

DATA MINING

Zoya Ansari
Maqbul Ansari
Dalya Khatun

Software Manual

1. To run the scripts, use Pandas Jupyter Notebook.
2. Download “CustomerComplaints.csv” from https://catalog.data.gov/dataset/consumer-complaint-database/resource/484bc506-82b0-4225-ab10-496cf7a1c2e9?inner_span=True
3. To import the dataset change the path in the first cell. The path should be the path to where you downloaded the dataset.
4. Run each cells or run all
5. Read the comments on top of each cells for information about projected output.

In [176]:

```
import pandas as pd
import matplotlib.pyplot as plt

# This is used for fast string concatenation
from io import StringIO

# Use nltk for valid words
import nltk
# Need to make hash 'dictionaries' from nltk for fast processing
import collections as co

import warnings # current version of seaborn generates a bunch of warnings that we'.
warnings.filterwarnings("ignore")
import seaborn as sns
import matplotlib.pyplot as plt
sns.set(style="white", color_codes=True)

# change the directory path for the database according to the location of the csv f.
data = pd.read_csv("/Users/dalyakhatun/Desktop/GSU/Data Mining/consumer_complaints.c
```

In [19]:

```
# Find the number of rows and columns
data.shape
```

Out[19]:

```
(746402, 18)
```

In [187]:

```
# display the dataset.  
data.head()
```

Out[187]:

	Date received	Product	Sub-product	Issue	Sub-issue	Consumer complaint narrative	Company public response	Company
0	07/29/2013	Consumer Loan	Vehicle loan	Managing the loan or lease	NaN	NaN	NaN	Wells F & Com
1	07/29/2013	Bank account or service	Checking account	Using a debit or ATM card	NaN	NaN	NaN	Wells F & Com
2	07/29/2013	Bank account or service	Checking account	Account opening, closing, or management	NaN	NaN	NaN	Santan Bank U
3	07/29/2013	Bank account or service	Checking account	Deposits and withdrawals	NaN	NaN	NaN	Wells F & Com
4	07/29/2013	Mortgage	Conventional fixed mortgage	Loan servicing, payments, escrow account	NaN	NaN	NaN	Franklir Credit Manag

In [27]:

```
# Find the types of data for each column  
data.dtypes
```

Out[27]:

```
Date received          object  
Product                object  
Sub-product            object  
Issue                  object  
Sub-issue              object  
Consumer complaint narrative  object  
Company public response  object  
Company                object  
State                  object  
ZIP code               object  
Tags                   object  
Consumer consent provided? object  
Submitted via          object  
Date sent to company    object  
Company response to consumer object  
Timely response?        object  
Consumer disputed?      object  
Complaint ID            int64  
dtype: object
```

In [57]:

```
# Display the header of each column in the dataset.  
data.columns
```

Out[57]:

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

In [53]:

```
# different products in the dataset
data.Product.unique()
```

Out[53]:

```
array(['Consumer Loan', 'Bank account or service', 'Mortgage',
      'Debt collection', 'Credit card', 'Credit reporting',
      'Student loan', 'Money transfers', 'Payday loan', 'Prepaid card',
      'Other financial service', 'Virtual currency'], dtype=object)
```

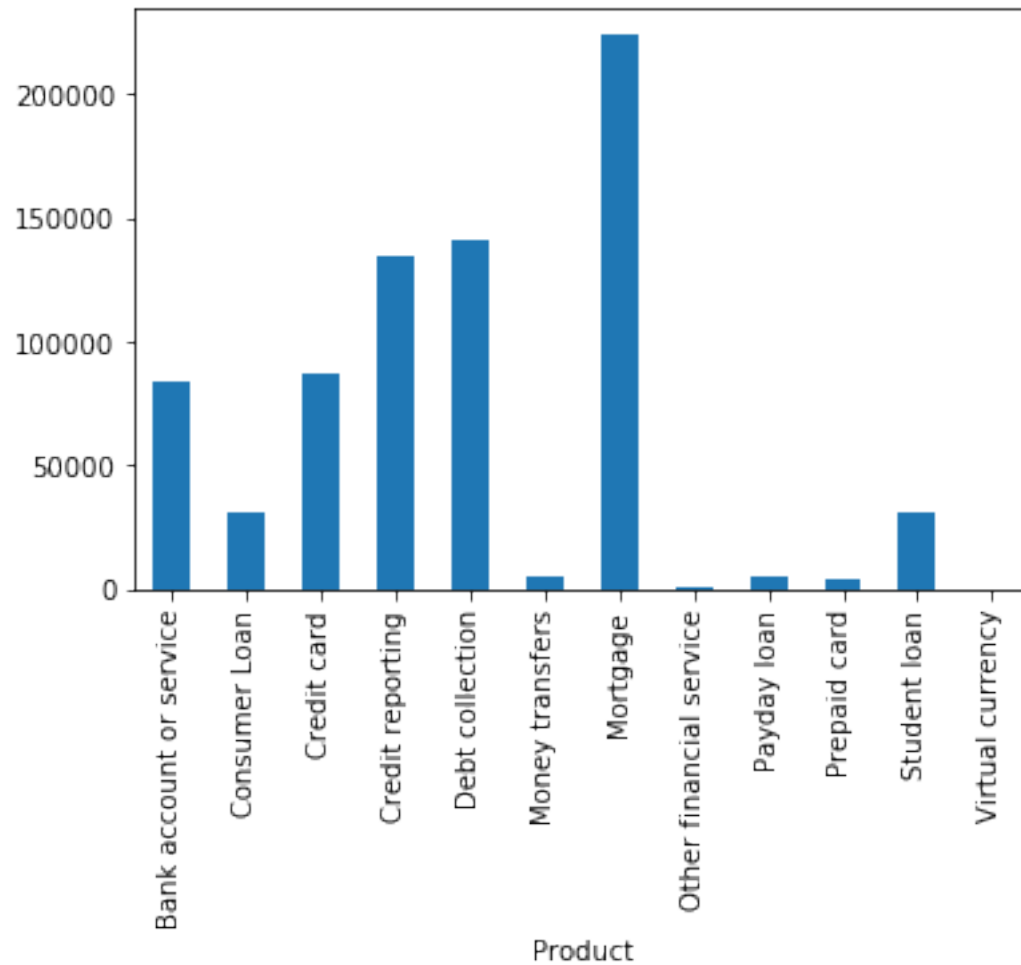
In [79]:

```
# How many complaints raised against each product?
productAgainstComplaints = data.groupby('Product').size()
print(productAgainstComplaints)
```

```
Product
Bank account or service      84071
Consumer Loan                30446
Credit card                  86938
Credit reporting            134970
Debt collection              140678
Money transfers               5187
Mortgage                     223642
Other financial service       984
Payday loan                   5366
Prepaid card                  3672
Student loan                  30431
Virtual currency              17
dtype: int64
```

In [83]:

```
# This graphs shows the number of complaint against each product.
productAgainstComplaints.plot(kind='bar')
plt.show()
```



In [85]:

```
# different companies in the dataset
data.Company.unique()
```

Out[85]:

```
array(['Wells Fargo & Company', 'Santander Bank US',
      'Franklin Credit Management', ..., 'Benn Realty Services, Inc.',
      'Sibbett Auto Sales LLC', 'LN Capital Corporation'], dtype=object)
```

In [94]:

```
# How many complaints raised against each Company?
CompanyAgainstComplaints = data.groupby('Company').size()
print(CompanyAgainstComplaints)
```

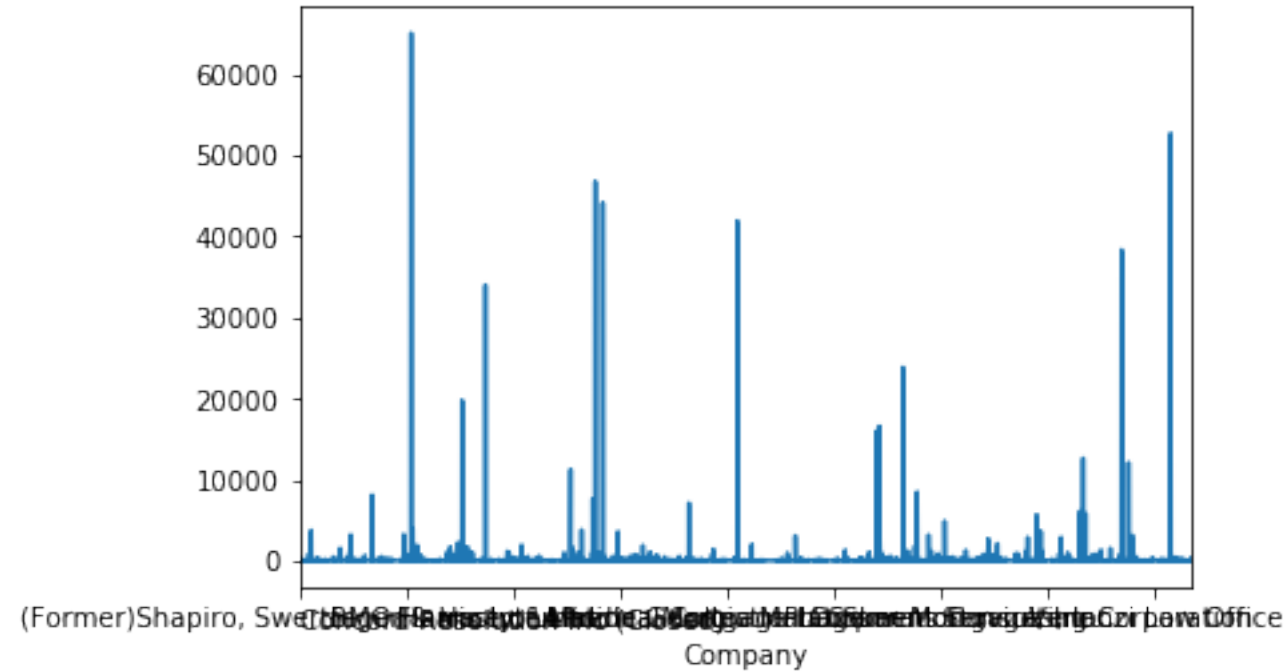
```
Company
(Former)Shapiro, Swertfeger & Hasty, LLP    4
1 Stop Money Centers, LLC                  1
1st 2nd Mortgage Company Of NJ, Inc.       1
```

1st Alliance Lending	17
1st Capital Finance of South Carolina, Inc.	1
1st Capital Mortgage, LLC	1
1st Fidelity Loan Servicing	5
1st Franklin Financial Corporation	54
1st Maryland Mortgage Corporation	1
1st Midwest Mortgage Corp	2
1st Money Center, Inc.	2
1st Portfolio Holding Corporation	1
1st Preference Mortgage	2
1st Priority Mortgage, Inc.	1
21st Mortgage Corporation	311
2233 Paradise Road LLC	8
2288984 Ontario Inc.	30
24 Asset Management Corp	5
245 Holdings LLC	1
360 Mortgage	47
3rd Generation, Inc.	19
4 Star Resolution LLC (Closed)	29
4M Collections, LLC	10
99th Floor LLC	1
A & E Futures LLC	7
A & O Recovery Solutions, LLC	5
A & S Collection Associates, Inc.	5
A+ Financial Services, Inc.	2
A-1 Fast Cash of Tucker, Inc	1
A.C.S. Companies, Inc. I	2
...	
Xtreme Products LLC	8
YKS Acceptance, Inc.	26
Yakima County Credit Service, INC	2
Yale Mortgage Corporation	18
Young America Holdings, LLC	2
ZENCO COLLECTION, LLC	58
ZINC Financial, Inc.,	2
ZOA, LLC	6
Zakheim Law Group, P.A.	26
Zarvad III S.A.	55
Zarzaur and Schwartz, PC	9
Zealandia Holding Company, Inc.	5
Zeidman's Jewelry & Loan of Michigan	1
Zeller & Associates, LLC	9
Zenith Financial & Insurance Solutions Inc.	1
Zenith Financial Mortgage Company, Inc.	1
Zenith Financial Network Inc	46
ZestFinance	7
Zeus Mortgage, Ltd.	2
Zingo Cash Illinois LLC	3
Zions Bancorporation	292
Zip Capital Group, LLC	1
Zwicker & Associates	179
eCon Credit LP	2
eMoneyUSA Holdings, LLC	2

i3 Lending, Inc	3
iFreedom Direct Corporation	15
iQuantified Management Services, LLC	5
iServe Trust	7
reekside Recovery And Capital Management	24
dtype: int64	

In [96]:

```
# This graphs show the number of complaint against each Company.
CompanyAgainstComplaints.plot()
plt.show()
```



In [99]:

```
# How many complaints have received timely response?  
# Almost 97.5% of the complaints have been addressed timely.
```

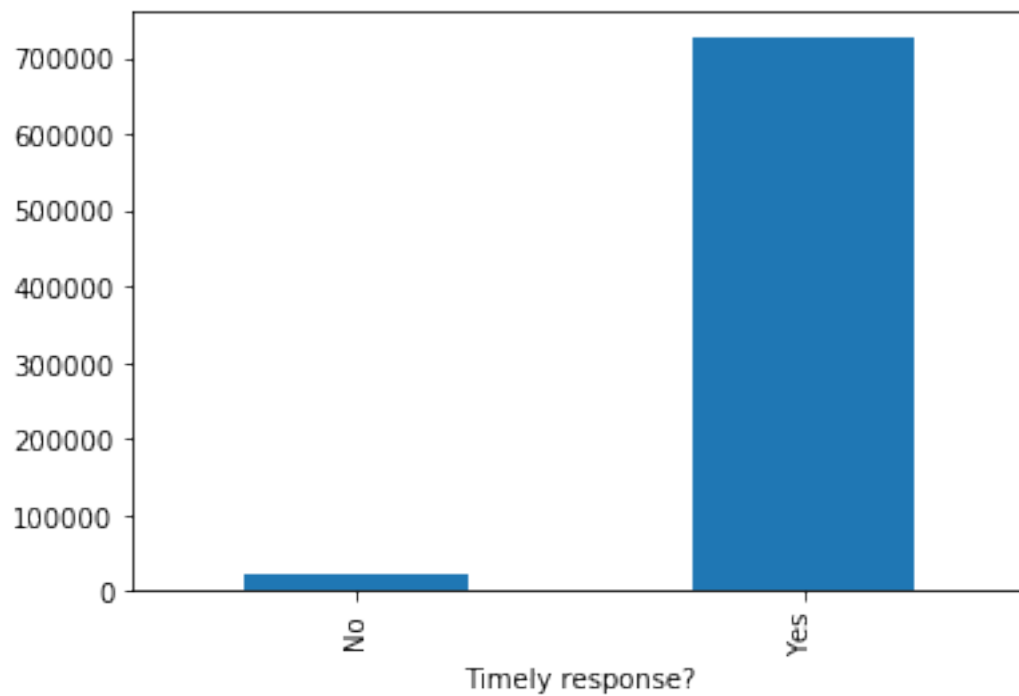
```
ReceivedTimelyResponse = data.groupby('Timely response?').size()  
print(ReceivedTimelyResponse)  
ReceivedTimelyResponse.plot(kind='bar')  
plt.show()
```

Timely response?

No 20926

Yes 725476

dtype: int64



In [101]:

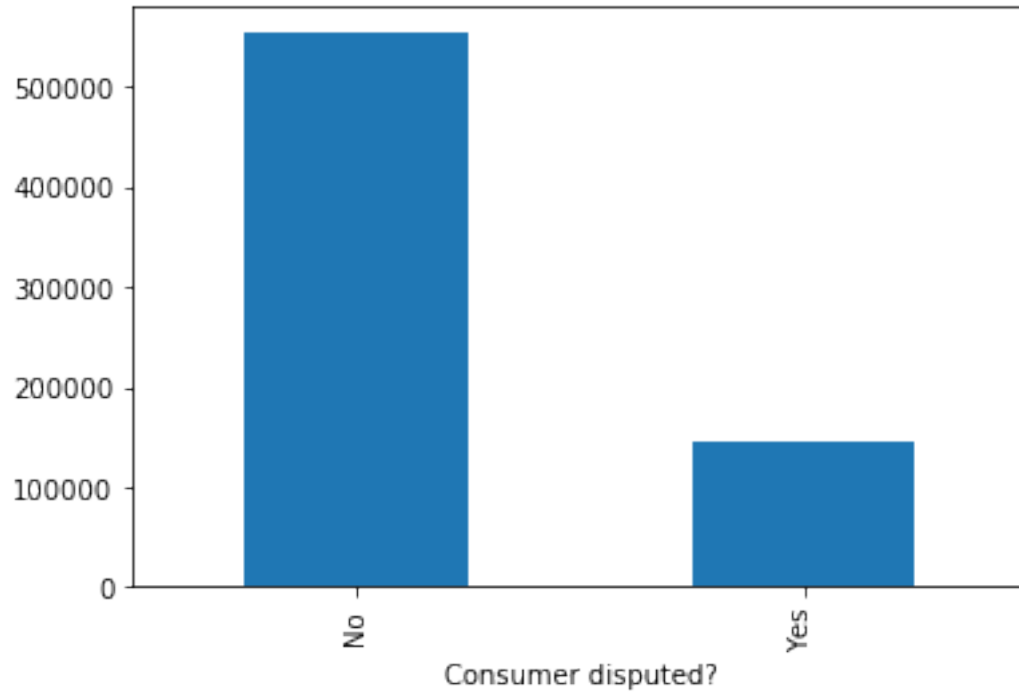
```
# How many consumer disputed?  
# Over 20% of the users disputed.  
disputed = data.groupby('Consumer disputed?').size()  
print(disputed)  
disputed.plot(kind='bar')  
plt.show()
```

Consumer disputed?

No 553302

Yes 145659

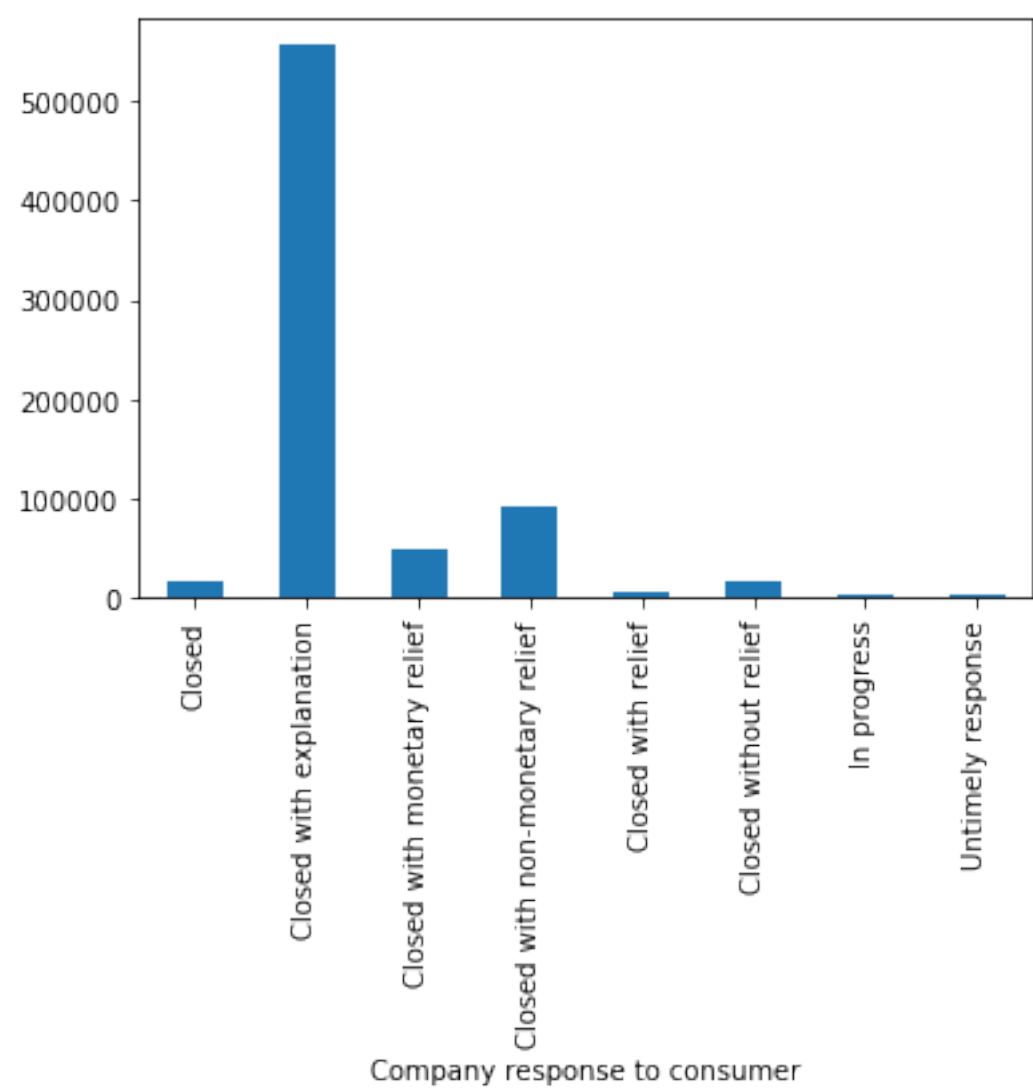
dtype: int64



In [103]:

```
# What are the different responses users get from the companies?
DifferentResponses = data.groupby('Company response to consumer').size()
print(DifferentResponses)
DifferentResponses.plot(kind='bar')
plt.show()
```

Company response to consumer	
Closed	17170
Closed with explanation	555668
Closed with monetary relief	49769
Closed with non-monetary relief	92404
Closed with relief	5298
Closed without relief	17863
In progress	4182
Untimely response	4048
dtype: int64	



In [107]:

```
# Analyzing text in consumer complaints descriptions
```

In [110]:

```
# We are only interested in data with consumer complaints.  
# Only using the data where 'Consumer complaint narrative' is not null  
  
data=data[data['Consumer complaint narrative'].notnull()]  
  
st=StringIO()  
data['Consumer complaint narrative'].apply(lambda x: st.write(x))  
  
k=st.getvalue()  
st.close()  
k=k.lower()  
k=k.split()
```

In [115]:

```
# Next only want valid strings  
words = co.Counter(nltk.corpus.words.words())  
stopWords =co.Counter( nltk.corpus.stopwords.words() )  
k=[i for i in k if i in words and i not in stopWords]  
st=" ".join(k)  
c = co.Counter(k)
```

In [144]:

```
# 30 most common words
c.most_common(30)
```

Out[144]:

```
[('credit', 178714),
 ('account', 133512),
 ('would', 104821),
 ('loan', 89573),
 ('payment', 82255),
 ('bank', 79684),
 ('told', 79669),
 ('received', 66514),
 ('debt', 60481),
 ('information', 57717),
 ('never', 56824),
 ('mortgage', 56814),
 ('pay', 52764),
 ('get', 52073),
 ('company', 51348),
 ('card', 50917),
 ('report', 50731),
 ('sent', 49832),
 ('time', 49195),
 ('said', 45661),
 ('call', 45457),
 ('could', 45353),
 ('letter', 44021),
 ('back', 42456),
 ('made', 39802),
 ('due', 38320),
 ('phone', 37934),
 ('since', 37570),
 ('amount', 37012),
 ('money', 35419)]
```

In [120]:

```
st[0:100]
```

Out[120]:

```
'one owe loan address box know longer use never even told credit repor
t thats still address one vehic'
```

In [122]:

```
print(k[0:10],"\n\nLength of k %s" % len(k))
```

```
['one', 'owe', 'loan', 'address', 'box', 'know', 'longer', 'use', 'never', 'even']
```

Length of k 8133025

In [129]:

```
# Let's get some text involving identity theft
searchS='victim of identity theft'
vi = data[data['Consumer complaint narrative'].str.find(searchS) >= 0]
data['victim']=None
data['e']=1
data['m']=None # This will be for 'Closed with monetary relief'
data['victim'] = data[data['Consumer complaint narrative'].str.find(searchS) >= 0]
data['m']=data[data['Company response to consumer'] == 'Closed with monetary relief']

# Take a look at some sample stories mindex to mindex_inc
# Adjust this, to see different stories
mindex=20
mindex_inc=5+mindex
si=StringIO()
vi['Consumer complaint narrative'].iloc[mindex:mindex_inc].apply(lambda x: si.write(x))

t=si.getvalue()
si.close()
print(t)
```

I have been victim of identity theft XXXX times now in about XXXX years time. I have filed complaints with Consumer Fraud and with the Credit Bureaus. Latest issue is so many hard inquiries on my file by auto finance companies I did NOT give written or oral permission to. Wrote them all and asked for copy of credit authorization form they have that gave them permission. None have been able to send any because I never gave them permission. Only XXXX dealerships were authorized, XXXX XXXX in writing, and XXXX XXXX XXXX orally over the phone. Yet there are XXXX inquiries total on my file with XXXX credit bureau alone. I know one has a XXXX day window to shop around for auto financing, and it only shows as XXXX Hard Inquiry. This has always been the case all my life, yet this time with XXXX XXXX, almost all their companies they dealt with show as Hard Inquiries on my credit. My file is supposed to be flagged due to the identity thefts, and yet they are allowing all these other entities to check my credit WITHOUT a credit authorization form signed by me. I have wrote Equifax and the others asking they remove those unauthorized inquiries and got back a form letter with a response that I can not understand at all '' inquiries are a matter of record of all companies that have accessed your credit file ''. The inquiries are still there and not removed and it has affected my credit

score drastically. I am so tired of these automated responses that get absolutely nothing done whatsoever to protect the consumer ... ME! They have allowed all sorts of entries on my credit that I never authorized, and then I get calls from collection agencies for loans I never made, and the vicious circle and cycle continues. At my age of XXXX I should NOT have to worry about this kind of criminal activity on my credit but I am worn out and tormented with this constant struggle to clear this garbage off my credit files, over and over and over and over again, with the Bureaus, from MY point of view are NOT co-operating. After all, its MY CREDIT INFORMATION, and yet they treat me like I am the culprit by trying to get it cleaned up and corrected. This Bureau system sucks in my opinion and needs to be done away with. I will be long dead if that ever happens, so meanwhile I am tormented with having to deal with thefts, illegal inquiries, incorrect entries and Identity issuesgrowing old SUCKS!!!!
XXXX XXXX

I am putting a complaint on Chexsystems cause I am having ongoing issues with them and they are not trying to correct the issue. In XXXX I had requested a copy of my chex systems file. I received a letter from them dated XXXX XXXX XXXX that my file was mixed with another file with similar information. The reason for this is because I have been a victim of identity theft. They requested for a copy of my social security number to correct the issue. I sent them in a copy of the letter and social security card front and back along with a police report that I had filed. They sent me a copy of my chex systems file at that point. After reviewing the file there are multiple inquiries on my Chexsystem's file that were a result of the identity theft. The original police report I sent in dated XXXX/XXXX/XXXX I sent to chexsystems and now I am attaching to my complaint the new police report that I have filed dated XXXX/XXXX/XXXX. The list of inquiries on the report are extensive the only inquiry that should be reporting on my report is XXXX XXXX XXXX on XXXX/XXXX/XXXX I have not requested or authorized anyone to access my chexsystems since then. The rest of the inquiries are the result of the identity theft.

I found that Wells Fargo reported to a credit reporting agency that I owed them a debt of {\$420.00} on XXXX 2012 on a credit card account. I never opened a credit card account with Wells Fargo and have no knowledge of any credit card charges made on a Wells Fargo credit card in my name. I wrote to Wells Fargo and informed them of these facts. I also told Wells Fargo that I believed that these charges were fraudulent and suspected that I had been a victim of identity theft. I requested that Wells Fargo conduct an investigation into the matter and that Wells Fargo provide to me all account documents and billing statements related to this claimed debt so that I could determine whether I had been a victim of identity theft as I suspected. Wells Fargo responded but refused my request for documentation and its response does not indicate that it conducted any investigation into my claim that I had been the v

ictim of credit card fraud and identity theft.

My identity was stolen back in XXXX 2008 when my purse was stolen from my person. I have the police report with the statement of theft. Today , I received a phone call at my place of work regarding a collection of debt. The company is Performant Financial Corporation and the caller was XXXX XXXX. I explained to XXXX that I am a victim of identity theft and her calling me at work was illegal under section 806 (5) of the Collection Practices Act.

Virtuoso Agency located in XXXX Colorado. The customer service agent was very rude. I had sent them a police report and other items verifying that I was a victim of identity theft. The original creditor direct energy cleared the balance realizing it was identity theft. I asked the debt collector to send paperwork to XXXX to clear my name. Virtuoso is reporting that it is in open collections to the credit agencies. The agent was rude and said I am not your friend and I do n't care and I do n't have to help you. All I asked was them to send a letter to XXXX that the account was recalled by XXXX because of fraud. They are reporting to the credit agencies even with the fact that they know its fraudulent debt and wont do anything to clear it.

```
from wordcloud import WordCloud

# Read the whole text.
text = s

# Generate a word cloud image
wordcloud = WordCloud().generate(text)

# Display the generated image:
import matplotlib.pyplot as plt

# take relative word frequencies into account, lower max_font_size
wordcloud = WordCloud(background_color="gray",max_words=len(k),max_font_size=50, rel
plt.figure()
plt.imshow(wordcloud)
plt.axis("off")
plt.show()
```

