# Big Data Analytics

# US Consumer Finance Complaints

## Bachelor of Engineering

## IN

## COMPUTER SCIENCE AND ENGINEERING (AI)

***Submitted By:***

| | | |
|---|---|---|
| **Jiya Lamba** | **01FE21BCI027** | **122** |
| **Shuchith GC** | **01FE21BCI029** | **123** |
| **Mehwish Nidgundi** | **01FE21BCI057** | **148** |
| **Ayesha Mulla** | **01FE22BCI405** | **157** |

## Under the guidance of

Dr. Suvarna Kanakraddi

**SCHOOL OF COMPUTER SCIENCE & ENGINEERING (AI)**

**HUBLI–580 031 (India)**

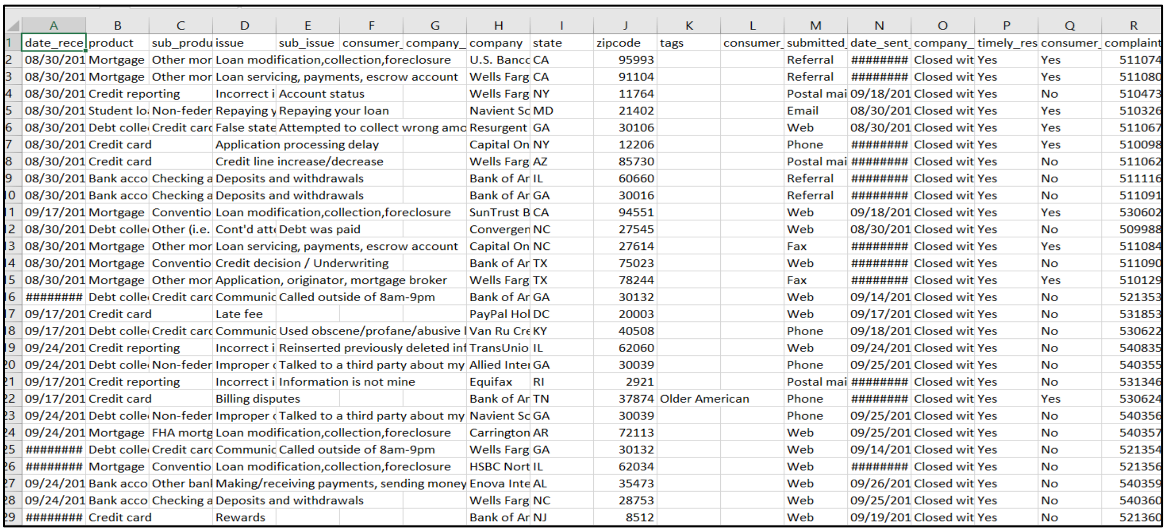
**Academic year 2024-25**

# Table of Content

# Chapter 1
# INTRODUCTION

The US Consumer Finance Complaints dataset, published by the Consumer Financial Protection Bureau, is a comprehensive collection of consumer grievances related to various financial products and services, including loans, credit cards, mortgages, and banking. This dataset offers a unique opportunity to analyze trends, patterns, and behaviors in consumer interactions with financial institutions, shedding light on common issues and areas requiring improvement. The goal of this Big Data Analysis project is to extract actionable insights from this large and complex dataset, enabling stakeholders to better understand customer concerns and enhance service quality. Using Hadoop's distributed storage (HDFS) and processing capabilities (MapReduce), the dataset is managed and analyzed efficiently. This analysis not only highlights systemic issues in the financial industry but also provides recommendations to companies for improving customer satisfaction and regulatory compliance.

# Chapter 2
# DATASET DESCRIPTION



| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | date_rece | product | sub_produ | issue | sub_issue | consumer_ | company_ | company | state | zipcode | tags | consumer_ | submitted_ | date_sent_ | company_ | timely_res | consumer_ | complaint |
| 2 | 08/30/201 | Mortgage | Other mor | Loan modification,collection,foreclosure | | | | U.S. Bancc | CA | 95993 | | | Referral | ######## | Closed wit | Yes | Yes | 511074 |
| 3 | 08/30/201 | Mortgage | Other mor | Loan servicing, payments, escrow account | | | | Wells Farg | CA | 91104 | | | Referral | ######## | Closed wit | Yes | Yes | 511080 |
| 4 | 08/30/201 | Credit reporting | | Incorrect i | Account status | | | Wells Farg | NY | 11764 | | | Postal mai | 09/18/201 | Closed wit | Yes | No | 510473 |
| 5 | 08/30/201 | Student lo | Non-feder | Repaying y | Repaying your loan | | | Navient Sc | MD | 21402 | | | Email | 08/30/201 | Closed wit | Yes | Yes | 510326 |
| 6 | 08/30/201 | Debt colle | Credit carc | False state | Attempted to collect wrong amc | | | Resurgent | GA | 30106 | | | Web | 08/30/201 | Closed wit | Yes | Yes | 511067 |
| 7 | 08/30/201 | Credit card | | Application processing delay | | | | Capital On | NY | 12206 | | | Phone | ######## | Closed wit | Yes | Yes | 510098 |
| 8 | 08/30/201 | Credit card | | Credit line increase/decrease | | | | Wells Farg | AZ | 85730 | | | Postal mai | ######## | Closed wit | Yes | No | 511062 |
| 9 | 08/30/201 | Bank acco | Checking a | Deposits and withdrawals | | | | Bank of Ar | IL | 60660 | | | Referral | ######## | Closed wit | Yes | No | 511116 |
| 10 | 08/30/201 | Bank acco | Checking a | Deposits and withdrawals | | | | Bank of Ar | GA | 30016 | | | Referral | ######## | Closed wit | Yes | No | 511091 |
| 11 | 09/17/201 | Mortgage | Conventio | Loan modification,collection,foreclosure | | | | SunTrust B | CA | 94551 | | | Web | 09/18/201 | Closed wit | Yes | Yes | 530602 |
| 12 | 08/30/201 | Debt colle | Other (i.e. | Cont'd att | Debt was paid | | | Convergen | NC | 27545 | | | Web | 08/30/201 | Closed wit | Yes | No | 509988 |
| 13 | 08/30/201 | Mortgage | Other mor | Loan servicing, payments, escrow account | | | | Capital On | NC | 27614 | | | Fax | ######## | Closed wit | Yes | Yes | 511084 |
| 14 | 08/30/201 | Mortgage | Conventio | Credit decision / Underwriting | | | | Bank of Ar | TX | 75023 | | | Web | ######## | Closed wit | Yes | No | 511090 |
| 15 | 08/30/201 | Mortgage | Other mor | Application, originator, mortgage broker | | | | Wells Farg | TX | 78244 | | | Fax | ######## | Closed wit | Yes | Yes | 510129 |
| 16 | ######## | Debt colle | Credit carc | Communic | Called outside of 8am-9pm | | | Bank of Ar | GA | 30132 | | | Web | 09/14/201 | Closed wit | Yes | No | 521353 |
| 17 | 09/17/201 | Credit card | | Late fee | | | | PayPal Hol | DC | 20003 | | | Web | 09/17/201 | Closed wit | Yes | No | 531853 |
| 18 | 09/17/201 | Debt colle | Credit carc | Communic | Used obscene/profane/abusive l | | | Van Ru Cre | KY | 40508 | | | Phone | 09/18/201 | Closed wit | Yes | No | 530622 |
| 19 | 09/24/201 | Credit reporting | | Incorrect i | Reinserted previously deleted inf | | | TransUnio | IL | 62060 | | | Web | 09/24/201 | Closed wit | Yes | No | 540835 |
| 20 | 09/24/201 | Debt colle | Non-feder | Improper c | Talked to a third party about my | | | Allied Inte | GA | 30039 | | | Phone | 09/25/201 | Closed wit | Yes | No | 540355 |
| 21 | 09/17/201 | Credit reporting | | Incorrect i | Information is not mine | | | Equifax | RI | 2921 | | | Postal mai | ######## | Closed wit | Yes | No | 531346 |
| 22 | 09/17/201 | Credit card | | Billing disputes | | | | Bank of Ar | TN | 37874 | Older American | | Phone | ######## | Closed wit | Yes | Yes | 530624 |
| 23 | 09/24/201 | Debt colle | Non-feder | Improper c | Talked to a third party about my | | | Navient Sc | GA | 30039 | | | Phone | 09/25/201 | Closed wit | Yes | No | 540356 |
| 24 | 09/24/201 | Mortgage | FHA mortg | Loan modification,collection,foreclosure | | | | Carrington | AR | 72113 | | | Web | 09/25/201 | Closed wit | Yes | No | 540357 |
| 25 | ######## | Debt colle | Credit carc | Communic | Called outside of 8am-9pm | | | Wells Farg | GA | 30132 | | | Web | 09/14/201 | Closed wit | Yes | No | 521354 |
| 26 | ######## | Mortgage | Conventio | Loan modification,collection,foreclosure | | | | HSBC Nort | IL | 62034 | | | Web | ######## | Closed wit | Yes | No | 521356 |
| 27 | 09/24/201 | Bank acco | Other banl | Making/receiving payments, sending money | | | | Enova Inte | AL | 35473 | | | Web | 09/26/201 | Closed wit | Yes | No | 540359 |
| 28 | 09/24/201 | Bank acco | Checking a | Deposits and withdrawals | | | | Wells Farg | NC | 28753 | | | Web | 09/25/201 | Closed wit | Yes | No | 540360 |
| 29 | ######## | Credit card | | Rewards | | | | Bank of Ar | NJ | 8512 | | | Web | 09/19/201 | Closed wit | Yes | No | 521360 |

The **US Consumer Finance Complaints Dataset**, maintained by the Consumer Financial Protection Bureau (CFPB), contains detailed information about consumer complaints related to financial products and services. Below are its key characteristics:

- **Source**: Collected and published by the CFPB, ensuring high reliability and relevance to consumer finance issues.
- **Time Span**: Includes complaints recorded over several years, providing longitudinal data for trend analysis.
- **Volume**: Comprises millions of records, making it ideal for Big Data processing and analysis.

## Key Fields in the Dataset

- **Complaint ID**: Unique identifier for each consumer complaint.
- **Product**: The type of financial product or service (e.g., credit card, mortgage, student loan).
- **Issue**: Description of the specific problem faced by the consumer (e.g., billing errors, fraud).
- **Sub-Product**: Further classification of the product (e.g., FHA mortgage, private student loan).

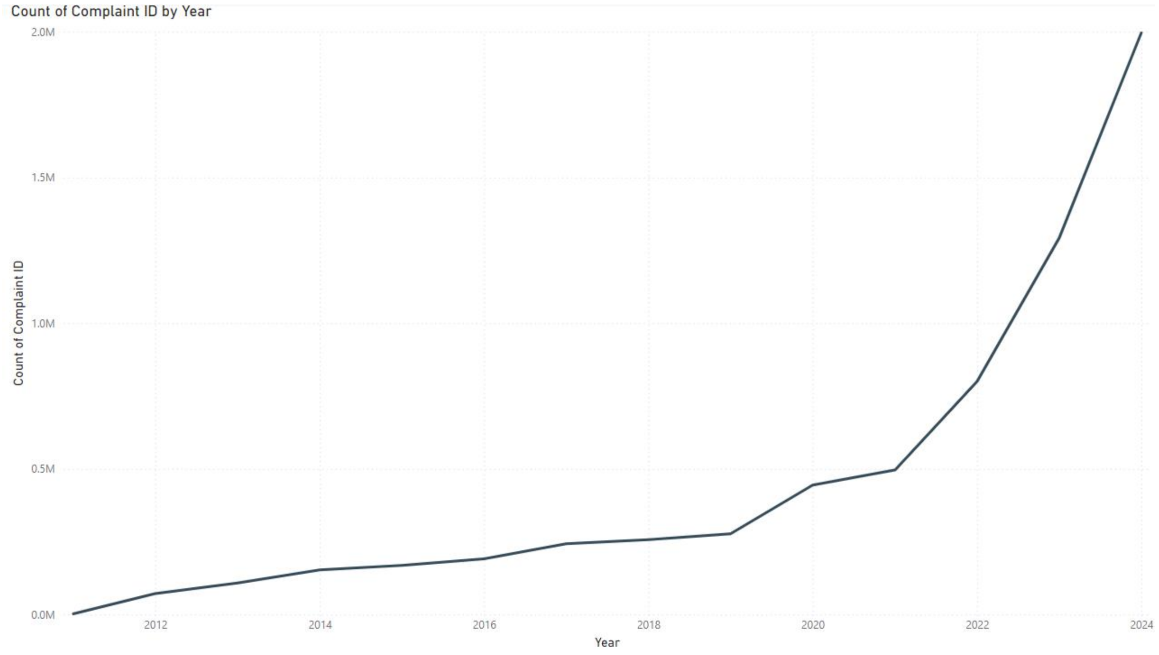
- **Company**: The financial institution or service provider associated with the complaint.
- **Consumer Complaint Narrative**: A free-text description of the consumer's issue (optional, as not all consumers provide narratives).
- **State and Zip Code**: Location details of the consumer.
- **Submission Method**: How the complaint was submitted (e.g., web, phone, email).
- **Company Response**: The company’s resolution or action taken regarding the complaint.
- **Timeliness**: Whether the company provided a timely response to the complaint.

Link: https://catalog.data.gov/dataset/consumer-complaint-database

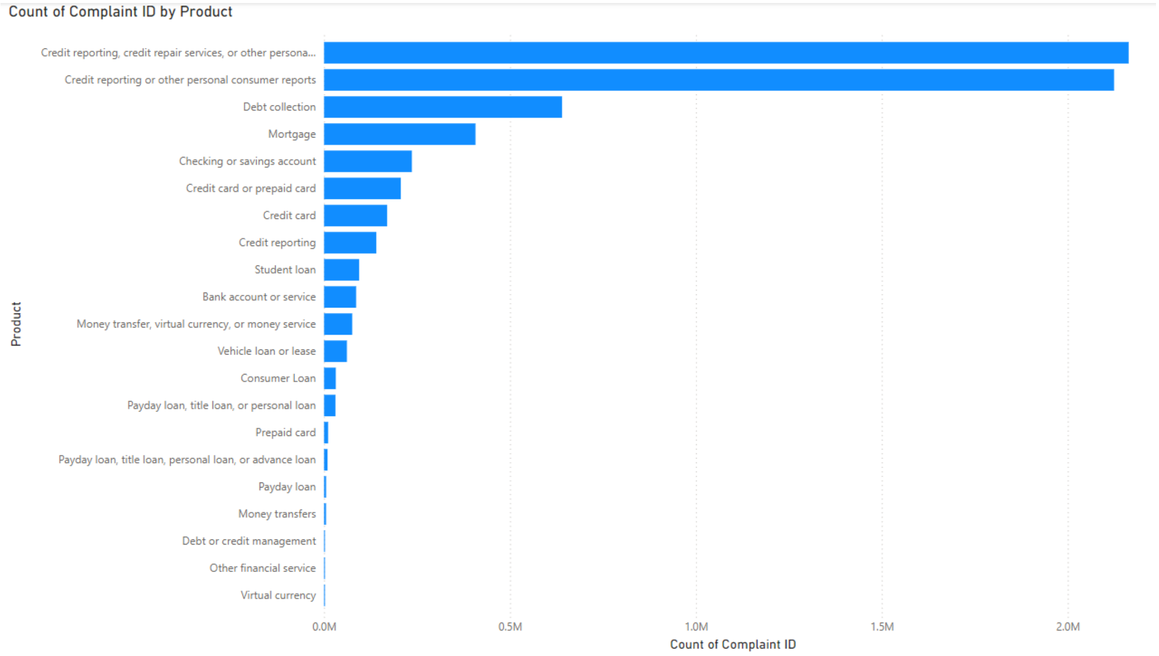
# Chapter 3

# INPUT DATA ANALYSIS USING BIG DATA SANDBOX TOOL

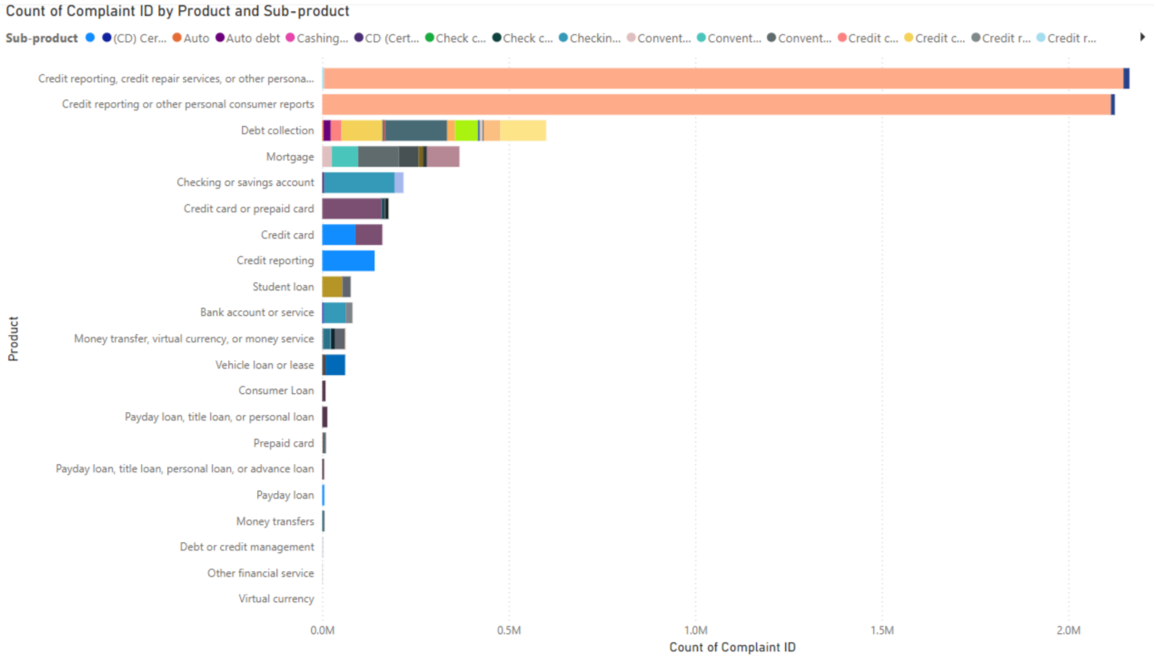
## Complaints Over Time (Line Chart)



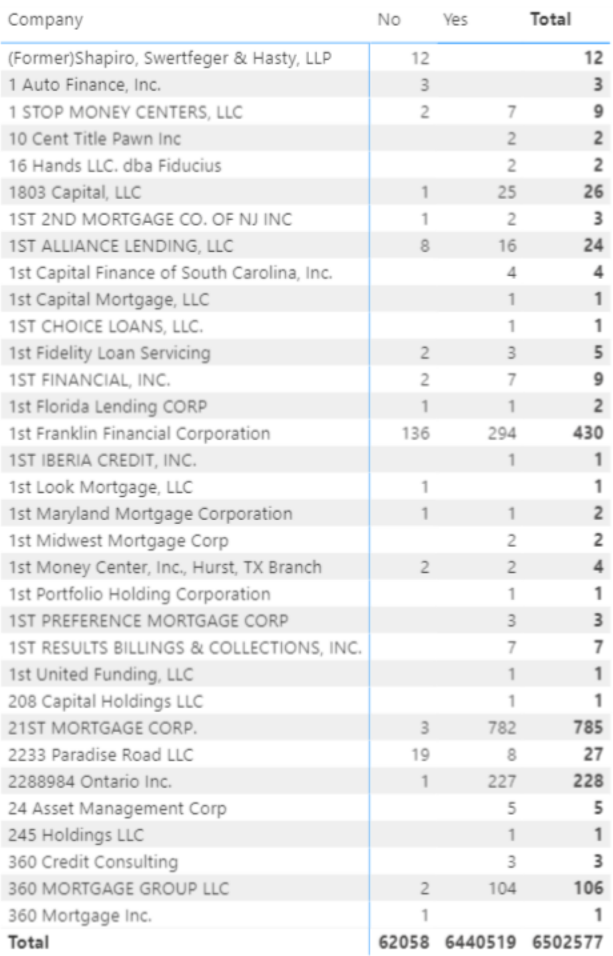
## Complaints by Product (Bar Chart)



## Sub-product Breakdown (Stacked Bar Chart)

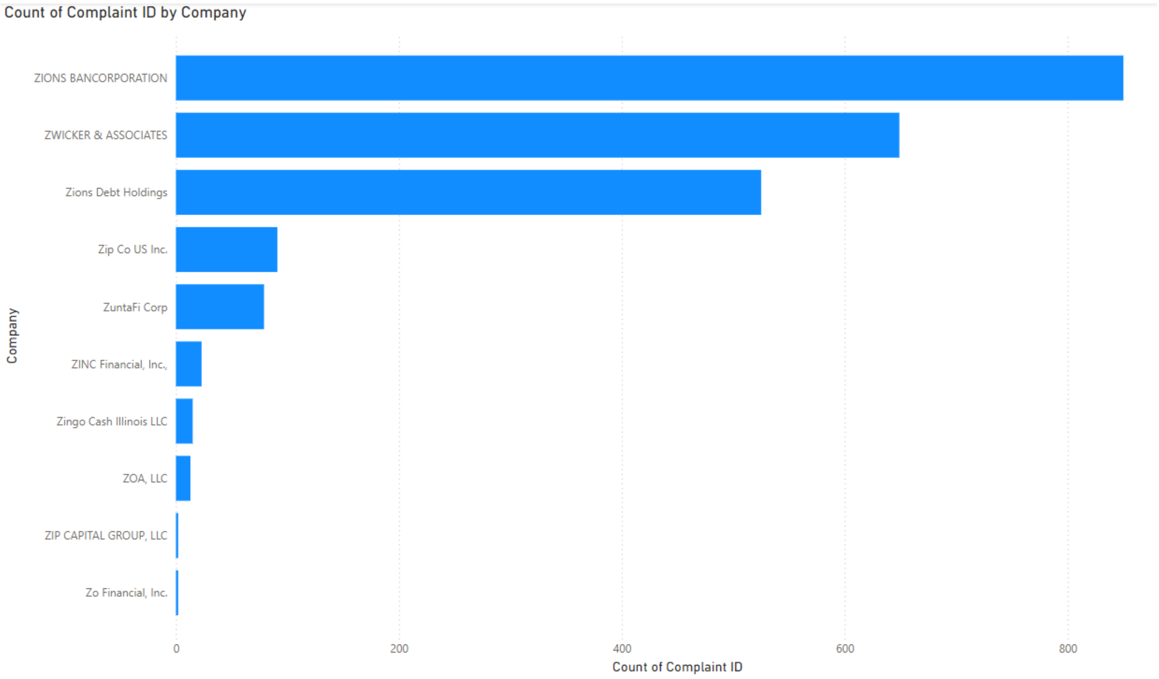


## Timely Responses by Company (Matrix)

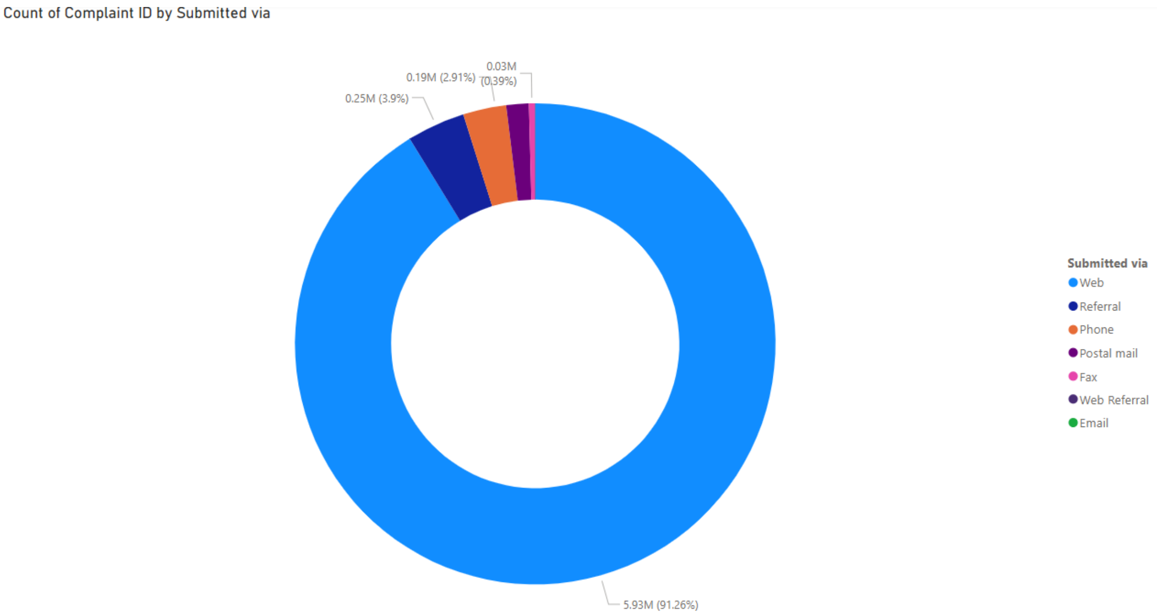


| Company | No | Yes | Total |
|---|---|---|---|
| (Former)Shapiro, Swertfeger & Hasty, LLP | 12 | | **12** |
| 1 Auto Finance, Inc. | 3 | | **3** |
| 1 STOP MONEY CENTERS, LLC | 2 | 7 | **9** |
| 10 Cent Title Pawn Inc | | 2 | **2** |
| 16 Hands LLC. dba Fiducius | | 2 | **2** |
| 1803 Capital, LLC | 1 | 25 | **26** |
| 1ST 2ND MORTGAGE CO. OF NJ INC | 1 | 2 | **3** |
| 1ST ALLIANCE LENDING, LLC | 8 | 16 | **24** |
| 1st Capital Finance of South Carolina, Inc. | | 4 | **4** |
| 1st Capital Mortgage, LLC | | 1 | **1** |
| 1ST CHOICE LOANS, LLC. | | 1 | **1** |
| 1st Fidelity Loan Servicing | 2 | 3 | **5** |
| 1ST FINANCIAL, INC. | 2 | 7 | **9** |
| 1st Florida Lending CORP | 1 | 1 | **2** |
| 1st Franklin Financial Corporation | 136 | 294 | **430** |
| 1ST IBERIA CREDIT, INC. | | 1 | **1** |
| 1st Look Mortgage, LLC | 1 | | **1** |
| 1st Maryland Mortgage Corporation | 1 | 1 | **2** |
| 1st Midwest Mortgage Corp | | 2 | **2** |
| 1st Money Center, Inc., Hurst, TX Branch | 2 | 2 | **4** |
| 1st Portfolio Holding Corporation | | 1 | **1** |
| 1ST PREFERENCE MORTGAGE CORP | | 3 | **3** |
| 1ST RESULTS BILLINGS & COLLECTIONS, INC. | | 7 | **7** |
| 1st United Funding, LLC | | 1 | **1** |
| 208 Capital Holdings LLC | | 1 | **1** |
| 21ST MORTGAGE CORP. | 3 | 782 | **785** |
| 2233 Paradise Road LLC | 19 | 8 | **27** |
| 2288984 Ontario Inc. | 1 | 227 | **228** |
| 24 Asset Management Corp | | 5 | **5** |
| 245 Holdings LLC | | 1 | **1** |
| 360 Credit Consulting | | 3 | **3** |
| 360 MORTGAGE GROUP LLC | 2 | 104 | **106** |
| 360 Mortgage Inc. | 1 | | **1** |
| **Total** | **62058** | **6440519** | **6502577** |

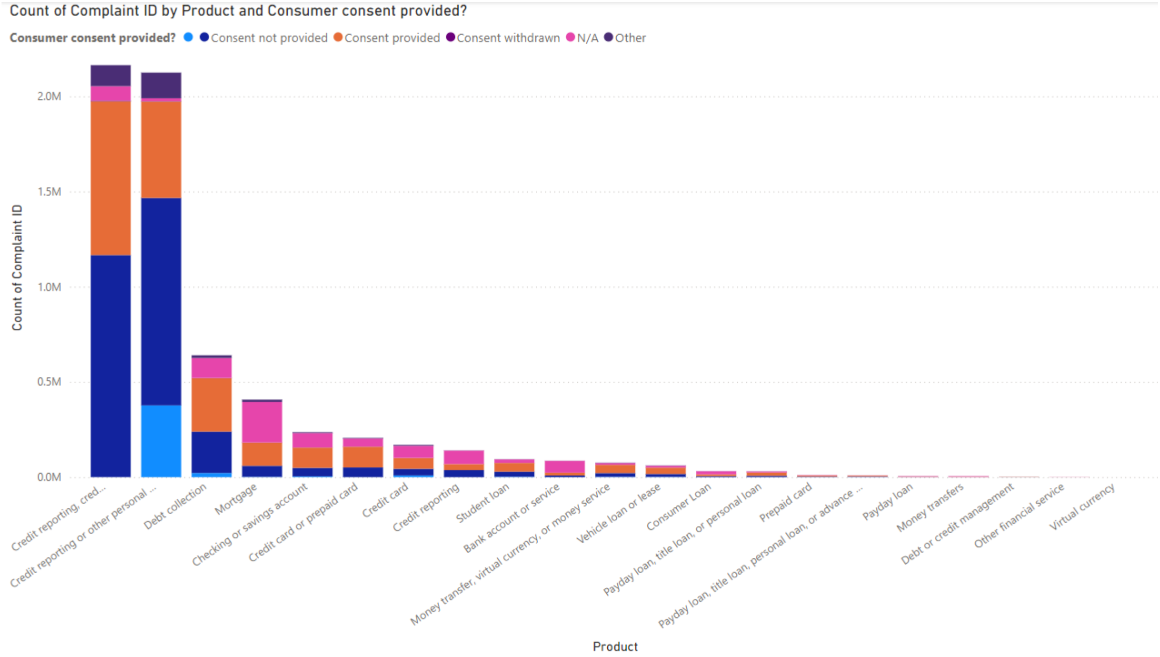
# Top 10 Companies with Most Complaints (Horizontal Bar Chart)



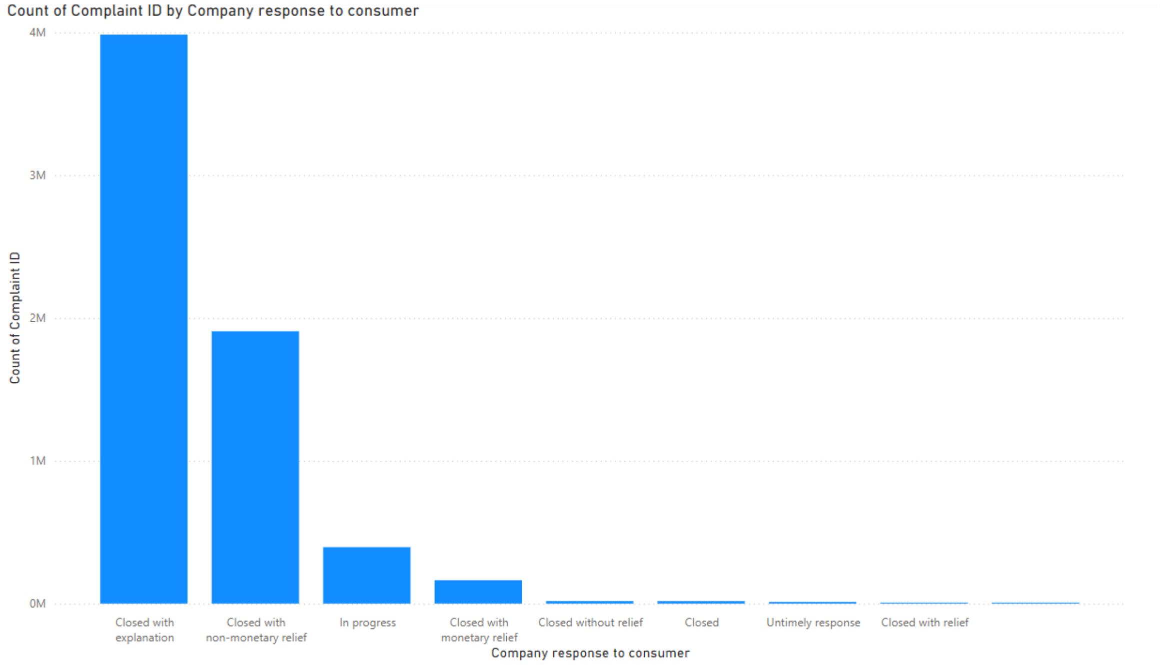
# Submission methods(Pie/Donut Chart)



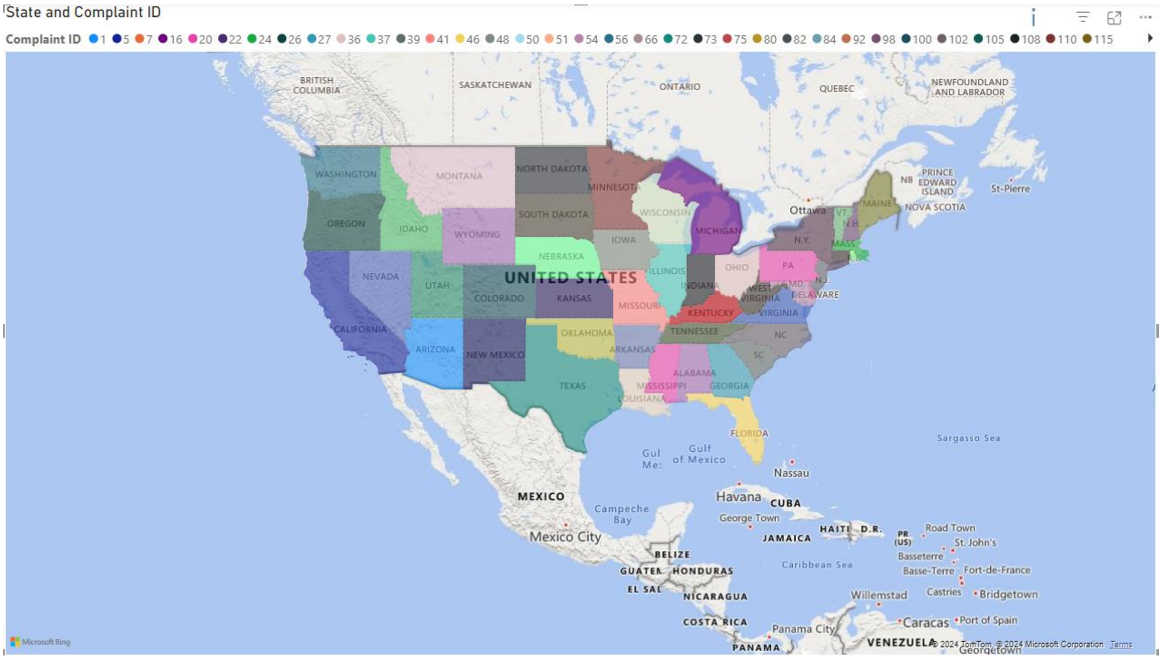
# Consumer Consent Analysis (Stacked Column Chart)



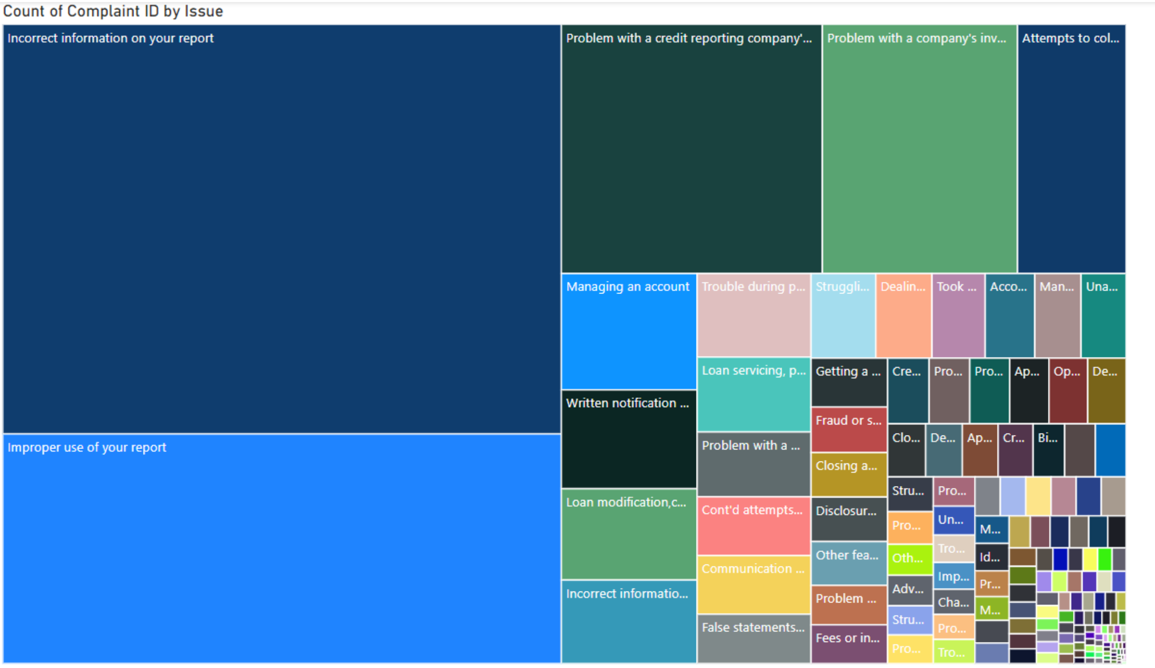
# Complaints by State (Map)



## Company Responses (Clustered Column Chart)



## Complaint Issues Breakdown (Tree Map)



# Chapter 4
# MODEL SELECTION

**Selected Model:** Logistic Regression

**Model Selection Explanation:**

Logistic Regression is a suitable choice for this project due to its effectiveness in handling classification tasks, such as predicting whether complaints are resolved on time or categorizing complaint severity. The model’s interpretability allows us to understand the influence of various features like product type, company, and submission method on outcomes.

It is computationally efficient, scalable for large datasets, and works well with features derived from distributed systems like Hadoop and Hive. Logistic Regression balances simplicity and performance, making it ideal for deriving actionable insights from the dataset while aligning with the project’s goals.

# Chapter 5

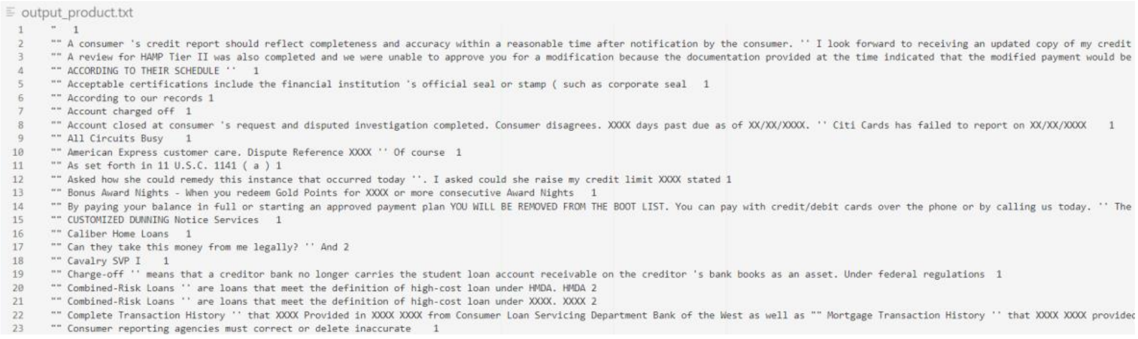
# HADOOP MAP REDUCE IMPLEMENTATION IN JAVA

## Query 1: Complaints per Product

```java
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

import java.io.IOException;

public class ProductMapper extends Mapper<Object, Text, Text, IntWritable> {

    private final static IntWritable one = new IntWritable(1);
    private Text product = new Text();

    public void map(Object key, Text value, Context context) throws IOException, InterruptedException {
        // Split the input line into fields using a comma as the delimiter
        String[] fields = value.toString().split(",", -1);

        // Check if the record has the expected number of fields and is not a header
        if (fields.length > 1 && !fields[1].equalsIgnoreCase("product")) {
            String productName = fields[1].trim(); // Extract the "product" field
            if (!productName.isEmpty()) {
                product.set(productName);
                context.write(product, one);
            }
        }
    }
}
```

```java
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

import java.io.IOException;

public class ProductReducer extends Reducer<Text, IntWritable, Text, IntWritable> {

    private IntWritable result = new IntWritable();

    public void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException,
InterruptedException {
        int sum = 0;
        for (IntWritable val : values) {
            sum += val.get(); // Sum all values for each product
        }
        result.set(sum);
        context.write(key, result); // Emit (Product, Total Complaints)
    }
}
```

## Output:



output_product.txt

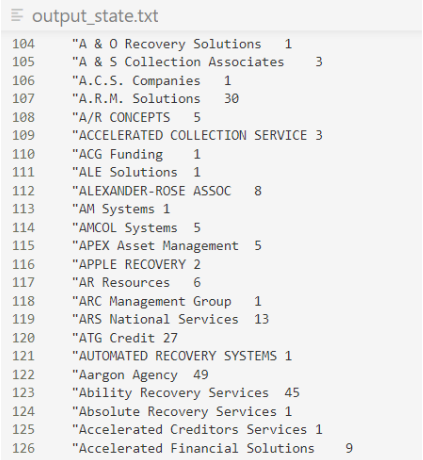
```
"   1
"" A consumer 's credit report should reflect completeness and accuracy within a reasonable time after notification by the consumer. '' I look forward to receiving an updated copy of my credit
"" A review for HAMP Tier II was also completed and we were unable to approve you for a modification because the documentation provided at the time indicated that the modified payment would be
"" ACCORDING TO THEIR SCHEDULE ''   1
"" Acceptable certifications include the financial institution 's official seal or stamp ( such as corporate seal   1
"" According to our records 1
"" Account charged off  1
"" Account closed at consumer 's request and disputed investigation completed. Consumer disagrees. XXXX days past due as of XX/XX/XXXX. '' Citi Cards has failed to report on XX/XX/XXXX    1
"" All Circuits Busy    1
"" American Express customer care. Dispute Reference XXXX '' Of course  1
"" As set forth in 11 U.S.C. 1141 ( a ) 1
"" Asked how she could remedy this instance that occurred today ''. I asked could she raise my credit limit XXXX stated 1
"" Bonus Award Nights - When you redeem Gold Points for XXXX or more consecutive Award Nights   1
"" By paying your balance in full or starting an approved payment plan YOU WILL BE REMOVED FROM THE BOOT LIST. You can pay with credit/debit cards over the phone or by calling us today. '' The
"" CUSTOMIZED DUNNING Notice Services   1
"" Caliber Home Loans   1
"" Can they take this money from me legally? '' And 2
"" Cavalry SVP I    1
"" Charge-off '' means that a creditor bank no longer carries the student loan account receivable on the creditor 's bank books as an asset. Under federal regulations  1
"" Combined-Risk Loans '' are loans that meet the definition of high-cost loan under HMDA. HMDA 2
"" Combined-Risk Loans '' are loans that meet the definition of high-cost loan under XXXX. XXXX 2
"" Complete Transaction History '' that XXXX Provided in XXXX XXXX from Consumer Loan Servicing Department Bank of the West as well as "" Mortgage Transaction History '' that XXXX XXXX provided
"" Consumer reporting agencies must correct or delete inaccurate    1
```

# Query 2: Complaints per State

```java
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

import java.io.IOException;

public class StateMapper extends Mapper<Object, Text, Text, IntWritable> {

    private final static IntWritable one = new IntWritable(1);
    private Text state = new Text();

    public void map(Object key, Text value, Context context) throws IOException, InterruptedException {
        // Split the input line into fields using a comma as the delimiter
        String[] fields = value.toString().split(",", -1);

        // Check if the record has the expected number of fields and is not a header
        if (fields.length > 9 && !fields[9].equalsIgnoreCase("state")) {
            String stateName = fields[9].trim(); // Extract the "state" field
            if (!stateName.isEmpty()) {
                state.set(stateName);
                context.write(state, one);
            }
        }
    }
}
```

```java
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

import java.io.IOException;

public class StateReducer extends Reducer<Text, IntWritable, Text, IntWritable> {

    private IntWritable result = new IntWritable();

    @Override
    public void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException,
InterruptedException {
        int sum = 0;

        // Sum all values for the given key (state)
        for (IntWritable val : values) {
            sum += val.get();
        }

        result.set(sum);
        context.write(key, result);
    }
}
```

# Output:



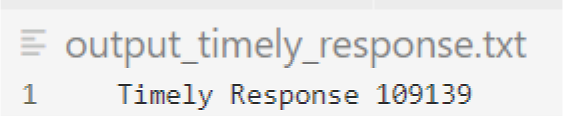
```
"A & O Recovery Solutions   1
"A & S Collection Associates    3
"A.C.S. Companies   1
"A.R.M. Solutions   30
"A/R CONCEPTS   5
"ACCELERATED COLLECTION SERVICE 3
"ACG Funding    1
"ALE Solutions  1
"ALEXANDER-ROSE ASSOC   8
"AM Systems 1
"AMCOL Systems  5
"APEX Asset Management  5
"APPLE RECOVERY 2
"AR Resources   6
"ARC Management Group   1
"ARS National Services  13
"ATG Credit 27
"AUTOMATED RECOVERY SYSTEMS 1
"Aargon Agency  49
"Ability Recovery Services  45
"Absolute Recovery Services 1
"Accelerated Creditors Services 1
"Accelerated Financial Solutions    9
```

# Query 3: Number of Complaints with Timely Response

```java
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

import java.io.IOException;

public class TimelyResponseMapper extends Mapper<Object, Text, Text, IntWritable> {

    private final static IntWritable one = new IntWritable(1);
    private Text timelyResponse = new Text("Timely Response");

    public void map(Object key, Text value, Context context) throws IOException, InterruptedException {
        // Split the input line into fields using a comma as the delimiter
        String[] fields = value.toString().split(",", -1);

        // Check if the record has the expected number of fields and is not a header
        if (fields.length > 16 && !fields[16].equalsIgnoreCase("timely_response")) {
            String response = fields[16].trim(); // Extract the "timely_response" field
            if (response.equalsIgnoreCase("yes")) { // Check if the response was "yes"
                context.write(timelyResponse, one);
            }
        }
    }
}
```

```java
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;
import java.io.IOException;

public class TimelyResponseReducer extends Reducer<Text, IntWritable, Text, IntWritable> {

    private IntWritable result = new IntWritable();

    @Override
    public void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException,
InterruptedException {
        int sum = 0;

        // Sum all values for the given key (Timely or Not Timely)
        for (IntWritable val : values) {
            sum += val.get();
        }

        result.set(sum);
        context.write(key, result);
    }
}
```

## Output:



output_timely_response.txt

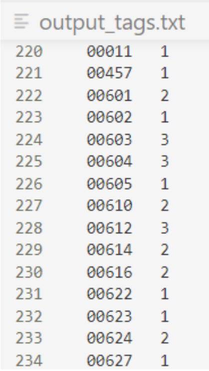
```
1      Timely Response 109139
```

# Query 4: Complaints by Tags

```java
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

import java.io.IOException;

public class TagsMapper extends Mapper<Object, Text, Text, IntWritable> {

    private final static IntWritable one = new IntWritable(1);
    private Text tag = new Text();

    public void map(Object key, Text value, Context context) throws IOException, InterruptedException {
        // Split the input line into fields using a comma as the delimiter
        String[] fields = value.toString().split(",", -1);

        // Check if the record has the expected number of fields and is not a header
        if (fields.length > 10 && !fields[10].equalsIgnoreCase("tags")) {
            String tagValue = fields[10].trim(); // Extract the "tags" field
            if (!tagValue.isEmpty()) {
                tag.set(tagValue);
                context.write(tag, one);
            }
        }
    }
}
```

```java
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

import java.io.IOException;

public class TagsReducer extends Reducer<Text, IntWritable, Text, IntWritable> {

    private IntWritable result = new IntWritable();

    @Override
    public void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException,
InterruptedException {
        int sum = 0;

        // Sum all values for the given key (Tag)
        for (IntWritable val : values) {
            sum += val.get();
        }

        result.set(sum);
        context.write(key, result);
    }
}
```

# Output:



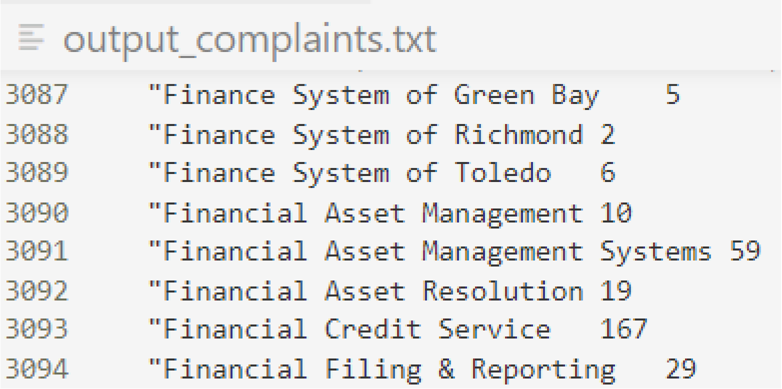
```
00011   1
00457   1
00601   2
00602   1
00603   3
00604   3
00605   1
00610   2
00612   3
00614   2
00616   2
00622   1
00623   1
00624   2
00627   1
```

# Query 5: Complaints per Company

```java
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Mapper;

import java.io.IOException;

public class ComplaintsMapper extends Mapper<LongWritable, Text, Text, IntWritable> {

    private Text company = new Text();
    private final static IntWritable one = new IntWritable(1);

    @Override
    protected void map(LongWritable key, Text value, Context context) throws IOException,
InterruptedException {
        String[] fields = value.toString().split(",", -1); // Split by comma
        if (fields.length > 7) { // Ensure enough fields exist
            String companyName = fields[7].trim(); // Assuming the 8th column is 'company'
            if (!companyName.isEmpty() && !companyName.equals("company")) { // Skip header row
                company.set(companyName);
                context.write(company, one);
            }
        }
    }
}
```

```java
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

import java.io.IOException;

public class ComplaintsReducer extends Reducer<Text, IntWritable, Text, IntWritable> {

    @Override
    protected void reduce(Text key, Iterable<IntWritable> values, Context context) throws IOException,
InterruptedException {
        int sum = 0;
        for (IntWritable value : values) {
            sum += value.get();
        }
        context.write(key, new IntWritable(sum));
    }
}
```

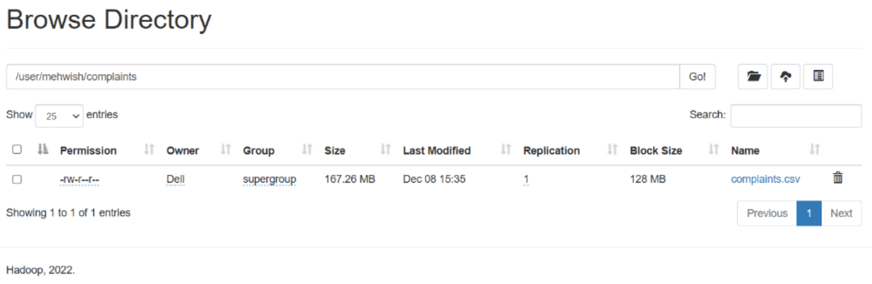
# Output:



output_complaints.txt

```
"Finance System of Green Bay	5
"Finance System of Richmond	2
"Finance System of Toledo	6
"Financial Asset Management	10
"Financial Asset Management Systems	59
"Financial Asset Resolution	19
"Financial Credit Service	167
"Financial Filing & Reporting	29
```

# Input Directory:



Browse Directory

/user/mehwish/complaints Go!

Show 25 entries Search:

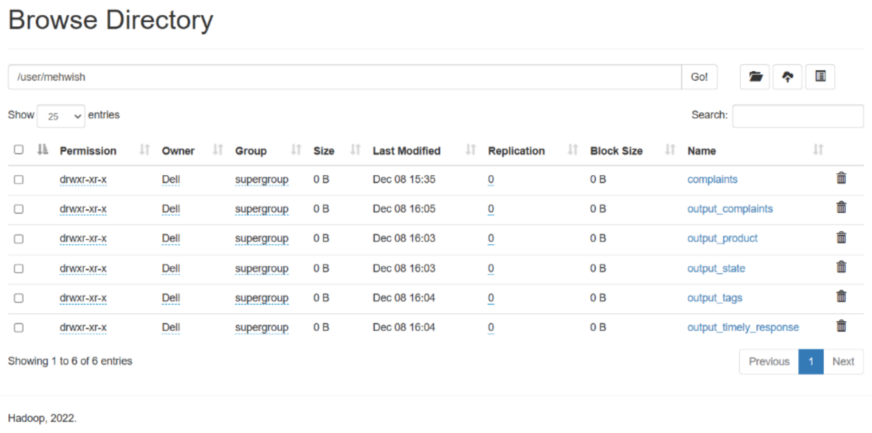
| □ | Permission | Owner | Group | Size | Last Modified | Replication | Block Size | Name |
|---|---|---|---|---|---|---|---|---|
| □ | -rw-r--r-- | Dell | supergroup | 167.26 MB | Dec 08 15:35 | 1 | 128 MB | complaints.csv |

Showing 1 to 1 of 1 entries

Previous 1 Next

Hadoop, 2022.

# Output Directory:



Browse Directory

/user/mehwish Go!

Show 25 entries Search:

| □ | Permission | Owner | Group | Size | Last Modified | Replication | Block Size | Name |
|---|---|---|---|---|---|---|---|---|
| □ | drwxr-xr-x | Dell | supergroup | 0 B | Dec 08 15:35 | 0 | 0 B | complaints |
| □ | drwxr-xr-x | Dell | supergroup | 0 B | Dec 08 16:05 | 0 | 0 B | output_complaints |
| □ | drwxr-xr-x | Dell | supergroup | 0 B | Dec 08 16:03 | 0 | 0 B | output_product |
| □ | drwxr-xr-x | Dell | supergroup | 0 B | Dec 08 16:03 | 0 | 0 B | output_state |
| □ | drwxr-xr-x | Dell | supergroup | 0 B | Dec 08 16:04 | 0 | 0 B | output_tags |
| □ | drwxr-xr-x | Dell | supergroup | 0 B | Dec 08 16:04 | 0 | 0 B | output_timely_response |

Showing 1 to 6 of 6 entries

Previous 1 Next

Hadoop, 2022.

# Chapter 6
# RESULT ANALYSIS IN HADOOP

## Common Trends in Consumer Complaints

1. **Frequent Financial Products Complained About**:
   - **Debt collection** is often the most complained-about category.
   - **Mortgages** come next, primarily due to foreclosure issues, payment disputes, and escrow problems.
   - **Credit reports** are also heavily reported, with inaccuracies being a recurring issue.
2. **Geographical Distribution**:
   - Larger states like **California**, **Texas**, and **New York** often have the highest number of complaints.
   - Smaller states or rural regions report fewer complaints, partly due to population differences and access to reporting systems.
3. **Submission Methods**:
   - Most complaints are submitted **online**, showing how accessible digital platforms have become for consumers.
   - Fewer complaints are submitted via phone or postal mail, particularly in recent years.
4. **Response Timeliness**:
   - Many companies respond in a "timely" manner as indicated by the timely_response field.
   - However, complaints with narratives or escalations to federal regulators often take longer to resolve.

## Potential Insights from Analysis

1. **Consumer Demographics**:
   - Tags like "Older Americans" and "Servicemember" reveal targeted consumer populations who might face unique financial challenges.
2. **Trends in Company Behavior**:

    - Certain companies consistently appear in complaints, often due to systemic issues (e.g., poor customer service, hidden fees, or inadequate dispute resolution).
    - Companies with the highest complaint counts often face legal and reputational risks.
3. **Regional Financial Issues**:
    - Foreclosure complaints are higher in regions with housing crises (e.g., after the 2008 housing crash).
    - Payday lending complaints are higher in states where such practices are prevalent.

## Key Analytical Results

### 1. Product Analysis:

- **Debt Collection** (top complaint): Consumers frequently report harassment and failure to validate debts.
- **Mortgage** (second-highest): Escrow management and improper foreclosure proceedings dominate.
- **Credit Reporting**: Disputes related to errors in credit scores and identity theft are common.

### 2. Company Analysis:

- Major financial institutions, such as **Equifax**, **Experian**, and **TransUnion**, often top complaint lists for credit reporting issues.
- **Bank of America**, **Wells Fargo**, and similar institutions are commonly reported for banking and mortgage issues.

### 3. State Analysis:

- States with larger populations (California, Texas) often see higher complaints.
- Financial literacy and local financial practices also play a role in the volume and type of complaints.

### 4. Timeliness:

- Most complaints get timely responses, but escalated or narrative-heavy complaints take longer to resolve.

# Chapter 7
# CONCLUSION

The analysis of the US Consumer Complaints Dataset highlights key trends and challenges within the financial sector, notably the rise in credit card complaints since 2011, reflecting growing consumer dissatisfaction. Addressing home loan objections requires regular audits of bank records by regulatory authorities and increased transparency through activity reports while safeguarding consumer privacy. Verifying documents and improving customer support systems can further reduce dissatisfaction.

This project enhanced our analytical skills through the use of Hadoop Distributed File System, Hive Query Language, and Power BI for visualizing trends. We found California had the highest number of complaints, with credit card issues being the most frequent. Most complaints were submitted online, and companies generally provided timely responses. These insights emphasize the need to address recurring complaints and implement corrective actions to improve customer satisfaction, offering valuable learnings for better service delivery in the financial sector.

# Chapter 8
# REFERENCES

- Java JDK 8
- Hadoop 3.2.4
- Hive 3.1.2
- Hadoop installation
- Dataset