

## ✓ Load data

- Train data as `train_data`
- Test data as `test_data`

```
import pandas as pd

TRAIN_FILE_PATH = '/content/train.csv'
TEST_FILE_PATH = '/content/test.csv'

train_data = pd.read_csv(TRAIN_FILE_PATH)
test_data = pd.read_csv(TEST_FILE_PATH)

# Remove null/empty rows
train_data = train_data[~train_data['crimeadditionalinfo'].isnull()]

# List few entries from data
train_data.head()
```

|   | category                              | sub_category                    | crimeadditionalinfo                               |
|---|---------------------------------------|---------------------------------|---|
| 0 | Online and Social Media Related Crime | Cyber Bullying Stalking Sexting | I had continue received random calls and abusi... |
| 1 | Online Financial Fraud                | Fraud CallVishing               | The above fraudster is continuously messaging ... |
| 2 | Online Gambling Betting               | Online Gambling Betting         | He is acting like a police and demanding for m... |
| 3 | Online and Social Media Related Crime | Online Job Fraud                | In apna Job I have applied for job interview f... |
| 4 | Online Financial Fraud                | Fraud CallVishing               | I received a call from lady stating that she w... |

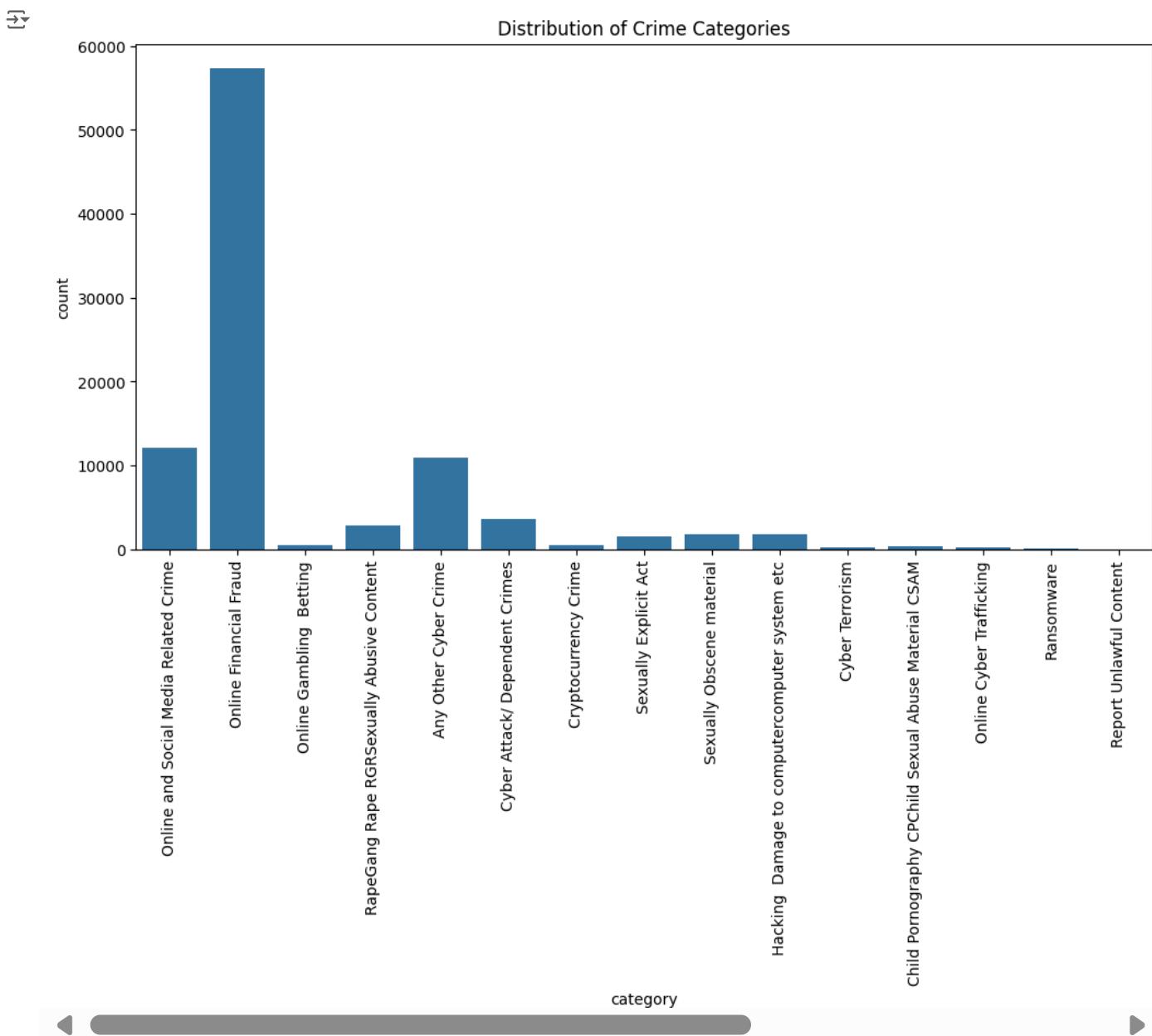
## ✓ Exploratory Data Analysis

### ✓ 1. Category Distribution: A bar chart to visualize the frequency of each crime category.

This helps understand which categories are most prevalent.

```
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(12, 6))
sns.countplot(x='category', data=train_data)
plt.xticks(rotation=90)
plt.title('Distribution of Crime Categories')
plt.show()
```

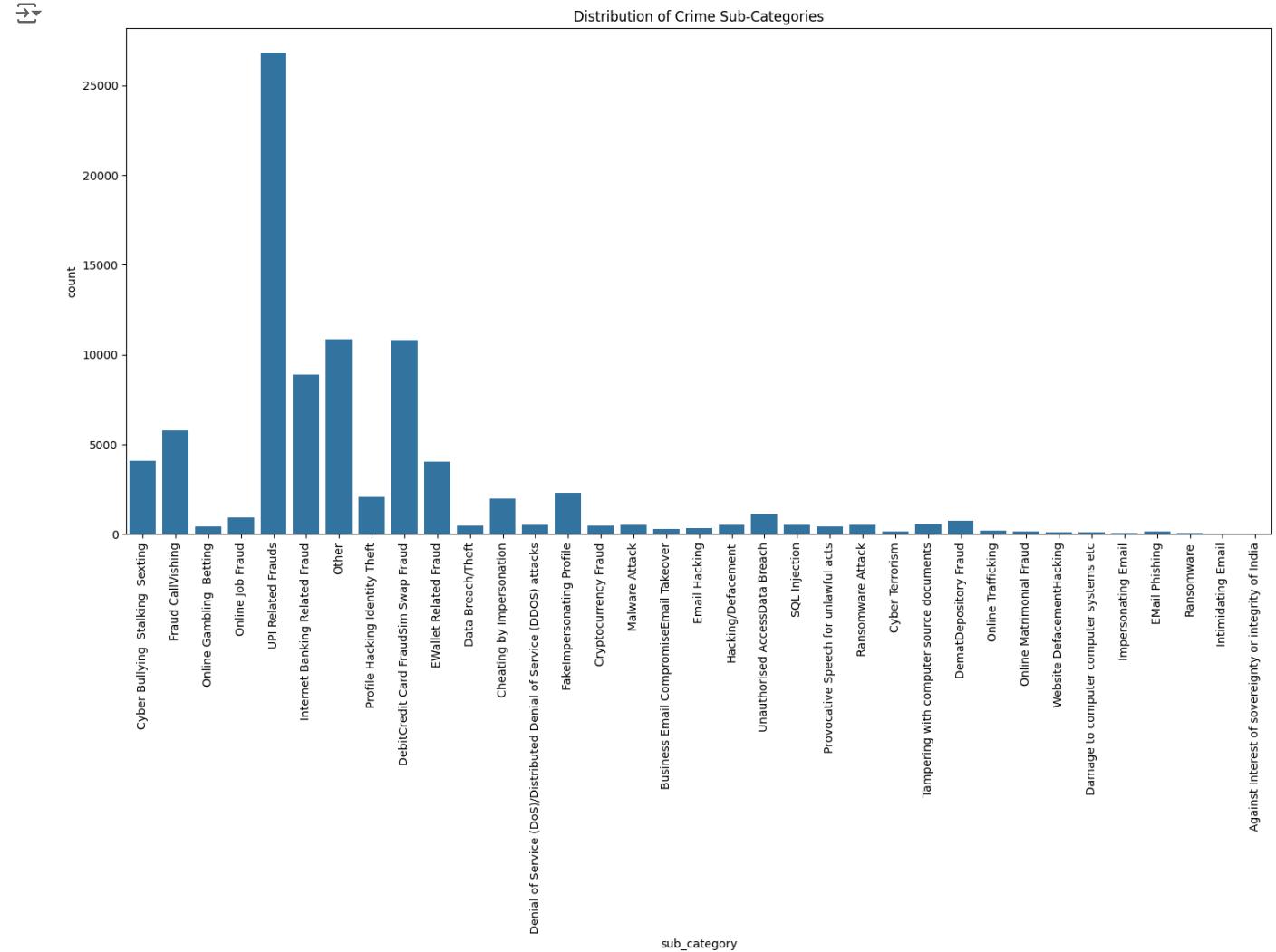


⌄ 2. Sub-Category Distribution: A bar chart to visualize the frequency of each sub-category.

This provides more granular insights into specific types of crimes within each category.

```
import matplotlib.pyplot as plt
import seaborn as sns

plt.figure(figsize=(18, 8))
sns.countplot(x='sub_category', data=train_data)
plt.xticks(rotation=90)
plt.title('Distribution of Crime Sub-Categories')
plt.show()
```

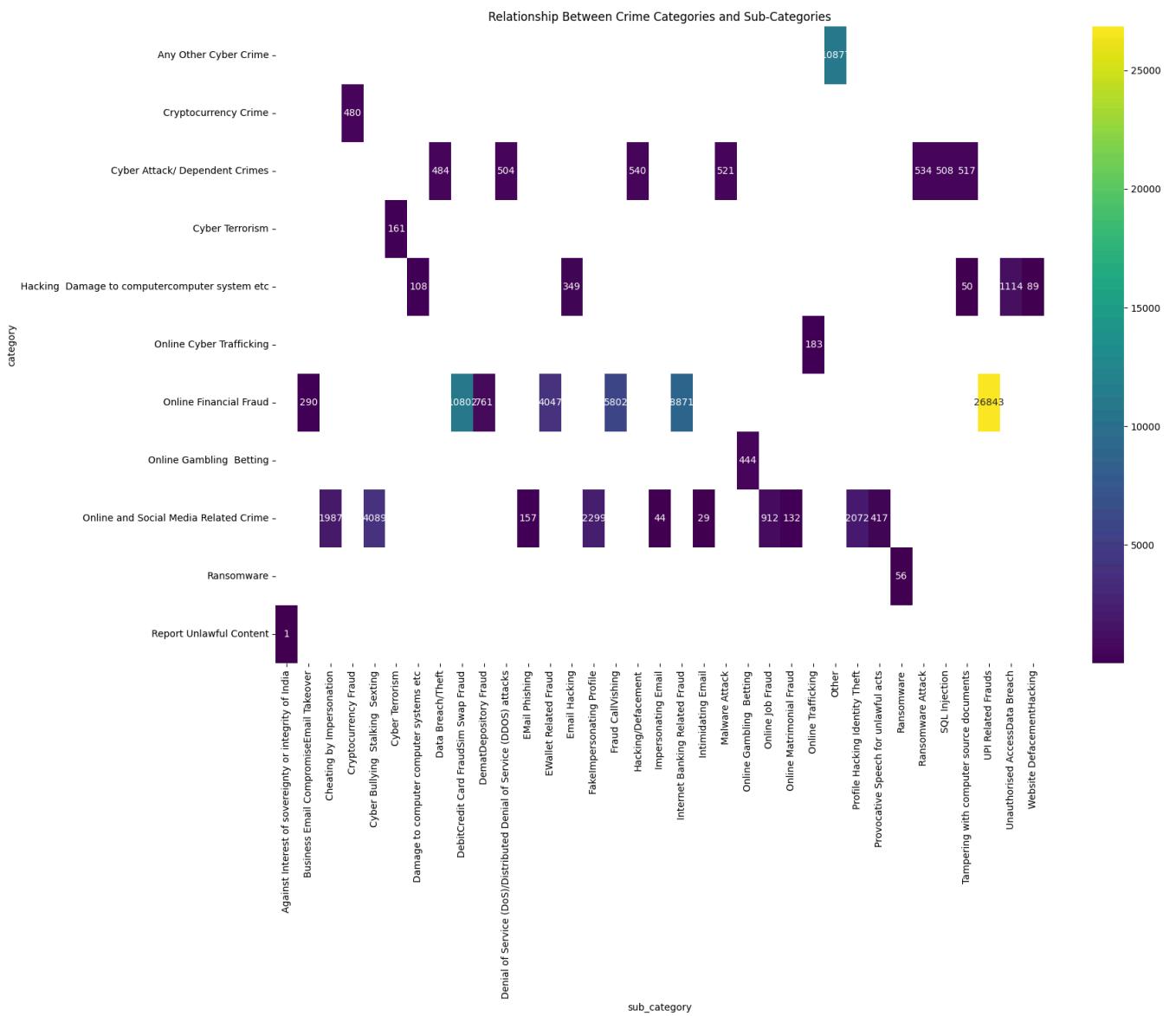


- 3. Category vs Sub-category: A heatmap to visualize the relationship between categories and sub-categories.

This reveals which sub-categories are most common within each category.

```
import matplotlib.pyplot as plt
import seaborn as sns

category_sub_category_counts = train_data.groupby(['category', 'sub_category'])['crimeadditionalinfo'].count().unstack()
plt.figure(figsize=(18, 12))
sns.heatmap(category_sub_category_counts, annot=True, fmt='0f', cmap='viridis')
plt.title('Relationship Between Crime Categories and Sub-Categories')
plt.show()
```



#### 4. word cloud of most frequently used words with the category and sub-category

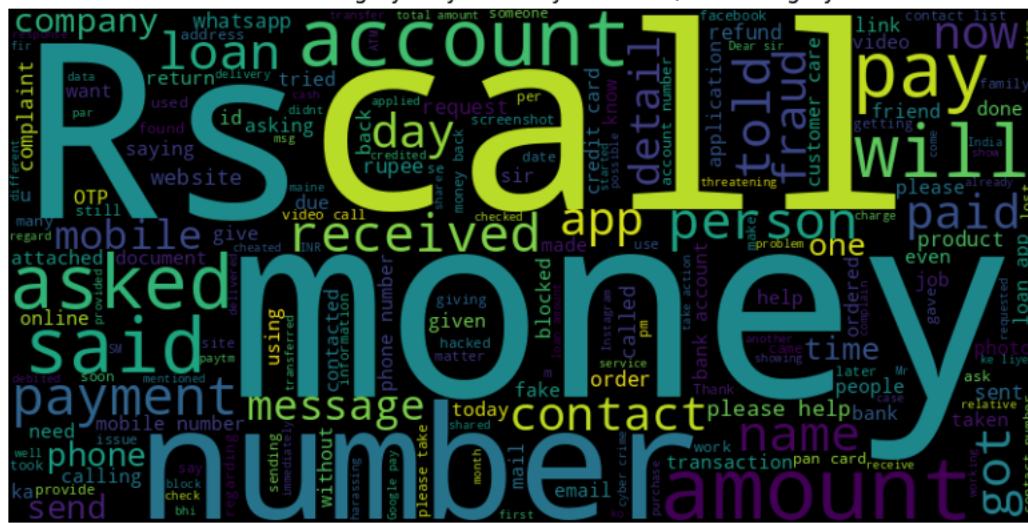
Helps determine prominently used words in each category and sub-category

```
from wordcloud import WordCloud
import matplotlib.pyplot as plt

# Combine all crimeadditionalinfo into a single string for each category and sub-category
category_sub_text = train_data.groupby(['category', 'sub_category'])['crimeadditionalinfo'].apply(lambda x: ' '.join(x)).reset_index()

# Function to generate and display word cloud for each group
def generate_wordcloud(category, sub_category, text):
    wordcloud = WordCloud(width=800, height=400, background_color='black').generate(text)
    plt.figure(figsize=(10, 5))
    plt.imshow(wordcloud, interpolation='bilinear')
    plt.axis('off')
    plt.title(f'Word Cloud for Category: {category}, Sub-category: {sub_category}')
    plt.show()
```

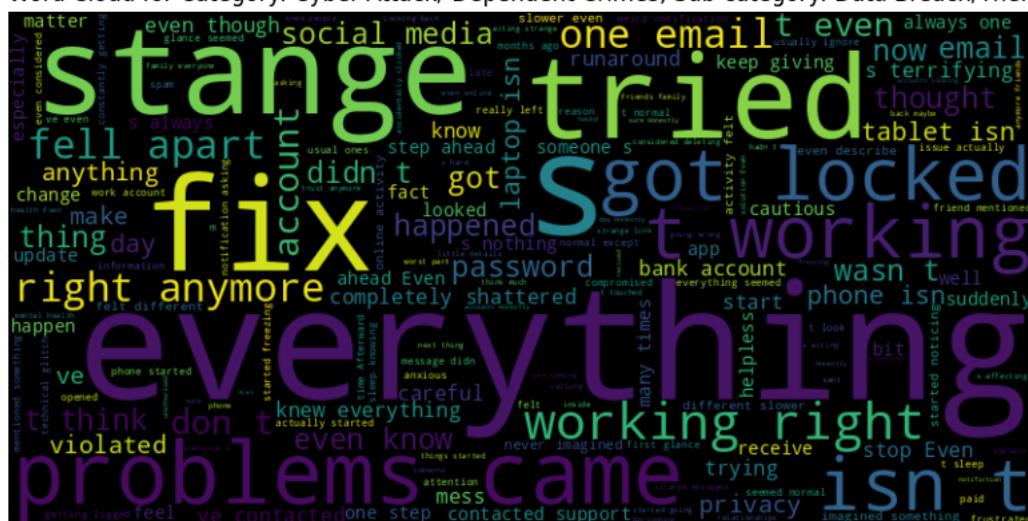
```
# Generate word cloud for each group
for index, row in category_sub_text.iterrows():
    generate_wordcloud(row['category'], row['sub_category'], row['crimeadditionalinfo'])
```



## Word Cloud for Category: Cryptocurrency Crime, Sub-category: Cryptocurrency Fraud

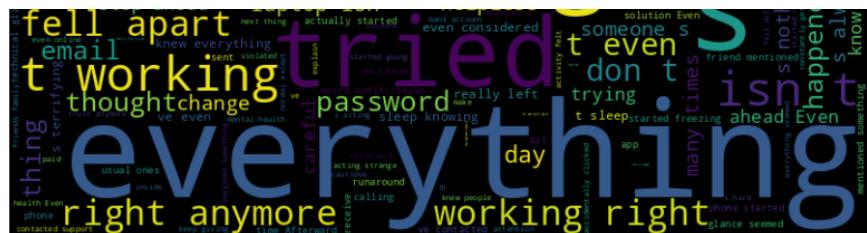


## Word Cloud for Category: Cyber Attack/ Dependent Crimes, Sub-category: Data Breach/Theft



## Word Cloud for Category: Cyber Attack/ Dependent Crimes, Sub-category: Denial of Service (DoS)/Distributed Denial of Service (DDOS) attacks

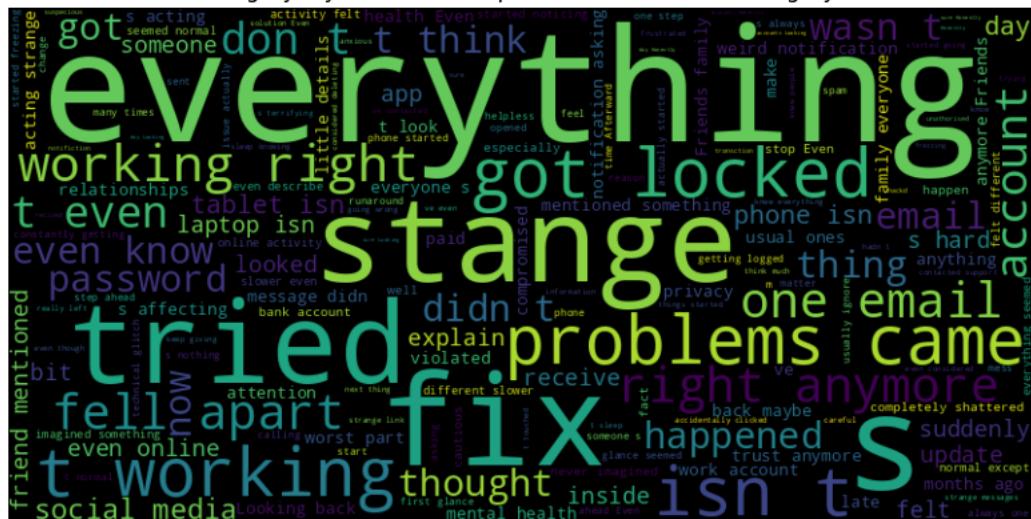




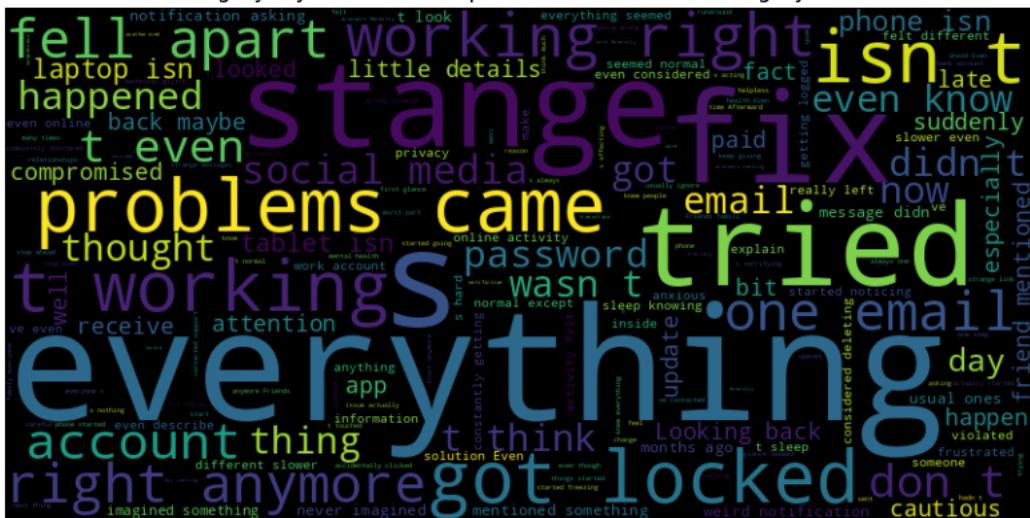
## Word Cloud for Category: Cyber Attack/ Dependent Crimes, Sub-category: Hacking/Defacement



## Word Cloud for Category: Cyber Attack/ Dependent Crimes, Sub-category: Malware Attack



## Word Cloud for Category: Cyber Attack/ Dependent Crimes, Sub-category: Ransomware Attack



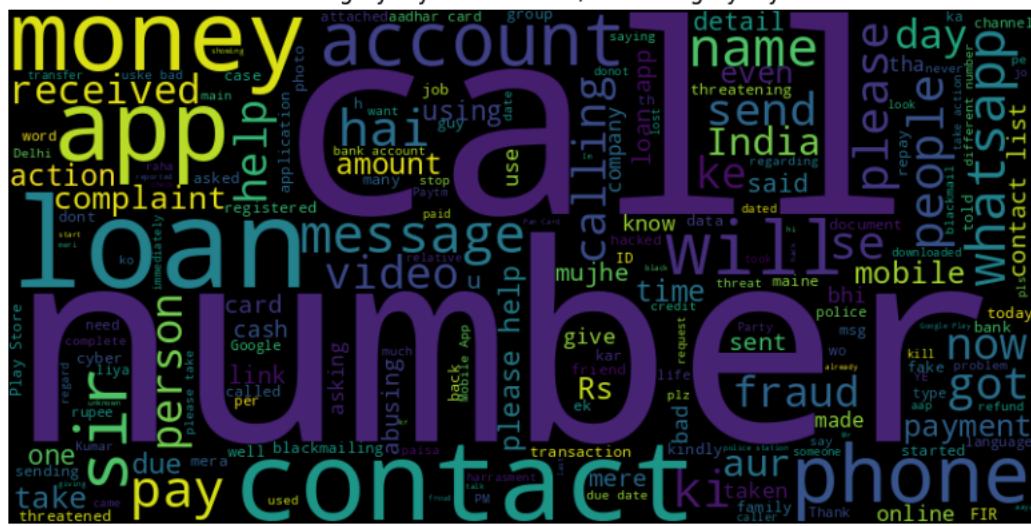
Word Cloud for Category: Cyber Attack/ Dependent Crimes, Sub-category: SQL Injection



Word Cloud for Category: Cyber Attack/ Dependent Crimes, Sub-category: Tampering with computer source documents

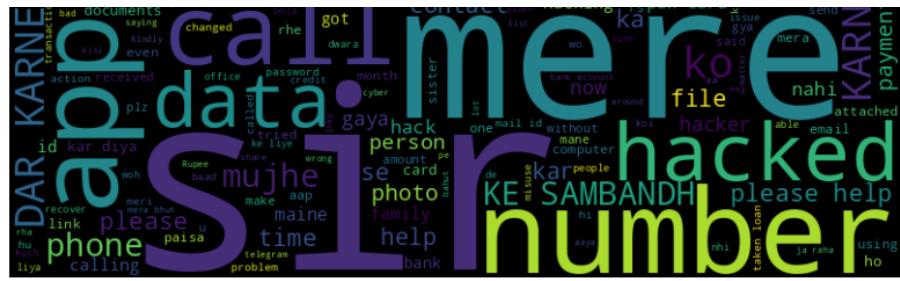


Word Cloud for Category: Cyber Terrorism, Sub-category: Cyber Terrorism

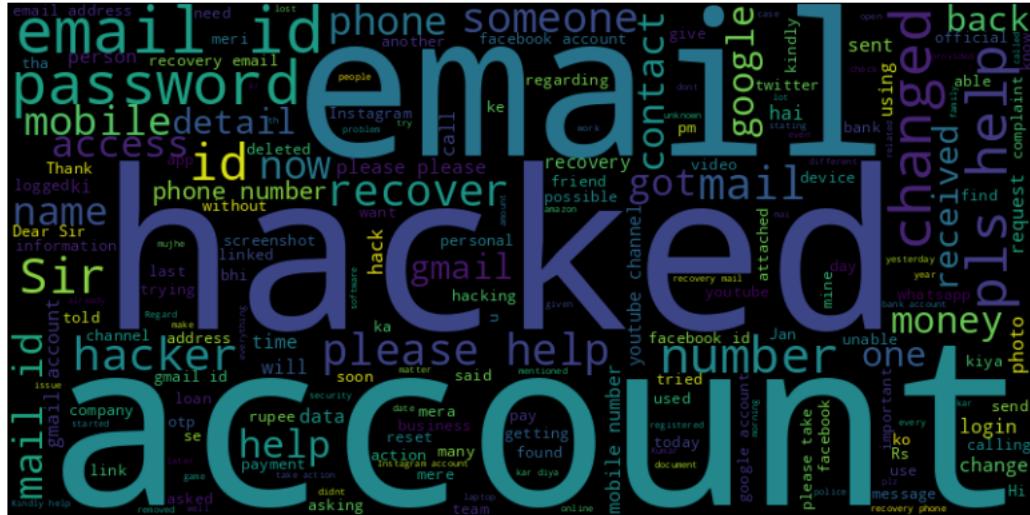


Word Cloud for Category: Hacking Damage to computercomputer system etc, Sub-category: Damage to computer computer systems etc

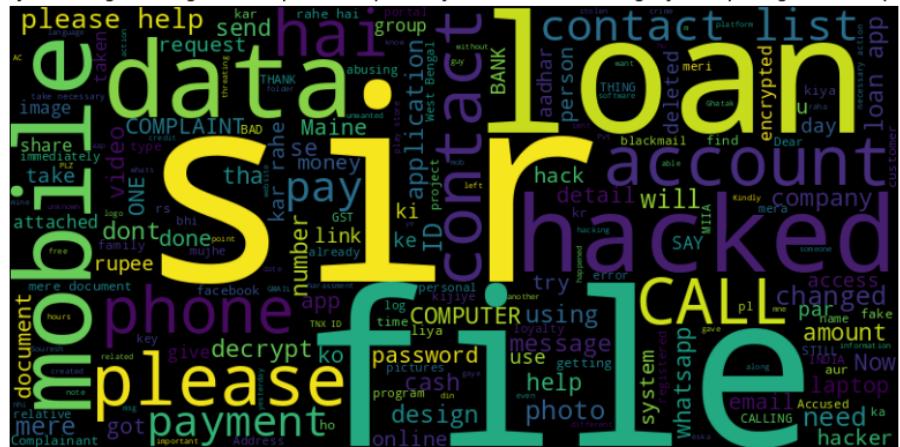




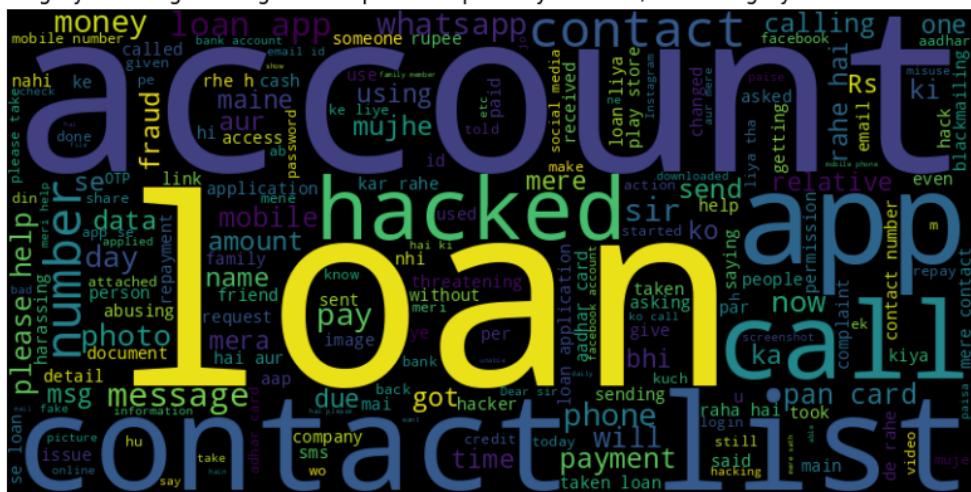
Word Cloud for Category: Hacking Damage to computercomputer system etc, Sub-category: Email Hacking



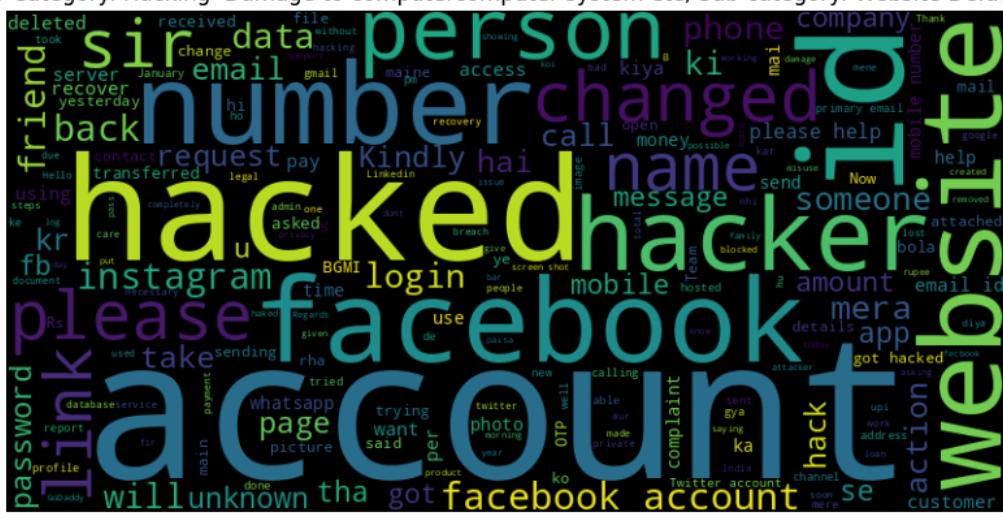
## Word Cloud for Category: Hacking Damage to computercomputer system etc, Sub-category: Tampering with computer source documents



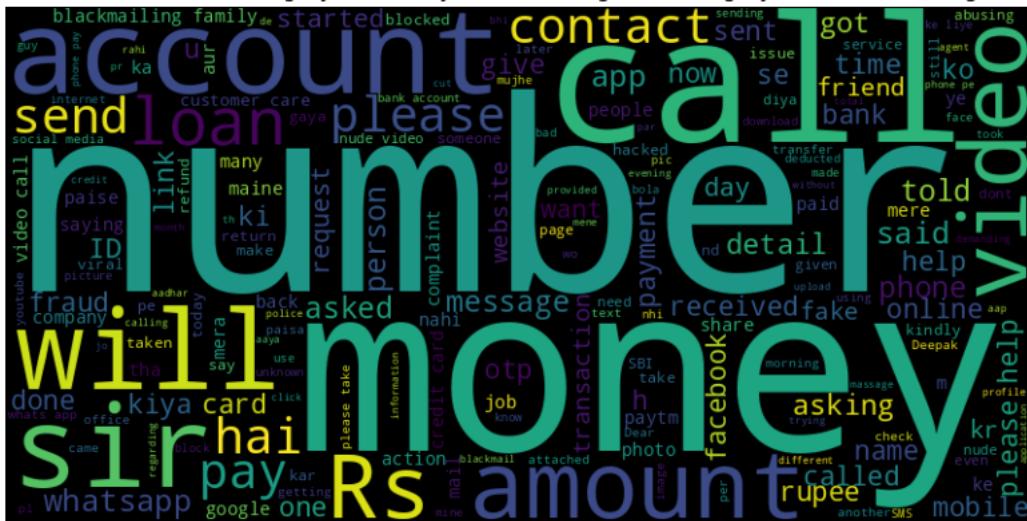
Word Cloud for Category: Hacking Damage to computercomputer system etc, Sub-category: Unauthorised AccessData Breach



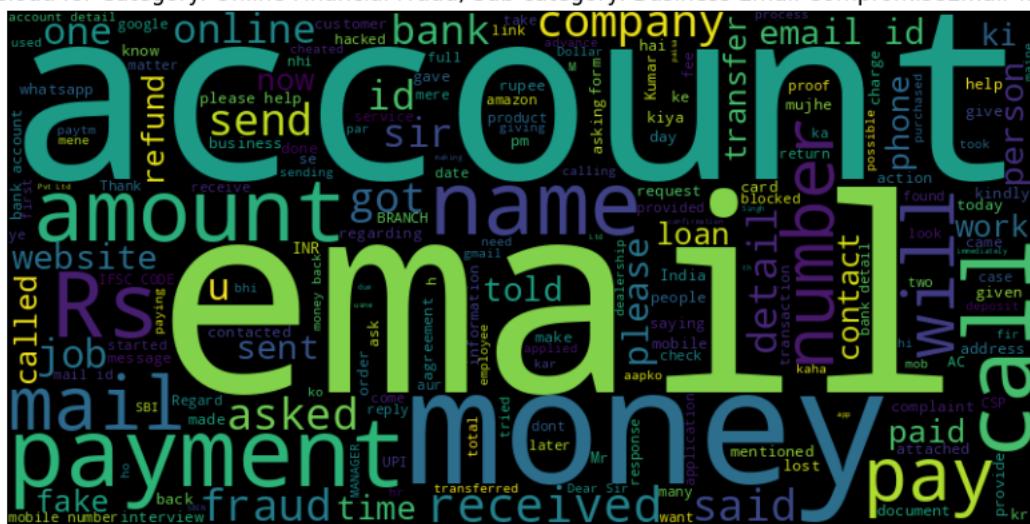
Word Cloud for Category: Hacking Damage to computercomputer system etc. Sub-category: Website DefacementHacking



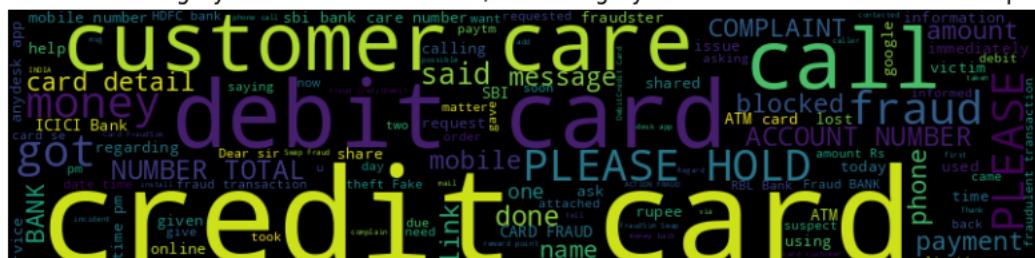
## Word Cloud for Category: Online Cyber Trafficking, Sub-category: Online Trafficking

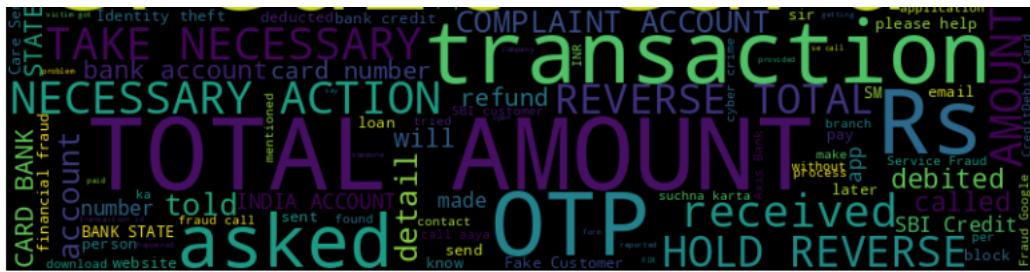


## Word Cloud for Category: Online Financial Fraud, Sub-category: Business Email CompromiseEmail Takeover

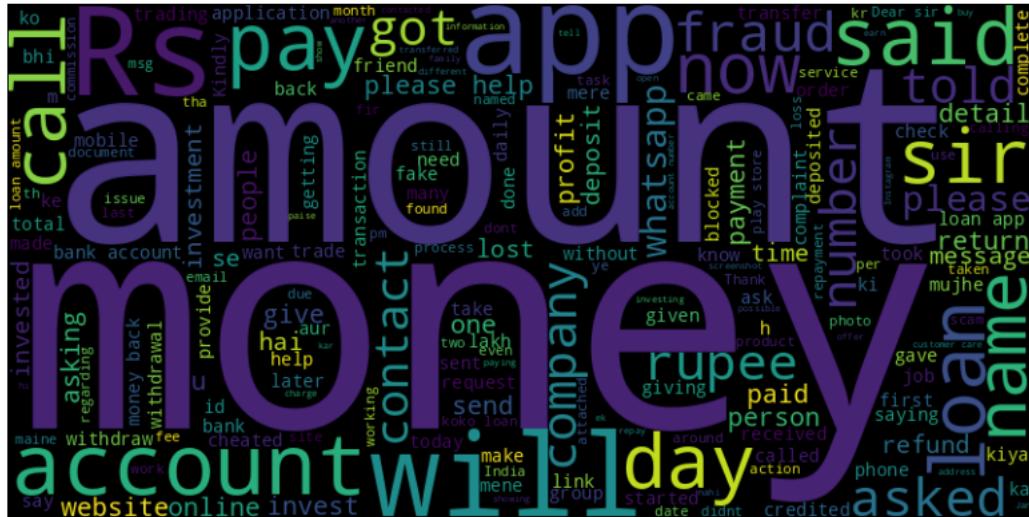


Word Cloud for Category: Online Financial Fraud, Sub-category: DebitCredit Card FraudSim Swap Fraud

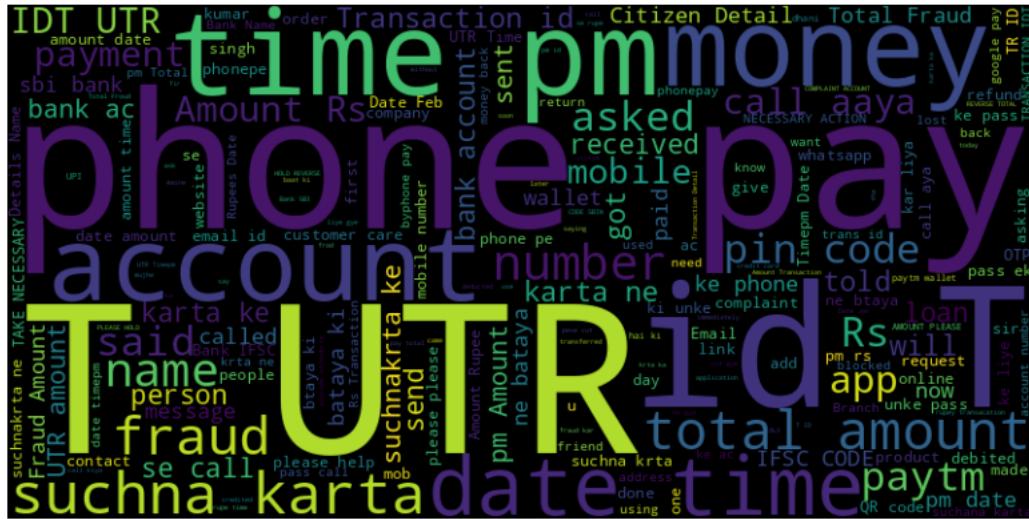




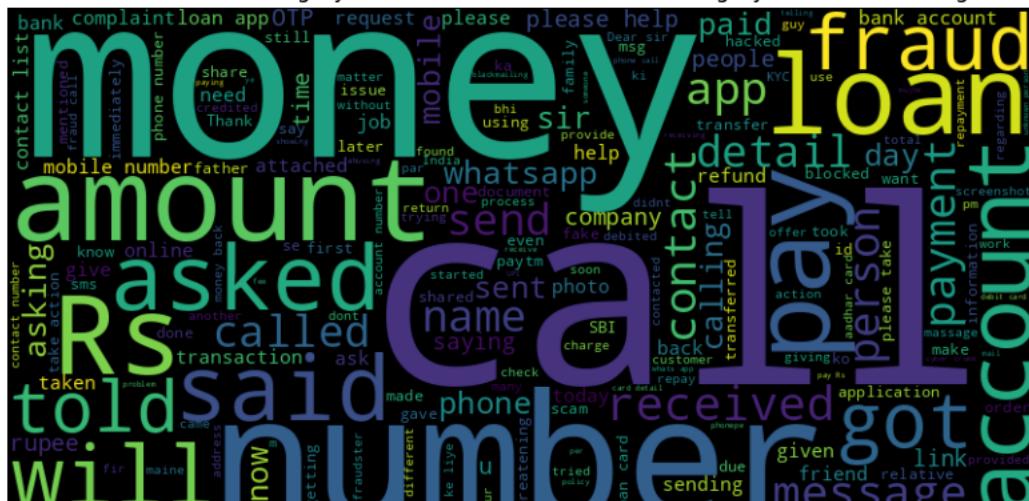
Word Cloud for Category: Online Financial Fraud, Sub-category: DematDepository Fraud



Word Cloud for Category: Online Financial Fraud, Sub-category: EWallet Related Fraud

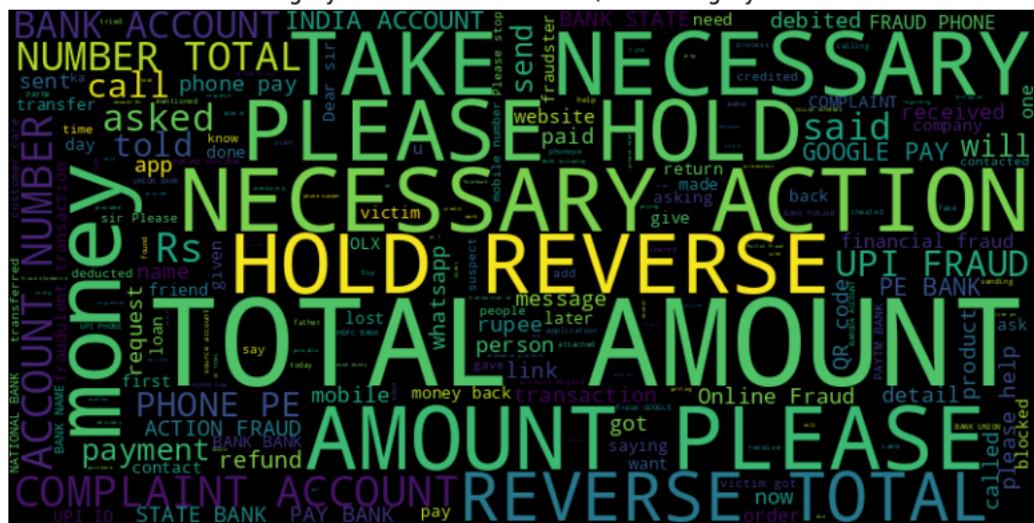


Word Cloud for Category: Online Financial Fraud, Sub-category: Fraud CallVishing

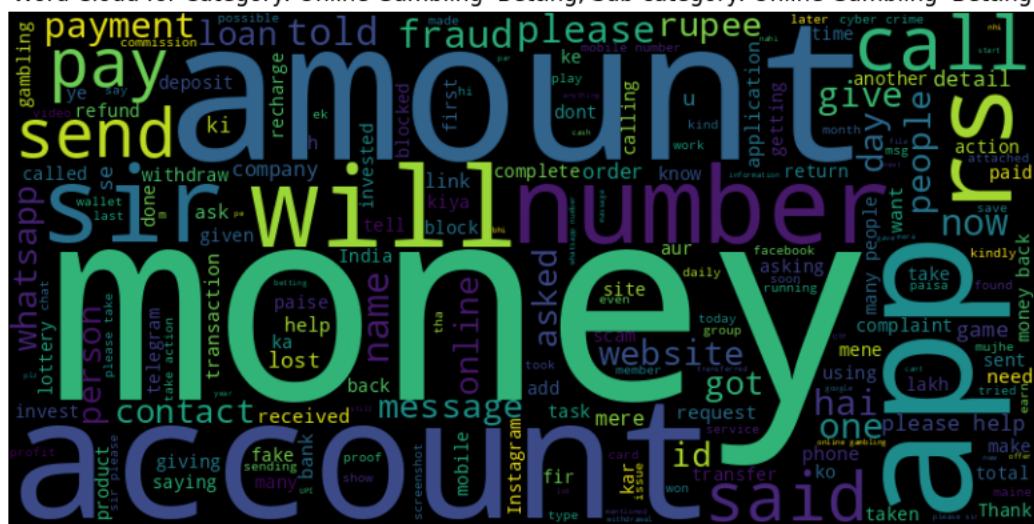




## Word Cloud for Category: Online Financial Fraud, Sub-category: UPI Related Frauds



## Word Cloud for Category: Online Gambling Betting, Sub-category: Online Gambling Betting

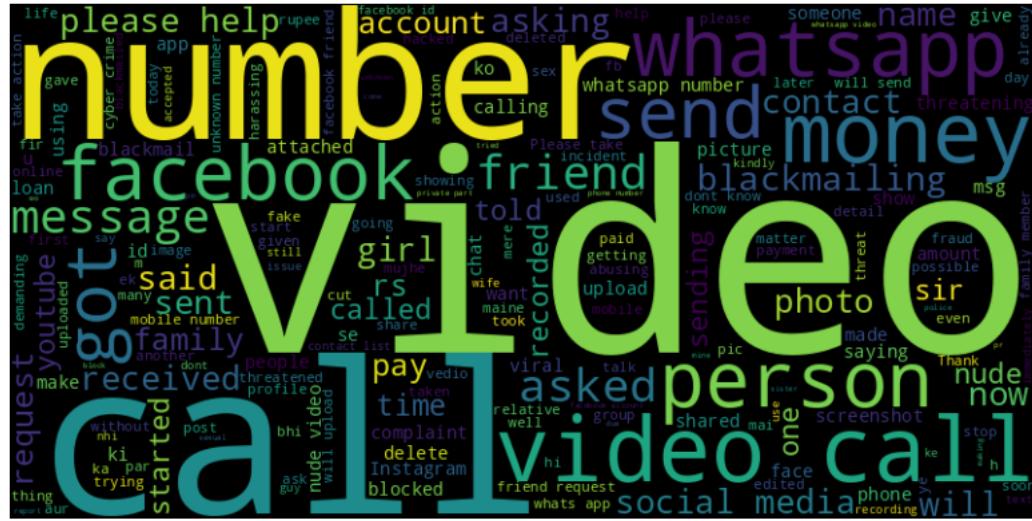


## Word Cloud for Category: Online and Social Media Related Crime, Sub-category: Cheating by Impersonation

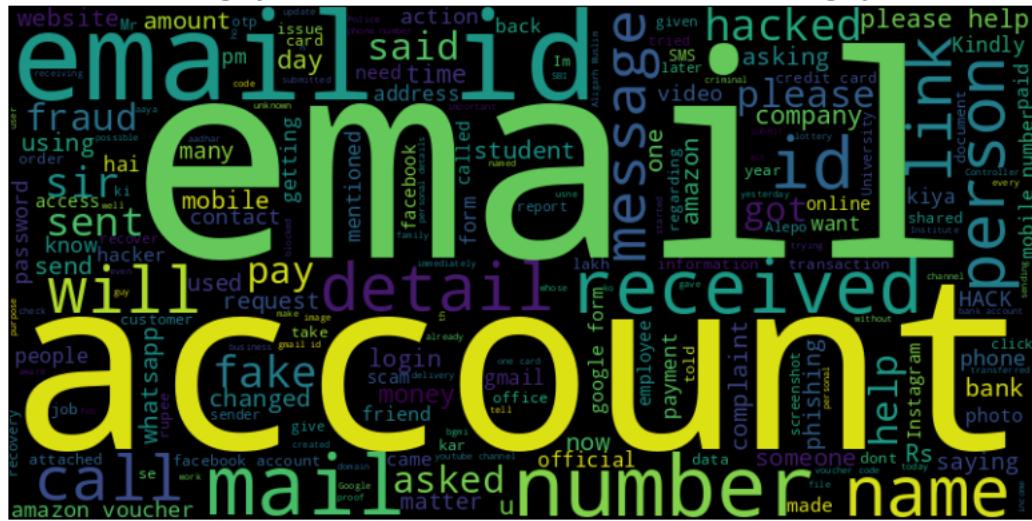




## Word Cloud for Category: Online and Social Media Related Crime, Sub-category: Cyber Bullying Stalking Sexting



## Word Cloud for Category: Online and Social Media Related Crime, Sub-category: EMail Phishing

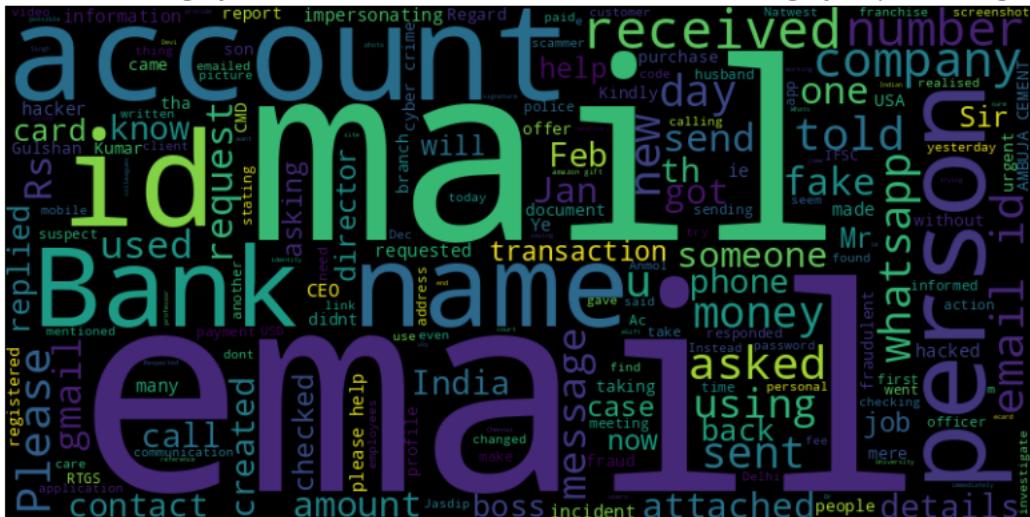


## Word Cloud for Category: Online and Social Media Related Crime, Sub-category: FakeImpersonating Profile

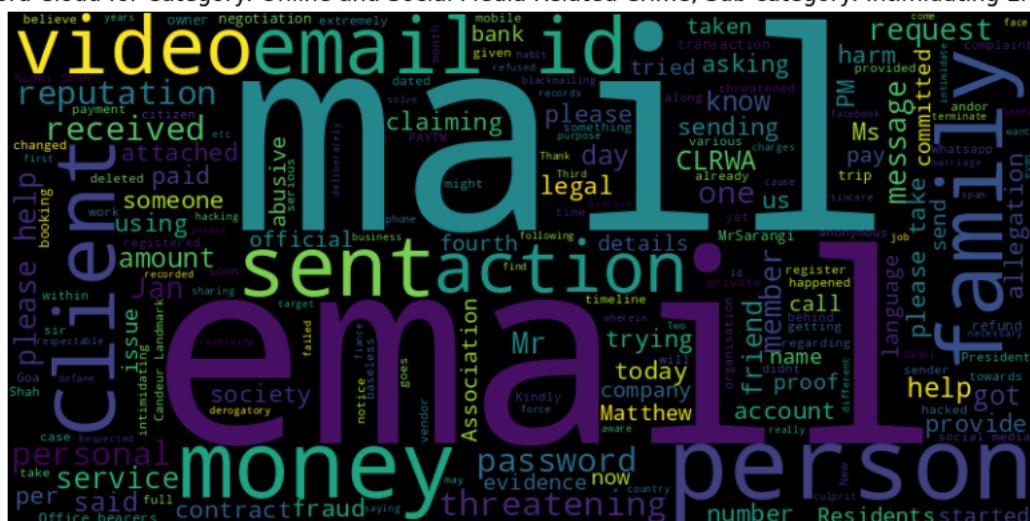




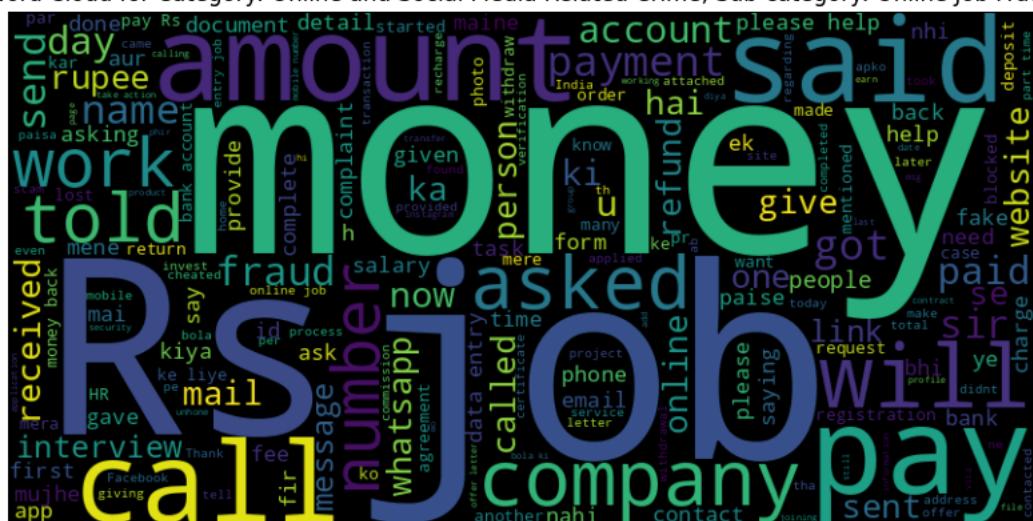
## Word Cloud for Category: Online and Social Media Related Crime, Sub-category: Impersonating Email



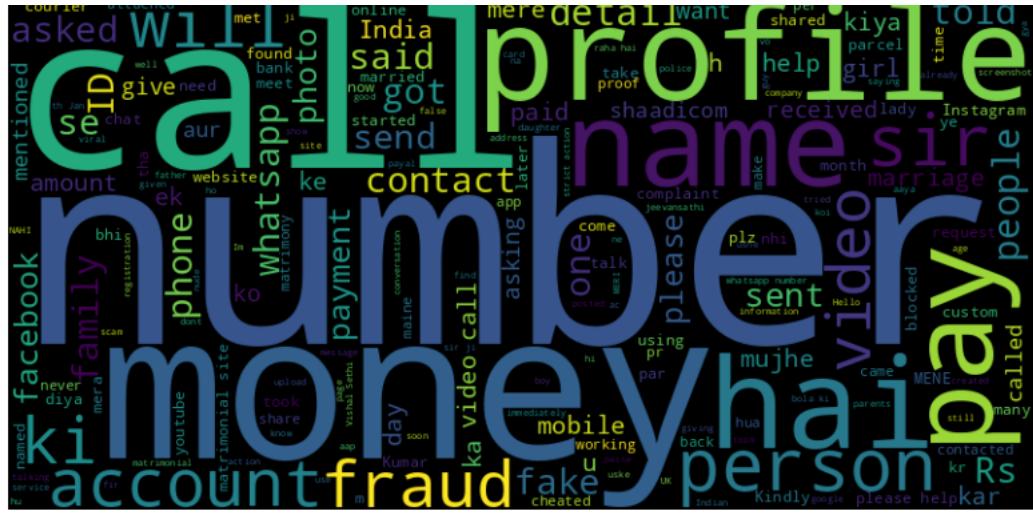
Word Cloud for Category: Online and Social Media Related Crime, Sub-category: Intimidating Email



## Word Cloud for Category: Online and Social Media Related Crime, Sub-category: Online Job Fraud



Word Cloud for Category: Online and Social Media Related Crime, Sub-category: Online Matrimonial Fraud



## Word Cloud for Category: Online and Social Media Related Crime, Sub-category: Profile Hacking Identity Theft

