use anabig114231;

CREATE EXTERNAL TABLE Retail1

(

InvoiceNo INT,

Stockcode STRING,

Description STRING,

Quantity INT,

InvoiceDate STRING,

Unitprice int,

CustomerID INT,

Country STRING

)

ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t';

LOAD DATA INPATH '/user/anabig114231/OnlineRetail.txt' INTO TABLE retail1;

ALTER TABLE retail1

SET TBLPROPERTIES ("skip.header.line.count"="1");

select \* from retail1;

##1 Revenue Aggregate by Country for top 5 countries.

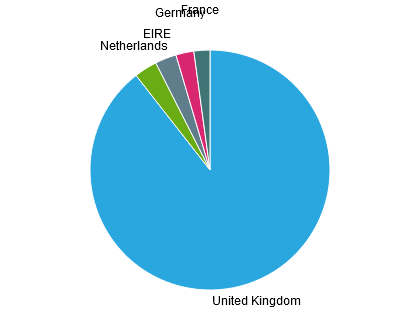
select country, sum(cast(Quantity as int )\* Unitprice) as total\_revenue from retail1

group by country

order by total\_revenue desc

limit 5;





##2 Sales Metrics like NumCustomers, NumTransctions, AvgNumItems, MinAmtperCustomer, MaxAmtperCustomer, AvgAmtperCustomer, StdDevAmtperCustomer etc by country for top 5 countries

WITH InvoiceAmount as (Select country, customerid, invoiceNo, count(distinct stockcode) as NumItems,

sum(cast(Quantity as int)\* Unitprice) as InvoiceTotal from retail1

where country in ('United Kingdom','Netherlands','EIRE','Germany','France')

group by country, invoiceno, customerid)

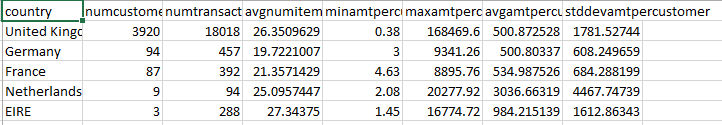
select country, count(distinct customerid)as Numcustomers, count(distinct invoiceno)as Numtransactions,

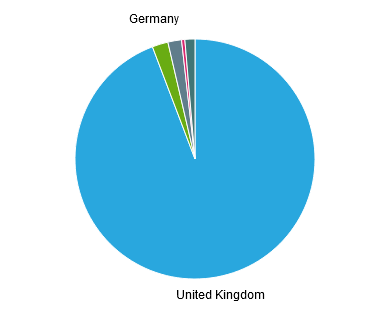
avg(NumItems) as AvgNumItems, min(InvoiceTotal) as MinAmtpercustomer, max(InvoiceTotal) as MaxAmtperCustomer,

avg(InvoiceTotal) as AvgAmtpercustomer, std(InvoiceTotal) as StdDevAmtperCustomer from InvoiceAmount

group by country

order by Numcustomers desc;





## 3. Daily Sales Activity like NumViists, TotalAmt, etc… per POSIX day of the year

select day(from\_unixtime(unix\_timestamp(invoicedate,'yyyy-MM-dd hh:mm'))) as dayofyear,

month(from\_unixtime(unix\_timestamp(invoicedate,'yyyy-MM-dd hh:mm'))) as monthofyear,

year(from\_unixtime(unix\_timestamp(invoicedate,'yyyy-MM-dd hh:mm'))) as yearofyear,

count(distinct invoiceno) as Numvisits, sum(cast(quantity as int) \* unitprice) as totalamount

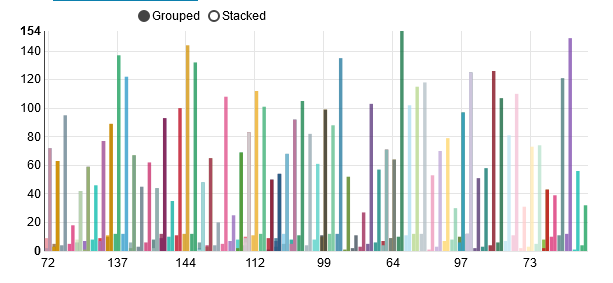
from retail1 where invoicedate > '2010-12-01' and invoicedate < '2011-12-09'

GROUP BY day(from\_unixtime(unix\_timestamp(invoicedate,'yyyy-MM-dd hh:mm'))),

month(from\_unixtime(unix\_timestamp(invoicedate,'yyyy-MM-dd hh:mm'))),

year(from\_unixtime(unix\_timestamp(invoicedate,'yyyy-MM-dd hh:mm')))

order by dayofyear, monthofyear, yearofyear;



## 4. Hourly Sales Activity like NumViists, TotalAmt, etc… per hour of day.

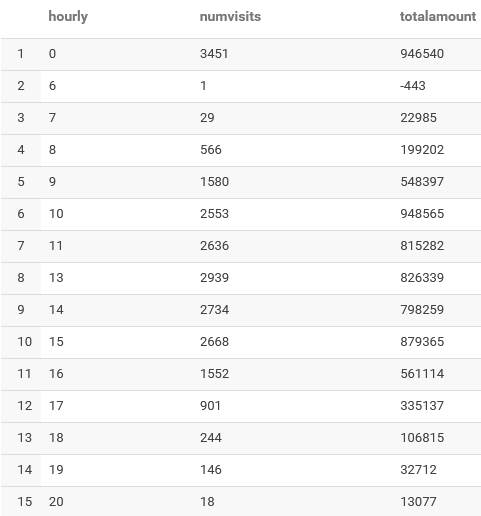
select hour(from\_unixtime(unix\_timestamp(invoicedate,'yyyy-MM-dd hh:mm'))) as hourly,

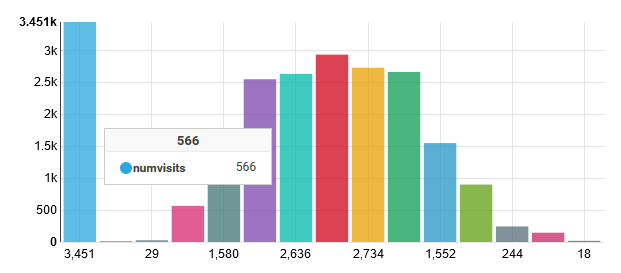
count(distinct invoiceno) as Numvisits, sum(cast(quantity as int) \* unitprice) as totalamount

from retail1 where invoicedate > '2010-12-01' and invoicedate < '2011-12-09'

GROUP BY hour(from\_unixtime(unix\_timestamp(invoicedate,'yyyy-MM-dd hh:mm')))

order by hourly;

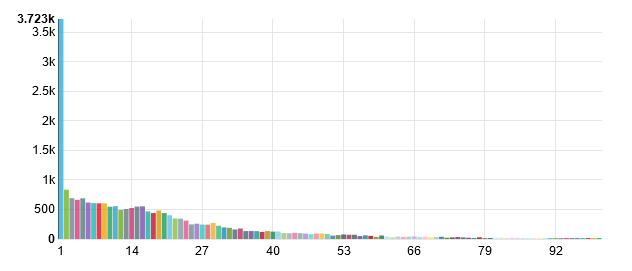




## 5. Basket size distribution (Note : Basket size= number of items in the transaction

With ItemDist as (Select invoiceno, count(distinct Stockcode) as Numitems from retail1 group by invoiceno)

select Numitems, count(Numitems) as CountNumitems from ItemDist group by Numitems order by Numitems asc;



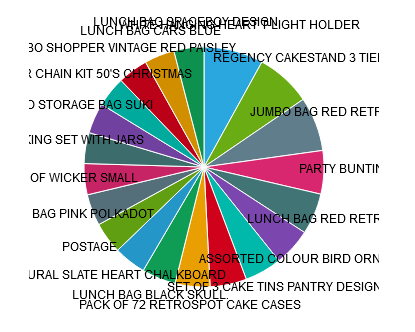
## 6. Top 20 items sold by frequency

Select Description, count(Stockcode) as ItemFrequency from retail1

group by Description

order by ItemFrequency desc limit 20;





## 7. Customer Lifetime Value distribution by intervals of 1000’s

With CLV as (select customerid, count(distinct invoiceno) as Numtransactions,

ceil(sum(cast(quantity as int) \* unitprice) /1000)\*1000 as CustomerLifeValue from retail1

where customerid is not NULL group by customerid)

select distinct CustomerLifeValue from CLV

order by CustomerLifeValue desc;

