-b.weekstartday)/7 as week_diff_val
from retention_table a inner join first_invoice_date b on a.customerid=b.customerid;
select * from week_diff order by customerid;
select weekstartday,
                  count(distinct case when week_diff_val =0 then customerid end) as week_0,
                  count(distinct case when week_diff_val =1 then customerid end) as week_1,
                  count(distinct case when week_diff_val =2 then customerid end) as week_2,
                  count(distinct case when week_diff_val =3 then customerid end) as week_3,
                  count(distinct case when week_diff_val =4 then customerid end) as week_4,
                  count(distinct case when week_diff_val =5 then customerid end) as week_5,
                  count(distinct case when week_diff_val =6 then customerid end) as week_6,
                  count(distinct case when week_diff_val =7 then customerid end) as week_7,
                  count(distinct case when week_diff_val =8 then customerid end) as week_8,
                  count(distinct case when week_diff_val =9 then customerid end) as week_9,
                  count(distinct case when week_diff_val =10 then customerid end) as week_10
           from week_diff group by 1 order by 1;
```

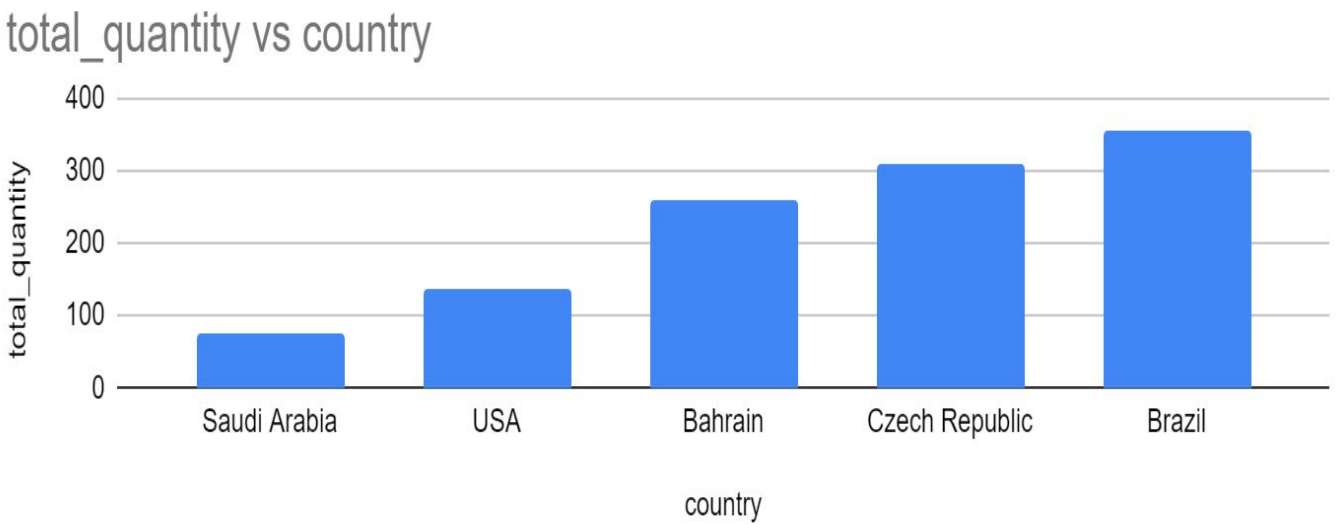
weekstartday	week_0	week_1	week_2	week_3	week_4	week_5	week_6	week_7	week_8	week_9	week_10
05-12-11	26	0	0	0	0	0	0	0	0	0	0
28-11-11	45	3	0	0	0	0	0	0	0	0	0
21-11-11	65	7	6	0	0	0	0	0	0	0	0
14-11-11	80	4	11	5	0	0	0	0	0	0	0
07-11-11	74	9	6	7	3	0	0	0	0	0	0
31-10-11	81	9	4	6	10	10	0	0	0	0	0
24-10-11	80	8	1	13	9	10	5	0	0	0	0
17-10-11	88	4	5	8	3	4	5	9	0	0	0
10-10-11	72	5	12	4	9	7	5	10	7	0	0
03-10-11	98	9	6	6	9	8	13	7	14	6	0
26-09-11	71	12	6	7	7	5	6	12	6	6	7
19-09-11	94	11	3	7	9	7	5	11	9	8	16
12-09-11	71	14	7	10	4	7	7	7	6	10	7
05-09-11	43	4	4	3	4	2	6	7	3	6	4
29-08-11	33	5	4	3	3	4	2	2	4	4	4
22-08-11	39	1	2	2	2	3	3	4	2	6	2
15-08-11	45	3	4	3	1	5	3	2	3	2	3
08-08-11	26	0	2	2	1	0	1	1	2	3	3
01-08-11	48	4	5	4	4	4	4	2	4	3	2
25-07-11	46	4	5	5	1	1	6	2	0	3	3
18-07-11	49	5	3	1	4	2	6	2	5	5	3
11-07-11	45	3	2	2	2	2	4	3	2	2	3
04-07-11	39	2	1	3	5	1	2	0	1	2	1
27-06-11	36	5	3	4	3	1	0	2	1	3	1
20-06-11	54	2	5	3	3	5	2	2	2	1	0
13-06-11	59	11	3	5	2	2	2	3	4	2	4
06-06-11	64	4	2	2	5	2	4	1	3	3	5
30-05-11	43	1	2	4	1	3	1	2	3	7	2
23-05-11	49	5	7	3	2	3	4	2	1	1	4
16-05-11	72	10	6	6	7	6	4	2	2	3	6
09-05-11	74	5	8	3	4	4	6	4	3	3	3
02-05-11	72	5	7	3	4	6	3	3	6	4	2
25-04-11	35	4	1	3	0	3	4	1	1	1	1
18-04-11	60	5	5	1	5	3	4	6	7	4	1
11-04-11	96	13	5	8	5	11	10	7	7	6	6
04-04-11	92	8	7	3	3	8	6	8	3	7	6
28-03-11	83	9	5	6	5	6	9	6	6	3	6
21-03-11	89	6	3	5	2	3	6	5	5	5	5


week_0, week_1, week_2, week_3, week_4…
100%
75%
50%
25%
0%
1-1-11
1-4-11
1-7-11
1-10-11
week_10
week_9
week_8
week_7
week_6
week_5
week_4
week_3
week_2
week_1
week_0

#Insights

- Total Sum of Quantity per country Bottom 5

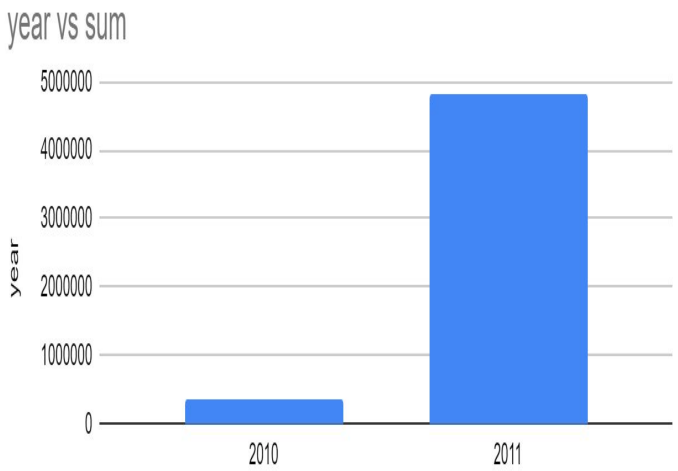
```
40  select country,sum(Quantity) as Total_Quantity from retail_store1 group by country order by sum(Quantity) limit 5;
```

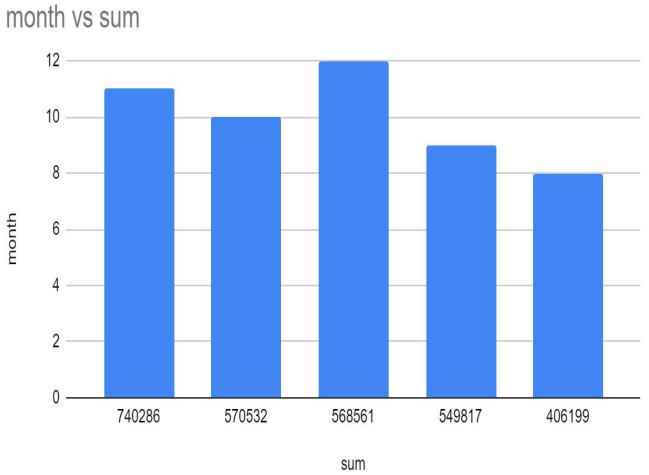


#Insights

- Extracting sum of quantity yearly and monthly

```
117  #YearWise/Monthwise Data
118  select sum(quantity), extract (Year FROM invoicedate) as Year from retail_store group by Year;
119  select sum(quantity), extract (Month FROM invoicedate) as Month from retail_store group by Month  order by sum(quantity) desc limit 5;
```





#Insights

- Most demanded description with stock-code in UK.

```
145 #Most demanded stockcode with description in UK.
146 select stockcode,description,country,count(stockcode) from retail_store where country='United Kingdom'  group by
147 stockcode,description,country order by count(stockcode) desc limit 5
```

- Most demanded description with stock-code other than UK.

```
146 select stockcode,description,country,count(stockcode) from retail_store where country!='United Kingdom'  group by
147 stockcode,description,country order by count(stockcode) desc limit 5
```

#Insights

- Revenue Calculation

```
154 select description,quantity,unitprice from retail_store;
155
156 drop table if exists Revenue1;
157 create temporary table Revenue1 as select *,(quantity*unitprice) as Revenue from retail_store;
158 select * from Revenue1;
159
160 select sum(Revenue) as Total_revenue from Revenue1
```

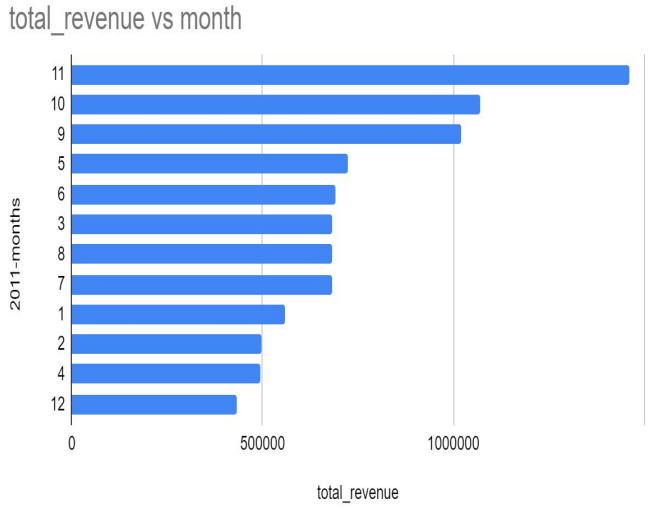
Total Revenue Amount : - Rs. 9747747.93

#Insights

- Monthly revenue (Only 2 months of 2010 hence calculated for 2011 year)

```
170 select extract(Year FROM invoicedate) as Year,extract(Month FROM invoicedate) as Month,sum(Revenue) as
171 Total_revenue from Revenue1 where extract (Year FROM invoicedate)='2011' group by Year,Month order by Total_revenue desc limit 5;
```

year	month	total_revenue
2011	11	1461756.25
2011	10	1070704.67
2011	9	1019687.622
2011	5	723333.51
2011	6	691123.12
2011	3	683267.08
2011	8	682680.51
2011	7	681300.111
2011	1	560000.26
2011	2	498062.65
2011	4	493207.121
2011	12	433668.01

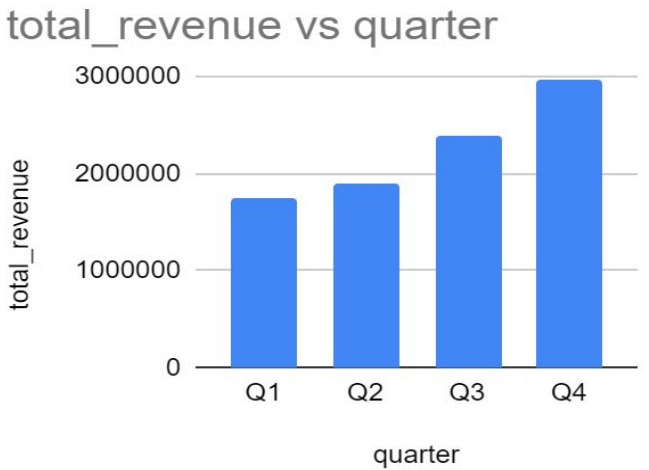


#Insights

- Quarter- wise Revenue

```
175  SELECT EXTRACT(YEAR FROM invoicedate) AS Year, CONCAT('Q', EXTRACT(QUARTER FROM invoicedate)) AS Quarter,SUM(Revenue) AS Total_revenue
176  FROM Revenue1 where extract (Year FROM invoicedate)='2011'GROUP BY Year, Quarter ORDER BY Year,Quarter;
```

year	quarter	total_revenue ▲
2011	Q1	1741329.99
2011	Q2	1907663.751
2011	Q3	2383668.243
2011	Q4	2966128.93



Thank You