BUAN 6346

Project Proposal- Complaint analysis

APRIL 5, 2019

MANISHA GUPTA

Mxg176230

Contents

Complaint analysis

Introduction
The Consumer Complaint Database is a collection of complaints on a range of consumer financial products and services, sent to companies for response. It was designed to connect consumers with financial companies to understand issues with their mortgages, fix errors on their credit reports, stop harassment from debt collectors, and get direct responses about problems with their credit cards, checking and savings accounts, student loans, and more.

The objective of this database is, by analyzing the data, to identify trends and problems in the marketplace and to help to do a better job in supervising companies, enforcing federal consumer financial laws and writing rules and regulations.

•   Each complaint is forwarded to the appropriate company for a response.

•   The complaint data is shared with state and federal agencies.

•   The complaint data is then analyzed to help with our work to supervise companies, enforce federal consumer financial laws, and write better rules and regulations.

Financial products and services sector have significant impact on our lives as they are directly related to our finances. It becomes need of hour to have proper understanding about how we can make correct decisions to buy a product or a service. Here in this project we analyze Customer complaints about various financial products to answer these questions.

Project Description
The objective of this project Complaint analysis is to track, categorize and handle customer complaints. When a customer registers a complaint, he or she is voicing a concern in relation to a product or service. However, not all complaints are to be treated equally and there are several questions to ask before taking action such as

- Has this happened before?
- Have the complaints been recorded?
- How often does the same compliant arise?
- Is there a pattern to this complaint in how it was received?
- Has the same customer reported this previously?

1. Analyzing the major customer concerns with banks in each state. We aim to gain an insight into the problems coming from retail banking in each state in the US. This can be derived from the available dataset and the goal is to use classifying the customer concerns based on the state, will give the major concerns in each state and we aim to find the top 3 concerns of each customer.
2. Derive business impact of customer concerns to banking institutions Classifying the available data based on whether customers' concerns were addressed or not will provide valuable insight into the effect of them on institutions' business. For example, a large number of unresolved complaints can be taken to mean that the specific customers have taken their business elsewhere.
3. Develop a performance metric based on time taken to resolve concern Depending on the type of concern and the time taken to resolve it, we can come up with a standardized performance metric, taking into account the various factors of the case and apply the metric across all future solutions. This will help pinpoint. those problems that take longer to resolve and therefore can be given a higher priority.
4. Perform the analysis to study the relationship of income with number of complaints.
5. Study the impact of age, gender or household on complaints datasets in detail.

Descriptive analysis

Data sets is available on https://www.consumerfinance.gov/data-research/consumer-complaints/ and its field description is https://www.consumerfinance.gov/complaint/data-use/.

Old Field Name	FIELD NAME	DESCRIPTION
Date received	datereceived	The date the CFPB received the complaint. For example, "05/25/2013."
Product	product	The type of product the consumer identified in the complaint. For example, "Checking or savings account" or "Student loan."
Sub-product	subproduct	The type of sub-product the consumer identified in the complaint. For example, "Checking account" or "Private student loan."
Issue	issue	The issue the consumer identified in the complaint. For example, "Managing an account" or "Struggling to repay your loan."
Sub-issue	subissue	The sub-issue the consumer identified in the complaint. For example, "Deposits and withdrawals" or "Problem lowering your monthly payments."
Consumer complaint narrative	narrative	Consumer complaint narrative is the consumer-submitted description of "what happened" from the complaint. Consumers must opt-in to share their narrative. We will not publish the narrative unless the consumer consents, and consumers can opt-out at any time. The CFPB takes reasonable steps to scrub personal information from each complaint that could be used to identify the consumer.
Company public response	publicresponse	The company's optional, public-facing response to a consumer's complaint. Companies can choose to select a response from a pre-set list of options that will be posted on the public database. For example, "Company believes complaint is the result of an isolated error."
Company	company	The complaint is about this company. For example, "ABC Bank."
State	state	The state of the mailing address provided by the consumer.
ZIP code	zipcode	The mailing ZIP code provided by the consumer. This field may: i) include the first five digits of a ZIP code; ii) include the first three digits of a ZIP code (if the consumer consented to publication of their complaint narrative); or iii) be blank (if ZIP codes have been submitted with non-numeric values, if there are less than 20,000 people in a given ZIP code, or if the complaint has an address outside of the United States).
Tags	tags	Data that supports easier searching and sorting of complaints submitted by or on behalf of consumers.
		For example, complaints where the submitter reports the age of the consumer as 62 years or older are tagged "Older American." Complaints submitted by or on behalf of a servicemember or the spouse or dependent of a servicemember are tagged "Servicemember." Servicemember includes anyone who is active duty, National Guard, or Reservist, as well as anyone who previously served and is a veteran or retiree.
Consumer consent provided?	consent	Identifies whether the consumer opted in to publish their complaint narrative. We do not publish the narrative unless the consumer consents, and consumers can opt-out at any time.
Submitted via	submitted	How the complaint was submitted to the CFPB. For example, "Web" or "Phone."
Date sent to company	datesent	The date the CFPB sent the complaint to the company.
Company response to consumer	responsetoconsumer	This is how the company responded. For example, "Closed with explanation."
Timely response?	timlyresponse	Whether the company gave a timely response. For example, "Yes" or "No."
Consumer disputed?	disputed	Whether the consumer disputed the company's response.
Complaint ID	complaintid	The unique identification number for a complaint.

All the fields except complaints ID, consumer complaint narrative are important as we can analyze this dataset by all such as Year, Product, Company, state, Complaints over time-

Most important fields are state, zipcode, product,issue, timlyresponse and how the complaints are submitting to the database to get the more valuable insights. we will be performing analysis by following-

- Total number of complaints
- Types of complaints
- Total complaints by year
- Total complaints by area/zip code
- Complaints by issue and year
- Percentage of issues by zip code
- Percentage of resolution (data field)
- Total percentage of resolution (data field- Customer Disputed) and by area/zip code to know where do most unresolved complaints occur?
- complaints Volume by company and product

Additional Datasets used

1. **IRS data by zip code**: A detailed description of the columns contained within the dataset can be found at: https://www.irs.gov/statistics/soi-tax-stats-individual-income-tax-statistics-2016-zip-code-data-soi
2. **2010-census-populations-by-zip-code-** Data can be found on https://catalog.data.gov/dataset/2010-census-populations-by-zip-code and its field names are self-explanatory.

Techniques used

- Hive aggregation queries
- HDFS
- Pyspark
- Graphs using aggregated data(optional)
- Segmentation analysis using ML(optional)

Performed Work

Python code to change the delimiter of the file

```python
import os as os

import pandas as pd

import datetime as dt

import numpy as np

import sys

import glob

import csv

import re

if __name__ == '__main__':

    filename = r"C:\Users\abhin\Documents\Manisha\utd\bigdata\complaints.csv"

    dftest = pd.read_csv(filename, header = 0,  quoting=csv.QUOTE_ALL )

    dfcolumns = dftest.columns

    cnt = 0

    for tp in dftest.dtypes:

        colname = dfcolumns[cnt]

        if 'object' in str(tp):

            if "date" not in colname.lower():

                dftest[colname] = dftest[colname].astype(str).replace('~', ' ').replace('""', '').map(lambda x: re.sub(r'\W+', ' ', x))

                dftest[colname] = dftest[colname].apply(lambda x: x if x!='nan' else '')

        cnt = cnt + 1

    dftest.to_csv(path_or_buf = "complaintsdata.csv", sep='~', index = False)
```

1. **Loading showing data**
   create database cc;
   use cc;

   ########## HIVE Script ####################3
   CREATE TABLE complaints(datereceived STRING,
   product STRING,
   subproduct STRING,
   issue STRING,
   subissue STRING,
   narrative STRING,
   publicresponse STRING,
   company STRING,
   state STRING,
   zipcode STRING,
   tags STRING,
   consent STRING,
   submitted STRING,
   datesent STRING,
   responsetoconsumer STRING,
   timlyresponse STRING,
   disputed STRING,
   complaintid STRING)
   ROW FORMAT DELIMITED
   FIELDS TERMINATED BY '~'
   ;

   LOAD DATA Local INPATH '/home/training/project/complaintsdata.csv' OVERWRITE INTO TABLE cc.complaints;

Total records

select count(*) from complaints;

2. **Find top 10 products with complaints**

select product, count(issue) as isu from complaints group by product order by isu desc limit 10;

```
Total MapReduce CPU Time Spent: 5 seconds 250 msec
OK
Credit reporting credit repair services or other personal consumer reports    18310
Debt collection 7063
Credit card or prepaid card    3706
Mortgage       3432
Checking or savings account    3126
Student loan    1510
Vehicle loan or lease  816
Money transfer virtual currency or money service      665
Payday loan title loan or personal loan 622
Product 1
Time taken: 38.915 seconds, Fetched: 10 row(s)
hive>
```

Percentage contribution

select t.product, isu/t1.tot from (select product, count(issue) as isu from complaints group by product order by isu desc limit 10) as t, (select count(*) as tot from complaints)t1;

```
Total MapReduce CPU Time Spent: 19 seconds 70 msec
OK
Credit reporting credit repair services or other personal consumer reports    4
6.65
Debt collection 17.99
Credit card or prepaid card    9.44
Mortgage       8.74
Checking or savings account    7.96
Student loan    3.85
Vehicle loan or lease  2.08
Money transfer virtual currency or money service      1.69
Payday loan title loan or personal loan 1.58
Product 0.0
Time taken: 126.137 seconds, Fetched: 10 row(s)
```

3. How many sub product are there and their % contribution

select t.subproduct, isu, round((isu/t1.tot)*100,2) from (select subproduct, count(issue) as isu from complaints where subproduct <> 'Sub-product'group by subproduct order by isu desc limit 10) as t, (select count(*) as tot from complaints)t1;

```
Credit reporting          18012   45.89
General purpose credit card or charge card      2879    7.33
Checking account          2332    5.94
Other debt      2039    5.19
Conventional home mortgage      2005    5.11
Credit card debt          1520    3.87
I do not know   1509    3.84
Medical debt    1141    2.91
Federal student loan servicing  1025    2.61
Loan    658     1.68
Time taken: 126.601 seconds, Fetched: 10 row(s)
```

4. How many issues are there and their % contribution?

select t.issue, isu, round((isu/t1.tot)*100,2) from (select issue, count(issue) as isu from complaints group by issue order by isu desc) as t, (select count(*) as tot from complaints)t1;

```
Incorrect information on your report    11719   29.86
Problem with a credit reporting company s investigation into an existing problem    4247    10.82
Attempts to collect debt not owed       3352    8.54
Managing an account     1914    4.88
Improper use of your report     1828    4.66
Trouble during payment process  1780    4.53
Written notification about debt 1419    3.62
Struggling to pay mortgage      1122    2.86
Dealing with your lender or servicer    981     2.5
Problem with a purchase shown on your statement 942     2.4
Communication tactics   776     1.98
Took or threatened to take negative or legal action     692     1.76
False statements or representation      643     1.64
Other features terms or problems        511     1.3
Fees or interest        446     1.14
Problem when making payments    408     1.04
Struggling to repay your loan   408     1.04
Unable to get your credit report or credit score        376     0.96
Closing an account      366     0.93
Getting a credit card   326     0.83
Problem with fraud alerts or security freezes   323     0.82
Opening an account      308     0.78
Problem with a lender or other company charging your account     297     0.76
Managing the loan or lease      290     0.74
Struggling to pay your loan     287     0.73
Fraud or scam   278     0.71
Advertising and marketing including promotional offers  259     0.66
```

5. How many sub issues are there and their % contribution?

select t.subissue, isu, round((isu/t1.tot)*100,2) from (select subissue, count(issue) as isu from complaints where subissue <> 'Sub-issue' group by subissue order by isu desc) as t, (select count(*) as tot from complaints)t1;

```
Information belongs to someone else      6539    16.66
None    5126    13.06
Their investigation did not fix an error on your report 3126    7.96
Account information incorrect    1845    4.7
Account status incorrect         1750    4.46
Debt is not yours        1701    4.33
Credit inquiries on your report that you don t recognize      1415    3.61
Didn t receive enough information to verify debt         945     2.41
Debt was result of identity theft        869     2.21
Deposits and withdrawals         672     1.71
Debt was paid    655     1.67
Credit card company isn t resolving a dispute about a purchase on your statement       655     1.67
Attempted to collect wrong amount        530     1.35
Personal information incorrect 512       1.3
Frequent or repeated calls       432     1.1
Didn t receive notice of right to dispute        431     1.1
Public record information inaccurate     405     1.03
Old information reappears or never goes away     372     0.95
Received bad information about your loan          367     0.94
Reporting company used your report improperly    365     0.93
Company closed your account      342     0.87
Was not notified of investigation status or results      337     0.86
Trouble with how payments are being handled      319     0.81
Threatened or suggested your credit would be damaged     308     0.78
```

6. How many public responses are there and their % contribution?

select t.publicresponse, isu, round((isu/t1.tot)*100,2) from (select publicresponse , count(issue) as isu from complaints where publicresponse <> 'Company public response' group by publicresponse order by isu desc) as t, (select count(*) as tot from complaints)t1;

```
OK
None    20534   52.31
Company has responded to the consumer and the CFPB and chooses not to provide a public response 16077    4
0.96
Company believes it acted appropriately as authorized by contract or law         2026    5.16
Company believes the complaint is the result of a misunderstanding       149     0.38
Company disputes the facts presented in the complaint    137     0.35
Company believes complaint represents an opportunity for improvement to better serve consumers   103     0
.26
Company believes complaint is the result of an isolated error    99      0.25
Company believes complaint caused principally by actions of third party outside the control or direction
of the company  80      0.2
Company can t verify or dispute the facts in the complaint       45      0.11
Time taken: 127.118 seconds, Fetched: 9 row(s)
```

7. How many companies are there and their % contribution?

select t.company, isu, round((isu/t1.tot)*100,2) from (select company, count(issue) as isu from complaints group by company order by isu desc) as t, (select count(*) as tot from complaints)t1;

```
EQUIFAX INC     5099    12.99
Experian Information Solutions Inc      4965    12.65
TRANSUNION INTERMEDIATE HOLDINGS INC    4732    12.06
JPMORGAN CHASE CO        1356    3.45
WELLS FARGO COMPANY      1175    2.99
CAPITAL ONE FINANCIAL CORPORATION       1159    2.95
CITIBANK N A    1114    2.84
Navient Solutions LLC    903     2.3
BANK OF AMERICA NATIONAL ASSOCIATION    789     2.01
SYNCHRONY FINANCIAL      649     1.65
AMERICAN EXPRESS COMPANY         470     1.2
AES PHEAA       416     1.06
Alliance Data Card Services      394     1.0
PORTFOLIO RECOVERY ASSOCIATES INC       382     0.97
U S BANCORP     313     0.8
NATIONSTAR MORTGAGE      302     0.77
DISCOVER BANK   292     0.74
TD BANK US HOLDING COMPANY       263     0.67
PAYPAL HOLDINGS INC      247     0.63
ENCORE CAPITAL GROUP INC         225     0.57
PNC Bank N A    218     0.56
ERC     211     0.54
Ditech Financial LLC     205     0.52
```

8. How many states are there and their % contribution?

select t.state, isu, round((isu/t1.tot)*100,2) from (select state, count(issue) as isu from complaints group by state order by isu desc) as t, (select count(*) as tot from complaints)t1;

```
Total MapReduce CPU Time Spent: 18 seconds 570 msec
OK
CA     4992    12.72
FL     4810    12.25
TX     3794    9.67
NY     2573    6.56
GA     2268    5.78
IL     1346    3.43
PA     1287    3.28
NC     1285    3.27
NJ     1216    3.1
OH     1173    2.99
None   1092    2.78
VA     1020    2.6
MD     870     2.22
MI     829     2.11
AZ     698     1.78
MO     689     1.76
TN     681     1.73
SC     660     1.68
MA     566     1.44
WA     558     1.42
NV     533     1.36
LA     532     1.36
AL     530     1.35
CO     521     1.33
UT     454     1.16
```

9. How many tags are there and their % contribution?

select t.tags, isu, round((isu/t1.tot)*100,2) from (select tags, count(issue) as isu from complaints group by tags order by isu desc) as t, (select count(*) as tot from complaints)t1;

```
None     33691    85.83
Servicemember    3565    9.08
Older American   1500    3.82
Older American Servicemember     494     1.26
Tags    1        0.0
Time taken: 126.644 seconds, Fetched: 5 row(s)
hive> █
```

10. Consumer consent column

select t.consent, isu, round((isu/t1.tot)*100,2) from (select consent , count(issue) as isu from complaints where consent <> 'Consumer consent provided?' group by consent order by isu desc) as t, (select count(*) as tot from complaints)t1;

```
Total MapReduce CPU Time Spent: 19 seconds 570 msec
OK
None     12153    30.96
Consent provided         10422    26.55
Consent not provided     9462     24.11
         5104    13.0
Other    2071    5.28
Consent withdrawn        38       0.1
Time taken: 128.396 seconds, Fetched: 6 row(s)
```

11. Complaint submission channel and their % contribution

select t.submitted, isu, round((isu/t1.tot)*100,2) from (select submitted, count(issue) as isu from complaints where submitted <> 'Submitted via' group by submitted order by isu desc) as t, (select count(*) as tot from complaints)t1;

```
Web      34146    86.99
Referral         2095    5.34
Phone    2031    5.17
Postal mail      803     2.05
Fax      174     0.44
Email    1       0.0
Time taken: 128.162 seconds, Fetched: 6 row(s)
hive> █
```

12. Checking timely response and their % contribution- How many complaints have received timely response

select t.timlyresponse , isu, round((isu/t1.tot)*100,2) from (select timlyresponse , count(issue) as isu from complaints where timlyresponse <> 'Timely response?' group by timlyresponse order by isu desc) as t, (select count(*) as tot from complaints)t1;

```
Total MapReduce CPU Time Spent: 19 seconds 330 msec
OK
Yes     38620   98.39
No       630    1.61
Time taken: 127.156 seconds, Fetched: 2 row(s)
```

13. Consumer disputed – how many consumers disputed and their % contribution

select t.disputed , isu, round((isu/t1.tot)*100,2) from (select disputed , count(issue) as isu from complaints where disputed <> 'Consumer disputed?' group by disputed order by isu desc) as t, (select count(*) as tot from complaints)t1;

```
OK
        39250   100.0
Time taken: 124.052 seconds, Fetched: 1 row(s)
```

No consumer disputed.

14. Company response to consumer – what are the different response user get from company

select t.responsetoconsumer , isu, round((isu/t1.tot)*100,2) from (select responsetoconsumer , count(issue) as isu from complaints where responsetoconsumer <> 'Company response to consumer' group by responsetoconsumer order by isu desc) as t, (select count(*) as tot from complaints)t1;

```
Closed with explanation 29579   75.36
Closed with non monetary relief 4864    12.39
In progress     2836    7.23
Closed with monetary relief    1665    4.24
Untimely response        306    0.78
Time taken: 126.93 seconds, Fetched: 5 row(s)
hive> █
```

15. Time taken to sent the complaint to company and their percentile

select percentile(diff,array(0.01,0.05,0.25,0.5,0.75,0.95,0.99)) from (select datediff(to_date(from_unixtime(UNIX_TIMESTAMP(datesent,'MM/dd/yy'))),to_date(from_unixtime(UNIX_TIMESTAMP(datereceived,'MM/dd/yy')))) as diff from complaints) as t;

```
[0.0,0.0,0.0,0.0,0.0,8.0,16.0]
Time taken: 32.918 seconds, Fetched: 1 row(s)
```

No of days taken when response to consumer is closed. (in progress or untimly response are not taken into consideration)

select percentile(diff,array(0.01,0.05,0.25,0.5,0.75,0.95,0.99)) from (select datediff(to_date(from_unixtime(UNIX_TIMESTAMP(datesent,'MM/dd/yy'))),to_date(from_unixtime(UNIX_TIMESTAMP(datereceived,'MM/dd/yy')))) as diff from complaints where responsetoconsumer not in ('In progress','Untimely response')) as t;

```
[0.0,0.0,0.0,0.0,0.0,8.0,16.0]
Time taken: 37.702 seconds, Fetched: 1 row(s)
```

Min, Max, Stdev

select min(diff), max(diff), stddev_pop(diff) from (select datediff(to_date(from_unixtime(UNIX_TIMESTAMP(datesent,'MM/dd/yy'))),to_date(from_unixtime(UNIX_TIMESTAMP(datereceived,'MM/dd/yy')))) as diff from complaints where responsetoconsumer not in ('In progress','Untimely response')) as t;

```
OK
0        67        3.6876431514393095
Time taken: 36.42 seconds, Fetched: 1 row(s)
```

16. complaint by zip code to company and their percentile

select t.zipcode, isu, round((isu/t1.tot)*100,2) from (select zipcode , count(issue) as isu from complaints group by zipcode order by isu desc) as t, (select count(*) as tot from complaints)t1;

```
None     5459      13.91
330XX    590       1.5
331XX    466       1.19
300XX    348       0.89
334XX    313       0.8
770XX    293       0.75
333XX    282       0.72
606XX    235       0.6
302XX    233       0.59
945XX    216       0.55
303XX    215       0.55
```

Cross Frequency

1. To check timely response by product

select product, timlyresponse, count(*) as cn from complaints group by product,timlyresponse limit 10;

```
Checking or savings account      No     22
Checking or savings account      Yes    3104
Credit card or prepaid card      No     19
Credit card or prepaid card      Yes    3687
Credit reporting credit repair services or other personal consumer reports     No    179
Credit reporting credit repair services or other personal consumer reports     Yes   18131
Debt collection No     259
Debt collection Yes    6804
Money transfer virtual currency or money service      No     12
Money transfer virtual currency or money service      Yes    653
```

2. To check Product and their subproduct

select product, subproduct, count(*) as c from complaints group by product, subproduct order by c desc;

```
Credit reporting credit repair services or other personal consumer reports     Credit reporting     18012
Credit card or prepaid card      General purpose credit card or charge card     2879
Checking or savings account      Checking account      2332
Debt collection Other debt       2039
Mortgage        Conventional home mortgage      2005
Debt collection Credit card debt       1520
Debt collection I do not know    1509
Debt collection Medical debt     1141
Student loan     Federal student loan servicing  1025
Vehicle loan or lease    Loan      658
Credit card or prepaid card      Store credit card        608
Mortgage        FHA mortgage      556
Student loan     Private student loan     485
Checking or savings account      Other banking product or service        442
Mortgage        Other type of mortgage  402
Payday loan title loan or personal loan Installment loan        274
Debt collection Auto debt       263
Checking or savings account      Savings account 249
Mortgage        VA mortgage      221
Credit reporting credit repair services or other personal consumer reports     Other personal consumer report  220
```

3. Products against complaints submission method

select product, submitted, count(*) as c from complaints group by product, submitted order by c desc;

```
Credit reporting credit repair services or other personal consumer reports      Web      17238
Debt collection Web      6529
Credit card or prepaid card      Web      2903
Mortgage      Web      2518
Checking or savings account      Web      1827
Student loan      Web      1384
Checking or savings account      Referral      871
Vehicle loan or lease      Web      654
Money transfer virtual currency or money service      Web      572
Credit reporting credit repair services or other personal consumer reports      Phone      521
Payday loan title loan or personal loan Web      521
Mortgage      Referral      495
Checking or savings account      Phone      363
Credit card or prepaid card      Phone      349
Credit card or prepaid card      Referral      338
Credit reporting credit repair services or other personal consumer reports      Postal mail      322
Mortgage      Phone      276
Debt collection Phone      259
Debt collection Postal mail      143
Credit reporting credit repair services or other personal consumer reports      Referral      139
Mortgage      Postal mail      116
Debt collection Referral      108
Credit card or prepaid card      Postal mail      105
Credit reporting credit repair services or other personal consumer reports      Fax      90
Vehicle loan or lease      Phone      78
Student loan      Phone      74
```

4. Products against response received by user

select product, responsetoconsumer, count(*) as c from complaints group by product, responsetoconsumer order by c desc;

```
Credit reporting credit repair services or other personal consumer reports      Closed with explanation 13040
Debt collection Closed with explanation 5807
Credit reporting credit repair services or other personal consumer reports      Closed with non monetary relief 3243
Mortgage      Closed with explanation 3041
Credit card or prepaid card      Closed with explanation 2453
Checking or savings account      Closed with explanation 2207
Credit reporting credit repair services or other personal consumer reports      In progress      1883
Student loan      Closed with explanation 1314
Debt collection Closed with non monetary relief 765
Vehicle loan or lease      Closed with explanation 672
Checking or savings account      Closed with monetary relief      664
Credit card or prepaid card      Closed with monetary relief      648
Payday loan title loan or personal loan Closed with explanation 528
Money transfer virtual currency or money service      Closed with explanation 517
Credit card or prepaid card      Closed with non monetary relief 414
Debt collection In progress      305
Mortgage      In progress      188
Credit card or prepaid card      In progress      186
Debt collection Untimely response      138
Checking or savings account      In progress      136
Checking or savings account      Closed with non monetary relief 117
Mortgage      Closed with non monetary relief 101
```

Top 10 most occurring values

Top 10 Product with complaints

select t.product, isu, round((isu/t1.tot)*100,2) from (select product, count(issue) as isu from complaints group by product order by isu desc limit 10) as t, (select count(*) as tot from complaints)t1;

```
Credit reporting credit repair services or other personal consumer reports      18310      46.65
Debt collection 7063      17.99
Credit card or prepaid card      3706      9.44
Mortgage      3432      8.74
Checking or savings account      3126      7.96
Student loan      1510      3.85
Vehicle loan or lease      816      2.08
Money transfer virtual currency or money service      665      1.69
Payday loan title loan or personal loan 622      1.58
Product 1      0.0
Time taken: 125.861 seconds, Fetched: 10 row(s)
```

Top 10 sub Product

select t.subproduct, isu, round((isu/t1.tot)*100,2) from (select subproduct, count(issue) as isu from complaints where subproduct <> 'Sub-product'group by subproduct order by isu desc limit 10) as t, (select count(*) as tot from complaints)t1;

Top 10 sub Product

select t.subproduct, isu, round((isu/t1.tot)*100,2) from (select subproduct, count(issue) as isu from complaints where subproduct <> 'Sub-product'group by subproduct order by isu desc limit 10) as t, (select count(*) as tot from complaints)t1;

```
Credit reporting         18012    45.89
General purpose credit card or charge card        2879      7.33
Checking account          2332     5.94
Other debt      2039     5.19
Conventional home mortgage        2005      5.11
Credit card debt          1520     3.87
I do not know   1509     3.84
Medical debt    1141     2.91
Federal student loan servicing    1025      2.61
Loan      658      1.68
Time taken: 126.727 seconds, Fetched: 10 row(s)
```

Top 10 issue

select t.issue, isu, round((isu/t1.tot)*100,2) from (select issue, count(issue) as isu from complaints group by issue order by isu desc limit 10) as t, (select count(*) as tot from complaints)t1;

```
Incorrect information on your report    11719   29.86
Problem with a credit reporting company s investigation into an existing problem       4247      10.82
Attempts to collect debt not owed       3352     8.54
Managing an account      1914     4.88
Improper use of your report      1828     4.66
Trouble during payment process   1780     4.53
Written notification about debt 1419     3.62
Struggling to pay mortgage       1122     2.86
Dealing with your lender or servicer     981      2.5
Problem with a purchase shown on your statement 942       2.4
Time taken: 124.311 seconds, Fetched: 10 row(s)
```

Top 10 sub issue

select t.subissue, isu, round((isu/t1.tot)*100,2) from (select subissue, count(issue) as isu from complaints where subissue <> 'Sub-issue' group by subissue order by isu desc limit 10) as t, (select count(*) as tot from complaints)t1;

```
Information belongs to someone else      6539     16.66
None     5126     13.06
Their investigation did not fix an error on your report 3126      7.96
Account information incorrect     1845     4.7
Account status incorrect         1750     4.46
Debt is not yours        1701     4.33
Credit inquiries on your report that you don t recognize         1415     3.61
Didn t receive enough information to verify debt         945      2.41
Debt was result of identity theft        869      2.21
Deposits and withdrawals         672      1.71
Time taken: 128.3 seconds, Fetched: 10 row(s)
```

Top 10 Response to public

select t.publicresponse, isu, round((isu/t1.tot)*100,2) from (select publicresponse , count(issue) as isu from complaints where publicresponse <> 'Company public response' group by publicresponse order by isu desc limit 10) as t, (select count(*) as tot from complaints)t1;

```
Total MapReduce CPU Time Spent: 19 seconds 470 msec
OK
None     20534    52.31
Company has responded to the consumer and the CFPB and chooses not to provide a public response 16077    40.96
Company believes it acted appropriately as authorized by contract or law          2026     5.16
Company believes the complaint is the result of a misunderstanding        149      0.38
Company disputes the facts presented in the complaint    137      0.35
Company believes complaint represents an opportunity for improvement to better serve consumers  103       0.26
Company believes complaint is the result of an isolated error    99       0.25
Company believes complaint caused principally by actions of third party outside the control or direction of the compa
ny       80       0.2
Company can t verify or dispute the facts in the complaint        45       0.11
Time taken: 127.153 seconds, Fetched: 9 row(s)
```

Top 10 company with complaints

select t.company, isu, round((isu/t1.tot)*100,2) from (select company, count(issue) as isu from complaints group by company order by isu desc limit 10) as t, (select count(*) as tot from complaints)t1;

```
EQUIFAX INC      5099     12.99
Experian Information Solutions Inc      4965     12.65
TRANSUNION INTERMEDIATE HOLDINGS INC    4732     12.06
JPMORGAN CHASE CO        1356     3.45
WELLS FARGO COMPANY      1175     2.99
CAPITAL ONE FINANCIAL CORPORATION       1159     2.95
CITIBANK N A     1114     2.84
Navient Solutions LLC    903      2.3
BANK OF AMERICA NATIONAL ASSOCIATION    789      2.01
SYNCHRONY FINANCIAL      649      1.65
Time taken: 127.341 seconds, Fetched: 10 row(s)
```

Top 10 states with complaints

select t.state, isu, round((isu/t1.tot)*100,2) from (select state, count(issue) as isu from complaints group by state order by isu desc limit 10) as t, (select count(*) as tot from complaints)t1;

```
CA      4992     12.72
FL      4810     12.25
TX      3794     9.67
NY      2573     6.56
GA      2268     5.78
IL      1346     3.43
PA      1287     3.28
NC      1285     3.27
NJ      1216     3.1
OH      1173     2.99
Time taken: 125.388 seconds, Fetched: 10 row(s)
```

Top 10 Dates in which complaints received

select t.datereceived, isu, round((isu/t1.tot)*100,2) from (select datereceived, count(issue) as isu from complaints group by datereceived order by isu desc limit 10) as t, (select count(*) as tot from complaints)t1;

```
02/12/19         961     2.45
01/30/19         942     2.4
02/06/19         938     2.39
02/20/19         937     2.39
01/31/19         917     2.34
02/05/19         876     2.23
02/19/19         859     2.19
02/13/19         854     2.18
02/26/19         853     2.17
02/22/19         846     2.16
```

Top 10 Dates in which complaints sent to company

select t.datesent, isu, round((isu/t1.tot)*100,2) from (select datesent, count(issue) as isu from complaints group by datesent order by isu desc limit 10) as t, (select count(*) as tot from complaints)t1;

```
01/30/19         1032    2.63
02/19/19         1003    2.56
01/31/19         958     2.44
02/12/19         937     2.39
02/21/19         936     2.38
02/06/19         927     2.36
02/05/19         922     2.35
02/26/19         917     2.34
02/20/19         903     2.3
02/14/19         848     2.16
```

Top states with avg time lag in sending complaints to company

select t.state, round(avg(t.diff),0) as avgt from (select state, datediff(to_date(from_unixtime(UNIX_TIMESTAMP(datesent,'MM/dd/yy'))),to_date(from_unixtime(UNIX_TIMESTAMP(datereceived,'MM/dd/yy')))) as diff from complaints) as t group by t.state order by avgt desc;

```
Total MapReduce CPU Time Spent: 16 seconds 500 msec
OK
VI      7.0
IA      3.0
MT      3.0
None    3.0
GU      3.0
WY      2.0
WV      2.0
WI      2.0
VT      2.0
PR      2.0
PA      2.0
OR      2.0
OK      2.0
NM      2.0
NH      2.0
NE      2.0
ND      2.0
ME      2.0
KY      2.0
DE      2.0
AZ      2.0
AR      2.0
AE      2.0
DC      1.0
NJ      1.0
CT      1.0
CO      1.0
CA      1.0
NC      1.0
```

Virgin island takes 7 days on average in sending complaints to company.

Top Products with avg time lag

select t.product, round(avg(t.diff),0) as avgt from (select product, datediff(to_date(from_unixtime(UNIX_TIMESTAMP(datesent,'MM/dd/yy'))),to_date(from_unixtime(UNIX_TIMESTAMP(datereceived,'MM/dd/yy')))) as diff from complaints) as t group by t.product order by avgt desc;

```
Payday loan title loan or personal loan 3.0
Vehicle loan or lease   2.0
Mortgage        2.0
Money transfer virtual currency or money service       2.0
Debt collection 2.0
Checking or savings account     2.0
Student loan    1.0
Credit reporting credit repair services or other personal consumer reports      1.0
Credit card or prepaid card     1.0
Product NULL
Time taken: 66.496 seconds, Fetched: 10 row(s)
```