

```
In [65]: import numpy as np
import pandas as pd
!pip install ydata-profiling
from pandas_profiling import ProfileReport
import matplotlib.pyplot as plt
import seaborn as sns
```

```
Requirement already satisfied: ydata-profiling in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (4.6.4)
Requirement already satisfied: scipy<1.12,>=1.4.1 in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from ydata-profiling) (1.11.1)
Requirement already satisfied: pandas!=1.4.0,<3,>1.1 in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from ydata-profiling) (2.0.3)
Requirement already satisfied: matplotlib<3.9,>=3.2 in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from ydata-profiling) (3.7.2)
Requirement already satisfied: pydantic>=2 in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from ydata-profiling) (2.5.3)
Requirement already satisfied: PyYAML<6.1,>=5.0.0 in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from ydata-profiling) (6.0)
Requirement already satisfied: jinja2<3.2,>=2.11.1 in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from ydata-profiling) (3.1.2)
Requirement already satisfied: visions[type_image_path]==0.7.5 in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from ydata-profiling) (0.7.5)
Requirement already satisfied: numpy<1.26,>=1.16.0 in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from ydata-profiling) (1.24.3)
Requirement already satisfied: htmlmin==0.1.12 in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from ydata-profiling) (0.1.12)
```

```
In [ ]:
```

```
In [66]: data = pd.read_csv('complaints-2024-01-17_11_41.csv')
```

In [67]: `data.head()`

Out[67]:

	Date received	Product	Sub-product	Issue	Sub-issue	Consumer complaint narrative	Company public response	Company	State	ZIP code	Tags	Consumer consent provided?	Subr
0	06/13/23	Mortgage	Conventional home mortgage	Struggling to pay mortgage		NaN	Company has responded to the consumer and the ...	BANK OF AMERICA, NATIONAL ASSOCIATION	FL	33483	NaN	NaN	F
1	04/13/23	Mortgage	Conventional home mortgage	Struggling to pay mortgage		NaN	I am a XXXX XXXX XXXX XXXX veteran and I ...	21ST MORTGAGE CORP.	IN	474XX	Servicemember	Consent provided	
2	09/20/23	Mortgage	Conventional home mortgage	Applying for a mortgage or misleading advertising or refinancing an exist...	Confusing or misleading advertising or marketing	I was preapproved for a conventional mortgage ...	NaN	NATIONS RELIABLE LENDING LLC	KY	41018	NaN	Consent provided	
3	09/09/23	Mortgage	Conventional home mortgage	Trouble during payment process	Payment process	I paid {\$100000.00} to Freedom Mortgage on XX/...	Company has responded to the consumer and the ...	Freedom Mortgage Company	DC	20009	NaN	Consent provided	
4	11/06/23	Mortgage	Conventional home mortgage	Trouble during payment process	Escrow, taxes, or insurance	Contacted mortgage servicer XX/XX/2023 to requ...	Company believes it acted appropriately as aut...	SELECT PORTFOLIO SERVICING, INC.	TN	38017	Older American	Consent provided	

In [68]: `data.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 71720 entries, 0 to 71719
Data columns (total 18 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Date received    71720 non-null   object  
 1   Product          71720 non-null   object  
 2   Sub-product      71720 non-null   object  
 3   Issue            71720 non-null   object  
 4   Sub-issue         9881 non-null   object  
 5   Consumer complaint narrative  37813 non-null   object  
 6   Company public response    36110 non-null   object  
 7   Company          71720 non-null   object  
 8   State             71288 non-null   object  
 9   ZIP code          71386 non-null   object  
 10  Tags              19721 non-null   object  
 11  Consumer consent provided? 56955 non-null   object  
 12  Submitted via     71720 non-null   object  
 13  Date sent to company 71720 non-null   object  
 14  Company response to consumer 71719 non-null   object  
 15  Timely response?    71720 non-null   object  
 16  Consumer disputed?   0 non-null      float64 
 17  Complaint ID       71720 non-null   int64  
dtypes: float64(1), int64(1), object(16)
memory usage: 9.8+ MB
```

```
In [69]: data.isnull().sum()
```

```
Out[69]: Date received          0  
Product              0  
Sub-product          0  
Issue                0  
Sub-issue             61839  
Consumer complaint narrative 33907  
Company public response   35610  
Company               0  
State                 432  
ZIP code              334  
Tags                  51999  
Consumer consent provided? 14765  
Submitted via          0  
Date sent to company    0  
Company response to consumer 1  
Timely response?        0  
Consumer disputed?      71720  
Complaint ID           0  
dtype: int64
```

```
In [70]: profile = ProfileReport(data, title='Pandas Profiling Report', explorative=True)
```

```
# Save the report to an HTML file  
profile.to_file("output_report.html")
```

```
/Users/jahnnavireddyganesina/anaconda3/lib/python3.11/site-packages/ydata_profiling/profile_report.py:354: UserWarning: Try running command: 'pip install --upgrade Pillow' to avoid ValueError  
warnings.warn(
```

```
Summarize dataset: 0%|          | 0/5 [00:00<?, ?it/s]
```

```
Generate report structure: 0%|          | 0/1 [00:00<?, ?it/s]
```

```
Render HTML: 0%|          | 0/1 [00:00<?, ?it/s]
```

```
Export report to file: 0%|          | 0/1 [00:00<?, ?it/s]
```

```
In [72]: data.columns
```

```
Out[72]: 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 [73]: data.drop(['Sub-issue', 'Company public response', 'ZIP code', 'Tags', 'Consumer disputed?'], axis=1, inplace=True)
```

In [74]: `data.head()`

Out[74]:

	Date received	Product	Sub-product	Issue	Consumer complaint narrative	Company	State	Consumer consent provided?	Submitted via	Date sent to company	Company response to consumer	Timely response?	Complaint ID
0	06/13/23	Mortgage	Conventional home mortgage	Struggling to pay mortgage	NaN	BANK OF AMERICA, NATIONAL ASSOCIATION	FL	NaN	Phone	06/13/23	Closed with explanation	Yes	71115
1	04/13/23	Mortgage	Conventional home mortgage	Struggling to pay mortgage	I am a XXXX XXXX XXXX XXXX veteran and I ...	21ST MORTGAGE CORP.	IN	Consent provided	Web	05/18/23	Closed with explanation	Yes	68346
2	09/20/23	Mortgage	Conventional home mortgage	Applying for a mortgage or refinancing an exist...	I was preapproved for a conventional mortgage ...	NATIONS RELIABLE LENDING LLC	KY	Consent provided	Web	09/20/23	Closed with explanation	No	75806
3	09/09/23	Mortgage	Conventional home mortgage	Trouble during payment process	I paid {\$100000.00} to Freedom Mortgage on XX/...	Freedom Mortgage Company	DC	Consent provided	Web	09/09/23	Closed with explanation	Yes	75246
4	11/06/23	Mortgage	Conventional home mortgage	Trouble during payment process	Contacted mortgage servicer XX/XX/2023 to requ...	SELECT PORTFOLIO SERVICING, INC.	TN	Consent provided	Web	11/06/23	Closed with explanation	Yes	78065

Type Markdown and LaTeX: α^2

```
In [75]: data.isnull().sum()
```

```
Out[75]: Date received          0  
Product              0  
Sub-product          0  
Issue                0  
Consumer complaint narrative  33907  
Company              0  
State                432  
Consumer consent provided?   14765  
Submitted via         0  
Date sent to company    0  
Company response to consumer 1  
Timely response?        0  
Complaint ID          0  
dtype: int64
```

```
In [76]: #Data Cleaning
```

```
data['State'].fillna('Unknown', inplace=True)
```

```
In [77]: data.isnull().sum()
```

```
Out[77]: Date received          0  
Product              0  
Sub-product          0  
Issue                0  
Consumer complaint narrative  33907  
Company              0  
State                0  
Consumer consent provided?   14765  
Submitted via         0  
Date sent to company    0  
Company response to consumer 1  
Timely response?        0  
Complaint ID          0  
dtype: int64
```

```
In [78]: data.dropna(inplace=True)
```

```
In [79]: data.isnull().sum()
```

```
Out[79]: Date received      0  
Product          0  
Sub-product      0  
Issue            0  
Consumer complaint narrative 0  
Company          0  
State            0  
Consumer consent provided? 0  
Submitted via    0  
Date sent to company 0  
Company response to consumer 0  
Timely response?   0  
Complaint ID     0  
dtype: int64
```

```
In [80]: #Checking for duplicates  
duplicate_rows = data[data.duplicated()]  
  
# Display the duplicate rows  
print("Duplicate Rows:")  
print(duplicate_rows)
```

```
Duplicate Rows:  
Empty DataFrame  
Columns: [Date received, Product, Sub-product, Issue, Consumer complaint narrative, Company, State, Consumer consent provided?, Submitted via, Date sent to company, Company response to consumer, Timely response?, Complaint ID]  
Index: []
```

```
In [81]: data['Timely response?'] = data['Timely response?'].replace({'Yes': 1, 'No': 0})
data.head()
```

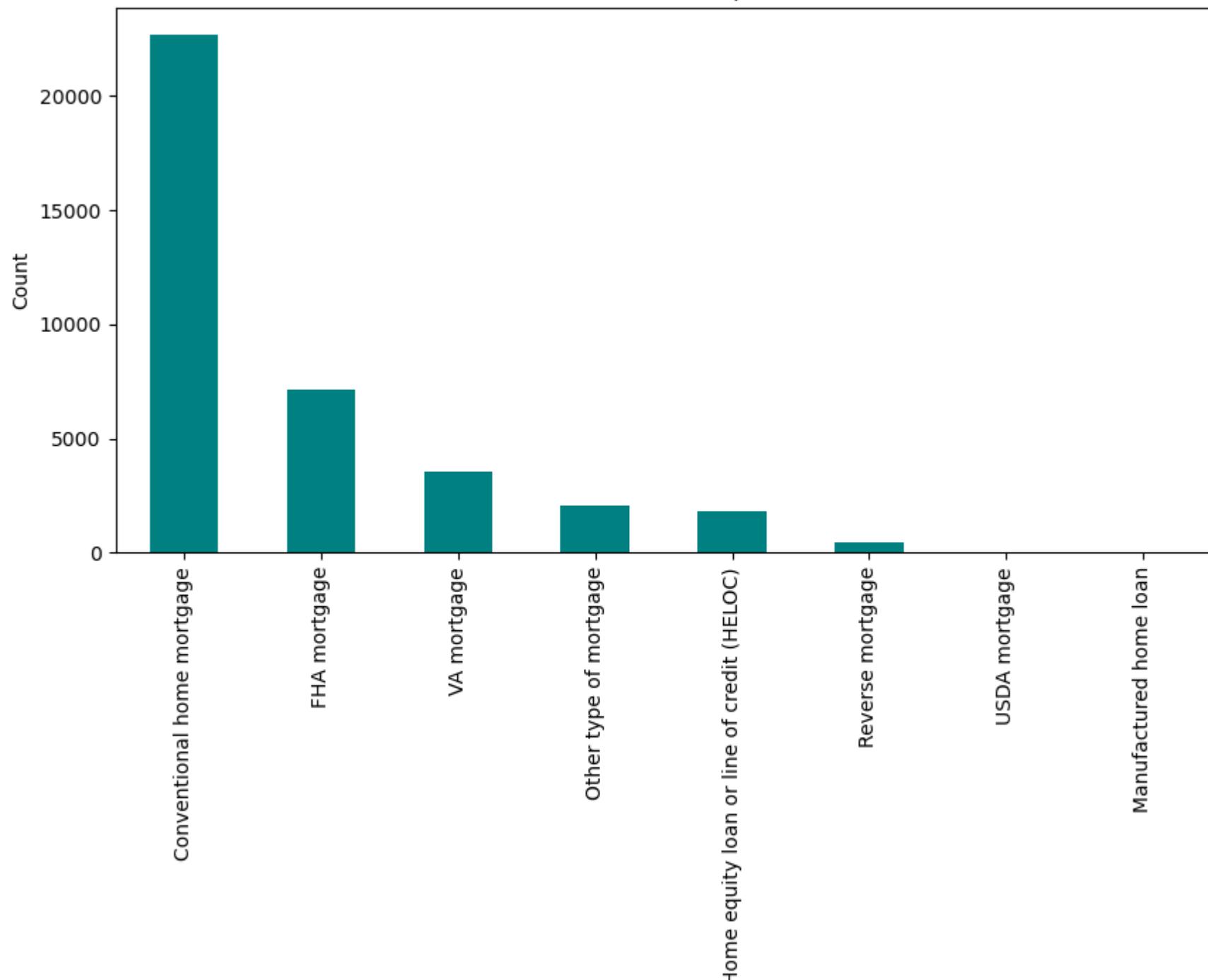
Out[81]:

	Date received	Product	Sub-product	Issue	Consumer complaint narrative	Company	State	Consumer consent provided?	Submitted via	Date sent to company	Company response to consumer	Timely response?	Complain ID
1	04/13/23	Mortgage	Conventional home mortgage	Struggling to pay mortgage	I am a XXXX XXXX XXXX XXXX veteran and I ...	21ST MORTGAGE CORP.	IN	Consent provided	Web	05/18/23	Closed with explanation	1	6834651
2	09/20/23	Mortgage	Conventional home mortgage	Applying for a mortgage or refinancing an exist...	I was preapproved for a conventional mortgage ...	NATIONS RELIABLE LENDING LLC	KY	Consent provided	Web	09/20/23	Closed with explanation	0	7580627
3	09/09/23	Mortgage	Conventional home mortgage	Trouble during payment process	I paid {\$100000.00} to Freedom Mortgage on XX/...	Freedom Mortgage Company	DC	Consent provided	Web	09/09/23	Closed with explanation	1	7524616
4	11/06/23	Mortgage	Conventional home mortgage	Trouble during payment process	Contacted mortgage servicer XX/XX/2023 to requ...	SELECT PORTFOLIO SERVICING, INC.	TN	Consent provided	Web	11/06/23	Closed with explanation	1	7806521
5	01/12/23	Mortgage	Conventional home mortgage	Trouble during payment process	After being told by Rocket Mortgage 's " Custo...	Rocket Mortgage, LLC	CA	Consent provided	Web	01/12/23	Closed with explanation	1	6431206

```
In [116]: data.to_excel('datafinalreqd.xlsx', index=False)
```

```
In [82]: #Data Visualization  
plt.figure(figsize=(10, 5))  
%matplotlib inline  
data['Sub-product'].value_counts().plot(kind='bar', color='teal')  
plt.title('Distribution of Sub-product')  
plt.xlabel('Sub-products')  
plt.ylabel('Count')  
plt.show()
```


Distribution of Sub-product



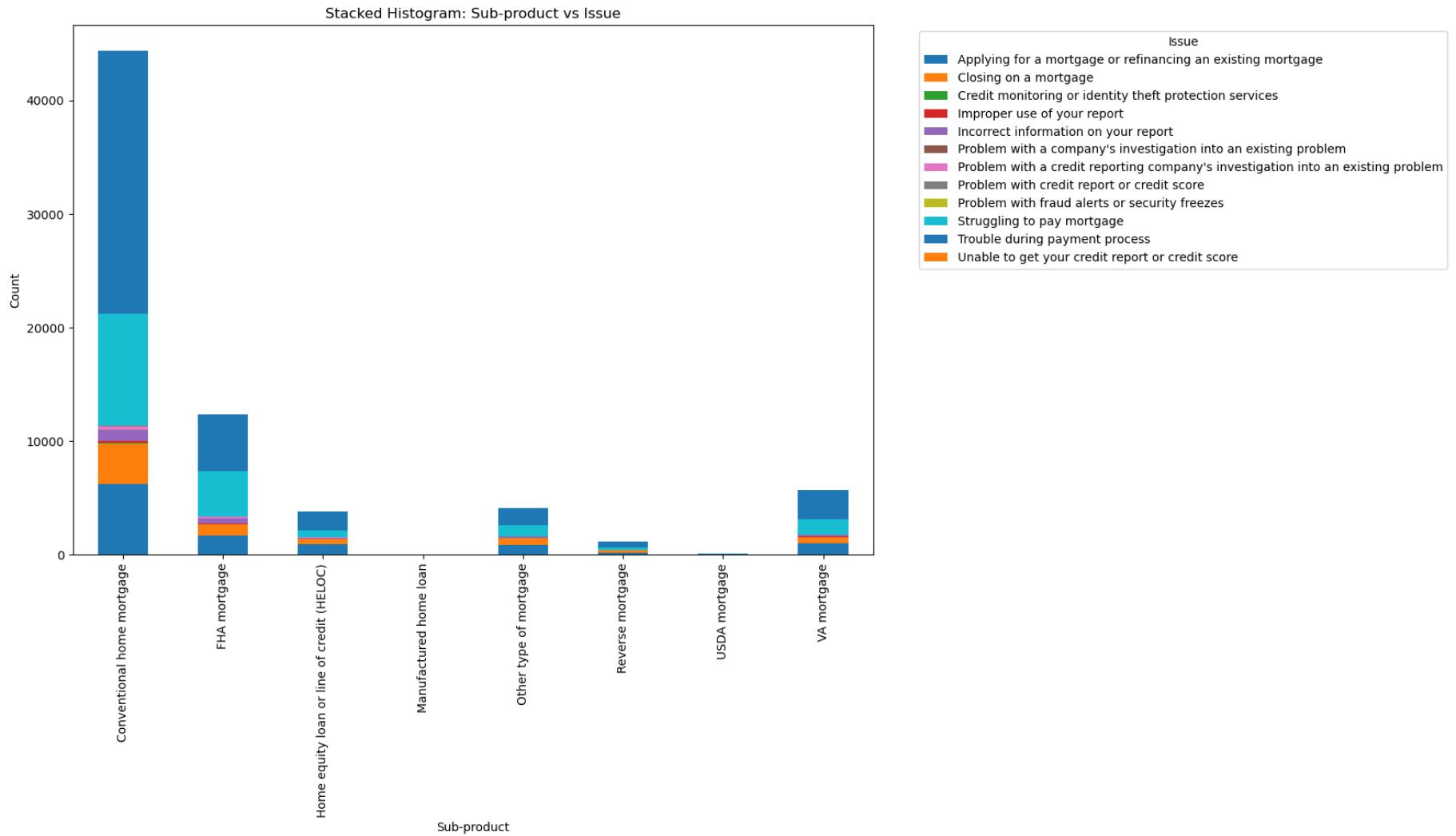
Sub-products

In [212]:

```
# Group by 'Sub-product' and 'Issue' and count occurrences
stacked_data = data.groupby(['Sub-product', 'Issue']).size().unstack(fill_value=0)

# Plot the stacked histogram
stacked_data.plot(kind='bar', stacked=True, figsize=(12, 8))
plt.title('Stacked Histogram: Sub-product vs Issue')
plt.xlabel('Sub-product')
plt.ylabel('Count')
plt.legend(title='Issue', bbox_to_anchor=(1.05, 1), loc='upper left')

plt.show()
```



```
In [86]: ! pip install pywaffle  
from pywaffle import Waffle
```

```
Requirement already satisfied: pywaffle in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (1.1.0)  
Requirement already satisfied: fontawesomefree in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from pywaffle) (6.5.1)  
Requirement already satisfied: matplotlib in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from pywaffle) (3.7.2)  
Requirement already satisfied: contourpy>=1.0.1 in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from matplotlib->pywaffle) (1.0.5)  
Requirement already satisfied: cycler>=0.10 in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from matplotlib->pywaffle) (0.11.0)  
Requirement already satisfied: fonttools>=4.22.0 in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from matplotlib->pywaffle) (4.25.0)  
Requirement already satisfied: kiwisolver>=1.0.1 in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from matplotlib->pywaffle) (1.4.4)  
Requirement already satisfied: numpy>=1.20 in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from matplotlib->pywaffle) (1.24.3)  
Requirement already satisfied: packaging>=20.0 in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from matplotlib->pywaffle) (23.1)  
Requirement already satisfied: pillow>=6.2.0 in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from matplotlib->pywaffle) (9.4.0)  
Requirement already satisfied: pyparsing<3.1,>=2.3.1 in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from matplotlib->pywaffle) (3.0.9)  
Requirement already satisfied: python-dateutil>=2.7 in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from matplotlib->pywaffle) (2.8.2)  
Requirement already satisfied: six>=1.5 in /Users/jahnavireddyganesina/anaconda3/lib/python3.11/site-packages (from python-dateutil>=2.7->matplotlib->pywaffle) (1.16.0)
```

```
In [ ]:
```

In [54]:

```
# Prepare Data
df = data.groupby('Submitted via').size().reset_index(name='counts')
n_categories = df.shape[0]
colors = [plt.cm.inferno_r(i/float(n_categories)) for i in range(n_categories)]

# Create Waffle chart
fig = plt.figure(figsize=(25, 12))
waffle_plot = Waffle(
    rows=9,
    columns=26,
    values=df['counts'],
    labels=["{} ({})".format(n[0], n[1]) for n in df[['Submitted via', 'counts']].itertuples()],
    legend={'loc': 'upper left', 'bbox_to_anchor': (1.05, 1), 'fontsize': 12},
    title={'label': '# Counts by Submitted via', 'loc': 'center', 'fontsize': 18},
    colors=colors
)
waffle_plot
```

Out [54]:

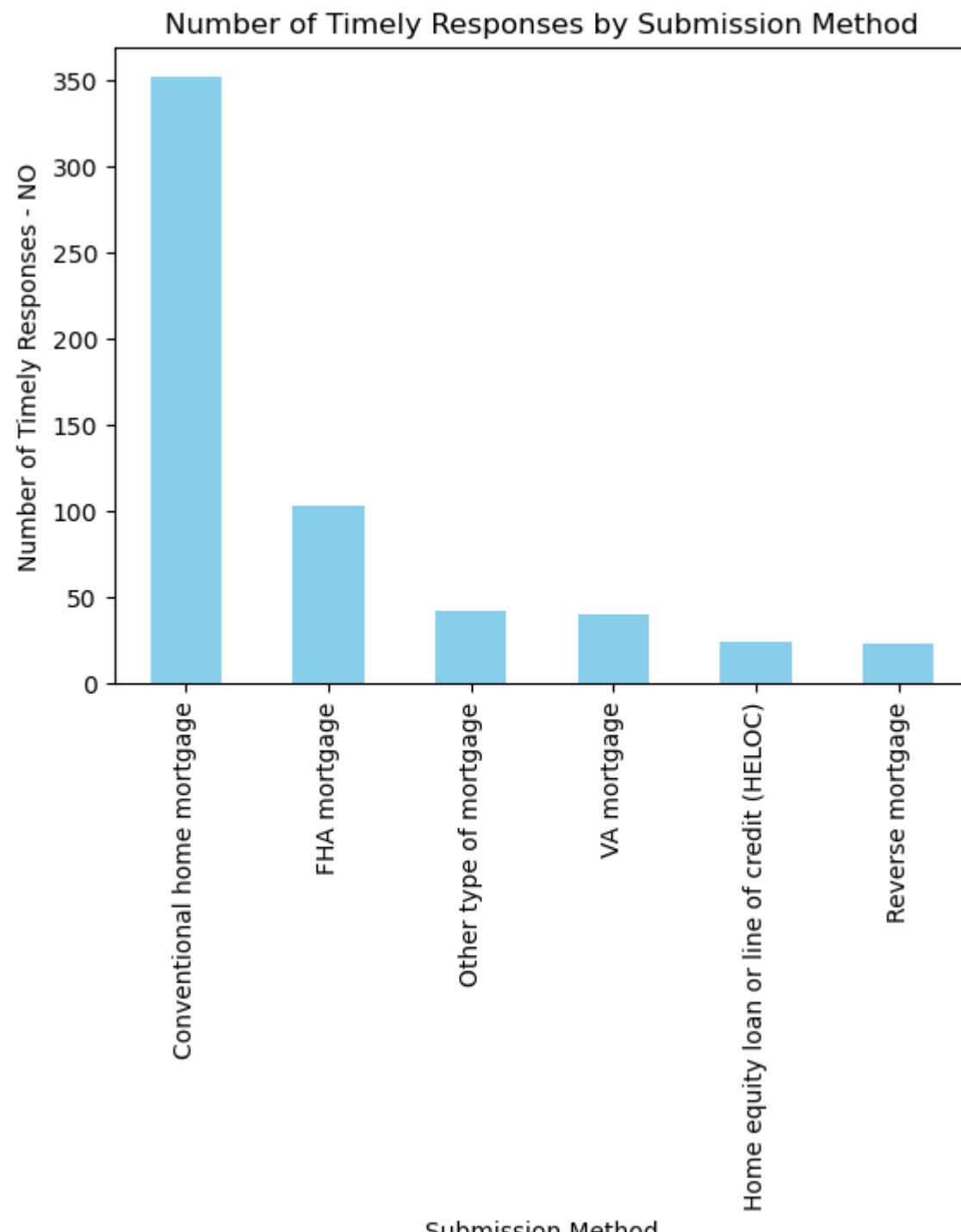


<Figure size 2500x1200 with 0 Axes>

```
In [88]: false_response = data[data['Timely response?'] == 0]
false_response

# Group by 'Submitted via' and count the number of True responses
submission_counts = false_response['Sub-product'].value_counts()

# Plot the results using matplotlib
submission_counts.plot(kind='bar', color='skyblue')
plt.title('Number of Timely Responses by Submission Method')
plt.xlabel('Submission Method')
plt.ylabel('Number of Timely Responses - NO')
plt.show()
```


```
In [102]: #Overall Summary of Inefficient banks
# Group by 'Company' and count the number of issues and timely responses
grouped_df = data.groupby('Company').agg({'Issue': 'count', 'Timely response?': lambda x: (x == 0).sum()})

# Rename the columns for clarity
grouped_df.rename(columns={'Issue': 'Number of Issues', 'Timely response?': 'Count of Timely Response 0'}, in

# Add a new column '% Distribution of Issues'
grouped_df['% Distribution of Issues'] = ((grouped_df['Count of Timely Response 0']) / grouped_df['Number of I

# Sort the table in descending order based on '% Distribution of Issues'
grouped_df = grouped_df.sort_values(by='Number of Issues', ascending=False)

# Create a table using Matplotlib
fig, ax = plt.subplots(figsize=(15, 5))

table_data = grouped_df.reset_index().values.tolist()

table = ax.table(cellText=table_data,
                  colLabels=['Company', 'Number of Issues', 'Count of Timely Response 0', '% Distribution of I
cellLoc='center',
loc='center')

table.auto_set_font_size(False)
table.set_fontsize(10)
table.scale(1.2, 1.2)

ax.axis('off') # Hide the axes

plt.show()
```

DHI Mortgage Company	32	4	12.5
HILLTOP HOLDINGS, INC	31	0	0.0
Cornerstone Home Lending Inc.	31	0	0.0
Click n' Close, Inc.	30	10	33.33
INTERCONTINENTAL CAPITAL GROUP	30	2	6.67
JFQ Lending, LLC	28	4	14.29
AMERICAN INTERNET MORTGAGE	28	0	0.0
Southwest Business Corporation	28	1	3.57
FIRST HOME MORTGAGE CORPORATION	27	4	14.81
CountryPlace Acceptance Corporation	26	3	11.54
NVR INC.	26	0	0.0
PulteGroup, Inc.	26	0	0.0
Mortgage Lenders Investment Trading Corp	26	0	0.0
EQUITY LOANS LLC	24	0	0.0
BANK OF THE WEST	24	0	0.0
TFS FINANCIAL CORPORATION	24	0	0.0
EQUIFAX, INC.	23	0	0.0
NATIONS LENDING CORPORATION	23	0	0.0
Security National Financial Corp	23	0	0.0
INTERFIRST MORTGAGE COMPANY	23	0	0.0
ZIONS BANCORPORATION	22	1	4.55
FIRST NATIONAL BANK OF PENNSYLVANIA	22	0	0.0
ALLIANT CREDIT UNION	22	0	0.0
SOUTHSTATE CORPORATION	21	0	0.0
ighborhood Assistance Corporation of America ("NA	21	1	4.76
FULTON FINANCIAL CORPORATION	21	0	0.0

```
In [103]: # Find average and median of 'Number of Issues'
average_number_of_issues = grouped_df['Number of Issues'].mean()
median_number_of_issues = grouped_df['Number of Issues'].median()

print(f'Average Number of Issues: {average_number_of_issues:.2f}')
print(f'Median Number of Issues: {median_number_of_issues}')

#skewed towards the left
```

Average Number of Issues: 40.88
Median Number of Issues: 2.0


```
In [2]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

data = pd.read_csv('/Users/vaishnavipoluru/Desktop/dataforproj.csv')

# Filter data for 'Mortgage'
product_data = data[data['Product'] == 'Mortgage']['Issue']

# Count occurrences of each issue
issue_counts = product_data.value_counts()

# Identify values to be grouped into 'Others' (e.g., values with less than 5%)
threshold_percentage = 2.5
mask = issue_counts / issue_counts.sum() < threshold_percentage / 100
values_to_combine = issue_counts.index[mask]

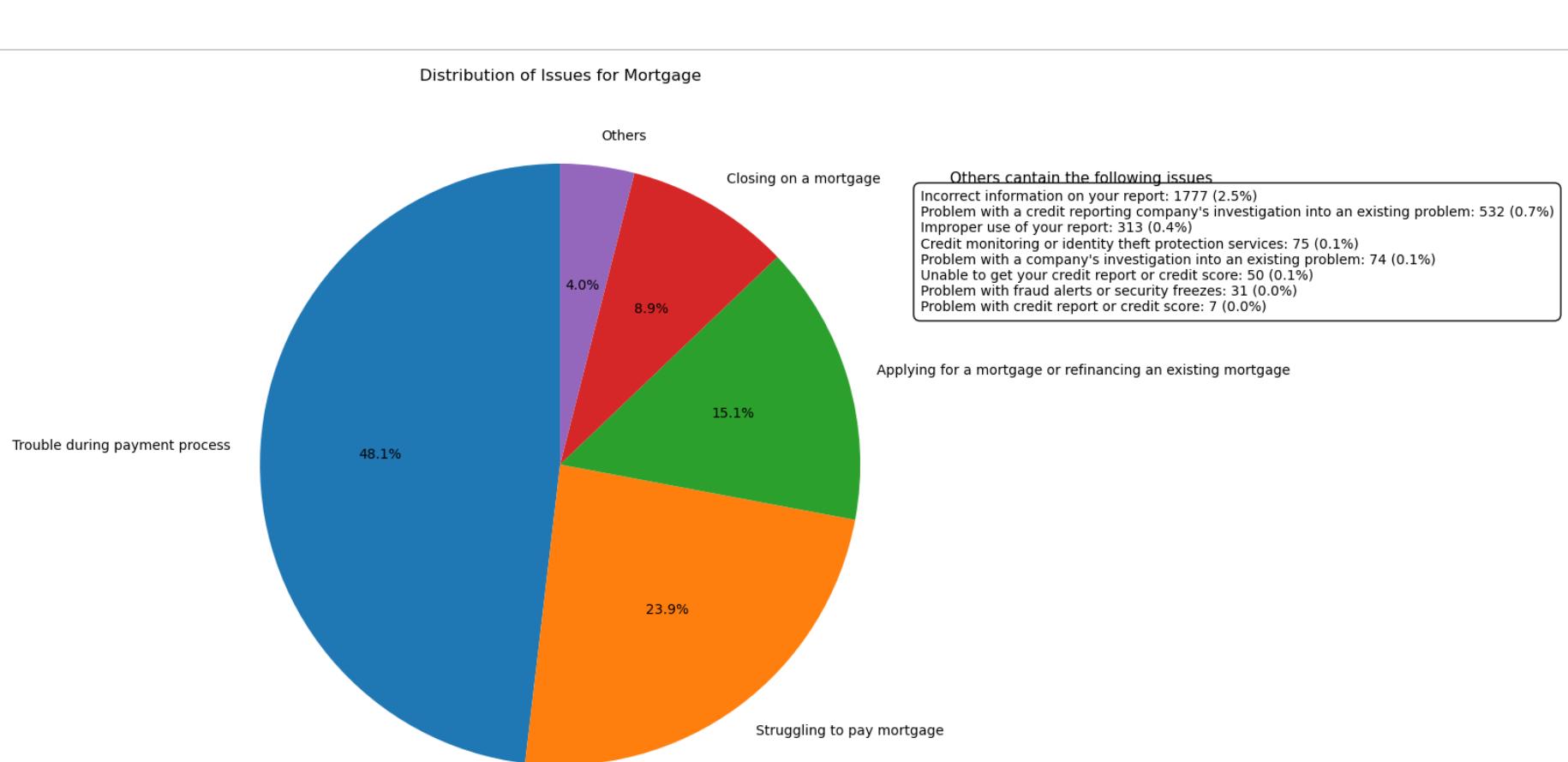
# Create a modified dataset with 'Others'
modified_issue_counts = issue_counts.copy()
others_count = issue_counts[values_to_combine].sum()
modified_issue_counts.loc['Others'] = others_count
modified_issue_counts.drop(index=values_to_combine, inplace=True)

# Create a pie chart
plt.figure(figsize=(10, 10))
wedges, texts, autotexts = plt.pie(modified_issue_counts, labels=modified_issue_counts.index, autopct='%.1f%')

# Add legend for 'Others'
legend_text = '\n'.join([f'{issue}: {count} ({percentage:.1f}%)' for issue, count, percentage in zip(values_to_combine, issue_counts, issue_counts / issue_counts.sum() * 100)])
plt.text(1.2, 0.5, legend_text, verticalalignment='bottom', fontsize=10, bbox=dict(facecolor='none', edgecolor='black', boxstyle='round', pad=10))

# Add title to the legend
legend_title = 'Others contain the following issues'
plt.text(1.3, 0.95, legend_title, fontsize=11, color='black', verticalalignment='center')

plt.title('Distribution of Issues for Mortgage')
plt.show()
```



In [215]:

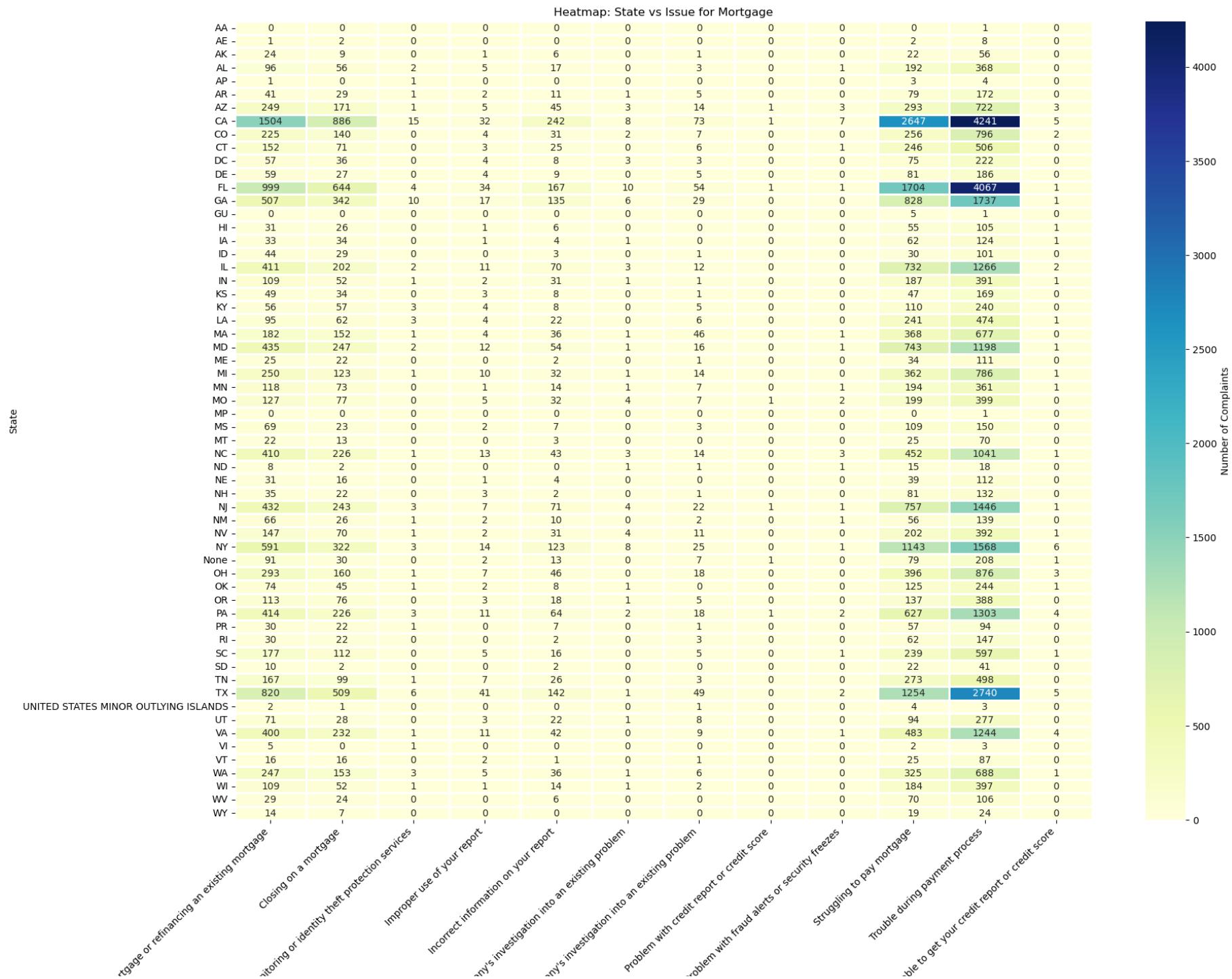
```
# Filter data for 'Mortgage'
product_data = data[data['Product'] == 'Mortgage']

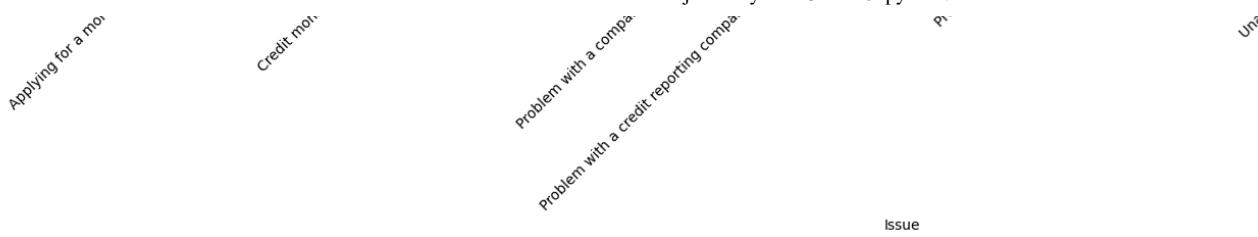
# Create a pivot table for "State" vs "Issue"
pivot_table = product_data.pivot_table(index='State', columns='Issue', aggfunc='size', fill_value=0)

# Create a heatmap
plt.figure(figsize=(20, 15))
sns.heatmap(pivot_table, cmap='YlGnBu', annot=True, fmt='d', linewidths=2, cbar_kws={'label': 'Number of Comp'})

# Wrap text on the y-axis
plt.yticks(rotation=0)
plt.xticks(rotation=45, ha='right') # Rotate x-axis labels for better readability

plt.title('Heatmap: State vs Issue for Mortgage')
plt.show()
```




```
In [6]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

data = pd.read_csv('/Users/vaishnavipoluru/Desktop/dataforproj.csv')

# Convert 'Date received' to datetime format with the specified format specifier
data['Date received'] = pd.to_datetime(data['Date received'], format='%d-%b-%y')

# Extract Month from 'Date received'
data['Month'] = data['Date received'].dt.strftime('%B')

# Define the chronological order of months
month_order = [
    'January', 'February', 'March', 'April', 'May', 'June',
    'July', 'August', 'September', 'October', 'November', 'December'
]

# Convert 'Month' to categorical with the defined order
data['Month'] = pd.Categorical(data['Month'], categories=month_order, ordered=True)

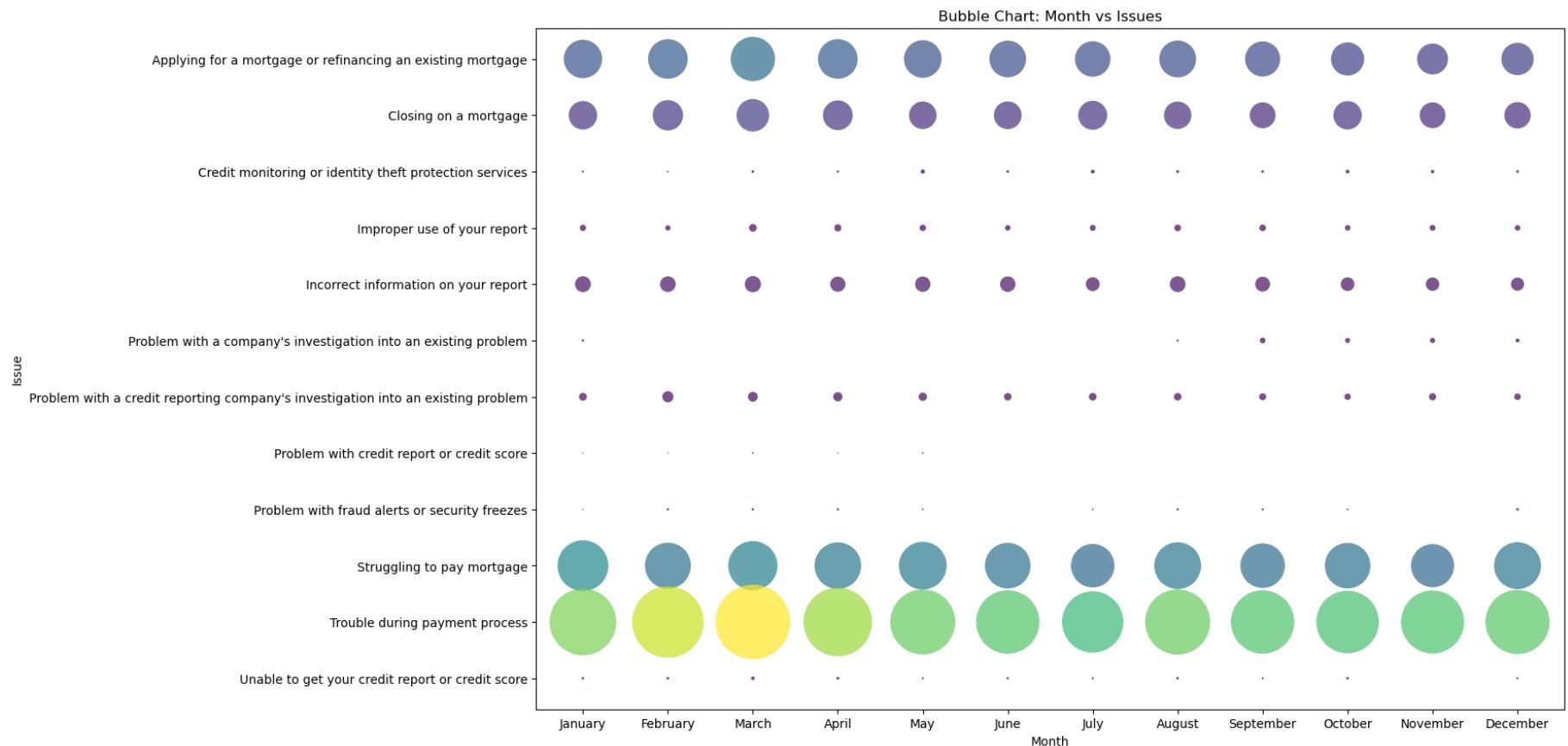
# Group by 'Month' and 'Issue' and count occurrences
bubble_data = data.groupby(['Month', 'Issue']).size().reset_index(name='Count')

# Define the size of the bubbles based on the count
bubble_data['Size'] = bubble_data['Count'] # Adjust the multiplier to change bubble size

# Set up the size range for the bubbles
min_size = bubble_data['Size'].min()
max_size = bubble_data['Size'].max()

# Create the bubble plot with x-axis as 'Month' and y-axis as 'Issue'
plt.figure(figsize=(15, 10)) # Increase the size of the chart
sns.scatterplot(data=bubble_data, x='Month', y='Issue', size='Size', hue='Count', sizes=(min_size, max_size),
plt.title('Bubble Chart: Month vs Issues')
plt.xlabel('Month')
plt.ylabel('Issue')

plt.show()
```



```
In [7]: # Data Preprocessing
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

data = pd.read_csv('/Users/vaishnavipoluru/Desktop/dataforproj.csv')

data['Date received'] = pd.to_datetime(data['Date received'])
data.dropna(subset=['Date received', 'Sub-product'], inplace=True)

# Create Time Series Data
data['Month'] = data['Date received'].dt.month
data['Sub-product'] = data['Sub-product'].astype(str) # Ensure 'Sub product' is treated as categorical
time_series_data = data.groupby(['Month', 'Sub-product']).size().unstack()

# Plotting
plt.figure(figsize=(12, 6))
time_series_data.plot(kind='line', marker='o')
plt.title('Complaints by Sub Product Across Months')
plt.xlabel('Month')
plt.ylabel('Number of Complaints')
plt.legend(title='Sub-product', bbox_to_anchor=(1, 1))
plt.grid(True)
plt.show()
```

<Figure size 1200x600 with 0 Axes>

