**HIVE QUERIES**

**Basic Query to create table in Hive and load data in hive table:**

**Create Table:**

CREATE TABLE Consumer\_Complaints(Date\_Received date, Product varchar(30), Subproduct varchar(100), Issue varchar(200), Sub\_Issue varchar(200), Consumer\_complaint\_narrative varchar(200), Company\_public\_response varchar(200), Company varchar(30), State varchar(10), ZIP\_code varchar(10), Submitted\_via varchar(10), Date\_sent\_to\_company date, Company\_response\_to\_consumer varchar(200), Timely\_response varchar(10), Consumer\_disputed varchar(10), Complaint\_ID int) row format serde 'com.bizo.hive.serde.csv.CSVSerde' stored as textfile;

L**oading Data in Hive Table:**

LOAD DATA inpath '/user/user01/Project/Consumer\_Complaints.csv' overwrite INTO TABLE Consumer\_Complaints;

**Query 1:**

INSERT OVERWRITE LOCAL DIRECTORY '/Project/Statewise\_Complaints' select count(complaint\_id) as Complaint\_Count , state from consumer\_complaints where state <> "" group by state order by Complaint\_Count desc;

The above query will give the list of states with number of complaints in those states in descending ordeR

**Output**

26866 CA

18431 FL

15817 TX

13072 NY

8474 GA

7670 NJ

6953 PA

6912 IL

6109 VA

5978 OH

5723 MD

5370 NC

4530 MI

4197 AZ

3990 WA

3583 MA

3211 CO

3146 TN

2652 SC

2593 NV

2487 MO

2309 OR

2204 AL

2173 IN

2156 MN

2151 WI

2095 LA

2073 CT

1410 KY

1382 OK

1152 DC

1047 UT

1017 MS

976 DE

954 KS

898 NM

828 AR

782 NH

778 IA

684 ME

676 ID

659 HI

624 RI

615 NE

510 WV

492 PR

317 VT

307 MT

283 SD

216 AK

187 ND

186 WY

91 AE

69 AP

49 VI

23 GU

12 FM

10 MH

6 AS

4 PW

4 MP

1 AA

**Query 2:**

INSERT OVERWRITE LOCAL DIRECTORY '/Project/Companies\_No\_Public\_Response' select count(complaint\_id) as Complaint\_Count , company from consumer\_complaints where company\_public\_response = "Company chooses not to provide a public response" group by company order by Complaint\_Count desc limit 10;

The above query will give the top ten companies which chose not to provide any public response.

**Output**

6288 Experian

5568 TransUnion

4556 Wells Fargo

4333 Bank of America

3042 Citibank

1762 U.S. Bancorp

681 SunTrust Bank

540 Barclays

455 BB&T Financial

399 Ally Financial Inc.

**Query 3:**

select company , count(Complaint\_ID) as counter from Consumer\_Complaints where Company\_response\_to\_consumer = "Closed with monetary relief" group by company order by counter desc limit 5;

The above query will give the top five companies which had closed the complaints with monetary relief

**Output**

Bank of America 1520

Citibank 1326

JPMorgan Chase 798

Wells Fargo 534

Capital One 324

**Query 4:**

select count(complaint\_id) as counter , substr(date\_received,1,2) from consumer\_complaints group by substr(date\_received,1,2) order by counter desc limit 12;

The above query will give the number of complaints by each month

**Output**

26746 09

22779 08

22310 10

15893 07

14573 06

14565 03

13768 04

13684 05

12707 02

12628 01

11707 12

11284 11

**Query 5:**

select product, count(complaint\_id) as complaints from consumer\_complaints group by product order by complaints desc;

The above query will give the products with most number of complaints.

**Output**

Mortgage 49879

Debt collection 46946

Credit reporting 38609

Credit card 18955

Bank account or service 18687

Consumer loan 8989

Student loan 5390

Payday loan 2037

Money transfers 1712

Prepaid card 1107

Other financial service 333

**Query 6:**

select state, count(company) as good\_companies from consumer\_complaints where company\_response\_to\_consumer='Closed with monetary relief' group by state order by good\_companies desc limit 15;

The above query will give the statewise Number of Companies that are best in providing response to complaints. These companies closed the complaints with Monetary Relief

**Output**

CA 1513

FL 934

NY 916

TX 721

PA 499

NJ 477

GA 448

IL 388

OH 346

MD 336

VA 328

MA 325

NC 311

MI 251

AZ 247

**Query 7:**

select issue, count(issue) as issues\_of\_thistype from consumer\_complaints group by issue order by issues\_of\_thistype desc limit 15;

The above query will give the common issues that are associated with large number of complaints.

**Output**

Incorrect information on credit report 28865

Loan modification,collection,foreclosure 22924

Cont'd attempts collect debt not owed 20300

Loan servicing, payments, escrow account 18964

Disclosure verification of debt 8290

Communication tactics 8196

Account opening, closing, or management 7986

Deposits and withdrawals 4957

Credit reporting company's investigation 4296

Managing the loan or lease 4181

False statements or representation 3903

Application, originator, mortgage broker 3649

Other 3291

Improper contact or sharing of info 3280

Dealing with my lender or servicer 3195

**Query 8:**

select company\_public\_response, count(company\_public\_response) as types\_of\_publicresponse from consumer\_complaints group by company\_public\_response order by types\_of\_publicresponse desc limit 9;

The above query will give the number of companies that chooses not to provide a public response, If they choose to provide a public response, what are the most common public response provided by them.

**Output:**

Company chooses not to provide a public response 32404

Company believes it acted appropriately as authorized by contract or law 8448

Company believes complaint caused principally by actions of third party outside the control or direction of the company 950

Company disputes the facts presented in the complaint 874

Company believes the complaint is the result of a misunderstanding 737

Company can't verify or dispute the facts in the complaint 728

Company believes complaint is the result of an isolated error 495

Company believes complaint represents an opportunity for improvement to better serve consumers 285

**Query 9:**

select Company, count(Complaint\_id) as Complaint\_Count from Consumer\_Complaints group by Company order by Complaint\_Count desc limit 5;

The above query will give the distribution of complaints among top five companies.

**Output**

Equifax 13438

Experian 12417

Bank of America 11737

TransUnion 11360

Wells Fargo 10583

**Query 10 :**

select count(complaint\_id) as counter , company from consumer\_complaints where product='Debt collection' group by company order by counter desc limit 5;

The above query will give the top five companies with maximum complaints for the product Debt Collection

**Output**

2473 Encore Capital Group

1737 Enhanced Recovery Company, LLC

1520 Portfolio Recovery Associates, Inc.

865 Citibank

837 Transworld Systems Inc.

46946 total

40000 other

**Query 11:**

select Company, count(Complaint\_ID) as count from Consumer\_Complaints where Product = 'Mortgage' group by Company order by count desc Limit 5;

The above query will give the top five companies with maximum complaints for the product Mortgage

**Output**

Ocwen 6006

Bank of America 5868

Wells Fargo 5726

Nationstar Mortgage 4765

JPMorgan Chase 3797

other 23000

**Query 12:**

select Company, count(Complaint\_ID) as count from Consumer\_Complaints where Product = 'Credit reporting' group by Company order by count desc Limit 5;

The above query will give the top five companies with maximum complaints for the product Credit Reporting

**Output**

Equifax 13393

Experian 12372

TransUnion 11316

FNIS (Fidelity National Information Services, Inc.) 277

LexisNexis 97

total 38609

**Query 13:**

select Company, count(Complaint\_ID) as count from Consumer\_Complaints where Company\_response\_to\_consumer = 'Untimely response' group by Company order by count desc Limit 5;

**Output**

Mobiloans, LLC 71

Concord Resolution Inc (Closed) 33

LHR Inc. 29

Global Recovery Group 26

Pinnacle Asset Management, LLC 25

**Query 14:**

select Company, count(Complaint\_ID) as count from Consumer\_Complaints where Company\_response\_to\_consumer <> 'In progress' and Company\_response\_to\_consumer <> 'Untimely response' and Timely\_response = 'Yes' group by Company order by count desc Limit 5;

The above query will give the top five companies which has closed customer complaints in both timely manner and with either explanation or relief.

**Output**

Equifax 16262

Experian 12033

Bank of America 11392

TransUnion 8233

Wells Fargo 5327

**Query 15:**

select Submitted\_via, count(Complaint\_ID) as count from Consumer\_Complaints group by Submitted\_via order by count desc Limit 6;

The above query will give the distribution of complaints among the submission type.

**Output**

Web 141844

Referral 25162

Phone 11660

Postal mail 11357

Fax 2607

Email 14

**Query 16:**

Difference in the number of complaints in 2014 and 2015.

select product , count(complaint\_id) as counter, substr(date\_received,7,4) from consumer\_complaints group by product,substr(date\_received,7,4) order by product limit 50;

The above query will give the Comparison in the number of complaints in 2014 and 2015.

**Output**

Bank account or service 13398 2015

Bank account or service 5289 2014

Consumer loan 2574 2014

Consumer loan 6415 2015

Credit card 5100 2014

Credit card 13855 2015

Credit reporting 10526 2014

Credit reporting 28083 2015

Debt collection 13653 2014

Debt collection 33293 2015

Money transfers 1293 2015

Money transfers 419 2014

Mortgage 15113 2014

Mortgage 34766 2015

Other financial service 93 2014

Other financial service 240 2015

Payday loan 721 2014

Payday loan 1316 2015

Prepaid card 280 2014

Prepaid card 827 2015

Product 1 ecei

Student loan 1594 2014

Student loan 3796 2015

**Query 17:**

select issue, count(Complaint\_ID) as count from Consumer\_Complaints group by issue order by count desc Limit 5;

The above query will give the most common issues regarding which most complaints were registered.

**Output**

Incorrect information on credit report 28865

Loan modification,collection,foreclosure 22923

Cont'd attempts collect debt not owed 20300

Loan servicing, payments, escrow account 18964

Disclosure verification of debt 8290

**Query 18:**

select issue, count(Complaint\_ID) as count from Consumer\_Complaints where Product = 'Mortgage' group by issue order by count desc Limit 5;

The above query will give the number of complaints for each issue within product Mortgage

**Output**

Loan modification,collection,foreclosure 22923

Loan servicing, payments, escrow account 18964

Application, originator, mortgage broker 3649

Settlement process and costs 1965

Credit decision / Underwriting 1338

**Query 19:**

select issue, count(Complaint\_ID) as count from Consumer\_Complaints where Product = 'Debt collection' group by issue order by count desc Limit 5;

The above query will give the number of complaints for each issue within product Debt Collection

**Output**

Cont'd attempts collect debt not owed 20300

Disclosure verification of debt 8290

Communication tactics 8196

False statements or representation 3903

Improper contact or sharing of info 3280

**Query 20:**

select issue, count(Complaint\_ID) as count from Consumer\_Complaints where Product = 'Credit reporting' group by issue order by count desc Limit 5;

The above query will give the number of complaints for each issue within product Credit Reporting.

**Output**

Incorrect information on credit report 28865

Credit reporting company's investigation 4296

Unable to get credit report/credit score 3084

Improper use of my credit report 1259

Credit monitoring or identity protection 1105

**Query 21: To get output 000000\_0 for naive bayes**

INSERT OVERWRITE LOCAL DIRECTORY '/Project/OutPredictFinal' select product , issue , company , company\_public\_response from consumer\_complaints where product <> "" and issue <> "" and company <> "" and company\_public\_response<>"";