```
In [1]:  # Importing Required Libraries
         import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         import random
         import chart_studio
         import chart_studio.plotly as py
         import chart_studio.tools as tls
```

Data Cleaning and Data Analysis

```
In [2]:  # Reading the complaints data
         df = pd.read_csv("complaints.csv")
```

```
C:\Users\Goutham\AppData\Local\Temp\ipykernel_17944\1014270965.py:2: DtypeWarning: Columns (9,16) have mixed types.
Specify dtype option on import or set low_memory=False.
  df = pd.read_csv("complaints.csv")
```

```
In [3]:  # Printing the shape of data frame
         df.shape
```

```
Out[3]:  (3017511, 18)
```

In [4]: *# Printing the first 5 rows of the data frame*
`df.head()`

Out[4]:

	Date received	Product	Sub-product	Issue	Sub-issue	Consumer complaint narrative	Company public response	Company	State	ZIP code	Tags	Consumer consent provided?	Submitted via	D sen comp
0	2022-09-14	Debt collection	Credit card debt	Attempts to collect debt not owed	Debt was result of identity theft	To Whom It May Concern : I recently received a...	Company has responded to the consumer and the ...	TRANSUNION INTERMEDIATE HOLDINGS, INC.	FL	33063.0	NaN	Consent provided	Web	2022
1	2022-10-08	Credit card or prepaid card	General-purpose credit card or charge card	Getting a credit card	Card opened as result of identity theft or fraud	NaN	NaN	CAPITAL ONE FINANCIAL CORPORATION	NY	10033.0	NaN	NaN	Web	2022
2	2022-09-14	Payday loan, title loan, or personal loan	Personal line of credit	Problem when making payments	NaN	Hello, on my credit report you have XXXX XXXX ...	Company has responded to the consumer and the ...	TRANSUNION INTERMEDIATE HOLDINGS, INC.	TX	77047.0	NaN	Consent provided	Web	2022
3	2022-10-19	Credit reporting, credit repair services, or o...	Credit reporting	Problem with a credit reporting company's inve...	Their investigation did not fix an error on yo...	NaN	NaN	EQUIFAX, INC.	TX	75904.0	NaN	NaN	Web	2022
4	2022-10-05	Credit reporting, credit repair services, or o...	Credit reporting	Problem with a credit reporting company's inve...	Their investigation did not fix an error on yo...	NaN	NaN	AMERICAN EXPRESS COMPANY	GA	31093.0	NaN	NaN	Web	2022

In [5]:
```python
# Count of all null values across columns
df.isna().sum()
```

Out[5]:
```
Date received                          0
Product                                0
Sub-product                       235294
Issue                                  0
Sub-issue                         679336
Consumer complaint narrative     1934522
Company public response          1709580
Company                                0
State                              39946
ZIP code                           40340
Tags                             2673211
Consumer consent provided?        805546
Submitted via                          0
Date sent to company                   0
Company response to consumer           4
Timely response?                       0
Consumer disputed?               2249061
Complaint ID                           0
dtype: int64
```

In [6]:
```python
# Filtering the data based on the condition
df[df['State'].isna()==True]
```

Out[6]:

	Date received	Product	Sub-product	Issue	Sub-issue	Consumer complaint narrative	Company public response	Company	State	ZIP code	Tags	Consumer consent provided?
1438	2022-10-17	Credit reporting, credit repair services, or o...	Credit reporting	Improper use of your report	Reporting company used your report improperly	NaN	NaN	TRANSUNION INTERMEDIATE HOLDINGS, INC.	NaN	NaN	NaN	NaN
2650	2022-10-06	Checking or savings account	Other banking product or service	Managing an account	Problem making or receiving payments	NaN	NaN	JPMORGAN CHASE & CO.	NaN	NaN	NaN	NaN
9029	2022-09-22	Checking or savings account	Savings account	Closing an account	Can't close your account	I need help. I am an XXXX citizen living in XX...	Company has responded to the consumer and the ...	WELLS FARGO & COMPANY	NaN	NaN	NaN	Consent provided
10411	2022-09-15	Credit card or prepaid card	General-purpose credit card or charge card	Getting a credit card	Card opened as result of identity theft or fraud	NaN	NaN	CIBC BANCORP USA INC.	NaN	NaN	NaN	NaN
11493	2022-10-04	Credit reporting, credit repair services, or o...	Credit reporting	Incorrect information on your report	Information belongs to someone else	Hi. I checked my Credit Score using my bank ap...	Company has responded to the consumer and the ...	Experian Information Solutions Inc.	NaN	NaN	NaN	Consent provided
...	...	...	...	...	...	...	...	...	...	...	...	...
3017470	2019-03-04	Mortgage	Conventional home mortgage	Struggling to pay mortgage	NaN	NaN	NaN	NATIONSTAR MORTGAGE	NaN	NaN	Servicemember	NaN
3017471	2019-02-27	Checking or savings account	Checking account	Managing an account	Deposits and withdrawals	NaN	NaN	JPMORGAN CHASE & CO.	NaN	NaN	NaN	NaN

	Date received	Product	Sub-product	Issue	Sub-issue	Consumer complaint narrative	Company public response	Company	State	ZIP code	Tags	Consumer consent provided?
3017474	2018-12-19	Mortgage	Other type of mortgage	Struggling to pay mortgage	NaN	NaN	NaN	JPMORGAN CHASE & CO.	NaN	NaN	NaN	NaN
3017476	2018-11-21	Money transfer, virtual currency, or money ser...	Domestic (US) money transfer	Fraud or scam	NaN	NaN	NaN	MONEYGRAM PAYMENT SYSTEMS WORLDWIDE INC	NaN	NaN	NaN	NaN
3017477	2019-01-02	Mortgage	Other type of mortgage	Struggling to pay mortgage	NaN	NaN	NaN	Ocwen Financial Corporation	NaN	NaN	NaN	NaN

39946 rows × 18 columns

In [7]:
```python
# Dropping the null values based on state column
df.dropna(subset=['State'],inplace=True)
```

In [8]: ```python
# Count of all null values across columns
df.isna().sum()
```

Out[8]:
```
Date received   0
Product   0
Sub-product   234082
Issue   0
Sub-issue   664176
Consumer complaint narrative   1899168
Company public response   1687198
Company   0
State   0
ZIP code   907
Tags   2638725
Consumer consent provided?   774541
Submitted via   0
Date sent to company   0
Company response to consumer   4
Timely response?   0
Consumer disputed?   2214765
Complaint ID   0
dtype: int64
```

In [9]: ```python
# Printing the shape of data frame
df.shape
```

Out[9]: (2977565, 18)

In [10]: ```python
# Reading the state data
st_data = pd.read_csv(r"C:\Users\Goutham\Downloads\State Map_Full Data.csv")
```

In [11]: ```python
# Printing the shape of data frame
st_data.shape
```

Out[11]: (50, 2)

In [12]: ```
st_data
```

Out[12]:

| | State | Full Form |
|---|---|---|
| 0 | AL | Alabama |
| 1 | AK | Alaska |
| 2 | AZ | Arizona |
| 3 | AR | Arkansas |
| 4 | CA | California |
| 5 | CO | Colorado |
| 6 | CT | Connecticut |
| 7 | DE | Delaware |
| 8 | FL | Florida |
| 9 | GA | Georgia |
| 10 | HI | Hawaii |
| 11 | ID | Idaho |
| 12 | IL | Illinois |
| 13 | IN | Indiana |
| 14 | IA | Iowa |
| 15 | KS | Kansas |
| 16 | KY | Kentucky |
| 17 | LA | Louisiana |
| 18 | ME | Maine |
| 19 | MD | Maryland |
| 20 | MA | Massachusetts |
| 21 | MI | Michigan |
| 22 | MN | Minnesota |
| 23 | MS | Mississippi |
| 24 | MO | Missouri |
| 25 | MT | Montana |

|    | State | Full Form |
|----|-------|-----------|
| 26 | NE | Nebraska |
| 27 | NV | Nevada |
| 28 | NH | New Hampshire |
| 29 | NJ | New Jersey |
| 30 | NM | New Mexico |
| 31 | NY | New York |
| 32 | NC | North Carolina |
| 33 | ND | North Dakota |
| 34 | OH | Ohio |
| 35 | OK | Oklahoma |
| 36 | OR | Oregon |
| 37 | PA | Pennsylvania |
| 38 | RI | Rhode Island |
| 39 | SC | South Carolina |
| 40 | SD | South Dakota |
| 41 | TN | Tennessee |
| 42 | TX | Texas |
| 43 | UT | Utah |
| 44 | VT | Vermont |
| 45 | VA | Virginia |
| 46 | WA | Washington |
| 47 | WV | West Virginia |
| 48 | WI | Wisconsin |
| 49 | WY | Wyoming |

In [13]: 
```python
# Merging the two data frames based on "State" Columns
cln_data = pd.merge(df, st_data, on='State')
```

In [14]:
```python
# Saving the cleaned data as a .csv file
cln_data.to_csv("Cleaned_data.csv",index=False)
```

In [15]:
```python
# Reading the cleaned data
cln_df = pd.read_csv("Cleaned_data.csv")
```

```
C:\Users\Goutham\AppData\Local\Temp\ipykernel_17944\2152950401.py:2: DtypeWarning: Columns (9) have mixed types. Spe
cify dtype option on import or set low_memory=False.
  cln_df = pd.read_csv("Cleaned_data.csv")
```

```
In [16]: cln_df
```

Out[16]:

	Date received	Product	Sub-product	Issue	Sub-issue	Consumer complaint narrative	Company public response	Company	State	ZIP code	Tags
0	2022-09-14	Debt collection	Credit card debt	Attempts to collect debt not owed	Debt was result of identity theft	To Whom It May Concern : I recently received a...	Company has responded to the consumer and the ...	TRANSUNION INTERMEDIATE HOLDINGS, INC.	FL	33063.0	NaN
1	2022-10-05	Payday loan, title loan, or personal loan	Personal line of credit	Getting a line of credit	NaN	NaN	NaN	Happy Money, Inc.	FL	33138.0	NaN
2	2022-10-15	Credit reporting, credit repair services, or o...	Credit reporting	Problem with a credit reporting company's inve...	Their investigation did not fix an error on yo...	NaN	NaN	Experian Information Solutions Inc.	FL	33076.0	NaN
3	2022-10-15	Credit reporting, credit repair services, or o...	Credit reporting	Improper use of your report	Reporting company used your report improperly	NaN	NaN	Experian Information Solutions Inc.	FL	33617.0	NaN
4	2022-10-11	Credit reporting, credit repair services, or o...	Credit reporting	Improper use of your report	Reporting company used your report improperly	NaN	NaN	Experian Information Solutions Inc.	FL	33573.0	NaN
...	...	...	...	...	...	...	...	...	...	...	...
2953066	2016-11-03	Student loan	Non-federal student loan	Can't repay my loan	Can't get flexible payment options	This is my 2nd complaint I 'm submitting again...	Company has responded to the consumer and the ...	Vermont Student Assistance Corporation	VT	5477.0	NaN

	Date received	Product	Sub-product	Issue	Sub-issue	Consumer complaint narrative	Company public response	Company	State	ZIP code	Tags
2953067	2015-07-02	Mortgage	Other mortgage	Loan modification,collection,foreclosure	NaN	The loan was an 80/20 loan used to purchase th...	NaN	Ocwen Financial Corporation	VT	5091.0	NaN
2953068	2016-05-04	Credit card	NaN	Late fee	NaN	Target XXXX XXXX XXXX XXXX, XXXX. XXXX Attn. :...	NaN	TD BANK US HOLDING COMPANY	VT	5408.0	Olde American
2953069	2018-03-19	Mortgage	Other type of mortgage	Trouble during payment process	NaN	NaN	NaN	CITIZENS FINANCIAL GROUP, INC.	VT	5478.0	NaN
2953070	2017-06-06	Mortgage	Home equity loan or line of credit (HELOC)	Struggling to pay mortgage	NaN	NaN	NaN	PEOPLE'S UNITED BANK, NATIONAL ASSOCIATION	VT	5401.0	NaN

2953071 rows × 19 columns

```python
In [18]:  # Printing the Count of Product Categories
          cln_df['Product'].value_counts()
```

```
Out[18]:  Credit reporting, credit repair services, or other personal consumer reports    1348044
          Debt collection                                                                  440292
          Mortgage                                                                         354993
          Credit card or prepaid card                                                      155509
          Credit reporting                                                                 138417
          Checking or savings account                                                      130919
          Credit card                                                                       87371
          Bank account or service                                                           83583
          Student loan                                                                      68382
          Money transfer, virtual currency, or money service                               40267
          Vehicle loan or lease                                                             35031
          Consumer Loan                                                                     31121
          Payday loan, title loan, or personal loan                                         23795
          Payday loan                                                                        5479
          Money transfers                                                                    5087
          Prepaid card                                                                       3743
          Other financial service                                                            1021
          Virtual currency                                                                     17
          Name: Product, dtype: int64
```

```python
In [20]:  # Printing the Count of Modes of Submission
          cln_df['Submitted via'].value_counts()
```

```
Out[20]:  Web            2493666
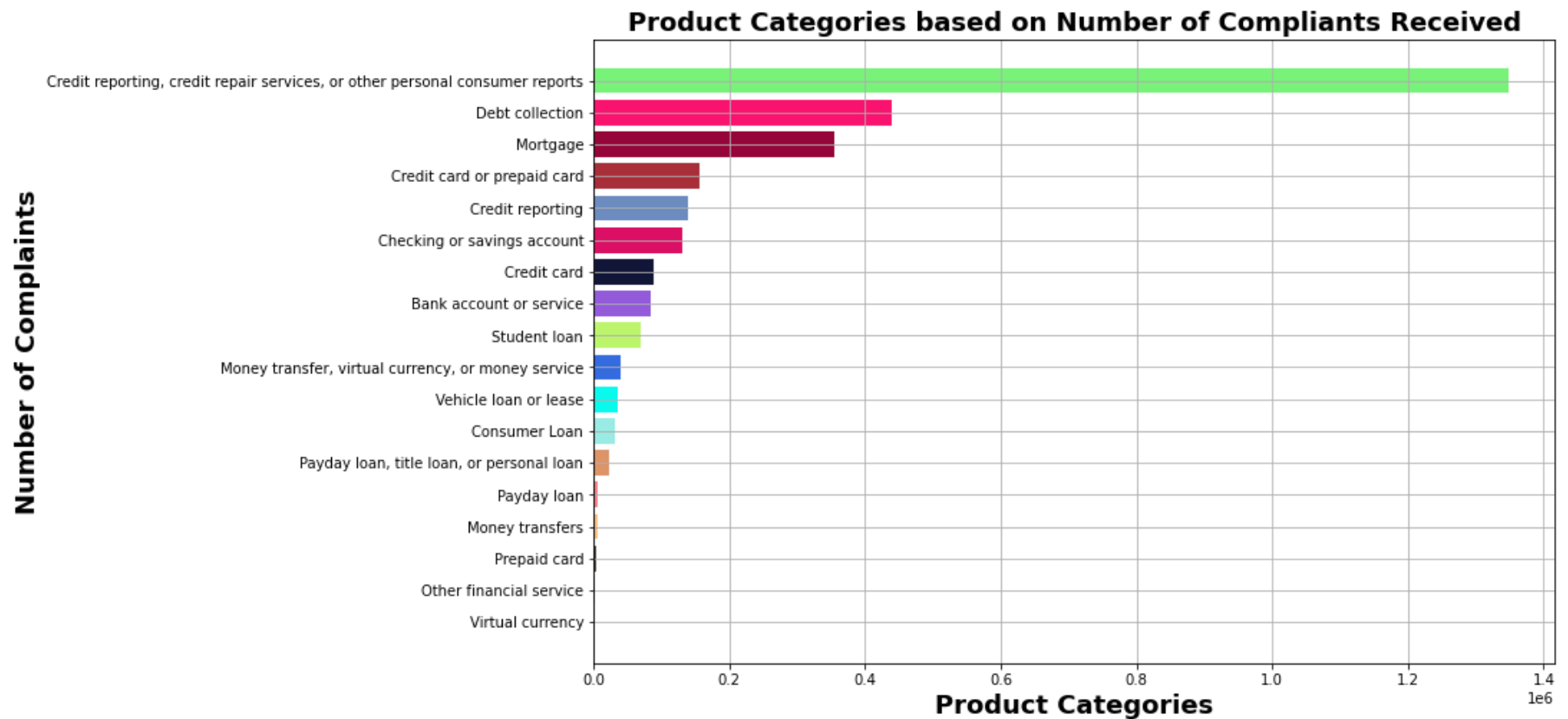          Referral        227974
          Phone           123167
          Postal mail      82930
          Fax              24583
          Web Referral       396
          Email              355
          Name: Submitted via, dtype: int64
```

In [21]:
```python
# Splitting the data to multiple excel files to load into the Tableau
cln_df.iloc[:1000000,:].to_excel("Cleaned_Data_sheet1.xlsx",index=False)
```

```
C:\Users\Goutham\anaconda3\lib\site-packages\xlsxwriter\worksheet.py:1170: UserWarning: Ignoring URL 'https://www.
fiserv.com/%20-%20Pursuant%20to%20my%20previous%20complaint%20number%20filed%20through%20the%20CFPB%20I%20would%20
like%20Fiserv%20to%20be%20aware%20of%20the%20following...When%20Fiserv%20acted%20in%20good%20faith%20and%20release
d%20those%20phone%20calls%20to%20XXXX%20XXXX...XXXX%20XXXX%20refused%20to%20give%20me%20a%20voice%20and%20no%20on
e%20at%20XXXX%20XXXX%20ever%20thought%20to%20ask%20if%20my%20son%20and%20I%20are%20doing%20ok...It%20was%20Fiserv%
20who%20saw%20fit%20to%20ask%20us%20if%20we%20are%20ok%20through%20the%20voice%20of%20Mr.%20XXXX%20XXXX%20of%20Fis
erv...Fiserv%20to%20their%20credit%20did%20give%20me%20a%20voice%20and%20I%20thank%20Fiserv%20for%20this...I%20wou
ld%20like%20Fiserv%20to%20be%20aware%20that%20I%20was%20not%20advised%20that%20those%20phone%20calls%20were%20rele
ased%20to%20XXXX%20XXXX%20which%20then%20set%20me%20up%20to%20fail%20as%20follows...Unlike%20Fiserv%20who%20gave%2
0me%20a%20voice...XXXX%20XXXX%20did%20not%20give%20me%20a%20voice...No%20one%20at%20XXXX%20XXXX%20upper%20level%20
management%20tried%20to%20hear%20my%20voice...XXXX%20XXXX%20at%20XXXX%20XXXX%20was%20the%20only%20one%20decent%20e
nough%20to%20reach%20out%20to%20me%20to%20let%20me%20know%20what%20was%20going%20on...I%20am%20relating%20this%20t
o%20Fiserv%20as%20part%20of%20my%20experience%20report%20which%20was%20a%20direct%20result%20of%20those%20phone%20
calls%20being%20released%20to%20XXXX%20XXXX...With%20Fiserv%20I%20acted%20in%20good%20faith%20to%20de%20escalate%2
0a%20situation%20that%20had%20already%20escalated%20more%20than%20I%20could%20have%20ever%20imagined...I%20thank%2
0Fiserv%20for%20giving%20me%20a%20voice%20and%20for%20seeing%20fit%20to%20offer%20me%20compensation%20through%20th
e%20voice%20of%20XXXX%20XXXX%20whose%20text%20messages%20indicating%20reasonable%20compensation%20be%20taken%20car
e%20of%20in%20my%20regard...Since%20then%20a%20series%20of%20events%20occurred%20that%20served%20to%20escalate%20t
his%20situation%20further%20and%20to%20add%20more%20stress%20to%20my%20already%20burdened%20life...On%20XX/XX/XXX
```

In [22]:
```python
# Splitting the data to multiple excel files to load into the Tableau
cln_df.iloc[1000000:2000000,:].to_excel("Cleaned_Data_sheet2.xlsx",index=False)
```

In [23]:
```python
# Splitting the data to multiple excel files to load into the Tableau
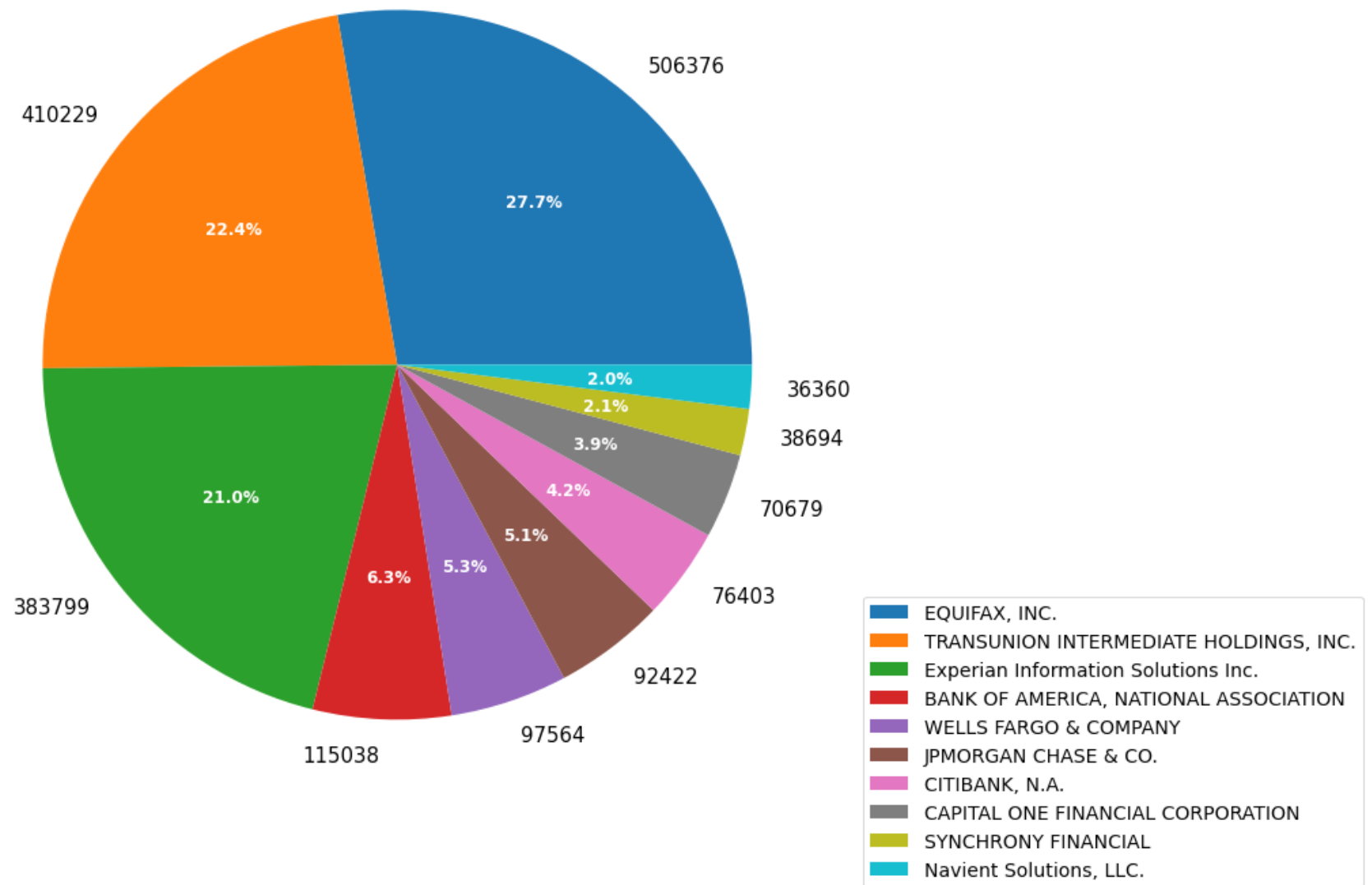cln_df.iloc[2000000:,:].to_excel("Cleaned_Data_sheet2.xlsx",index=False)
```

Exploratory Data Analysis

In [24]:
```python
# Bar Chart Showing "Top 10 Companies which received highest number of complaints"
colors = []
for j in range(18):
    colors.append("#"+''.join([random.choice('ABCDEF0123456789') for i in range(6)]))
plt.figure(figsize=(12,8))
plt.barh(cln_df['Product'].value_counts(ascending=True).index.tolist(), cln_df['Product'].value_counts(ascending=True)
plt.xlabel("Product Categories", fontweight="bold", fontsize=18)
plt.ylabel("Number of Complaints", fontweight="bold", fontsize=18)
plt.title("Product Categories based on Number of Compliants Received", fontweight="bold", fontsize=18)
plt.grid()
plt.show()
```

Product Categories based on Number of Compliants Received

Product Category	(approximate bar position)
Credit reporting, credit repair services, or other personal consumer reports	~1.35
Debt collection	~0.44
Mortgage	~0.36
Credit card or prepaid card	~0.16
Credit reporting	~0.14
Checking or savings account	~0.13
Credit card	~0.09
Bank account or service	~0.08
Student loan	~0.07
Money transfer, virtual currency, or money service	~0.04
Vehicle loan or lease	~0.03
Consumer Loan	~0.03
Payday loan, title loan, or personal loan	~0.02
Payday loan	small
Money transfers	small
Prepaid card	small
Other financial service	small
Virtual currency	small

X-axis label: **Product Categories** (scale 0.0 to 1.4, 1e6)
Y-axis label: **Number of Complaints**

In [25]:
```python
# Pie Chart showing "Top 10 Companies which received highest number of complaint"
plt.figure(figsize=(12,12))
patches, texts, pcts = plt.pie(cln_df['Company'].value_counts().values.tolist()[:10], labels=cln_df['Company'].value_c
plt.legend(cln_df['Company'].value_counts().index.tolist()[:10], bbox_to_anchor =(1.6,-0.10), loc='lower right', fonts
plt.setp(pcts, color='white', fontweight='bold', fontsize=12)
plt.setp(texts, fontsize=15)
plt.title("Top 10 Companies which received highest number of complaints", fontweight="bold", fontsize=20)
plt.grid()
plt.show()
```

Top 10 Companies which received highest number of complaints

Legend:
- EQUIFAX, INC.
- TRANSUNION INTERMEDIATE HOLDINGS, INC.
- Experian Information Solutions Inc.
- BANK OF AMERICA, NATIONAL ASSOCIATION
- WELLS FARGO & COMPANY
- JPMORGAN CHASE & CO.
- CITIBANK, N.A.
- CAPITAL ONE FINANCIAL CORPORATION
- SYNCHRONY FINANCIAL
- Navient Solutions, LLC.

In [26]:
```python
# Bar plot showing "Modes of submission used by Users to register Complaints"
plt.figure(figsize=(12,12))
plt.bar(cln_df['Submitted via'].value_counts().index.tolist(), cln_df['Submitted via'].value_counts().values.tolist(),
plt.xlabel("Modes of Submission of Complaints", fontweight="bold", fontsize=18)
plt.ylabel("Number of Complaints", fontweight="bold", fontsize=18)
plt.title("Modes of submission used by Users to register Complaints", fontweight="bold", fontsize=18)
plt.grid()
for i in range(len(df['Submitted via'].value_counts().values.tolist())):
  plt.text(i, cln_df['Submitted via'].value_counts().values.tolist()[i], cln_df['Submitted via'].value_counts().values
plt.show()
```

Modes of submission used by Users to register Complaints
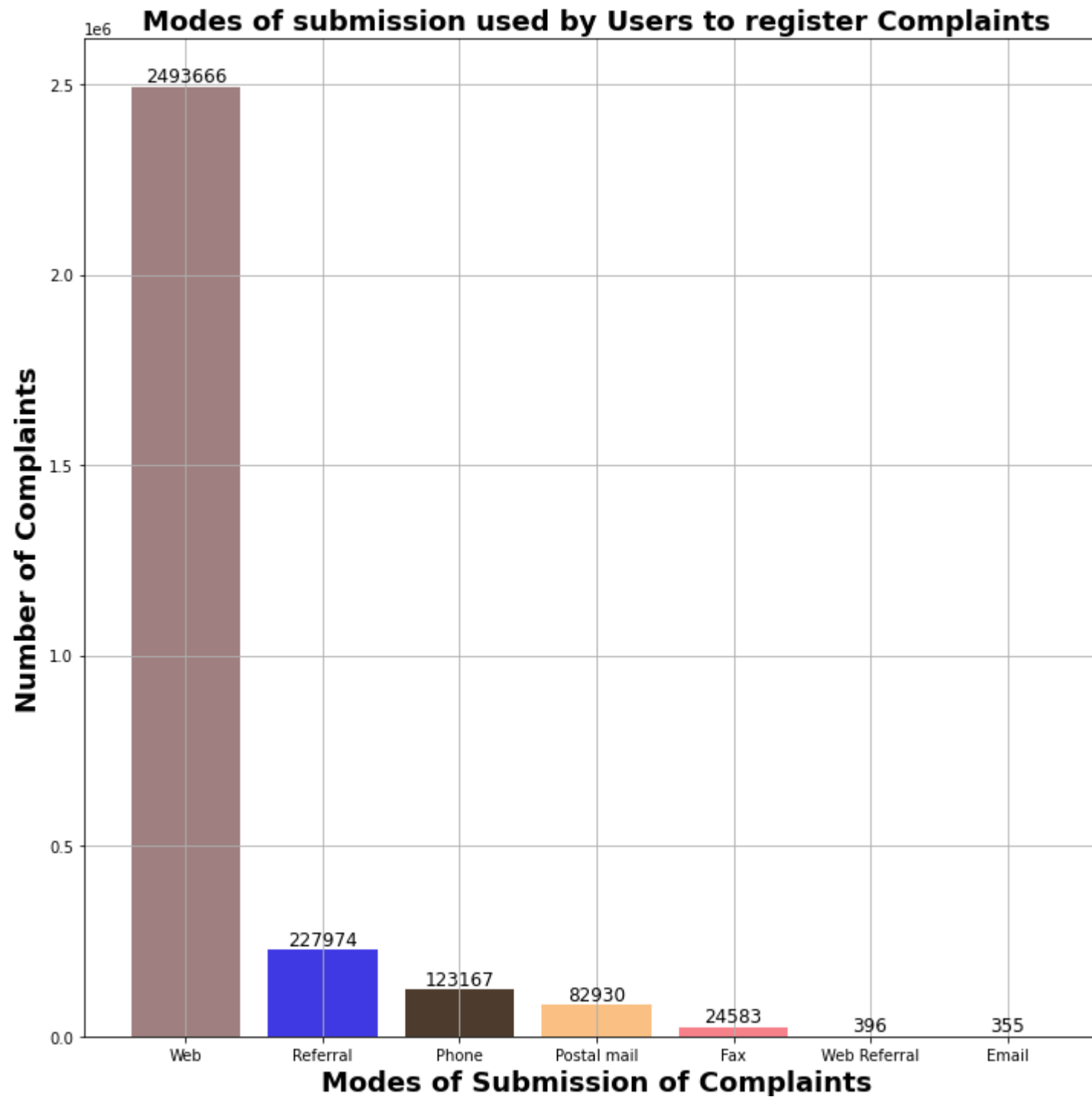
In [27]:
```python
temp_df = pd.DataFrame({
    'State' : cln_df['State'].value_counts().index,
    'Count' : cln_df['State'].value_counts().values
})
```

In [28]:
```python
# Choropleth Map Showing "State wise complaint Analysis"
import plotly.express as px
fig = px.choropleth(temp_df,
                    locations='State',
                    locationmode="USA-states",
                    scope="usa",
                    color='Count',
                    color_continuous_scale="inferno_r",
                    )


fig.update_layout(
    title_text = 'Number of Complaints by State',
    title_font_family="Times New Roman",
    title_font_size = 22,
    title_font_color="black",
    title_x=0.45,
        )
```

In [29]:
```python
# Pushing the Choropleth map to Plotly chart Studio

username = 'GSGoutham'
api_key = '4DjQpkNhJhb0LNUeIA52'
chart_studio.tools.set_credentials_file(username=username,api_key=api_key)
py.plot(fig,filename = "Complaints by State", auto_open = False)
```

Out[29]: 'https://plotly.com/~GSGoutham/1/'

In [ ]: