**Teleco Customer Churn Analysis**

**Used Hive,HUE and HDFS**

A telecommunications company is concerned about the number of customers leaving their landline business for cable competitors. They need to understand who is leaving. This data set provides info to help you predict behavior to retain customers. You can analyze all relevant customer data and develop focused customer retention programs.

**Data Set Description**

The data set includes information about:

* Customers who left within the last month – the column is called Churn
* Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
* Customer account information – how long they’ve been a customer, contract, payment method, paperless billing, monthly charges, and total charges
* Demographic info about customers – gender, age range, and if they have partners and dependents

**Exploration Ideas**

1. Display average bill (Total charges) paid by male and female.

2. Display the most frequently used payment method a male.

3. If the customer is not having any internet connection then what is the minimum bill that the customer has paid.

4. Count the number of senior citizen who have dependents (i.e. 1) and give them 20% discount on final bill and marked them as 'Sr Discount'.

5. Display customer id and their tenure whose tenure is more than 20 years and are using paperless billing method to save an environment(Note: Here the tenure given in the dataset is in months. You are expected to convert these months to years and proceed)

6. Which is the most preferred internet service used by the customer, is it for Fiber optic or for DSL?

7. Customer who are using Streaming movies option. Calculate its final bill by increasing the monthly bill by 9.5%

Create database bigdata\_foundation\_project\_stephy;

Use bigdata\_foundation\_project\_stephy;

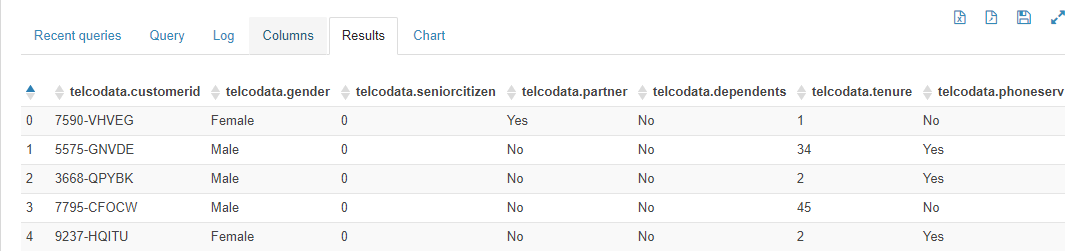
create table telcodata(customerid STRING, gender STRING, seniorcitizen INT,partner STRING,dependents STRING, tenure INT, phoneservice STRING, multiplelines STRING, internetservice STRING, onlinesecurity STRING,onlinebackup STRING,deviceprotection STRING, techsupport STRING,streamingtv STRING,streamingmovies STRING,contract STRING,paperlessbilling STRING, paymentmethod STRING, monthlycharges FLOAT, totalcharges FLOAT, churn FLOAT)

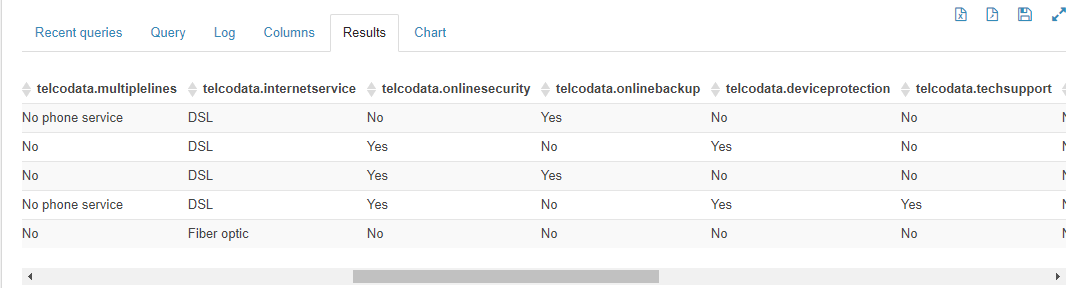
row format delimited fields terminated by ','

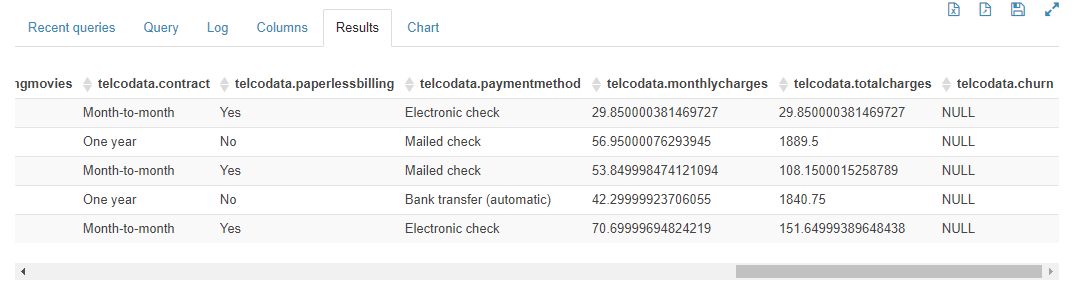
tblproperties ("skip.header.line.count"="1");

load data inpath '/user/mrstephyjacob4610/Churn.csv' into table telcodata;

select \* from telcodata limit 5;

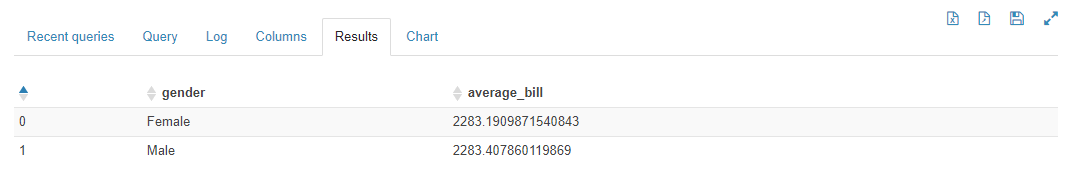


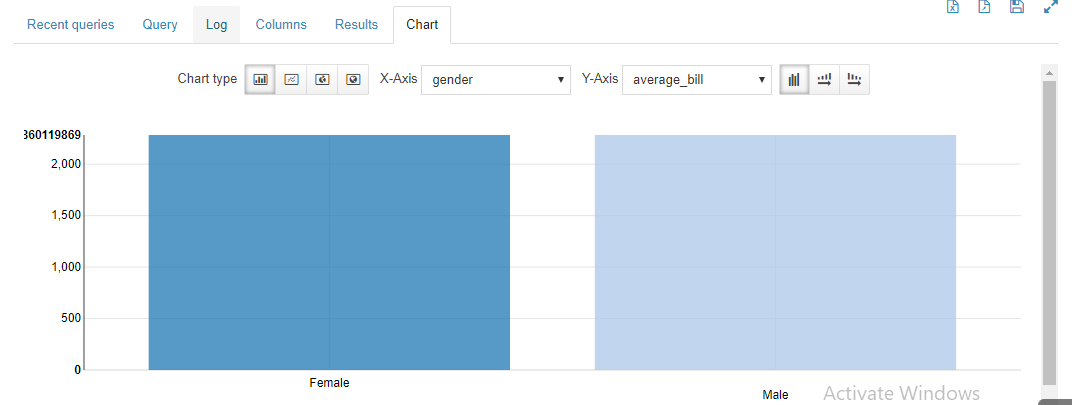




**/\* Display average bill (Total charges) paid by male and female\*/**

select gender,avg(totalcharges) as average\_bill from telcodata group by gender;

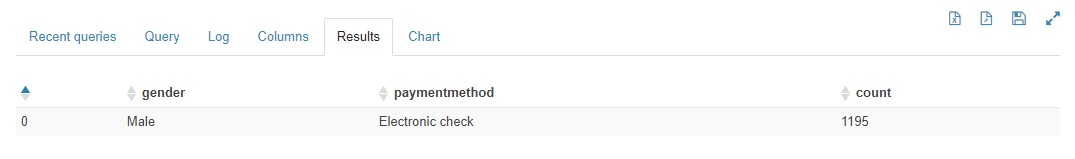




**/\* Display the most frequently used payment method a male.\*/**

select gender,paymentmethod,count(paymentmethod) as count from telcodata where gender =='Male' group by gender,paymentmethod

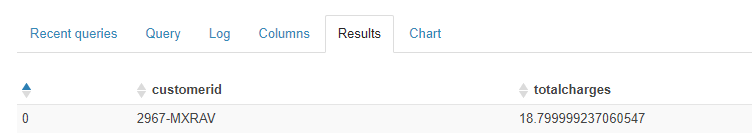
order by count desc limit 1;



**/\*If the customer is not having any internet connection then what is the minimum bill that the customer has paid.\*/**

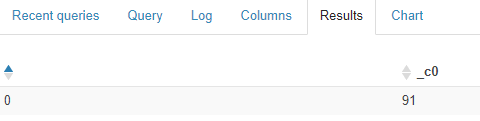
select customerid,totalcharges from telcodata where internetservice=='No' and totalcharges >=0

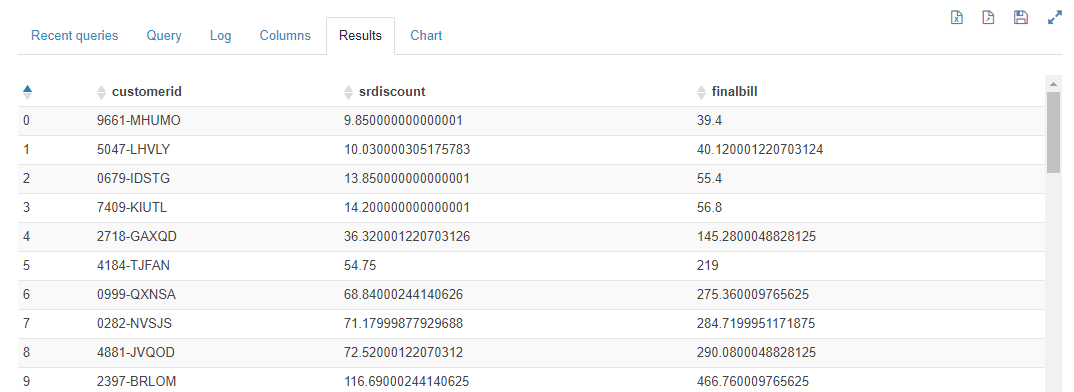
order by totalcharges limit 1;



**/\* Count the number of senior citizen who have dependents (i.e. 1) and give them 20% discount on final bill and marked them as 'Sr Discount'\*/**

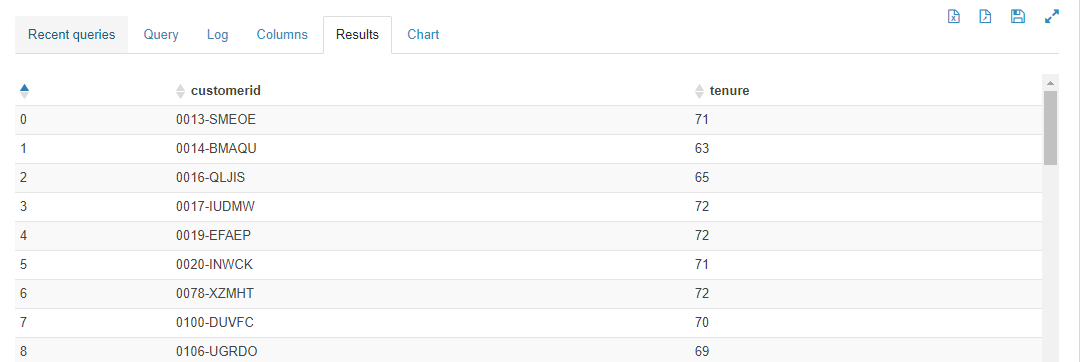
select count(customerid) from telcodata where seniorcitizen=1 and dependents=='Yes';





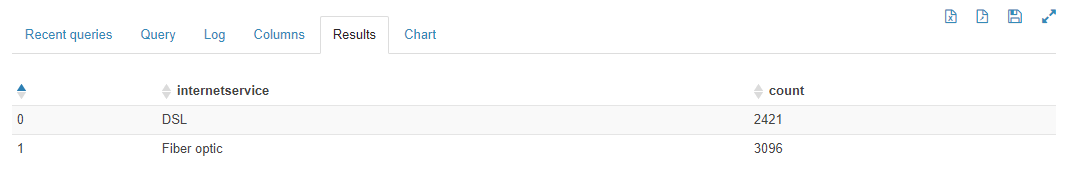
**/\* Display customer id and their tenure whose tenure is more than 20 years and are using paperless billing method to save an environment(Note: Here the tenure given in the dataset is in months. You are expected to convert these months to years and proceed)\*/**

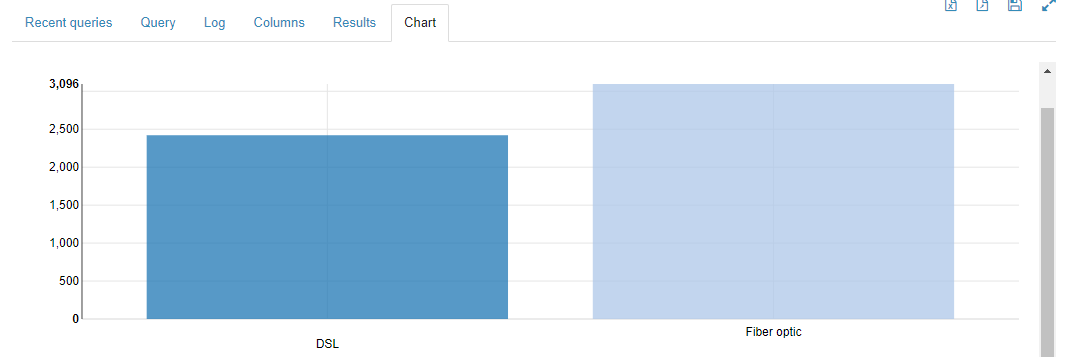
select customerid,tenure from telcodata where (tenure/12)>=5 and paperlessbilling =='Yes' order by customerid;



**/\* Which is the most preferred internet service used by the customer, is it for Fiber optic or for DSL?\*/**

select internetservice,count(internetservice) count from telcodata where internetservice=='DSL' or internetservice=='Fiber optic' group by internetservice;





**/\* Customer who are using Streaming movies option. Calculate its final bill by increasing the monthly bill by 9.5%\*/**

select customerid,monthlycharges as current\_monthly\_charges,(monthlycharges+monthlycharges\*0.095) as new\_monthly\_charges

from telcodata where streamingmovies=='Yes' order by customerid;

