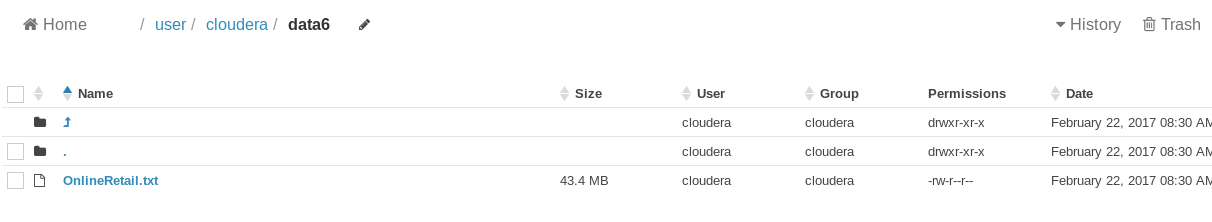
Project6: Retail Sales Analytics

1. saving table at appropriate HDFS location –



2 . Revenue Aggregate By Country for top 10 countries

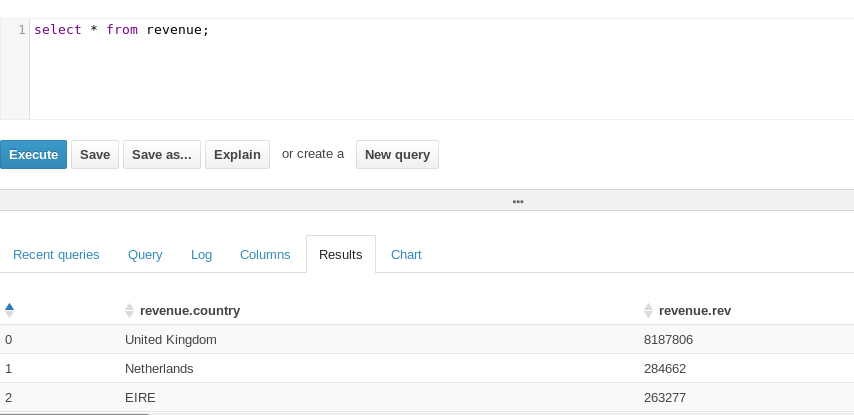
From command line pig –f /home/cloudera/Desktop/pig\_proj6.pig



|  |  |  |  |
| --- | --- | --- | --- |
| 3 . output : | |  | |
|  | |  | |
|  | |  | |

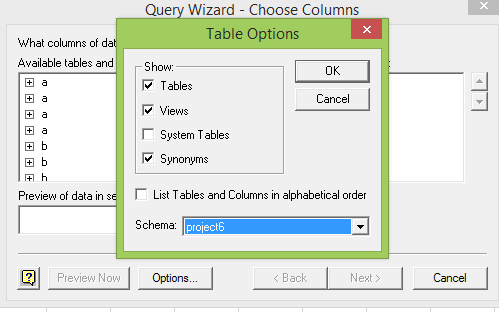
Creating hive table corresponding to pig file containing Revenue Aggregate By Country .

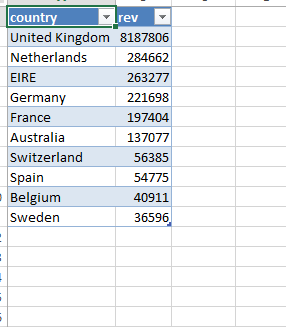


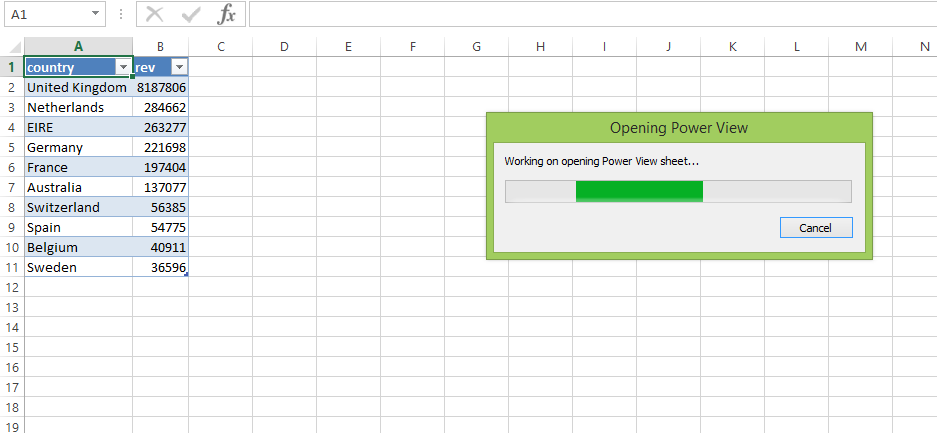


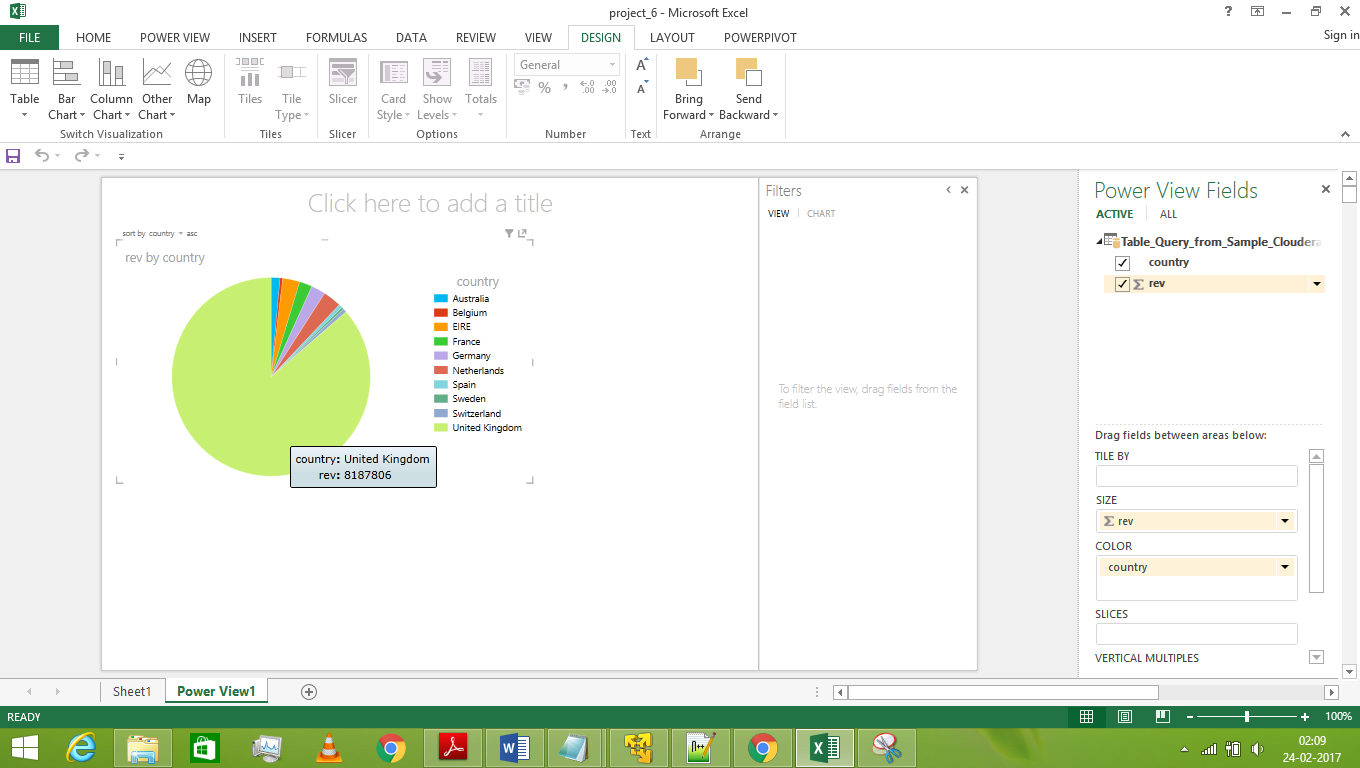
4 . Integration with excel 2013 and showing BI report

(process of integration with excel 2013 has been explained in project 1)

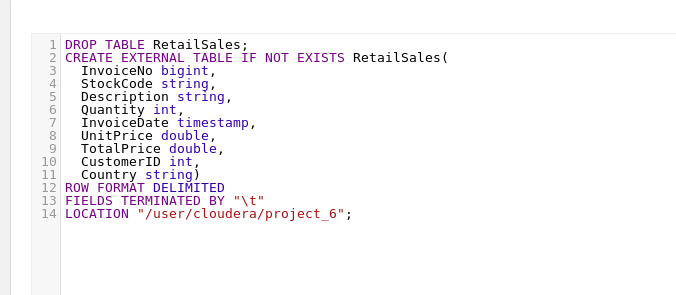


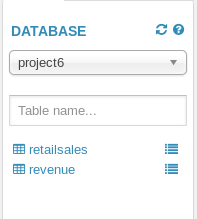






5 . Creating a hive table corresponding to previous pig cleaned data ‘RetailSalesClean’ and which was further saved at location /user/cloudera/project\_6





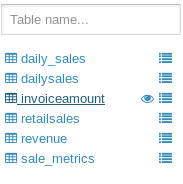
Ques :Sales Metrics like NumCustomers, NumTransactions, AvgNumItems, MinAmtperCustomer, MaxAmtperCustomer, AvgAmtperCustomer, StdDevAmtperCustomer etc. .. by country for top 5 countries

In command promt :

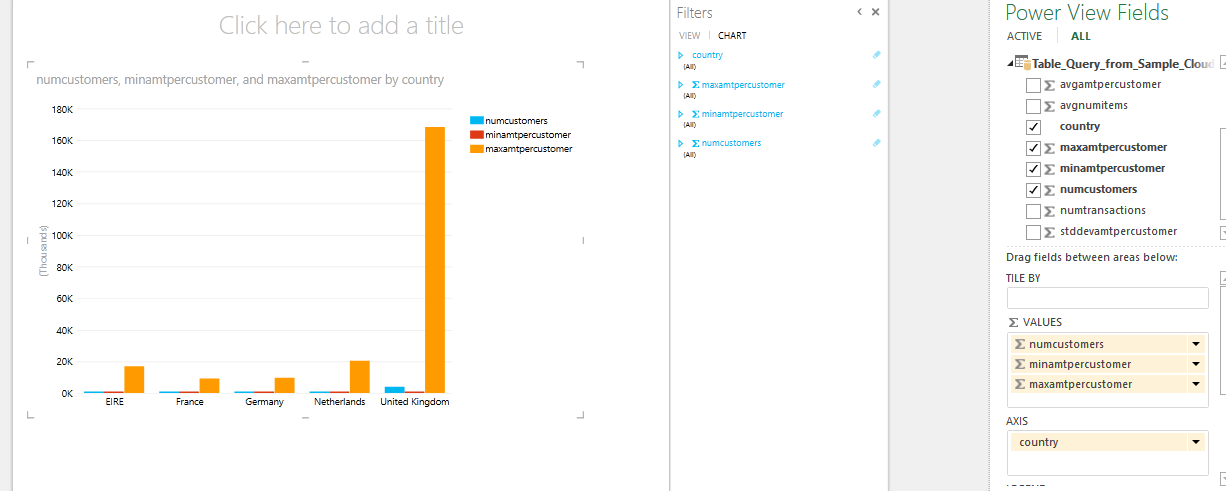
Hive –f /user/cloudera/Desktop/untitled/hive\_analysis\_1.hql



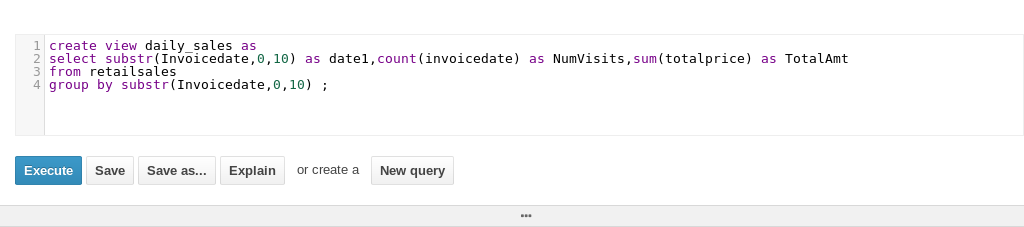
Output :

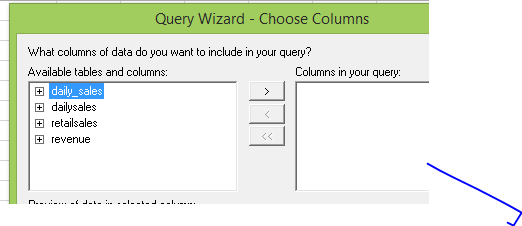


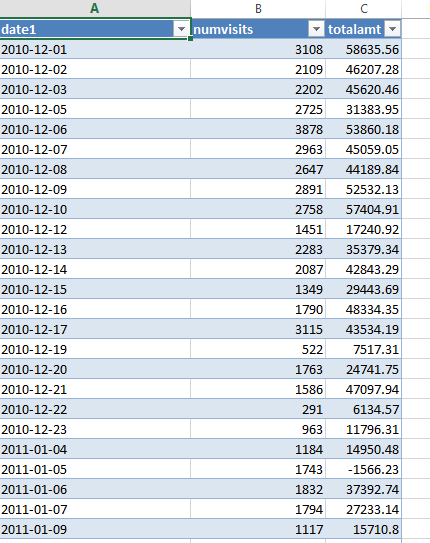
Analysis :

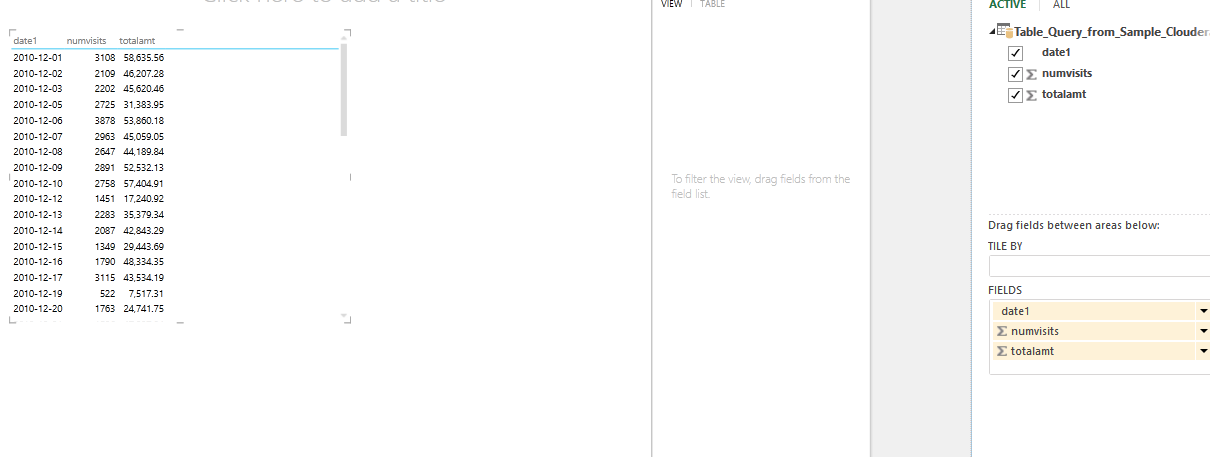


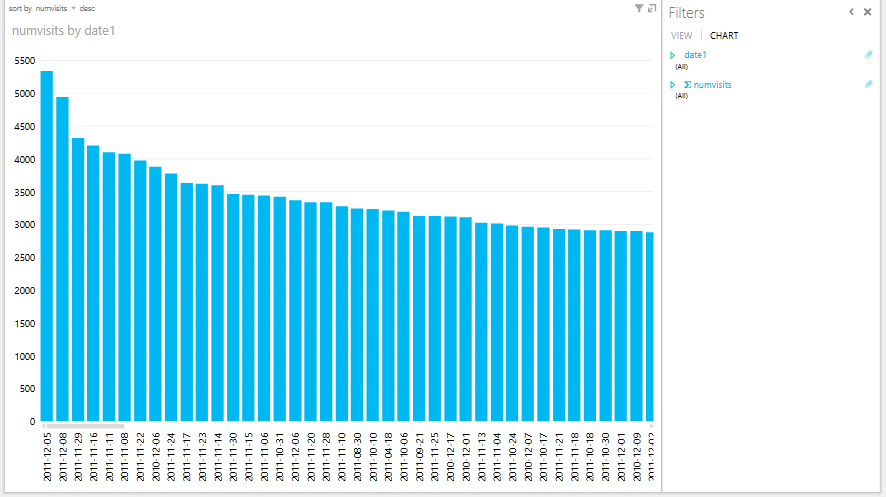
Ques 3 :Daily Sales Activity like NumVisits, TotalAmt etc… per POSIX day of the year



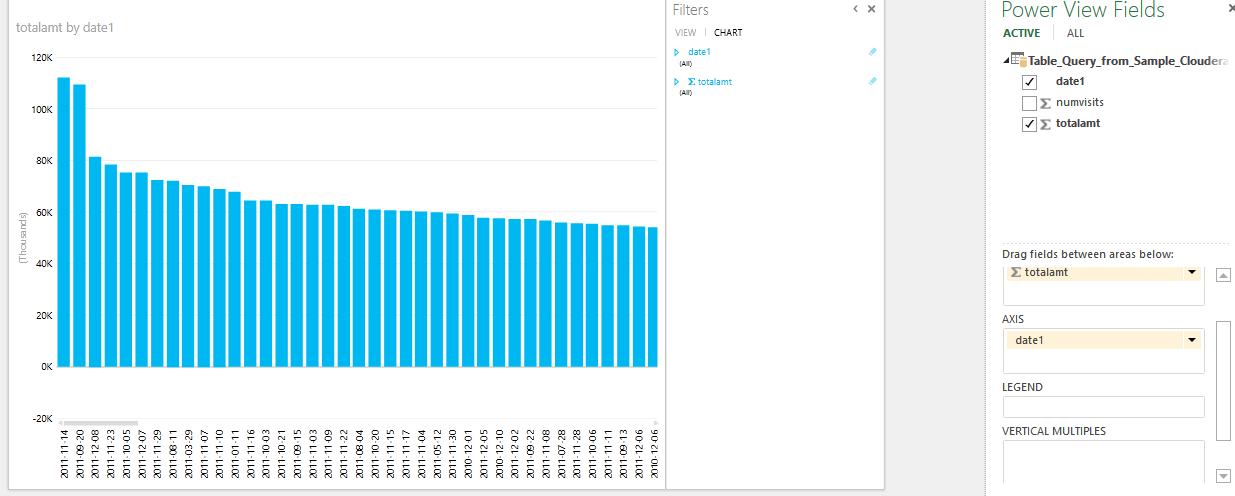






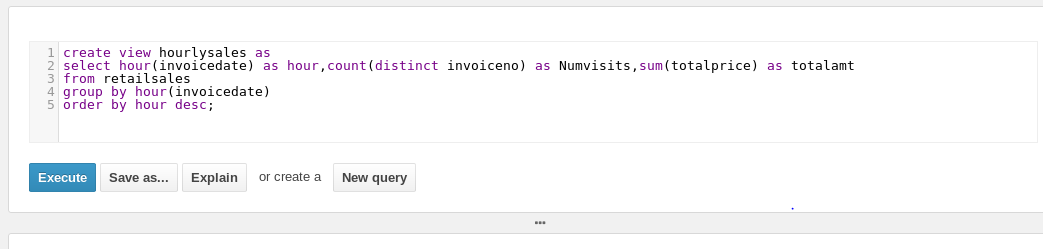


We see that most number of visits is on 05-12-2011

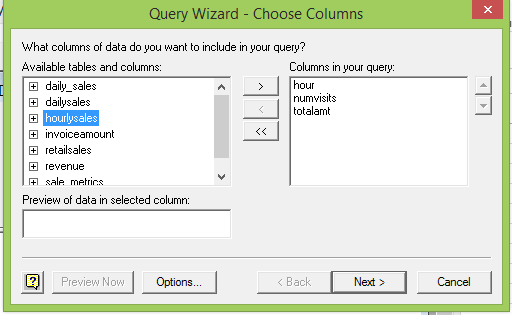


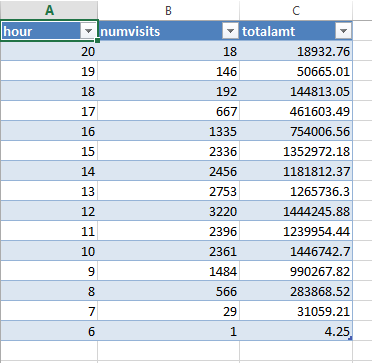
Max revenue generated on 14-11-2011 .

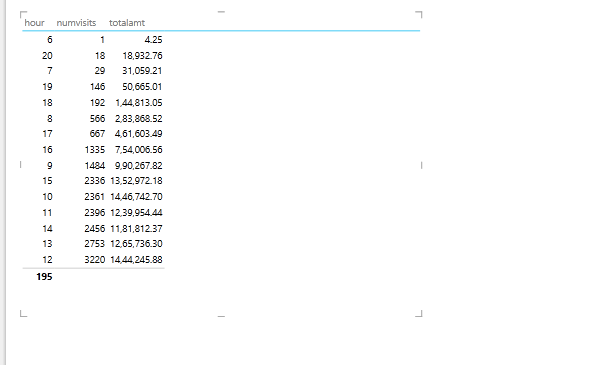
Ques 4 . Hourly sales Activity like NumVisits, TotalAmt etc… per hour of day



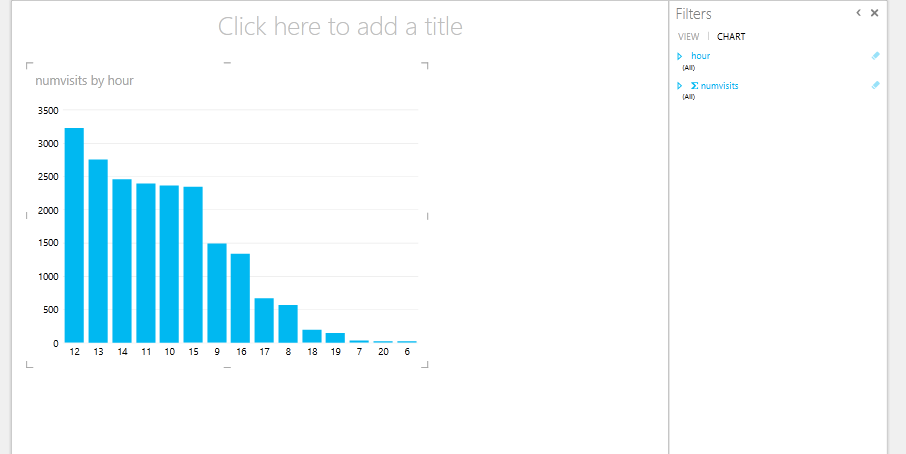
Analysis :



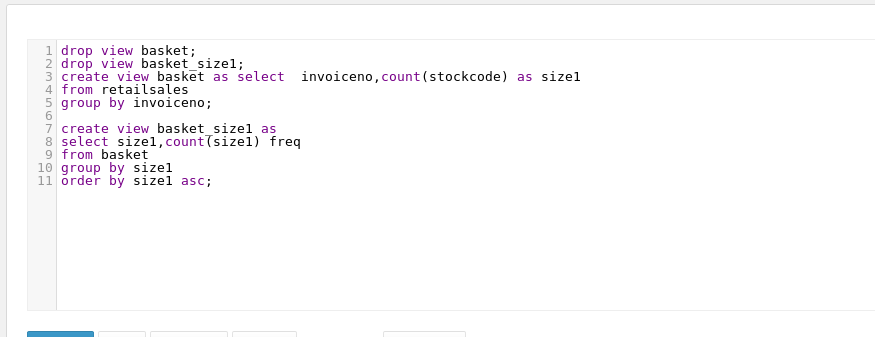




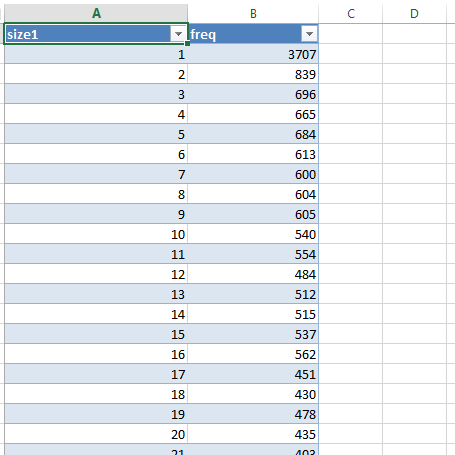
Hourly sales trend :

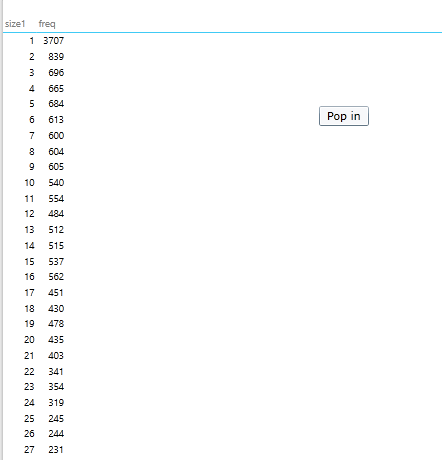


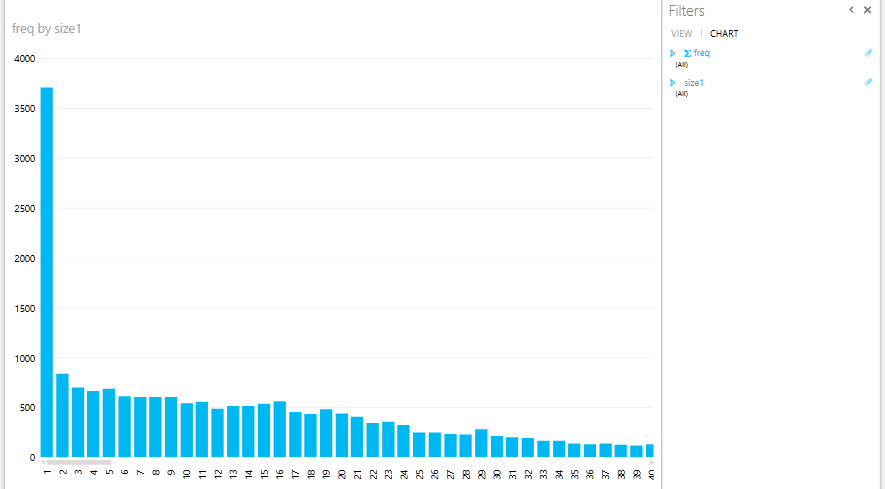
Ques 5 : Basket size distribution



Analysis:





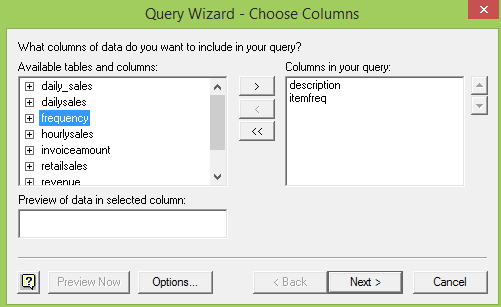


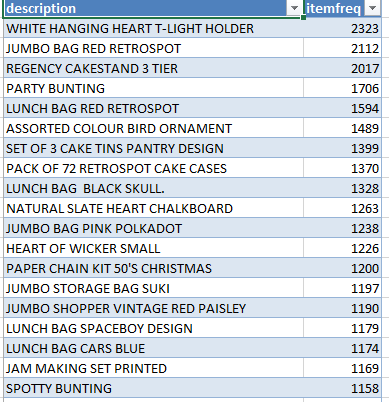
Hence we can see the basket size distribution . Most of the customer purchased a single product .

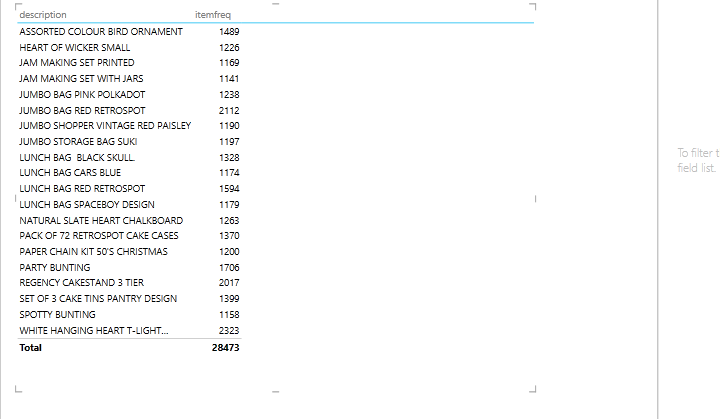
Ques 6 : Top 20 Items sold by frequency

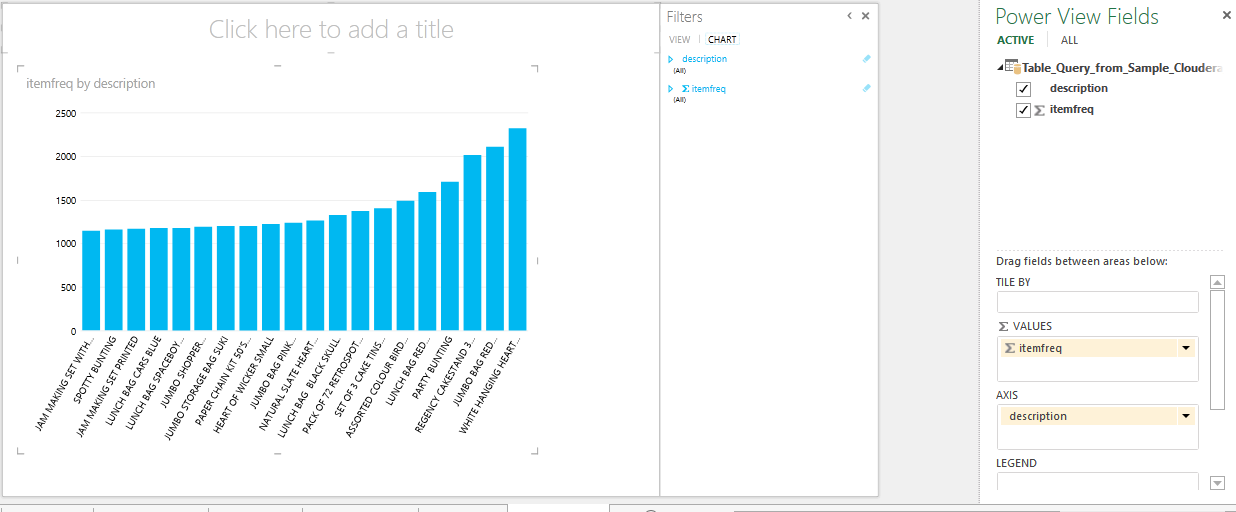


Analysis :





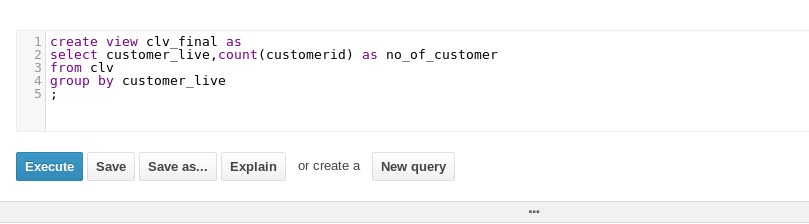




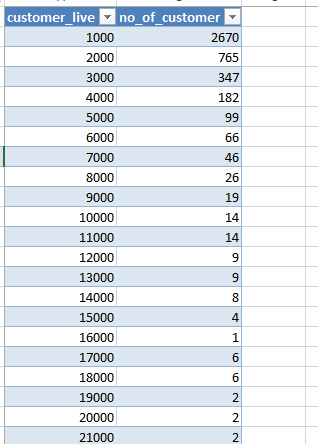
Hence we see the top 20 item sold by frequency .

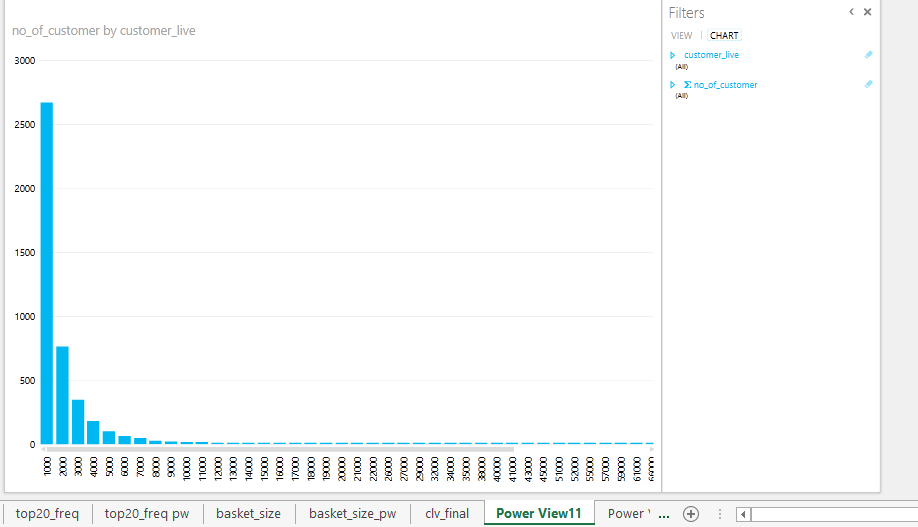
Ques 7 . Customer Lifetime Value distribution by intervals of 1000’s





Analysis :





Above is the desired result .

Excel 2013 analysis link ( pls enable power view mode in ms excel 2013):

<https://drive.google.com/open?id=0BxjYrCpAmIoqQjU4RDZqaXBnZk0>

Ques 8 .

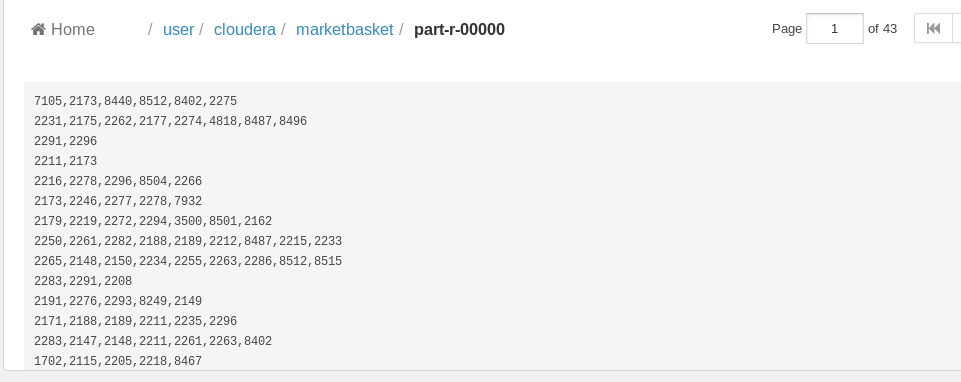
Create a “MarketBaskets.pig” code for data preparation for the market basket analysis using any predictive modeling software like spark MLLib

From command promt :

Pig –f /home/cloudera/Desktop/untitled/marketbasket.pig



Output :



Above file can be readily used for further analysis .