Import basic Pandas and Numpy Library

In [ ]:
```
import pandas as pd
import numpy as np
```

DATA LOADING

Let's Understand the data

In [ ]:
```
complaints_file = pd.read_csv('/Users/sarthak/Desktop/LinkedInTHA/complaints.csv')
complaints_file.head(5)
```

```
/Users/sarthak/opt/anaconda3/lib/python3.8/site-packages/IPython/core/interactiveshell.py:3418: DtypeWarning: Co
lumns (16) have mixed types.Specify dtype option on import or set low_memory=False.
  exec(code_obj, self.user_global_ns, self.user_ns)
```

Out[ ]:

	Date received	Product	Sub-product	Issue	Sub-issue	Consumer complaint narrative	Company public response	Company	State	ZIP code	Tags
0	2023-12-24	Credit reporting or other personal consumer re...	Credit reporting	Incorrect information on your report	Information belongs to someone else	XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX X...	Company has responded to the consumer and the ...	Experian Information Solutions Inc.	IN	46168	NaN
1	2024-01-31	Credit reporting or other personal consumer re...	Credit reporting	Problem with a company's investigation into an...	Was not notified of investigation status or re...	NaN	NaN	TRANSUNION INTERMEDIATE HOLDINGS, INC.	NC	28269	NaN
2	2024-01-31	Credit reporting or other personal consumer re...	Credit reporting	Incorrect information on your report	Personal information incorrect	NaN	NaN	EQUIFAX, INC.	LA	708XX	NaN
3	2024-01-31	Debt collection	Credit card debt	Took or threatened to take negative or legal a...	Seized or attempted to seize your property	NaN	NaN	NAVY FEDERAL CREDIT UNION	VA	237XX	Servicemember
4	2024-01-31	Credit reporting or other personal consumer re...	Credit reporting	Incorrect information on your report	Account information incorrect	NaN	NaN	EQUIFAX, INC.	GA	30080	NaN

In [ ]:
```
Company = complaints_file['Company'].unique().shape[0]
Company
# There are 7119 unique companies in the dataset
```

Out[ ]: 7119

In [ ]:
```
print(complaints_file.columns)
# Name of the columns present in the dataset
```

```
Index(['Date received', 'Product', 'Sub-product', 'Issue', 'Sub-issue',
       'Consumer complaint narrative', 'Company public response', 'Company',
       'State', 'ZIP code', 'Tags', 'Consumer consent provided?',
       'Submitted via', 'Date sent to company', 'Company response to consumer',
       'Timely response?', 'Consumer disputed?', 'Complaint ID'],
      dtype='object')
```

In [ ]:
```
complaints_file.values[0]
# This is a sample of one record and their values against the columns mentioned above
```

```
Out[ ]: array(['2023-12-24',
               'Credit reporting or other personal consumer reports',
               'Credit reporting', 'Incorrect information on your report',
               'Information belongs to someone else',
               'XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XX
XX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XX
XX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX Subject : Identity Theft and Una
uthorized Collections on Credit Report Dear [ Collection Agency Name ], I hope this letter finds you well. I a
m writing to bring your attention to five collections that have appeared on my credit report, which are a resu
lt of identity theft. This unfortunate incident occurred on XX/XX/XXXX, when my car was stolen along with pers
onal identification documents. \n\nI would like to clarify that I did not initiate or authorize any of the tra
nsactions associated with these collections. As the victim of identity theft, I have taken the necessary steps
to report the incident to the relevant authorities, including filing a police report and notifying the Federal
Trade Commission ( FTC ). \n\nEnclosed with this letter are copies of the police report and FTC complaint, pro
viding evidence of the identity theft and unauthorized transactions. I kindly request that you review the info
rmation and take appropriate action to rectify this matter promptly. \n\nI understand that your role as a coll
ection agency is to recover debts owed to your client. However, in this case, I am not responsible for the deb
ts in question, as they were incurred fraudulently without my consent. I request that you remove these collect
ions from my credit report immediately, as they are negatively impacting my financial standing. \n\nI have als
o taken additional steps to protect my identity and prevent further fraudulent activities. These actions inclu
de placing a fraud alert on my credit file, notifying my financial institutions, and monitoring my credit repo
rts regularly. I encourage you to cooperate fully in this matter and help resolve the situation as soon as pos
sible',
               'Company has responded to the consumer and the CFPB and chooses not to provide a public response',
               'Experian Information Solutions Inc.', 'IN', '46168', nan,
               'Consent provided', 'Web', '2023-12-24',
               'Closed with non-monetary relief', 'Yes', nan, 8053505],
              dtype=object)
```

```
In [ ]: complaints_file.shape[0]
        # It calculates the total number of records, wwe can also use len(complaints_file) for the same
```

```
Out[ ]: 4750727
```

```
In [ ]: complaints_file_v2 = complaints_file.drop_duplicates()
        complaints_file_v2.shape[0]
        # It shows that there are no duplicate records in the dataset
```

```
Out[ ]: 4750727
```

a. Date

```
In [ ]: max_date = max(complaints_file['Date received'])
        min_date = min(complaints_file['Date received'])
```

```
In [ ]: max_date
```

```
Out[ ]: '2024-02-26'
```

```
In [ ]: min_date
```

```
Out[ ]: '2011-12-01'
```

b. Product

```
In [ ]: complaints_file['Product'].describe()
        # There are 21 unique types of products in this dataset
```

```
Out[ ]: count                                                 4750727
        unique                                                     21
        top       Credit reporting, credit repair services, or o...
        freq                                                  2163883
        Name: Product, dtype: object
```

```
In [ ]: Product = complaints_file['Product'].unique()
        Product
```

```
Out[ ]: array(['Credit reporting or other personal consumer reports',
               'Debt collection', 'Debt or credit management', 'Credit card',
               'Checking or savings account', 'Mortgage', 'Vehicle loan or lease',
               'Payday loan, title loan, personal loan, or advance loan',
               'Credit reporting, credit repair services, or other personal consumer reports',
               'Money transfer, virtual currency, or money service',
               'Payday loan', 'Student loan', 'Bank account or service',
               'Credit card or prepaid card',
               'Payday loan, title loan, or personal loan', 'Prepaid card',
               'Consumer Loan', 'Credit reporting', 'Other financial service',
               'Money transfers', 'Virtual currency'], dtype=object)
```

c. Sub-Product

In [ ]:
```
complaints_file['Sub-product'].describe()
# There are 86 unique types of sub-products in this dataset
```

Out[ ]:
```
count              4515437
unique                  86
top       Credit reporting
freq                2753202
Name: Sub-product, dtype: object
```

In [ ]:
```
Sub_product = complaints_file['Sub-product'].unique()
Sub_product
```

Out[ ]:
```
array(['Credit reporting', 'Credit card debt', 'Debt settlement',
       'General-purpose credit card or charge card', 'Checking account',
       'Payday loan debt', 'I do not know', 'Other mortgage',
       'VA mortgage', 'Loan', 'Installment loan',
       'Federal student loan debt', 'Telecommunications debt',
       'Virtual currency', 'Domestic (US) money transfer', 'Other debt',
       'Rental debt', 'Store credit card', 'Medical debt',
       'Other personal consumer report', nan, 'Savings account',
       'Auto debt', 'Federal student loan servicing',
       'Personal line of credit', 'Conventional home mortgage',
       'Other banking product or service', 'Conventional fixed mortgage',
       'Mobile or digital wallet', 'Reverse mortgage', 'FHA mortgage',
       'USDA mortgage', 'Government benefit card', 'Mortgage debt',
       'Payday loan', 'Title loan', 'Other type of mortgage',
       'Private student loan', 'General-purpose prepaid card',
       'Credit repair services', 'Refund anticipation check',
       'International money transfer', 'Check cashing service', 'Lease',
       'Vehicle loan', 'Foreign currency exchange', 'Gift card',
       'Conventional adjustable mortgage (ARM)',
       "Money order, traveler's check or cashier's check",
       'Home equity loan or line of credit (HELOC)',
       'Other (i.e. phone, health club, etc.)',
       'Home equity loan or line of credit',
       'CD (Certificate of Deposit)',
       "Traveler's check or cashier's check",
       '(CD) Certificate of deposit', 'Payroll card',
       'Manufactured home loan', 'Other advances of future income',
       'Pawn loan', 'Earned wage access', 'Private student loan debt',
       'Mortgage modification or foreclosure avoidance',
       'Student loan debt relief',
       'Tax refund anticipation loan or check', 'Student prepaid card',
       'Other bank product/service', 'Non-federal student loan',
       'Mortgage', 'Money order', 'Credit card', 'Medical',
       'Federal student loan', 'Second mortgage', 'Auto', 'Vehicle lease',
       'General purpose card', 'Cashing a check without an account',
       'Credit repair', 'Mobile wallet',
       'Government benefit payment card', 'Gift or merchant card',
       'Traveler's/Cashier's checks', 'Other special purpose card',
       'Check cashing', 'ID prepaid card',
       'Electronic Benefit Transfer / EBT card', 'Transit card'],
      dtype=object)
```

d. Company public response

In [ ]:
```
complaints_file['Company public response'].describe()
# There are 11 unique types of responses that company provided to consumers
```

Out[ ]:
```
count                                               2247755
unique                                                   11
top       Company has responded to the consumer and the ...
freq                                                2006306
Name: Company public response, dtype: object
```

In [ ]:
```
complaints_file['Company public response'].value_counts()
# Here is the count distribution of each unique responses
```

Out[ ]:
```
Company has responded to the consumer and the CFPB and chooses not to provide a public response
2006306
Company believes it acted appropriately as authorized by contract or law
137986
Company chooses not to provide a public response
52473
Company believes the complaint is the result of a misunderstanding
12992
Company disputes the facts presented in the complaint
10785
Company believes complaint caused principally by actions of third party outside the control or direction of th
e company       7748
Company believes complaint is the result of an isolated error
6259
Company believes complaint represents an opportunity for improvement to better serve consumers
4766
Company can't verify or dispute the facts in the complaint
4383
Company believes the complaint provided an opportunity to answer consumer's questions
3947
Company believes complaint relates to a discontinued policy or procedure
110
Name: Company public response, dtype: int64
```

e. Consumer Consent provided?

In [ ]:
```
complaints_file['Consumer consent provided?'].describe()
# There are 4 unique types of valus in "consumer consent provided?" column
```

Out[ ]:
```
count                 3788651
unique                      4
top       Consent not provided
freq                  1858034
Name: Consumer consent provided?, dtype: object
```

In [ ]:
```
Consumer_consent_provided = complaints_file['Consumer consent provided?'].unique()
Consumer_consent_provided
```

Out[ ]:
```
array(['Consent provided', 'Other', nan, 'Consent not provided',
       'Consent withdrawn'], dtype=object)
```

f. Submitted Via

In [ ]:
```
complaints_file['Submitted via'].describe()
# There are 7 unique ways by which the complaint/issues were submitted
```

Out[ ]:
```
count     4750727
unique          7
top           Web
freq      4208891
Name: Submitted via, dtype: object
```

In [ ]:
```
complaints_file['Submitted via'].value_counts()
```

Out[ ]:
```
Web             4208891
Referral         246000
Phone            175748
Postal mail       92764
Fax               25658
Web Referral       1241
Email               425
Name: Submitted via, dtype: int64
```

g. Consumer Complaint narrative

In [ ]:
```
complaints_file['Consumer complaint narrative'].describe()
```

Out[ ]:
```
count                                             1700839
unique                                            1402648
top       In accordance with the Fair Credit Reporting a...
freq                                                 4470
Name: Consumer complaint narrative, dtype: object
```

In [ ]:
```
Consumer_complaint_narrative = complaints_file['Consumer complaint narrative'].unique()
Consumer_complaint_narrative
```

Out[ ]: array(['XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XX
XX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XX
XX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX Subject : Identity Theft and Una
uthorized Collections on Credit Report Dear [Collection Agency Name], I hope this letter finds you well. I a
m writing to bring your attention to five collections that have appeared on my credit report, which are a resu
lt of identity theft. This unfortunate incident occurred on XX/XX/XXXX, when my car was stolen along with pers
onal identification documents. \n\nI would like to clarify that I did not initiate or authorize any of the tra
nsactions associated with these collections. As the victim of identity theft, I have taken the necessary steps
to report the incident to the relevant authorities, including filing a police report and notifying the Federal
Trade Commission (FTC). \n\nEnclosed with this letter are copies of the police report and FTC complaint, pro
viding evidence of the identity theft and unauthorized transactions. I kindly request that you review the info
rmation and take appropriate action to rectify this matter promptly. \n\nI understand that your role as a coll
ection agency is to recover debts owed to your client. However, in this case, I am not responsible for the deb
ts in question, as they were incurred fraudulently without my consent. I request that you remove these collect
ions from my credit report immediately, as they are negatively impacting my financial standing. \n\nI have als
o taken additional steps to protect my identity and prevent further fraudulent activities. These actions inclu
de placing a fraud alert on my credit file, notifying my financial institutions, and monitoring my credit repo
rts regularly. I encourage you to cooperate fully in this matter and help resolve the situation as soon as pos
sible',
       nan,
       'This is my notice of liability and intent to sue. \nSince slavery and involuntary servitude is illegal
and unlawful I would have to give you written consent to cause injury to me. This is human trafficking by way
of debt slavery! You have imposed a tax on me without my written consent! Article 1 Section 8 clause 1 only CO
NGRESS can impose a tax! By the business charter of Wisconsin chapter 180 you are not Congress you are a corpo
rate entity! \nContract law states there has to be a meeting of the minds, consideration and value given! PROV
E written Consent! This is damaging my life and keep me in complete debt bondage. By way of involuntary servit
ude! I have not given this company written consent! \nRemove this account immediately! You have 15 days and th
is is a good faith attempt rectify this unlawful act. I am a private citizen! Remove XXXX XXXX XXXX XXXX XXXX
DEPT OF EDUCATIONXXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXX
X XXXX XXXX XXXX! \n-Debt bondage is focused on human trafficking crimes in which the traffickers primary mean
s of coercion is debt manipulation.',
       ...,
       'On XX/XX/XXXX, at approximately XXXX. EST, I received several alerts from XXXX (a company of PayPal)
suggesting that I had submitted a money transfer. Since I have never used XXXX nor ever submitted any money tr
ansfer I immediately reported to XXXX via XXXX that the following transactions were fraudulent and not authori
zed by me. \n\nThose transactions were as follows : 1. Transaction ID : XXXX, Amount : {$130.00} 2. Transactio
n ID : XXXX ; Amount {$130.00} 3. Transaction ID : XXXX ; Amount {$170.00} Since I reported this as fraudulent
at XXXX XXXX XXXX immediately I did not believe that any money would be withdrawn from my account. By XXXXXXXX
XXXX I discovered that those charges were withdrawn from my bank account. I immediately notified my bank of th
e fraudulent activity and they processed a stop payment. Thereafter, I received several emails from XXXX refle
cting that I " owed \'\' a debt. I notified them on XX/XX/XXXX, that those transactions were fraudulent, but c
ontinued to received payment requests between XX/XX/XXXX through XX/XX/XXXX. I received notice on XX/XX/XXXX o
r XXXX that they denied my claim based upon " account history \'\' however, I requested evidence of this accou
nt history (since I never opened a XXXX account) and XXXX responded on XX/XX/XXXX, that XXXX could not provi
de details of the fraudulent transactions without a law enforcement agency. Please note, the original denial b
y XXXX occurred without anyone ever responding via email or phone call to discuss the facts. \n\nApproximately
two weeks later, it was discovered that our home computer system had been hacked since the same hacker used my
credit card to make payments on some account. It should be noted that because these transactions were from Pay
Pal, they ultimately resolved these fraudulent claims in my favor because they recognized that the reason that
the " history \'\' appeared to be legitimate was because the hacker broke into my computer system and it appea
red as if the individual were in my home THEY WERE NOT. Therefore, I filed an identify theft report on XX/XX/X
XXX, with the government and I also filed a complaint with the Better Business Bureau. \n\nOnce I filed my com
plaint with the XXXX, the company sent responses wherein they continued to deny that the transactions were aut
horized and notified me that I could contact their executive offices. I did that on XX/XX/XXXX, to an individu
al named XXXX, wherein I provided a detailed response and evidence showing that my account had been hacked, bu
t received no response from this. On XX/XX/XXXX, I received a response from XXXX through XXXX yet again was d
enying my claim since they believed it to be accurate. On XX/XX/XXXX, I provided a response via XXXX to XXXX
and directly to the executives offices to an individual named XXXX yet again explaining that the authorized tr
ansactions appeared legitimate due to a hack and not because I authorized same. I also provided evidence regar
ding same. On XX/XX/XXXX, I also sent a follow-up identity theft report to the XXXX offices and still received
no reply. \n\nOn XX/XX/XXXX, I received yet another response from XXXX via XXXX saying they denied my claim a
fter doing extensive research, but yet again they have never once either emailed or called me to discuss this
matter, or explained how the evidence I provided does not support that the transactions were authorized by me.
I also responded via XXXX expressing my dissatisfaction with the response. \n\nI did not authorize these tran
sactions since the evidence I\'ve provided proves that a hacker gained access to my computer/home network and
used the IP address that XXXX claims is " proof \'\' that I authorized these transactions, when in fact that i
s not true. \n\nI am filing this claim since the identity theft website indicated that this was an appropriate
next step and since this company is contending that this debt is valid and that they will notify collections s
ince they believe same to be a valid debt. This is NOT a valid debt. If you research this company you will see
that these type of fraudulent transactions happen frequently and they quickly " deny \'\' the claims leaving t
he consumer responsible for their inept actions. Had XXXX simply not allowed the transactions not to go throug
h when I originally notified them of this fraud, no money would have exchanged hands and this would not be a p
roblem. However, they failed to do so. I will fight these charges through every legal means possible since I w
ill not be held responsible for something I did not do primarily because if I do not then the individual that
hacked my computer and then ultimately my PayPal account (since I never had XXXX) would be able to get away
with this dangerous behavior. XXXX denying my claims is resulting in the perpetuation of fraud and they too mu
st be stopped. \n\nI am attaching several documents which have been provided to XXXX via XXXX and directly to
their executive offices for consideration.',
       'I have an injury committed by GENERAL MOTORS FINANCIAL towards me. I sent GENERAL MOTORS FINANCIAL , I
NC. a letter CERTIFIED MAIL NUMBER # XXXX on XX/XX/2022 requesting my credit balance refund Pursuant to 15 USC
CH 41 1666 (d) also Pursuant to 12 CFR 226.11, GENERAL MOTORS FINANCIAL did not respond and now I am contact
ing the CFPB to help with the validation of debt, as GENERAL MOTORS FINANCIAL , INC. stated that I owe an alle
ged debt. I called XX/XX/XXXX to get a response as to why nothing was sent regarding the letter about the cred
it balance refund. I received an email on XX/XX/2022 thinking this email was regarding the response to my lett
er. It was not, I called on XX/XX/2022, responding to the email that was sent to me by XXXX XXXX. I called and

asked about the refund, and no one was able to respond. I asked if my letter was received and I was told it was received on XX/XX/2022, and no written communication as requested has been given as of yet. I called again XX/XX/2022 and spoke to XXXX who was professional and informed me that she couldnt assist me on the number I dialed in her department that a manager would need to handle the call. She stated someone would call me back and I agreed on a 1 time phone call. I received a call back from XXXX XXXX Manager He stated he didn\'t want the call recorded and I stated as well that I did not want the call recorded and he stated " we do this for quality assurance \'\' I stated that " I do my recording for quality assurance \'\' He stated he didn\'t know me and that he doesn\'t have to be recorded \'\' I stated that I didn\'t know him and I choose to not be recorded. \'\' He said " I will submit a request \'\' I stated " A request? \'\' I do not want the call recorded period. I asked to speak to someone else who didn\'t want to be combative, but actually move forward with the call. He stated " it looks like you are working with an attorney so you don\'t need to record the call \'\', I stated " who informed you that I was being represented by an attorney? \'\' He stated well it assumes that you are based on the paperwork. \'\' I asked to speak to another supervisor immediately. He refused several times to get me another manager. I asked him several times for clarity, " are you refusing to get me another supervisor? \'\' He said " you can call back \'\' and informed him that " I am already on the phone. I am not calling back, you can get me to another supervisor now that I am on the line already. \'\' Eventually he said " ok \'\' put me on hold and came back and said " no one else was available \'\' and asked if they could call me back I replied " I am authorizing a one time call back, yes \'\'. He said ok someone will call you back. At XXXX cst I finished an email and sent it to XXXX XXXX to inform her of my encounter with her agents and questions about the refund and that no one had returned the call after her agent XXXX XXXX practices deceptive acts and misleading me attempting to make me think someone would call me back. GM financial closed at XXXX cst. XXXX didn\'t put in a request to have someone call me back. This is unacceptable and goes against GENERAL MOTORS FINANCIAL values and ethics. I have yet to receive a phone call back about the matter nor has anyone responded to the email I sent to XXXX informing her agent didnt put any request in to call me back to handle the payment arrangement. Called GENERAL MOTOR FINANCIAL , INC. again on XX/XX/2022 and spoke to XXXX XXXX who didnt have any way of helping me with a response to the letter I sent in communicating the credit refund. She wanted to pass me over to another department that handles refunds in the customer service department. I asked to speak to a manager in that department. She stated she would have one give me a callback and in the meantime we would wait for a supervisor in the customer service department that helps with refunds. XXXX XXXX returned from hold stating, They had a long wait, and she submitted the request for the supervisor to call me back to talk about the assistance program. I received a call back from XXXX XXXX XXXX and informed him that I was calling pertaining to my refund ; however, I was informed by XXXX XXXX that this department could not help me. It would be best for us not to go over the refund, but lets go over the assistance program. XXXX XXXX XXXX stated the account was in default and I stated how?, when I sent an instrument payment on XX/XX/2022 and received CERTIFIED MAIL # XXXX. He stated the payment has not been put in the system. When informing him of the payment instrument I sent in he stated and I quote XXXX we will not be entertaining this payment you sent in he laughed you know good as well that sending in that amount will do nothing for your account. This was an attempt of XXXX XXXX XXXX XXXX attempting to disgrace me PURSUANT to 15 U. S. C. 1692 (e) (7). He also took the liberty to mention that We will discuss the contract you signed with us! Informed that the contract should be null and void due to the numerous violations during the application process, I am currently aware of my rights and the misconduct GM Financial is handling with consumers is erroneous and does need to be investigated. XXXX XXXX did not even attempt to go over the assistance program but instead said take us to court, I know you have already done the research! I responded, You must not be familiar with XXXX XXXX Article XXXX XXXX, I NEED VALIDATION OF DEBT! XXXX XXXX XXXX employed with GENERAL MOTORS FINANCIAL COMPANY , INC. attempted to disgrace me and make a joke out of my character, by me exercising my rights. This is the reason we are here today. I do not trust this company to respond without the CFPB involved and GENERAL MOTORS FINANCIAL COMPANY , INC. should be following the law and should be held accountable for not following the law. I would also like to point out another fact, my account was not in default when contacting GENERAL MOTORS FINANCIAL COMPANY , INC. on XX/XX/2022, XXXX XXXX instead of going over any programs he escalated the call and never mentioned and solutions or remedies. By agent XXXX lying about someone calling me back and no one calling me back after saying someone would and no response from the department manager, CAUSING MORE INJURY PUTTING THIS SAID ACCOUNT IN DEFAULT. I am now requesting GENERAL MOTORS FINANCIAL COMPANY , INC. validate the debt and original creditor. Its Federal Law Pursuant to 18 U. S. C. Title 8, that there is no debt present. I need the debt validated and I need a written response as I am allowing this company to indeed cure this situation before I take any legal action and I WILL be taking legal action. If the debt can not be validated I need my title and the lien released as well as any debt removed indefinitely AND the credit balance refund amount that I requested in my initial contact letter. I also would like to include that before I was not aware of my rights with GENERAL MOTORS FINANCIAL COMPANY , INC. and I am very aware now. During the contract signing I was not made aware of 15 USC 1605 I was not to be charged a down payment on my property. TRUTH AND LENDING ACT also backs this same law and GM financial allowed XXXX XXXX finance manager to sit there and add insult to injury by being deceptive and misleading. GENERAL MOTORS FINANCIAL , INCXXXX non-communication is in agreement to owing the refund.',

'I am tired of the run around from XXXX XXXX in trying to come out of forbearance (via the XXXX XXXX) which has been ongoing now for the last XXXX months. Where to begin? I was told up front I\'d be able to enter a deferral plan upon exiting forbearance only to be told other-wise. I then subsequently wasted MONTHS on XXXX loan modification requests which were both denied. Now there has been mis-representation on who actually owns my Note which has led me to believe there are greater mechanisms occurring in an attempt to foreclose on my home. Per the attachment, after asking several times, XXXX XXXX informed me that XXXX XXXX XXXXXXXX XXXX XXXX XXXX XXXX) holds my mortgage Note. However, I spoke to XXXX XXXX on XX/XX/XXXX and they told me that they DO NOT OWN MY MORTGAGE NOTE AND HAVENT SINCE XX/XX/XXXX. You can call XXXX ; (then select Option # XXXX) to verify this information. This therefore begs the question : Why would XXXX XXXX send me a letter fraudulently stating otherwise? (again, defer to attached) Now I most recently received a letter stating that XXXX XXXX XXXX XXXX XXXX XXXX, Series XXXX owns my Note. Yet its the exact same contact information per the XXXX XXXX XXXX XXXX. WHY CAN I NOT OBTAIN THEIR DIRECT PHONE NUMBER ASIDE FORM A FAKE, GENERIC MAILING ADDRESS? \nXXXX XXXX NEEDS TO ADVISE ME WHO OWNS MY LOAN NOTE IMMEDIATELY ALONG WITH THEIR PHONE CONTACT INFORMATION OR I WILL HAVE NO CHOICE BUT TO PURUSE LEGAL ACTION ON THIS MATTER DUE TO THE FOLLOWING REASONS : 1) Mis-representing the owner of my Note (for YEARS) as being XXXX XXXX . 2) Not allowing me a deferral option post-forbearance after being told otherwise up front. All XXXXm repeatedly told is Deferral is not allowed. \'\' SAYS WHO???? I would like to hear it directly from this mysterious owner and/or investor of my Note.\n\n3) Not allowing me to take a loan modification either : I was asked UP FRONT if Id accept a higher payment for the 2nd loan modification request. After saying Yes and then waiting for XXXX months, I was told my loan modification request was denied because the payment couldnt be LOWERED. Really?? Why ask me then if Im ok with a higher payment up front if it was already known raising the payment wasnt even an option for a loan modification? I was never advised by any of the many representatives I spoke to for either of my loan modification requests, that approval was based on the presumption of a lower monthly payment.\n\n4) The imposter investor named XXXX? Last time I called, I was told by the representative XXXX was the investor holding my Note whom allegedly does not allow deferral options. Interestingly, I can not find any company by that name. I AM TIRED OF THE LIES, MIS-INFORMATION AND RUN ARO

```
UND. I AM READY AND ABLE TO RESUME MY MORTGAGE PAYMENTS BUT XXXX XXXX HAS MADE IT IMPOSSIBLE IN MY MANY ATTEMP
TS TO DO SO AFTER MIS-LEADING ME UP FRONT. I feel that at this point, due to all of the mis-information, confu
sion and their unwillingness to offer any assistance except to tell me verbatim " Sell your house \'\' or " Ge
t a 2nd job \'\' that I have no recourse but to pursue legal remediation. I am beyond exasparated at this poin
t as to how things have gone thus far and hence, the reason I am filing this complaint. \n\nFEEDBACK FOR RESOL
UTION 1 ) What is the phone number for this entity? Again, I called XXXX XXXX and I sure couldnt speak to anyo
ne from this entity. \n\nXXXX XXXX XXXX XXXX XXXX, as XXXX for XXXX XXXX XXXX XXXX XXXX XXXX XXXX, Series XXXX
XXXX XXXX XXXX XXXX XXXX, VA XXXX \'\' XXXX ) Why am I unable to contact the investor / owner of my Note direc
tly? \n\n3 ) Why did XXXX XXXX ignore my original question of why does a loan modification presume a lower pay
ment amount? When in fact, they asked me up front prior to my 2nd loan modification request if I would be ok w
ith a higher payment? Again -- can I request a loan modification even if it\'s for a reasonably higher amount?
\n\n4 ) Lastly, I still do not have a clear answer or proof that the investor or owner of my Note does NOT all
ow the deferral of my forbearance amount ... which essentially doubled because I had to wait another 7 months
for them to tell me it was declined and during said period they told me not to make any payments. If I knew wh
o my actual investor was, I could contact them directly to confirm this information.'],
      dtype=object)
```

h. Company response to consumer

In [ ]:
```
complaints_file['Company response to consumer'].describe()
# There are 8 unique types of company's response to consumers
```

Out[ ]:
```
count                     4750719
unique                          8
top       Closed with explanation
freq                      3269758
Name: Company response to consumer, dtype: object
```

In [ ]:
```
complaints_file['Company response to consumer'].value_counts()
```

Out[ ]:
```
Closed with explanation            3269758
Closed with non-monetary relief    1098851
In progress                         182262
Closed with monetary relief         148356
Closed without relief                17868
Closed                               17611
Untimely response                    10709
Closed with relief                    5304
Name: Company response to consumer, dtype: int64
```

i. Consumer Disputed

In [ ]:
```
complaints_file['Consumer disputed?'].value_counts()
# Around 3% (148378/total) of consumers has disputed against the company
```

Out[ ]:
```
No     619938
Yes    148378
Name: Consumer disputed?, dtype: int64
```

Now we are going to calculate the overall percentage of different company's response to customers (against each unique companies):

In [ ]:
```
def calculate_percentages(df):
    total_responses = len(df)
    in_progress = (df['Company response to consumer'] == 'In progress').sum() / total_responses * 100
    closed_nonrelief = df['Company response to consumer'].isin(['Closed with non-monetary relief', 'Closed with
    closed_relief = df['Company response to consumer'].str.contains('Closed').sum() / total_responses * 100 - c
    untimely_response = (df['Company response to consumer'] == 'Untimely response').sum() / total_responses * 1
    return pd.Series([in_progress, closed_nonrelief, closed_relief, untimely_response])

response_percentages = complaints_file.groupby('Company').apply(calculate_percentages).reset_index()
response_percentages.columns = ['Company', 'In progress percentage', 'Closed non-relief percentage', 'Closed re
sorted_response_percentages = response_percentages.sort_values(by='In progress percentage', ascending=False)

sorted_response_percentages
```

Out[ ]:

	Company	In progress percentage	Closed non-relief percentage	Closed relief percentage	Untimely response percentage
1236	CNY Management Group Inc	100.0	0.000000	0.000000	0.0
3218	Home Base Mortgage Group	100.0	0.000000	0.000000	0.0
3925	Lincoln Capital, LLC	100.0	0.000000	0.000000	0.0
6300	TGA	50.0	0.000000	50.000000	0.0
2906	Get Financed LLC	50.0	0.000000	50.000000	0.0
...	...	...	...	...	...
2423	F&M Capital LLC	0.0	0.000000	100.000000	0.0
2422	F&L Marketing Enterprises LLC	0.0	20.000000	80.000000	0.0
2421	F&B Acquisition Group LLC	0.0	0.000000	100.000000	0.0
2420	F C Tucker Company, Inc	0.0	0.000000	100.000000	0.0
7118	Lippman Recupero, LLC	0.0	13.333333	86.666667	0.0

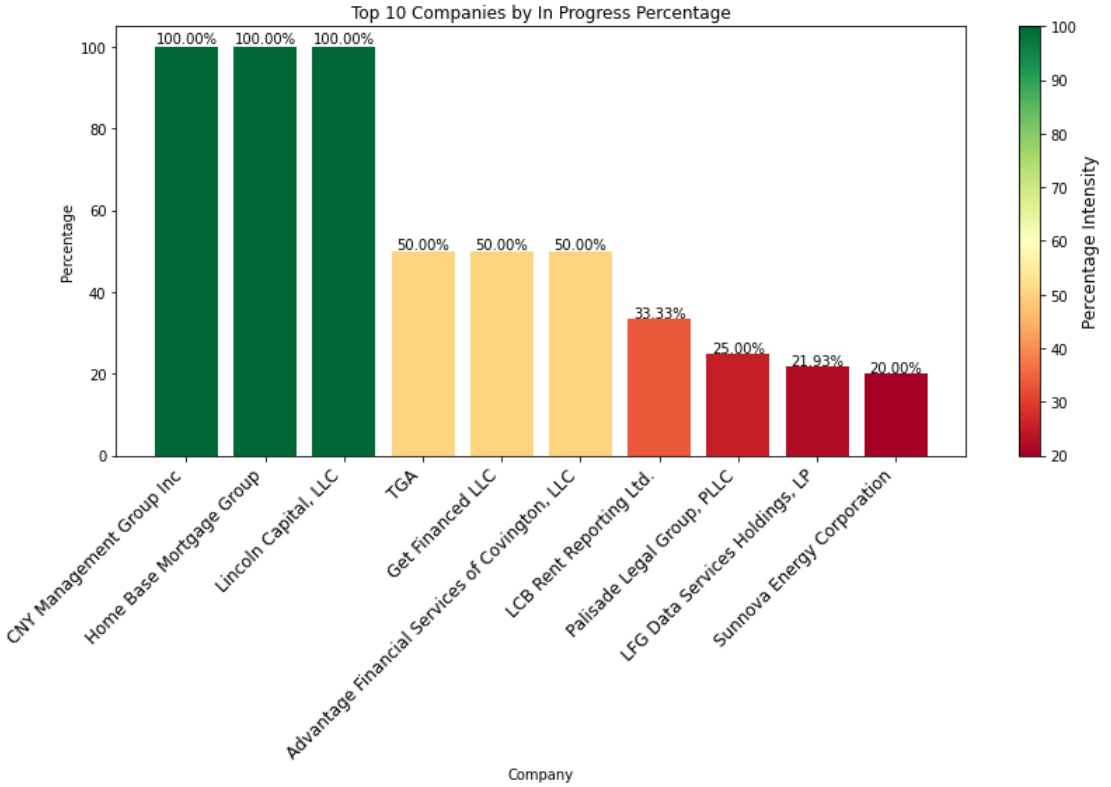
7119 rows × 5 columns

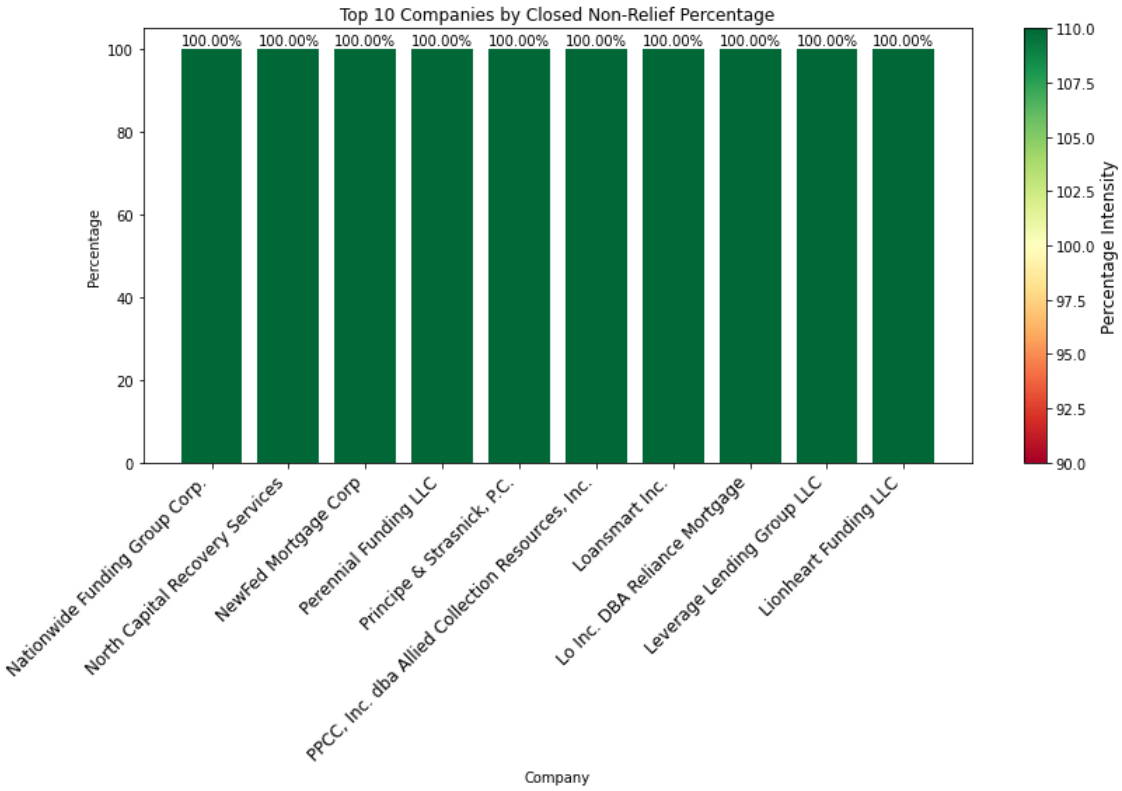
Now we will plot the graph: sorted_response_percentages

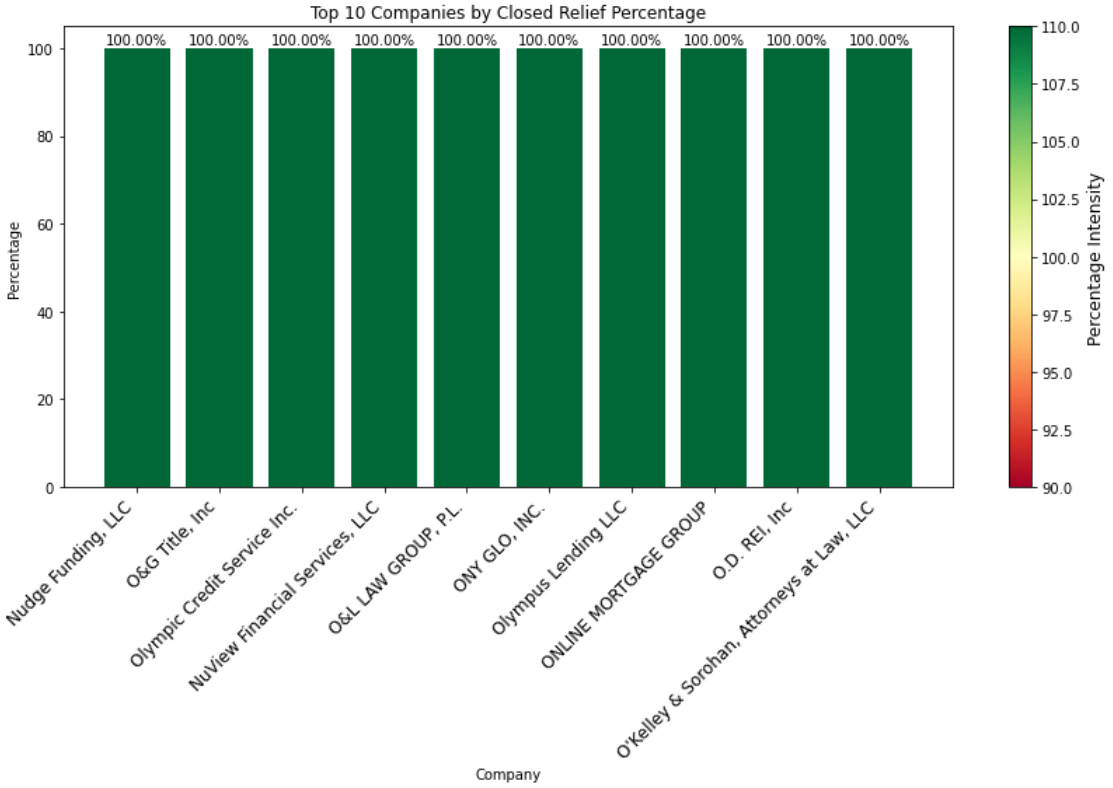
In [ ]:
```
import matplotlib.pyplot as plt
import numpy as np

def plot_top_10_companies_with_style(df, column, title):
    top_10 = df.nlargest(10, column)

    fig, ax = plt.subplots(figsize=(12, 8))  # Adjusted for a better fit for vertical bars
    ax.set_facecolor('white')  # Set background to white
    fig.patch.set_facecolor('white')

    # Use a custom color map from red to green
    cmap = plt.get_cmap('RdYlGn')
    colors = cmap(np.interp(top_10[column], (top_10[column].min(), top_10[column].max()), (0, 1)))

    bars = plt.bar(top_10['Company'], top_10[column], color=colors)

    plt.xticks(rotation=45, ha="right", fontsize=15)  # Rotate company names for clarity

    plt.ylabel('Percentage', color='black')
    plt.xlabel('Company', color='black')
    plt.title(title, color='black')

    ax.tick_params(axis='y', colors='black')
    ax.tick_params(axis='x', colors='black', labelsize=12)  # Make company names a bit larger

    for bar in bars:
        label_y_pos = bar.get_height() + abs(bar.get_height() * 0.01)  # Adjust label position based on bar hei
        ax.text(bar.get_x() + bar.get_width() / 2, label_y_pos, f'{bar.get_height():.2f}%', ha='center', color=

    # Adjust the color bar to match the red to green color scheme
    sm = plt.cm.ScalarMappable(cmap='RdYlGn', norm=plt.Normalize(vmin=min(top_10[column]), vmax=max(top_10[colu
    sm.set_array([])
    cbar = plt.colorbar(sm)
    cbar.ax.tick_params(labelsize=10, colors='black')
    cbar.set_label('Percentage Intensity', color='black', fontsize=12)

    plt.tight_layout()  # Adjust layout to make sure everything fits without overlapping
    plt.show()
```

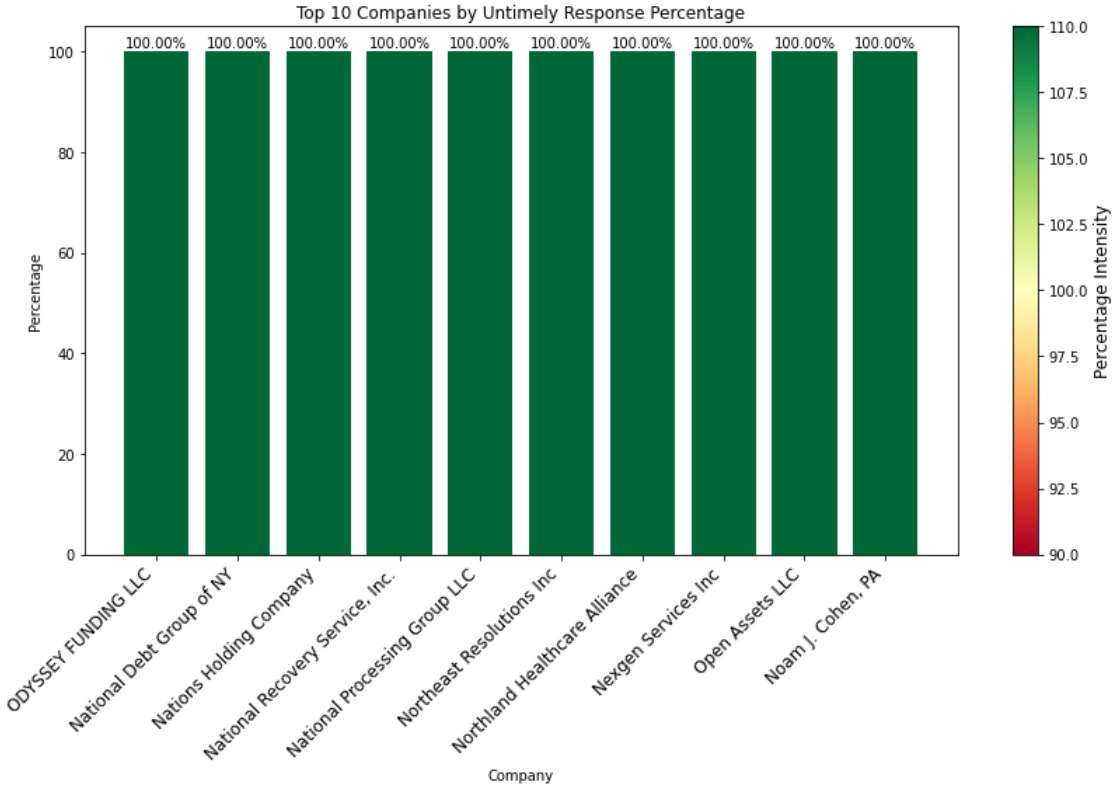
In [ ]:
```
plot_top_10_companies_with_style(sorted_response_percentages, 'In progress percentage', 'Top 10 Companies by In
plot_top_10_companies_with_style(sorted_response_percentages, 'Closed non-relief percentage', 'Top 10 Companies
plot_top_10_companies_with_style(sorted_response_percentages, 'Closed relief percentage', 'Top 10 Companies by
plot_top_10_companies_with_style(sorted_response_percentages, 'Untimely response percentage', 'Top 10 Companies
```

```
<ipython-input-181-14282cfa6da6>:33: MatplotlibDeprecationWarning: Unable to determine Axes to steal space for C
olorbar. Using gca(), but will raise in the future. Either provide the *cax* argument to use as the Axes for the
Colorbar, provide the *ax* argument to steal space from it, or add *mappable* to an Axes.
  cbar = plt.colorbar(sm)
```


Top 10 Companies by In Progress Percentage
Percentage
Company
Percentage Intensity
100.00%
100.00%
100.00%
50.00%
50.00%
50.00%
33.33%
25.00%
21.93%
20.00%
CNY Management Group Inc
Home Base Mortgage Group
Lincoln Capital, LLC
TGA
Get Financed LLC
Advantage Financial Services of Covington, LLC
LCB Rent Reporting Ltd.
Palisade Legal Group, PLLC
LFG Data Services Holdings, LP
Sunnova Energy Corporation


Top 10 Companies by Closed Non-Relief Percentage
Percentage
Company
Percentage Intensity
100.00%
Nationwide Funding Group Corp.
North Capital Recovery Services
NewFed Mortgage Corp
Perennial Funding LLC
Principe & Strasnick, P.C.
PPCC, Inc. dba Allied Collection Resources, Inc.
Loansmart Inc.
Lo Inc. DBA Reliance Mortgage
Leverage Lending Group LLC
Lionheart Funding LLC





Now we are going to calculate the respective percentages of customers who provided consent to the companies:

In [ ]:
```
def calculate_consent_percentages(df):
    total_responses = len(df)
    consent_provided = (df['Consumer consent provided?'] == 'Consent provided').sum() / total_responses * 100
    other = (df['Consumer consent provided?'] == 'Other').sum() / total_responses * 100
    nan_values = df['Consumer consent provided?'].isna().sum() / total_responses * 100
    consent_not_provided = (df['Consumer consent provided?'] == 'Consent not provided').sum() / total_responses
    consent_withdrawn = (df['Consumer consent provided?'] == 'Consent withdrawn').sum() / total_responses * 100
    return pd.Series([consent_provided, other, nan_values, consent_not_provided, consent_withdrawn])
```

In [ ]:
```
consent_percentages = complaints_file.groupby('Company').apply(calculate_consent_percentages).reset_index()
consent_percentages.columns = ['Company', 'Consent provided percentage', 'Other percentage', 'NaN percentage',
sorted_consent_percentages = consent_percentages.sort_values(by='Consent provided percentage', ascending=False)
```

```
sorted_consent_percentages
```

Out[ ]:

	Company	Consent provided percentage	Other percentage	NaN percentage	Consent not provided percentage	Consent withdrawn percentage
3559	KENT MORTGAGE CORPORATION	100.0	0.0	0.0	0.0	0.0
5027	Paragon Mortgage Services, Inc.	100.0	0.0	0.0	0.0	0.0
5037	Parker & Associates	100.0	0.0	0.0	0.0	0.0
5040	Parlanti & Cooperman, LLP	100.0	0.0	0.0	0.0	0.0
1690	Concorde Land Title Services, Inc.	100.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...
4766	O.D. REI, Inc	0.0	0.0	100.0	0.0	0.0
1926	Crossman Portfolio Management	0.0	0.0	100.0	0.0	0.0
4784	ONLINE MORTGAGE GROUP	0.0	0.0	100.0	0.0	0.0
4785	ONTARIO AUTO LENDING	0.0	0.0	100.0	0.0	0.0
3223	Home Lending Pal, Inc.	0.0	100.0	0.0	0.0	0.0

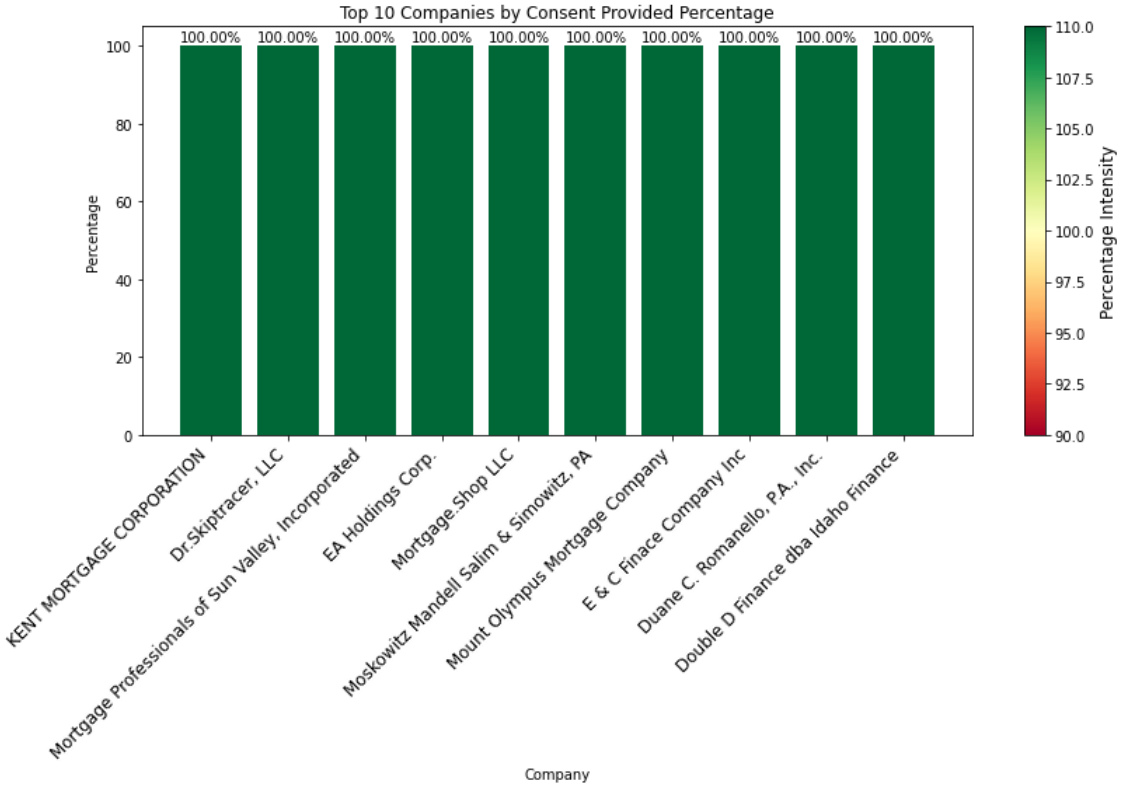
7119 rows × 6 columns

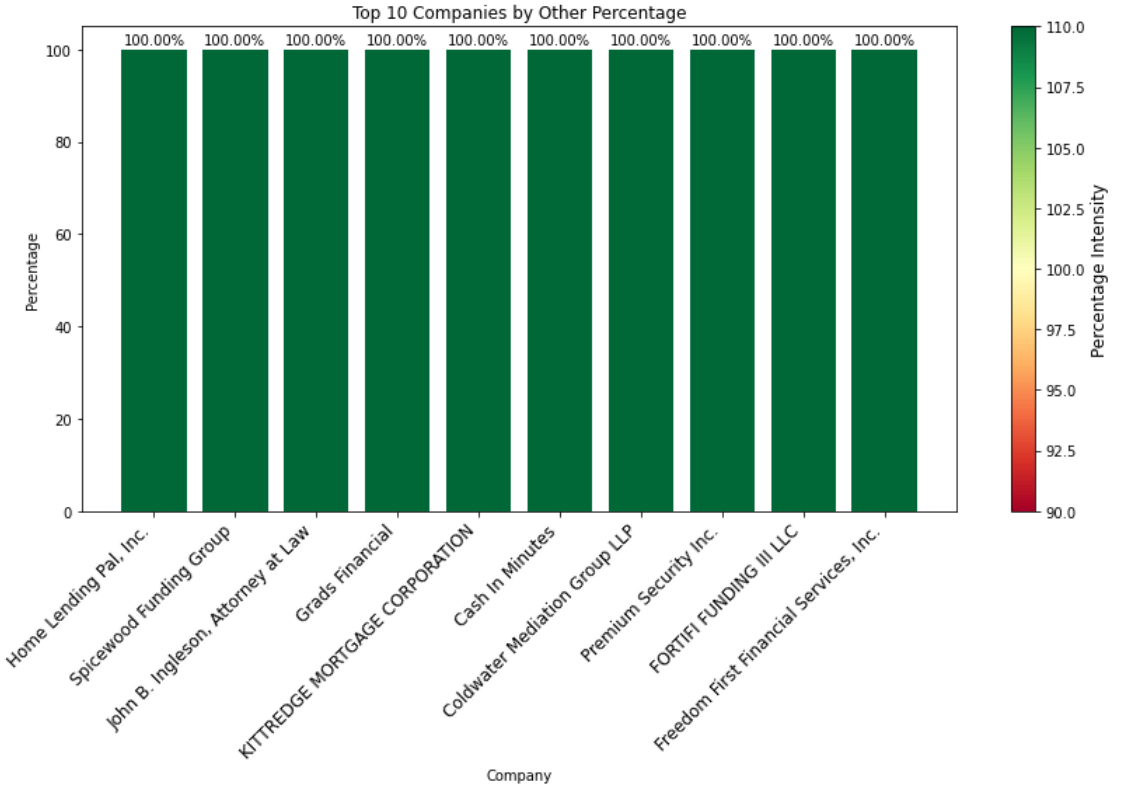
Now we will plot the graph: sorted_consent_percentages

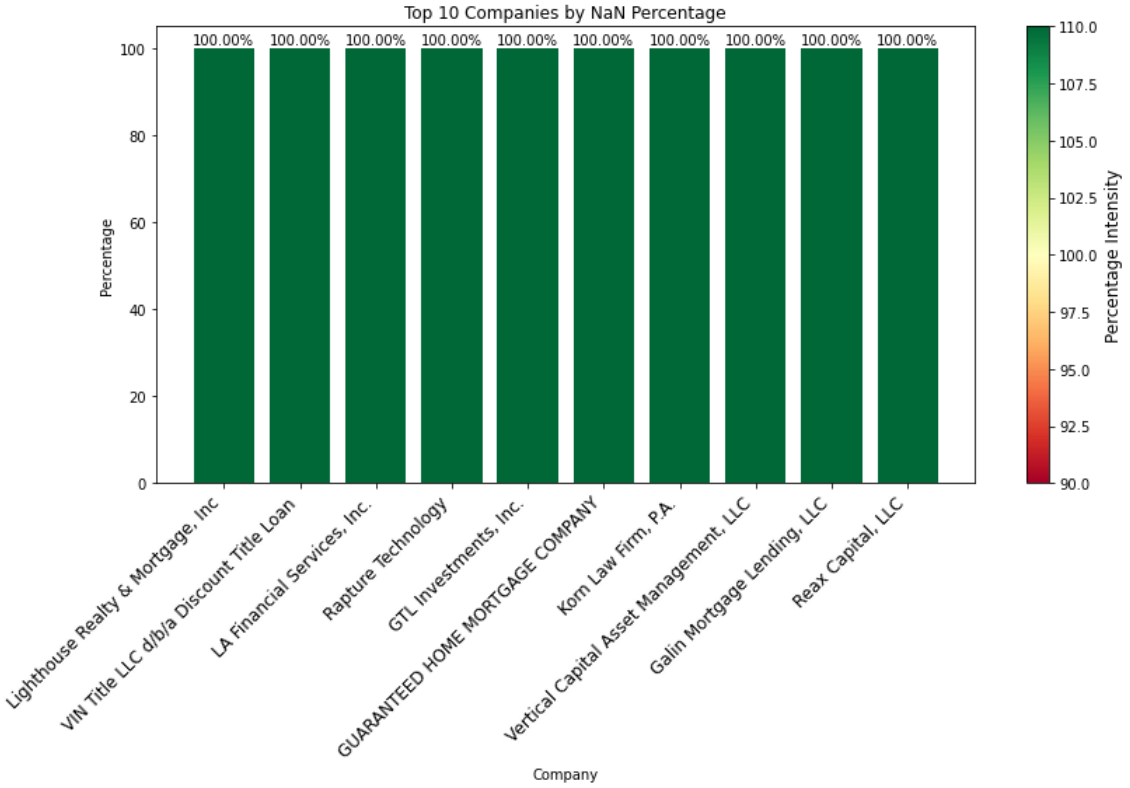
In [ ]:
```
def plot_top_10_companies_consent_with_style(df, column_name, title):
    sorted_df = df.sort_values(by=column_name, ascending=False)
    top_10 = sorted_df.head(10)

    fig, ax = plt.subplots(figsize=(12, 8))
    ax.set_facecolor('white')
    fig.patch.set_facecolor('white')

    cmap = plt.get_cmap('RdYlGn')
    colors = cmap(np.interp(top_10[column_name], (top_10[column_name].min(), top_10[column_name].max()), (0, 1)

    bars = plt.bar(top_10['Company'], top_10[column_name], color=colors)

    plt.xticks(rotation=45, ha="right", fontsize=20)

    plt.ylabel('Percentage', color='black')
    plt.xlabel('Company', color='black')
    plt.title(title, color='black')

    ax.tick_params(axis='y', colors='black')
    ax.tick_params(axis='x', colors='black', labelsize=12)

    for bar in bars:
        label_y_pos = bar.get_height() + abs(bar.get_height() * 0.01)
        ax.text(bar.get_x() + bar.get_width() / 2, label_y_pos, f'{bar.get_height():.2f}%', ha='center', color=

    sm = plt.cm.ScalarMappable(cmap='RdYlGn', norm=plt.Normalize(vmin=min(top_10[column_name]), vmax=max(top_10
    sm.set_array([])
    cbar = plt.colorbar(sm)
    cbar.ax.tick_params(labelsize=10, colors='black')
    cbar.set_label('Percentage Intensity', color='black', fontsize=12)

    plt.tight_layout()
    plt.show()
```

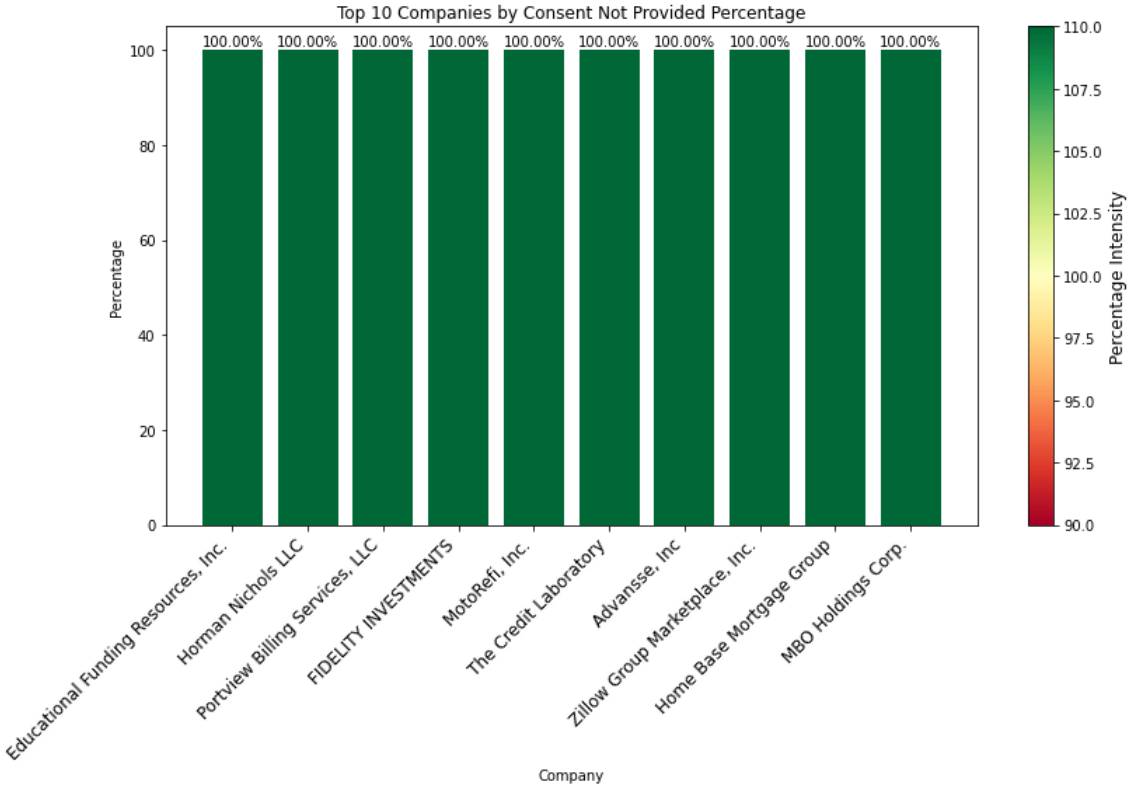
In [ ]:
```
plot_top_10_companies_consent_with_style(sorted_consent_percentages, 'Consent provided percentage', 'Top 10 Com
plot_top_10_companies_consent_with_style(sorted_consent_percentages, 'Other percentage', 'Top 10 Companies by O
plot_top_10_companies_consent_with_style(sorted_consent_percentages, 'NaN percentage', 'Top 10 Companies by NaN
plot_top_10_companies_consent_with_style(sorted_consent_percentages, 'Consent not provided percentage', 'Top 10
plot_top_10_companies_consent_with_style(sorted_consent_percentages, 'Consent withdrawn percentage', 'Top 10 Co
```

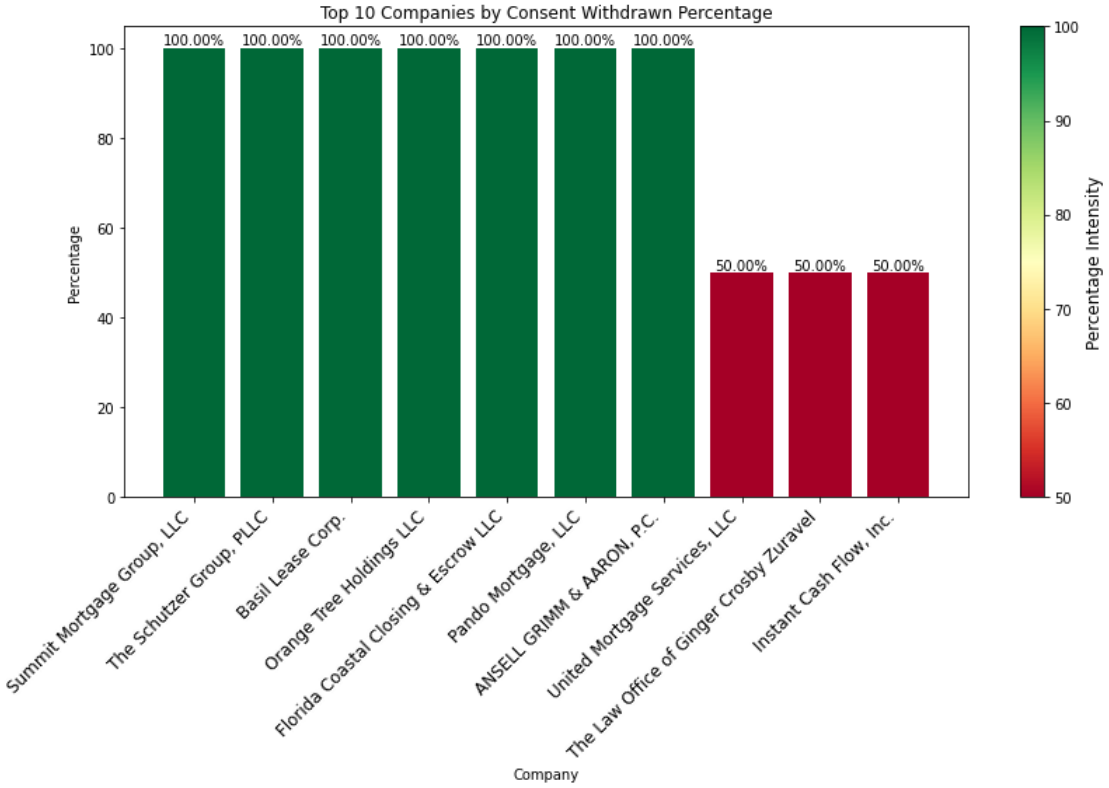
```
<ipython-input-192-ecdd72c075f9>:34: MatplotlibDeprecationWarning: Unable to determine Axes to steal space for C
olorbar. Using gca(), but will raise in the future. Either provide the *cax* argument to use as the Axes for the
Colorbar, provide the *ax* argument to steal space from it, or add *mappable* to an Axes.
  cbar = plt.colorbar(sm)
```


Top 10 Companies by Consent Provided Percentage
100.00%
Percentage
0
20
40
60
80
100
KENT MORTGAGE CORPORATION
Dr.Skiptracer, LLC
Mortgage Professionals of Sun Valley, Incorporated
EA Holdings Corp.
Mortgage Shop LLC
Moskowitz Mandell Salim & Simowitz, PA
Mount Olympus Mortgage Company
E & C Finace Company Inc
Duane C. Romanello, P.A., Inc.
Double D Finance dba Idaho Finance
Company
Percentage Intensity
90.0
92.5
95.0
97.5
100.0
102.5
105.0
107.5
110.0


Top 10 Companies by Other Percentage
100.00%
Percentage
0
20
40
60
80
100
Home Lending Pal, Inc.
Spicewood Funding Group
John B. Ingleson, Attorney at Law
Grads Financial
KITTREDGE MORTGAGE CORPORATION
Cash In Minutes
Coldwater Mediation Group LLP
Premium Security Inc.
FORTIFI FUNDING III LLC
Freedom First Financial Services, Inc.
Company
Percentage Intensity
90.0
92.5
95.0
97.5
100.0
102.5
105.0
107.5
110.0


Top 10 Companies by NaN Percentage
100.00%
Percentage
0
20
40
60
80
100
Lighthouse Realty & Mortgage, Inc
VIN Title LLC d/b/a Discount Title Loan
LA Financial Services, Inc.
Rapture Technology
GTL Investments, Inc.
GUARANTEED HOME MORTGAGE COMPANY
Korn Law Firm, P.A.
Vertical Capital Asset Management, LLC
Galin Mortgage Lending, LLC
Reax Capital, LLC
Company
Percentage Intensity
90.0
92.5
95.0
97.5
100.0
102.5
105.0
107.5
110.0


Top 10 Companies by Consent Not Provided Percentage
100.00%
Percentage
0
20
40
60
80
100
Educational Funding Resources, Inc.
Horman Nichols LLC
Portview Billing Services, LLC
FIDELITY INVESTMENTS
MotoRefi, Inc.
The Credit Laboratory
Advansse, Inc
Zillow Group Marketplace, Inc.
Home Base Mortgage Group
MBO Holdings Corp.
Company
Percentage Intensity
90.0
92.5
95.0
97.5
100.0
102.5
105.0
107.5
110.0



Let's perform data cleaning process

In [ ]:
```
complaints = pd.DataFrame(complaints_file)
complaints.head(5)
```

Out[ ]:

	Date received	Product	Sub-product	Issue	Sub-issue	Consumer complaint narrative	Company public response	Company	State	ZIP code	Tags
0	2023-12-24	Credit reporting or other personal consumer re...	Credit reporting	Incorrect information on your report	Information belongs to someone else	XXXX XXXX XXXX XXXX XXXX XXXX XXXX XXXX X...	Company has responded to the consumer and the ...	Experian Information Solutions Inc.	IN	46168	NaN
1	2024-01-31	Credit reporting or other personal consumer re...	Credit reporting	Problem with a company's investigation into an...	Was not notified of investigation status or re...	NaN	NaN	TRANSUNION INTERMEDIATE HOLDINGS, INC.	NC	28269	NaN
2	2024-01-31	Credit reporting or other personal consumer re...	Credit reporting	Incorrect information on your report	Personal information incorrect	NaN	NaN	EQUIFAX, INC.	LA	708XX	NaN
3	2024-01-31	Debt collection	Credit card debt	Took or threatened to take negative or legal a...	Seized or attempted to seize your property	NaN	NaN	NAVY FEDERAL CREDIT UNION	VA	237XX	Servicemember
4	2024-01-31	Credit reporting or other personal consumer re...	Credit reporting	Incorrect information on your report	Account information incorrect	NaN	NaN	EQUIFAX, INC.	GA	30080	NaN

In [ ]:
```
complaints.columns
```

Out[ ]:
```
Index(['Date received', 'Product', 'Sub-product', 'Issue', 'Sub-issue',
       'Consumer complaint narrative', 'Company public response', 'Company',
       'State', 'ZIP code', 'Tags', 'Consumer consent provided?',
       'Submitted via', 'Date sent to company', 'Company response to consumer',
       'Timely response?', 'Consumer disputed?', 'Complaint ID'],
      dtype='object')
```

In [ ]:
```
# This will remove leading and trailing whitespace (including spaces, tabs, and newline characters) from a stri
complaints['Issue'] = complaints['Issue'].str.strip()
complaints['Sub-issue'] = complaints['Sub-issue'].str.strip()
complaints['Consumer complaint narrative'] = complaints['Consumer complaint narrative'].str.strip()
```

In [ ]:
```
complaints_final = complaints.dropna()
len(complaints_final)

# THIS IS IMPORTANT: It shows that we can't proceed with dropping any records with even a single NaN values as
# Hence we will not proceed with this methodologies
# Quick insight: only 3120 records are there with no NaN values in any columns, ie, clean data.
```

Out[ ]:
```
3120
```

In [ ]:
```
complaints_final.head()
# The cleaned data with 3120 records looks like this, although we will not prefer this data for analysis.
```

Out[ ]:

	Date received	Product	Sub-product	Issue	Sub-issue	Consumer complaint narrative	Company public response	Company	State	Z coc
374786	2016-11-01	Debt collection	Medical	Cont'd attempts collect debt not owed	Debt was paid	I am being brought to court for a debt that ha...	Company disputes the facts presented in the co...	Overton, Russell, Doerr and Donovan, LLP	NY	125X
459728	2016-09-09	Debt collection	I do not know	Cont'd attempts collect debt not owed	Debt is not mine	I have never had a bank account with Wells Far...	Company has responded to the consumer and the ...	WELLS FARGO & COMPANY	GA	XXXX
566014	2017-01-17	Debt collection	Credit card	Cont'd attempts collect debt not owed	Debt resulted from identity theft	We completed a previous dispute under XXXX. \n...	Company believes complaint is the result of an...	UNITED SERVICES AUTOMOBILE ASSOCIATION	NC	2725
590723	2017-02-17	Debt collection	Credit card	Cont'd attempts collect debt not owed	Debt is not mine	Letter from First National Collection Bureau, ...	Company believes it acted appropriately as aut...	First National Collection Bureau, Inc.	CA	920X
590745	2017-04-08	Debt collection	Credit card	Communication tactics	Threatened to take legal action	departmentReceived phone call from representat...	Company believes it acted appropriately as aut...	Monterey Financial Services LLC	NC	274(

AIM: Filter the data based on three column making sure that we are focusing on important issues and subissues for our analysis

In [ ]:
```
filtered_complaints = complaints.loc[
    complaints['Company public response'].isin([
        "Company can't verify or dispute the facts in the complaint",
        'Company chooses not to provide a public response',
        'Company believes complaint is the result of an isolated error',
        'Company believes complaint represents an opportunity for improvement to better serve consumers',
        "Company believes the complaint provided an opportunity to answer consumer's questions"
    ]) |
    complaints['Company response to consumer'].isin([
        'In progress',
        'Untimely response'
    ])
    |
    complaints['Consumer disputed?'].isin([
        'Yes'
    ])
]
```

```
len(filtered_complaints)
#Here is the total nuumber of records in the cleaned dataset
```

Out[ ]: 402666

In [ ]:
```
filtered_complaints.head()
```

Out[ ]:

	Date received	Product	Sub-product	Issue	Sub-issue	Consumer complaint narrative	Company public response	Company	State	ZIP code	Tags
1	2024-01-31	Credit reporting or other personal consumer re...	Credit reporting	Problem with a company's investigation into an...	Was not notified of investigation status or re...	NaN	NaN	TRANSUNION INTERMEDIATE HOLDINGS, INC.	NC	28269	NaN
2	2024-01-31	Credit reporting or other personal consumer re...	Credit reporting	Incorrect information on your report	Personal information incorrect	NaN	NaN	EQUIFAX, INC.	LA	708XX	NaN
3	2024-01-31	Debt collection	Credit card debt	Took or threatened to take negative or legal a...	Seized or attempted to seize your property	NaN	NaN	NAVY FEDERAL CREDIT UNION	VA	237XX	Servicemember
4	2024-01-31	Credit reporting or other personal consumer re...	Credit reporting	Incorrect information on your report	Account information incorrect	NaN	NaN	EQUIFAX, INC.	GA	30080	NaN
5	2024-01-31	Credit reporting or other personal consumer re...	Credit reporting	Incorrect information on your report	Information belongs to someone else	NaN	NaN	EQUIFAX, INC.	IL	60651	NaN

Now that we have the cleaned and structured data, we can now proceed with the semantic sentiment analysisn of the three textual columns (Issue, Sub-isue, Consumer complaint narrative) columns using different methods

1. TextBlob

In [ ]:
```
pip install TextBlob
```

```
Requirement already satisfied: TextBlob in /Users/sarthak/opt/anaconda3/lib/python3.8/site-packages (0.18.0.post
0)
Requirement already satisfied: nltk>=3.8 in /Users/sarthak/opt/anaconda3/lib/python3.8/site-packages (from TextB
lob) (3.8.1)
Requirement already satisfied: click in /Users/sarthak/opt/anaconda3/lib/python3.8/site-packages (from nltk>=3.8
->TextBlob) (8.1.7)
Requirement already satisfied: joblib in /Users/sarthak/opt/anaconda3/lib/python3.8/site-packages (from nltk>=3.
8->TextBlob) (0.17.0)
Requirement already satisfied: regex>=2021.8.3 in /Users/sarthak/opt/anaconda3/lib/python3.8/site-packages (from
nltk>=3.8->TextBlob) (2023.12.25)
Requirement already satisfied: tqdm in /Users/sarthak/opt/anaconda3/lib/python3.8/site-packages (from nltk>=3.8-
>TextBlob) (4.66.1)
DEPRECATION: pyodbc 4.0.0-unsupported has a non-standard version number. pip 24.1 will enforce this behaviour ch
ange. A possible replacement is to upgrade to a newer version of pyodbc or contact the author to suggest that th
ey release a version with a conforming version number. Discussion can be found at https://github.com/pypa/pip/is
sues/12063
Note: you may need to restart the kernel to use updated packages.
```

In [ ]:
```
# Sample example
text = "I loved working in the take home asignment for LinkedIn"
blob = TextBlob(text)
sentiment = blob.sentiment
print("Polarity:", sentiment.polarity)
```

```
Polarity: 0.7
```

In [ ]:
```
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
```

```
from collections import Counter
from textblob import TextBlob
```

In [ ]:
```
def calculate_sentiment(text):
    try:
        return TextBlob(text).sentiment.polarity
    except:
        return None
```

In [ ]:
```
filtered_complaints['issue_sentiment'] = filtered_complaints["Issue"].apply(calculate_sentiment)
filtered_complaints['sub_issue_sentiment'] = filtered_complaints["Sub-issue"].apply(calculate_sentiment)
filtered_complaints['consumer_complaint_narrative_sentiment'] = filtered_complaints["Consumer complaint narrati

filtered_complaints.head(5)

# You will see that the scores are populated for the three columns mentioned above and it has aded three new co
```

```
<ipython-input-44-82d04271a676>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#retu
rning-a-view-versus-a-copy
  filtered_complaints['issue_sentiment'] = filtered_complaints["Issue"].apply(calculate_sentiment)
<ipython-input-44-82d04271a676>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#retu
rning-a-view-versus-a-copy
  filtered_complaints['sub_issue_sentiment'] = filtered_complaints["Sub-issue"].apply(calculate_sentiment)
<ipython-input-44-82d04271a676>:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#retu
rning-a-view-versus-a-copy
  filtered_complaints['consumer_complaint_narrative_sentiment'] = filtered_complaints["Consumer complaint narrat
ive"].apply(calculate_sentiment)
```

Out[ ]:

	Date received	Product	Sub-product	Issue	Sub-issue	Consumer complaint narrative	Company public response	Company	State	ZIP code	...	Consumer consent provided?
1	2024-01-31	Credit reporting or other personal consumer re...	Credit reporting	Problem with a company's investigation into an...	Was not notified of investigation status or re...	NaN	NaN	TRANSUNION INTERMEDIATE HOLDINGS, INC.	NC	28269	...	Other
2	2024-01-31	Credit reporting or other personal consumer re...	Credit reporting	Incorrect information on your report	Personal information incorrect	NaN	NaN	EQUIFAX, INC.	LA	708XX	...	Other
3	2024-01-31	Debt collection	Credit card debt	Took or threatened to take negative or legal a...	Seized or attempted to seize your property	NaN	NaN	NAVY FEDERAL CREDIT UNION	VA	237XX	...	NaN
4	2024-01-31	Credit reporting or other personal consumer re...	Credit reporting	Incorrect information on your report	Account information incorrect	NaN	NaN	EQUIFAX, INC.	GA	30080	...	Other
5	2024-01-31	Credit reporting or other personal consumer re...	Credit reporting	Incorrect information on your report	Information belongs to someone else	NaN	NaN	EQUIFAX, INC.	IL	60651	...	NaN

5 rows × 21 columns

2. VADER sentiment score

In [ ]:
```
pip install nltk
```

```
Requirement already satisfied: nltk in /Users/sarthak/opt/anaconda3/lib/python3.8/site-packages (3.8.1)
Requirement already satisfied: click in /Users/sarthak/opt/anaconda3/lib/python3.8/site-packages (from nltk) (8.
1.7)
Requirement already satisfied: joblib in /Users/sarthak/opt/anaconda3/lib/python3.8/site-packages (from nltk)
(0.17.0)
Requirement already satisfied: regex>=2021.8.3 in /Users/sarthak/opt/anaconda3/lib/python3.8/site-packages (from
nltk) (2023.12.25)
Requirement already satisfied: tqdm in /Users/sarthak/opt/anaconda3/lib/python3.8/site-packages (from nltk) (4.6
6.1)
DEPRECATION: pyodbc 4.0.0-unsupported has a non-standard version number. pip 24.1 will enforce this behaviour ch
ange. A possible replacement is to upgrade to a newer version of pyodbc or contact the author to suggest that th
ey release a version with a conforming version number. Discussion can be found at https://github.com/pypa/pip/is
sues/12063
Note: you may need to restart the kernel to use updated packages.
```

In [ ]:
```
import nltk
nltk.download('vader_lexicon')
from nltk.sentiment import SentimentIntensityAnalyzer
sia = SentimentIntensityAnalyzer()
```

```
[nltk_data] Downloading package vader_lexicon to
[nltk_data]     /Users/sarthak/nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!
```

In [ ]:
```
# Sample example (Positive Sentiment)
sia.polarity_scores('I am elated to interview for LinkedIn')
```

Out[ ]:
```
{'neg': 0.0, 'neu': 0.543, 'pos': 0.457, 'compound': 0.6369}
```

In [ ]:
```
# Sample example (Neutral Sentiment)
sia.polarity_scores('This assignment was really challenging !!')
```

Out[ ]:
```
{'neg': 0.0, 'neu': 0.669, 'pos': 0.331, 'compound': 0.3563}
```

In [ ]:
```
# Sample example (Negative Sentiment)

sia.polarity_scores('This assignment was really tough !!')
```

Out[ ]:
```
{'neg': 0.322, 'neu': 0.678, 'pos': 0.0, 'compound': -0.335}
```

In [ ]:
```
import concurrent.futures
```

In [ ]:
```
def get_polarity_scores(text):
    if pd.isnull(text):  # Check if the text is NaN
        return None
    score = sia.polarity_scores(text)
    return score['compound']
```

In [ ]:
```
def apply_parallel(df, column):
    with concurrent.futures.ProcessPoolExecutor() as executor:
        return list(executor.map(get_polarity_scores, df[column]))
```

In [ ]:
```
filtered_complaints['Issue'].apply(lambda x: sia.polarity_scores(x))
```

Out[ ]:
```
1          {'neg': 0.474, 'neu': 0.526, 'pos': 0.0, 'comp...
2          {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound...
3          {'neg': 0.472, 'neu': 0.423, 'pos': 0.106, 'co...
4          {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound...
5          {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound...
                                ...
4750041    {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound...
4750094    {'neg': 0.333, 'neu': 0.667, 'pos': 0.0, 'comp...
4750165    {'neg': 0.333, 'neu': 0.667, 'pos': 0.0, 'comp...
4750359    {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound...
4750475    {'neg': 0.474, 'neu': 0.526, 'pos': 0.0, 'comp...
Name: Issue, Length: 402666, dtype: object
```

In [ ]:
```
filtered_complaints['Issue_PolarityScore_Sentiment'] = filtered_complaints['Issue'].apply(lambda x: sia.polarit
filtered_complaints['SubIssue_PolarityScore_Sentiment'] = filtered_complaints['Sub-issue'].apply(lambda x: sia.
filtered_complaints['ConsumerComplaintNarrative_PolarityScore_Sentiment'] = filtered_complaints['Consumer compl

filtered_complaints.head(5)
```

```
<ipython-input-54-90c4dc948cf2>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  filtered_complaints['Issue_PolarityScore_Sentiment'] = filtered_complaints['Issue'].apply(lambda x: sia.polarity_scores(str(x))['compound'] if pd.notnull(x) else 0)
<ipython-input-54-90c4dc948cf2>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  filtered_complaints['SubIssue_PolarityScore_Sentiment'] = filtered_complaints['Sub-issue'].apply(lambda x: sia.polarity_scores(str(x))['compound'] if pd.notnull(x) else 0)
<ipython-input-54-90c4dc948cf2>:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  filtered_complaints['ConsumerComplaintNarrative_PolarityScore_Sentiment'] = filtered_complaints['Consumer complaint narrative'].apply(lambda x: sia.polarity_scores(str(x))['compound'] if pd.notnull(x) else 0)
```

Out[ ]:

	Date received	Product	Sub-product	Issue	Sub-issue	Consumer complaint narrative	Company public response	Company	State	ZIP code	...	Company response to consumer
1	2024-01-31	Credit reporting or other personal consumer re...	Credit reporting	Problem with a company's investigation into an...	Was not notified of investigation status or re...	NaN	NaN	TRANSUNION INTERMEDIATE HOLDINGS, INC.	NC	28269	...	In progress
2	2024-01-31	Credit reporting or other personal consumer re...	Credit reporting	Incorrect information on your report	Personal information incorrect	NaN	NaN	EQUIFAX, INC.	LA	708XX	...	In progress
3	2024-01-31	Debt collection	Credit card debt	Took or threatened to take negative or legal a...	Seized or attempted to seize your property	NaN	NaN	NAVY FEDERAL CREDIT UNION	VA	237XX	...	In progress
4	2024-01-31	Credit reporting or other personal consumer re...	Credit reporting	Incorrect information on your report	Account information incorrect	NaN	NaN	EQUIFAX, INC.	GA	30080	...	In progress
5	2024-01-31	Credit reporting or other personal consumer re...	Credit reporting	Incorrect information on your report	Information belongs to someone else	NaN	NaN	EQUIFAX, INC.	IL	60651	...	In progress

5 rows × 24 columns

Insights By Company

a. Issue

We are going to find top companies (worst/best performing) in terms of dealing with issues, that was captured in both the NLP techniques

In [ ]:
```
Issue_textblob_average_scores_company = filtered_complaints[filtered_complaints['issue_sentiment'] != 0].groupb
Issue_textblob_average_scores_company_sorted = Issue_textblob_average_scores_company.sort_values(ascending=True
Issue_textblob_average_scores_company_sorted.head(10)
```

Out[ ]:
```
Company
Friendly Finance Corporation                          -0.5
National Consumer Telecom & Utilities Exchange, Inc.  -0.5
Go Capital Holdings, LLC                              -0.5
Turner Acceptance Corp.                               -0.5
LoanNow, LLC  (Closed)                                -0.5
LOAN TO LEARN                                         -0.5
OKCoin USA Inc.                                       -0.5
PROSPERITY BANCSHARES, INC.                           -0.5
Regional Management Corporation                       -0.5
Glennon Law Firm, LLC                                 -0.5
Name: issue_sentiment, dtype: float64
```

In [ ]:
```
Issue_textblob_average_scores_company_sorted.tail(10)
```

Out[ ]:
```
Company
Gateway Recovery Solutions              9.251859e-18
Select Resource Group                   9.251859e-18
Jeffrey R. Lankenau, Attorney at Law    9.251859e-18
Boone & Associates Group LLC            9.251859e-18
Strategic Financial Solutions, LLC      6.666667e-02
National Capital Solutions              1.000000e-01
Saint Services LLC                      1.000000e-01
BID SOLUTIONS LLC                       1.000000e-01
CALM, Inc.                              1.000000e-01
Michel Law, LLC d/b/a Level One Law     1.000000e-01
Name: issue_sentiment, dtype: float64
```

In [ ]:
```
Issue_textblob_average_scores_company_vader = filtered_complaints[filtered_complaints['Issue_PolarityScore_Sent
Issue_textblob_average_scores_company_sorted_vader = Issue_textblob_average_scores_company_vader.sort_values(as
Issue_textblob_average_scores_company_sorted_vader.head(10)
```

Out[ ]:
```
Company
BELING & ASSOCIATES                -0.8176
Fast Track Servicing               -0.8176
THE STUDENT LOAN HELP CENTER       -0.8176
Apple Law Group, Inc.              -0.8176
Relay Financial (US), Corp.        -0.8176
Legacy Credits                     -0.8176
Trinity Enterprises Partners, LLC  -0.8176
BANKERS HEALTHCARE GROUP LLC       -0.8176
BAM Management US Holdings Inc.    -0.8176
National Settlement Services, Inc  -0.8176
Name: Issue_PolarityScore_Sentiment, dtype: float64
```

In [ ]:
```
Issue_textblob_average_scores_company_sorted.tail(10)
```

Out[ ]:
```
Company
Dwight Financial, Inc.            0.421500
CONTRACT RESOLVE GROUP LLC        0.421500
Verisk Analytics Inc              0.424591
One Technologies, LP              0.432820
Hertzmark, Crean & Lahey, LLP.    0.458800
CITY NATIONAL BANK                0.476700
A-L FINANCIAL CORP                0.636900
Gabriel FInancial Group, Inc.     0.636900
Neighborhood Autos, LLC           0.636900
AMERICAN FINANCING CORPORATION    0.636900
Name: Issue_PolarityScore_Sentiment, dtype: float64
```

In [ ]:
```
top_180_textblob_companies = Issue_textblob_average_scores_company_sorted.head(180).index.tolist()
top_180_vader_companies = Issue_textblob_average_scores_company_sorted_vader.head(180).index.tolist()
common_companies_in_top_180 = set(top_180_textblob_companies) & set(top_180_vader_companies)
common_companies_in_top_180
```

Out[ ]:
```
{'Coldwater Mediation Group LLP',
 'Money Service Centers of Hawaii',
 'RABOBANK, NATIONAL ASSOCIATION',
 'The Money Company'}
```

In [ ]:
```
top_100_textblob_companies = Issue_textblob_average_scores_company_sorted.head(100).index.tolist()
top_100_vader_companies = Issue_textblob_average_scores_company_sorted_vader.head(100).index.tolist()
common_companies_in_top_100 = set(top_100_textblob_companies) & set(top_100_vader_companies)
common_companies_in_top_100
```

Out[ ]:
```
{'Money Service Centers of Hawaii'}
```

In [ ]:
```
bottom_230_textblob_companies = Issue_textblob_average_scores_company_sorted.tail(230).index.tolist()
bottom_230_vader_companies = Issue_textblob_average_scores_company_sorted_vader.tail(230).index.tolist()
common_companies_in_bottom_230 = set(bottom_230_textblob_companies) & set(bottom_230_vader_companies)
common_companies_in_bottom_230
```

Out[ ]:
```
{'Advantage One Credit, LLC',
 'IMC Capital, LLC',
 'Money Tree Lending',
 'Southwest Business Corporation'}
```

In [ ]:
```
bottom_170_textblob_companies = Issue_textblob_average_scores_company_sorted.tail(170).index.tolist()
bottom_170_vader_companies = Issue_textblob_average_scores_company_sorted_vader.tail(170).index.tolist()
common_companies_in_bottom_170 = set(bottom_170_textblob_companies) & set(bottom_170_vader_companies)
common_companies_in_bottom_170
```

Out[ ]:
```
{'Money Tree Lending'}
```

b. Sub-Issue

We are going to find top companies (worst/best performing) in terms of dealing with issues, that was captured in both the NLP techniques

In [ ]:
```
filtered_complaints_non_zero = filtered_complaints[(filtered_complaints['sub_issue_sentiment'] != 0) & (~filter
subissue_textblob_average_scores_company = filtered_complaints_non_zero.groupby('Company')['sub_issue_sentiment
subissue_textblob_average_scores_company_sorted = subissue_textblob_average_scores_company.sort_values(ascendin
subissue_textblob_average_scores_company_sorted.head(10)
```

Out[ ]:
```
Company
Apex Processing                        -0.7
Strada Education Network, Inc.         -0.7
Rhode Island Student Loan Authority    -0.7
Express Enrollment LLC                 -0.7
Brelvis Consulting, LLC                -0.7
Student Loan Discharge Options LLC     -0.7
Student Loan Freedom, LLC              -0.7
Student Loan Care, LLC                 -0.7
Federal Student Loan Doc Prep LLC      -0.7
Minnesota Office of Higher Education   -0.7
Name: sub_issue_sentiment, dtype: float64
```

In [ ]:
```
subissue_textblob_average_scores_company_sorted.tail(10)
```

Out[ ]:
```
Company
Auto Buyers Credit LLC                          0.5
IPAC'S Inc.                                     0.5
Executive Acquisitions Group, LLC               0.5
Goldcar Lending Inc.                            0.5
Portfolio Investment Exchange, Inc.             0.5
Ascendium Education Group                       0.5
Rental Kharma, Inc                              0.5
Sterling Credit Corp.                           0.5
Expedite Financial, Inc                         0.5
Residential Mortgage Services Holdings, Inc.    0.5
Name: sub_issue_sentiment, dtype: float64
```

In [ ]:
```
filtered_complaints_non_zero_polarity = filtered_complaints[(filtered_complaints['SubIssue_PolarityScore_Sentim
subissue_polarity_average_scores_company = filtered_complaints_non_zero_polarity.groupby('Company')['SubIssue_P
subissue_polarity_average_scores_company_sorted = subissue_polarity_average_scores_company.sort_values(ascendin
subissue_polarity_average_scores_company_sorted.head(10)
```

Out[ ]:
```
Company
PAY-O-MATIC                                                -0.8402
Accounts Interchange Group LLC                             -0.8402
Debt Direct Portfolio Management, LLC                      -0.8402
First Direct Mediation                                     -0.8402
Central Atlantic Sales Inc. T/A Priority Buy Here Pay Here -0.8402
Lacy Katzen LLP                                            -0.8402
RUPP AND ASSOCIATES, INC.                                  -0.8402
Smith & Smith Investors LTD                                -0.7783
Presto Loan Center, LLC                                    -0.7783
Best Loan Service, LLC.                                    -0.7783
Name: SubIssue_PolarityScore_Sentiment, dtype: float64
```

In [ ]:
```
subissue_polarity_average_scores_company_sorted.tail(10)
```

Out[ ]:
```
Company
LUMINATE HOME LOANS, INC.        0.3818
LL Services Parent LLC           0.3818
LHM Financial                    0.3818
First Help Financial, LLC        0.3818
MEDALLION MORTGAGE COMPANY       0.3818
First Premier Financial, Inc.    0.3968
Block, Inc.                      0.3968
Earnest Inc.                     0.4404
NATIONAL ASSET MORTGAGE, LLC     0.4588
Gabriel FInancial Group, Inc.    0.4939
Name: SubIssue_PolarityScore_Sentiment, dtype: float64
```

In [ ]:
```
top_112_textblob_companies_subissue = subissue_textblob_average_scores_company_sorted.head(112).index.tolist()
top_112_vader_companies_subissue = subissue_polarity_average_scores_company_sorted.head(112).index.tolist()
common_companies_in_top_112_subissue = set(top_112_textblob_companies_subissue) & set(top_112_vader_companies_s
common_companies_in_top_112_subissue
```

Out[ ]:
```
{'BANK OF HAWAII CORPORATION',
 'CATALYST LENDING, INC.',
 'Grace Period, Inc.',
 'Greyhound Capital Group LLC',
 'LICTY, INC.'}
```

In [ ]:
```
top_100_textblob_companies_subissue = subissue_textblob_average_scores_company_sorted.head(100).index.tolist()
top_100_vader_companies_subissue = subissue_polarity_average_scores_company_sorted.head(100).index.tolist()
common_companies_in_top_100_subissue = set(top_100_textblob_companies_subissue) & set(top_100_vader_companies_s
common_companies_in_top_100_subissue
```

Out[ ]:
```
{'BANK OF HAWAII CORPORATION', 'Grace Period, Inc.'}
```

In [ ]:
```
bottom_60_textblob_companies_subissue = subissue_textblob_average_scores_company_sorted.tail(60).index.tolist()
bottom_60_vader_companies_subissue = subissue_polarity_average_scores_company_sorted.tail(60).index.tolist()
common_companies_in_bottom_60_subissue = set(bottom_60_textblob_companies_subissue) & set(bottom_60_vader_compa
common_companies_in_bottom_60_subissue
```

Out[ ]:
```
{'ACRANET INC',
 'Earnest Inc.',
 'Gabriel FInancial Group, Inc.',
 'RUOFF MORTGAGE'}
```

In [ ]:
```
bottom_50_textblob_companies_subissue = subissue_textblob_average_scores_company_sorted.tail(50).index.tolist()
bottom_50_vader_companies_subissue = subissue_polarity_average_scores_company_sorted.tail(50).index.tolist()
common_companies_in_bottom_50_subissue = set(bottom_50_textblob_companies_subissue) & set(bottom_50_vader_compa
common_companies_in_bottom_50_subissue
```

Out[ ]:
```
{'Earnest Inc.', 'Gabriel FInancial Group, Inc.'}
```

c. Consumer Complaint Narrative

We are going to find top companies (worst/best performing) in terms of dealing with issues, that was captured in both the NLP techniques

In [ ]:
```
filtered_complaints_non_zero_consumer = filtered_complaints[(filtered_complaints['consumer_complaint_narrative_
subissue_textblob_average_scores_company_consumer = filtered_complaints_non_zero_consumer.groupby('Company')['c
subissue_textblob_average_scores_company_sorted_consumer = subissue_textblob_average_scores_company_consumer.so
subissue_textblob_average_scores_company_sorted_consumer.head(10)
```

Out[ ]:
```
Company
Newman, Hesse & Associates, P. A.   -0.850000
Allarin Professional Group LLC      -0.714286
Capital Recovery, LLC (Closed)      -0.600000
The Shindler Law Firm               -0.500000
The Judgment Group                  -0.500000
Woodward Capital LLC                -0.500000
Granite Bay Acceptance Inc.         -0.500000
Affiliated Acceptance Corporation   -0.500000
Stonebridge Lending Inc.            -0.500000
Partridge Snow & Hahn LLP           -0.483333
Name: consumer_complaint_narrative_sentiment, dtype: float64
```

In [ ]:
```
subissue_textblob_average_scores_company_sorted_consumer.tail(10)
```

Out[ ]:
```
Company
ADVANCE MORTGAGE & INVESTMENT                0.441667
Ray Skillman Ford Inc                        0.492857
Filaport Acquisitions                        0.500000
Gotham Collection Services Corp.             0.500000
Greater Collection Agency Bureau, LLC        0.500000
The Best Credit Group                        0.500000
Open Dealer Exchange, LLC                    0.500000
CrediautoUSA Financial Company LLC           0.509189
MASON MCDUFFIE MORTGAGE CORPORATION          0.575000
Clarfield, Okon, Salomone and Pincus, P.L.   0.700000
Name: consumer_complaint_narrative_sentiment, dtype: float64
```

In [ ]:
```
filtered_complaints_non_zero_consumer_polarity = filtered_complaints[(filtered_complaints['ConsumerComplaintNar
polarity_average_scores_company_consumer = filtered_complaints_non_zero_consumer_polarity.groupby('Company')['C
polarity_average_scores_company_sorted_consumer = polarity_average_scores_company_consumer.sort_values(ascendin
polarity_average_scores_company_sorted_consumer.head(10)
```

Out[ ]:
```
Company
National Payment Solutions of New York LLC   -0.9981
PRESTAMOS RAPIDITOS                          -0.9971
Credit Bureau of Jonesboro, Inc.             -0.9969
Credit Resolutions, LLC                      -0.9968
Amplicon LLC                                 -0.9958
Bayview Investment Group LLC                 -0.9956
The Accounts Retrievable System, Inc.        -0.9950
Sullivan & Terranova                         -0.9950
Indie Technology DBA Found                   -0.9942
WCS LENDING, LLC                             -0.9933
Name: ConsumerComplaintNarrative_PolarityScore_Sentiment, dtype: float64
```

In [ ]:
```
polarity_average_scores_company_sorted_consumer.tail(10)
```

Out[ ]:
```
Company
PINNACLE FINANCIAL PARTNERS, INC.    0.9942
FIRST CONSUMER, LLC                  0.9946
First Texas Auto Credit, Inc.        0.9948
Point Boosters                       0.9968
A-L FINANCIAL CORP                   0.9968
Klarna AB                            0.9970
Great State Mortgage, LLC            0.9977
Udren Law Offices, P.C.              0.9977
AMALGAMATED TOKEN SERVICES, INC.     0.9989
Bright Capital Inc                   0.9989
Name: ConsumerComplaintNarrative_PolarityScore_Sentiment, dtype: float64
```

In [ ]:
```
top_100_textblob_companies_subissue = subissue_textblob_average_scores_company_sorted_consumer.head(100).index.
top_100_vader_companies_subissue = polarity_average_scores_company_sorted_consumer.head(100).index.tolist()
common_companies_in_top_100_subissue = set(top_100_textblob_companies_subissue) & set(top_100_vader_companies_s
common_companies_in_top_100_subissue
```

Out[ ]:
```
{'Account Liquidation Services Inc.',
 'Harbor Front Acquisitions LLC (Closed)',
 'Michael Haynes & Associates, LLC',
 'Partridge Snow & Hahn LLP',
 'Skrill USA, Inc.',
 'The Law Offices of Gerald E Moore & Associates, PC'}
```

In [ ]:
```
top_50_textblob_companies_subissue = subissue_textblob_average_scores_company_sorted_consumer.head(50).index.to
top_50_vader_companies_subissue = polarity_average_scores_company_sorted_consumer.head(50).index.tolist()
common_companies_in_top_50_subissue = set(top_50_textblob_companies_subissue) & set(top_50_vader_companies_subi
common_companies_in_top_50_subissue
```

Out[ ]:
```
{'Partridge Snow & Hahn LLP'}
```

In [ ]:
```
bottom_80_textblob_companies_subissue = subissue_textblob_average_scores_company_sorted_consumer.tail(70).index
bottom_100_vader_companies_subissue = polarity_average_scores_company_sorted_consumer.tail(70).index.tolist()
common_companies_in_bottom_100_subissue = set(bottom_100_textblob_companies_subissue) & set(bottom_100_vader_co
common_companies_in_bottom_100_subissue
```

Out[ ]:
```
{'Bright Capital Inc',
 'Brightwater Capital, LLC',
 'Credit Bureau of Northern California Collections Services Inc',
 'LoanLeaders of America, Inc.',
 'William D. Meeker Enterprises, Inc.'}
```

In [ ]:
```
bottom_60_textblob_companies_subissue = subissue_textblob_average_scores_company_sorted_consumer.tail(60).index
bottom_60_vader_companies_subissue = polarity_average_scores_company_sorted_consumer.tail(60).index.tolist()
common_companies_in_bottom_60_subissue = set(bottom_60_textblob_companies_subissue) & set(bottom_60_vader_compa
common_companies_in_bottom_60_subissue
```

Out[ ]:
```
{'LoanLeaders of America, Inc.'}
```

Insights By State

We are going to find top 5 states with high negative sentiment and further assess their products sentiment and common issues found in these states

In [ ]:
```
Issue_textblob_average_scores = filtered_complaints[filtered_complaints['issue_sentiment'] != 0].groupby('State
Issue_textblob_average_scores_sorted = Issue_textblob_average_scores.sort_values(ascending=True)
Issue_textblob_average_scores_sorted
```

Out[ ]:
```
State
NM   -0.323843
AP   -0.323333
OK   -0.318351
AK   -0.317045
MO   -0.314123
PW   -0.312500
AL   -0.312246
AZ   -0.311681
IN   -0.309387
MS   -0.308558
AR   -0.307203
VT   -0.305851
PR   -0.305458
WV   -0.303711
TX   -0.298222
KS   -0.296759
OH   -0.295034
OR   -0.294161
RI   -0.293310
NV   -0.292835
MI   -0.292621
DE   -0.292500
DC   -0.290966
MD   -0.290678
PA   -0.289703
MA   -0.288814
ID   -0.286218
ND   -0.286058
SC   -0.285392
KY   -0.284016
IA   -0.283911
WI   -0.282914
AE   -0.282143
IL   -0.281525
VA   -0.280027
FL   -0.280011
LA   -0.279434
NC   -0.279404
CT   -0.278692
CA   -0.277706
WA   -0.275821
GA   -0.273928
ME   -0.273881
NY   -0.273500
UT   -0.272986
SD   -0.271667
HI   -0.271348
CO   -0.270952
TN   -0.269916
NJ   -0.269559
NE   -0.268280
MT   -0.264352
NH   -0.258416
MN   -0.254800
WY   -0.253704
VI   -0.212500
AA   -0.200000
GU   -0.125000
FM   -0.125000
Name: issue_sentiment, dtype: float64
```

In [ ]:
```
SubIssue_textblob_average_scores = filtered_complaints[filtered_complaints['sub_issue_sentiment'] != 0].groupby
SubIssue_textblob_average_scores_sorted = Issue_textblob_average_scores.sort_values(ascending=True)
SubIssue_textblob_average_scores_sorted
```

Out[ ]:
```
State
NM   -0.323843
AP   -0.323333
OK   -0.318351
AK   -0.317045
MO   -0.314123
PW   -0.312500
AL   -0.312246
AZ   -0.311681
IN   -0.309387
MS   -0.308558
AR   -0.307203
VT   -0.305851
PR   -0.305458
WV   -0.303711
TX   -0.298222
KS   -0.296759
OH   -0.295034
OR   -0.294161
RI   -0.293310
NV   -0.292835
MI   -0.292621
DE   -0.292500
DC   -0.290966
MD   -0.290678
PA   -0.289703
MA   -0.288814
ID   -0.286218
ND   -0.286058
SC   -0.285392
KY   -0.284016
IA   -0.283911
WI   -0.282914
AE   -0.282143
IL   -0.281525
VA   -0.280027
FL   -0.280011
LA   -0.279434
NC   -0.279404
CT   -0.278692
CA   -0.277706
WA   -0.275821
GA   -0.273928
ME   -0.273881
NY   -0.273500
UT   -0.272986
SD   -0.271667
HI   -0.271348
CO   -0.270952
TN   -0.269916
NJ   -0.269559
NE   -0.268280
MT   -0.264352
NH   -0.258416
MN   -0.254800
WY   -0.253704
VI   -0.212500
AA   -0.200000
GU   -0.125000
FM   -0.125000
Name: issue_sentiment, dtype: float64
```

In [ ]:
```
consumer_narrative_textblob_average_scores = filtered_complaints[filtered_complaints['consumer_complaint_narrat
consumer_narrative_textblob_average_scores_sorted = Issue_textblob_average_scores.sort_values(ascending=True)
consumer_narrative_textblob_average_scores_sorted
```

Out[ ]:
```
State
NM   -0.323843
AP   -0.323333
OK   -0.318351
AK   -0.317045
MO   -0.314123
PW   -0.312500
AL   -0.312246
AZ   -0.311681
IN   -0.309387
MS   -0.308558
AR   -0.307203
VT   -0.305851
PR   -0.305458
WV   -0.303711
TX   -0.298222
KS   -0.296759
OH   -0.295034
OR   -0.294161
RI   -0.293310
NV   -0.292835
MI   -0.292621
DE   -0.292500
DC   -0.290966
MD   -0.290678
PA   -0.289703
MA   -0.288814
ID   -0.286218
ND   -0.286058
SC   -0.285392
KY   -0.284016
IA   -0.283911
WI   -0.282914
AE   -0.282143
IL   -0.281525
VA   -0.280027
FL   -0.280011
LA   -0.279434
NC   -0.279404
CT   -0.278692
CA   -0.277706
WA   -0.275821
GA   -0.273928
ME   -0.273881
NY   -0.273500
UT   -0.272986
SD   -0.271667
HI   -0.271348
CO   -0.270952
TN   -0.269916
NJ   -0.269559
NE   -0.268280
MT   -0.264352
NH   -0.258416
MN   -0.254800
WY   -0.253704
VI   -0.212500
AA   -0.200000
GU   -0.125000
FM   -0.125000
Name: issue_sentiment, dtype: float64
```

In [ ]:
```
consumer_narrative_textblob_average_scores_df = consumer_narrative_textblob_average_scores_sorted.reset_index(n
consumer_narrative_textblob_average_scores_df
```

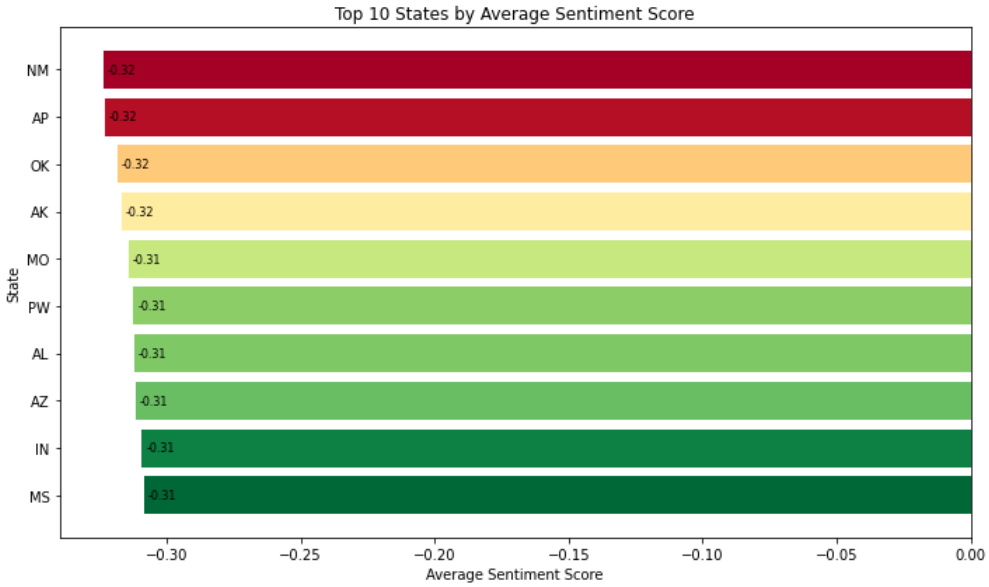
Out[ ]:

	State	Average Sentiment Score
0	NM	-0.323843
1	AP	-0.323333
2	OK	-0.318351
3	AK	-0.317045
4	MO	-0.314123
5	PW	-0.312500
6	AL	-0.312246
7	AZ	-0.311681
8	IN	-0.309387
9	MS	-0.308558
10	AR	-0.307203
11	VT	-0.305851
12	PR	-0.305458
13	WV	-0.303711
14	TX	-0.298222
15	KS	-0.296759
16	OH	-0.295034
17	OR	-0.294161
18	RI	-0.293310
19	NV	-0.292835
20	MI	-0.292621
21	DE	-0.292500
22	DC	-0.290966
23	MD	-0.290678
24	PA	-0.289703
25	MA	-0.288814
26	ID	-0.286218
27	ND	-0.286058
28	SC	-0.285392
29	KY	-0.284016
30	IA	-0.283911
31	WI	-0.282914
32	AE	-0.282143
33	IL	-0.281525
34	VA	-0.280027
35	FL	-0.280011
36	LA	-0.279434
37	NC	-0.279404
38	CT	-0.278692
39	CA	-0.277706
40	WA	-0.275821
41	GA	-0.273928
42	ME	-0.273881
43	NY	-0.273500
44	UT	-0.272986
45	SD	-0.271667
46	HI	-0.271348
47	CO	-0.270952
48	TN	-0.269916

	State	Average Sentiment Score
49	NJ	-0.269559
50	NE	-0.268280
51	MT	-0.264352
52	NH	-0.258416
53	MN	-0.254800
54	WY	-0.253704
55	VI	-0.212500
56	AA	-0.200000
57	GU	-0.125000
58	FM	-0.125000

In [ ]:
```
import matplotlib.pyplot as plt
import numpy as np

def plot_top_10_state_sentiment_scores_with_style(sentiment_scores_sorted, title):
    top_10_sentiment_scores_df = sentiment_scores_sorted.nsmallest(10, 'Average Sentiment Score').sort_values(b

    fig, ax = plt.subplots(figsize=(10, 6))
    ax.set_facecolor('white')
    fig.patch.set_facecolor('white')

    cmap = plt.get_cmap('RdYlGn')
    colors = cmap(np.interp(top_10_sentiment_scores_df['Average Sentiment Score'],
                            (top_10_sentiment_scores_df['Average Sentiment Score'].min(), top_10_sentiment_scor

    bars = plt.barh(top_10_sentiment_scores_df['State'], top_10_sentiment_scores_df['Average Sentiment Score'],

    plt.xlabel('Average Sentiment Score', color='black')
    plt.ylabel('State', color='black')
    plt.title(title, color='black')

    ax.tick_params(axis='x', colors='black')
    ax.tick_params(axis='y', colors='black', labelsize=10)

    for bar in bars:
        width = bar.get_width()
        ax.annotate(f'{width:.2f}',
                    xy=(width, bar.get_y() + bar.get_height() / 2),
                    xytext=(3, 0),  # 3 points horizontal offset
                    textcoords="offset points",
                    ha='left', va='center', color='black', fontsize=8)

    plt.tight_layout()
    plt.show()
```

In [ ]:
```
# Ensure consumer_narrative_textblob_average_scores_sorted is your sorted Series or DataFrame
plot_top_10_state_sentiment_scores_with_style(consumer_narrative_textblob_average_scores_df, 'Top 10 States by
```



In [ ]:
```
filtered_issues = filtered_complaints.loc[
    (filtered_complaints['State'].isin(['MN', 'AP', 'OK', 'AK', 'MO'])) &
    (filtered_complaints['issue_sentiment'] != 0)
```

```
]

# We have further captured dataset where only top 5 states are there, mentioned above
```

In [ ]:
```
filtered_issues.head(5)
```

Out[ ]:

	Date received	Product	Sub-product	Issue	Sub-issue	Consumer complaint narrative	Company public response	Company	State	ZIP code	...	Con resp cons
2281	2024-01-19	Checking or savings account	Checking account	Problem with a lender or other company chargin...	Transaction was not authorized	NaN	NaN	BANK OF AMERICA, NATIONAL ASSOCIATION	OK	731XX	...	pro
16241	2024-02-06	Credit reporting or other personal consumer re...	Other personal consumer report	Identity theft protection or other monitoring ...	Problem canceling credit monitoring or identif...	NaN	NaN	Experian Information Solutions Inc.	OK	74137	...	pro
17168	2024-01-31	Mortgage	FHA mortgage	Trouble during payment process	Payment process	NaN	NaN	NEW YORK COMMUNITY BANCORP INC	MO	63137	...	pro
19060	2024-01-22	Mortgage	Conventional home mortgage	Trouble during payment process	Trying to communicate with the company to fix ...	NaN	NaN	PNC Bank N.A.	MO	63303	...	pro
21877	2024-01-19	Checking or savings account	Checking account	Problem with a lender or other company chargin...	Money was taken from your account on the wrong...	NaN	NaN	BANK OF AMERICA, NATIONAL ASSOCIATION	MN	55057	...	pro

5 rows × 24 columns

Whatare the key issues in the top 5 states with highest negative sentiment

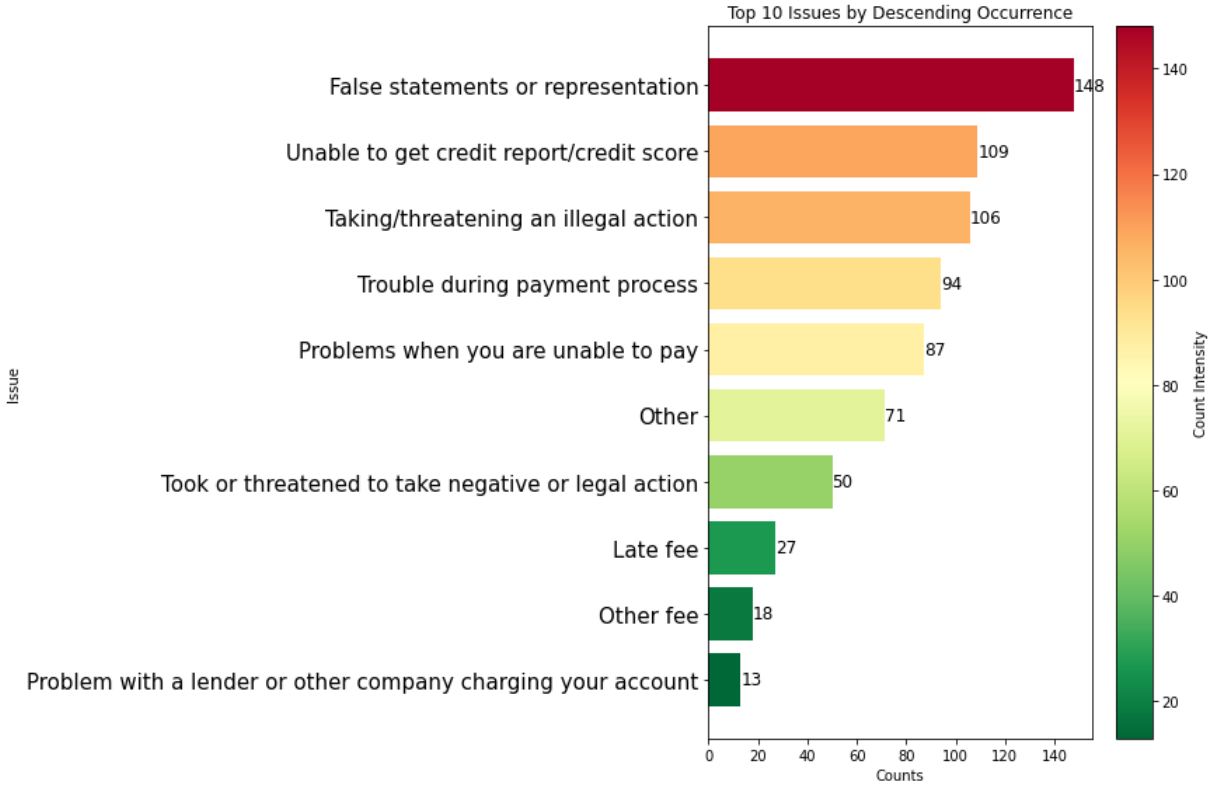
In [ ]:
```
mn_issue_counts = filtered_issues.groupby('Issue').size().sort_values(ascending=False)
mn_issue_counts
```

Out[ ]:
```
Issue
False statements or representation                              148
Unable to get credit report/credit score                        109
Taking/threatening an illegal action                            106
Trouble during payment process                                   94
Problems when you are unable to pay                              87
Other                                                            71
Took or threatened to take negative or legal action              50
Late fee                                                         27
Other fee                                                        18
Problem with a lender or other company charging your account     13
Unable to get your credit report or credit score                  9
Other transaction issues                                          8
Money was not available when promised                             8
Trouble using the card                                            6
Charged bank acct wrong day or amt                                4
Other transaction problem                                         4
Other service issues                                              4
Other features, terms, or problems                                4
Advertising and marketing, including promotional offers           3
Trouble using your card                                           2
Unauthorized transactions or other transaction problem            2
Wrong amount charged or received                                  2
Identity theft protection or other monitoring services            1
Confusing or missing disclosures                                  1
Confusing or misleading advertising or marketing                  1
dtype: int64
```

In [ ]:
```
import matplotlib.pyplot as plt
import numpy as np

mn_issue_counts_df = mn_issue_counts.reset_index(name='Counts').nlargest(10, 'Counts')
```

```
fig, ax = plt.subplots(figsize=(12, 8))
ax.set_facecolor('white')
fig.patch.set_facecolor('white')

colors = plt.cm.RdYlGn_r(np.interp(mn_issue_counts_df['Counts'], (mn_issue_counts_df['Counts'].min(), mn_issue_

bars = plt.barh(mn_issue_counts_df['Issue'], mn_issue_counts_df['Counts'], color=colors)

plt.gca().invert_yaxis()

plt.xlabel('Counts', color='black')
plt.ylabel('Issue', color='black')
plt.title('Top 10 Issues by Descending Occurrence', color='black')

ax.tick_params(axis='x', colors='black')
ax.tick_params(axis='y', colors='black', labelsize=15)

for bar in bars:
    ax.text(bar.get_width(), bar.get_y() + bar.get_height() / 2,
            f'{bar.get_width():.0f}',
            va='center', ha='left', color='black', fontsize=12)

sm = plt.cm.ScalarMappable(cmap='RdYlGn_r', norm=plt.Normalize(vmin=mn_issue_counts_df['Counts'].min(), vmax=mn
sm.set_array([])
cbar = plt.colorbar(sm)
cbar.ax.tick_params(labelsize=10, colors='black')
cbar.set_label('Count Intensity', color='black', fontsize=10)

plt.tight_layout()
plt.show()
```

```
<ipython-input-238-a3b73202b85a>:41: MatplotlibDeprecationWarning: Unable to determine Axes to steal space for C
olorbar. Using gca(), but will raise in the future. Either provide the *cax* argument to use as the Axes for the
Colorbar, provide the *ax* argument to steal space from it, or add *mappable* to an Axes.
  cbar = plt.colorbar(sm)
```

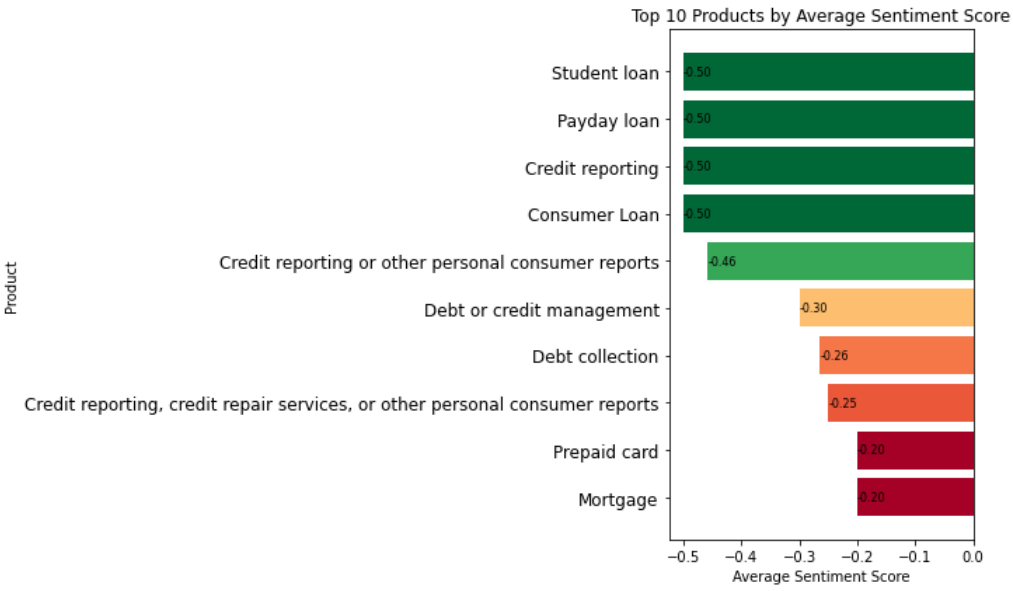


Whatare the key products in the top 5 states with highest negative sentiment

In [ ]:
```
Issue_textblob_average_scores = filtered_issues[filtered_issues['issue_sentiment'] != 0].groupby('Product')['is
Issue_textblob_average_scores_sorted = Issue_textblob_average_scores.sort_values(ascending=True)
Issue_textblob_average_scores_sorted
```

Out[ ]:
```
Product
Consumer Loan                                                                -0.500000
Credit reporting                                                             -0.500000
Payday loan                                                                  -0.500000
Student loan                                                                 -0.500000
Credit reporting or other personal consumer reports                          -0.458333
Debt or credit management                                                    -0.300000
Debt collection                                                              -0.264474
Credit reporting, credit repair services, or other personal consumer reports -0.250000
Prepaid card                                                                 -0.200000
Mortgage                                                                     -0.200000
Money transfers                                                              -0.191667
Credit card                                                                  -0.161875
Money transfer, virtual currency, or money service                           -0.143750
Checking or savings account                                                  -0.125000
Credit card or prepaid card                                                  -0.125000
Virtual currency                                                             -0.125000
Name: issue_sentiment, dtype: float64
```

In [ ]:
```
import matplotlib.pyplot as plt
import numpy as np

top_10_product_sentiment = Issue_textblob_average_scores_sorted.head(10).reset_index(name='Average Sentiment Sc
top_10_product_sentiment.sort_values(by='Average Sentiment Score', ascending=False, inplace=True)

fig, ax = plt.subplots(figsize=(10, 6))
ax.set_facecolor('white')
fig.patch.set_facecolor('white')

colors = plt.cm.RdYlGn_r(np.interp(top_10_product_sentiment['Average Sentiment Score'],
                                   (top_10_product_sentiment['Average Sentiment Score'].min(), top_10_product_se

bars = plt.barh(top_10_product_sentiment['Product'], top_10_product_sentiment['Average Sentiment Score'], color

plt.xlabel('Average Sentiment Score', color='black')
plt.ylabel('Product', color='black')
plt.title('Top 10 Products by Average Sentiment Score', color='black')

for bar in bars:
    width = bar.get_width()
    plt.text(width, bar.get_y() + bar.get_height() / 2, f"{width:.2f}", ha='left', va='center', color='black',

ax.tick_params(axis='x', colors='black')
ax.tick_params(axis='y', colors='black', labelsize=12)

plt.tight_layout()
plt.show()
```



THANK YOU !!!