Lojain Idris

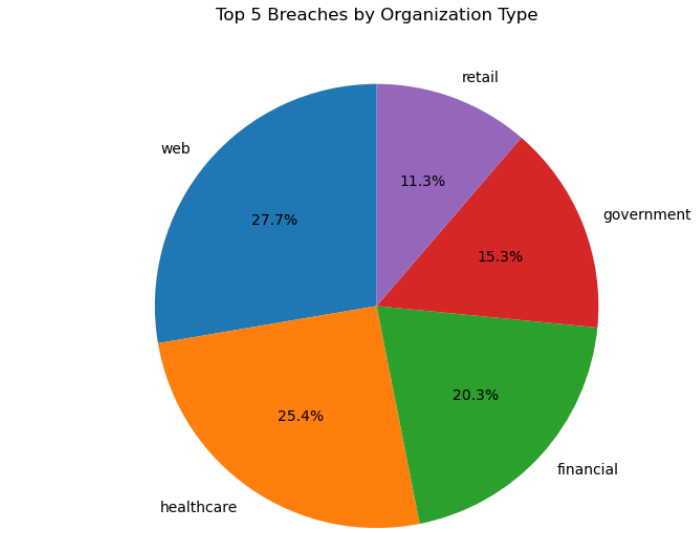
CTEC298

Six Plots of Top Data Breaches 2004-2021

1. Pie Plot

```
•[8]: #creating fig and axes
fig, axes = plt.subplots(figsize=(6, 6))

#creating or plotting the Pie chart
org_counts = df['Organization type'].value_counts().nlargest(5)# top five
axes.pie(org_counts, labels=org_counts.index, autopct='%1.1f%%', startangle=90)
axes.set_title("Top 5 Breaches by Organization Type")
plt.tight_layout()
plt.show()
```

Top 5 Breaches by Organization Type

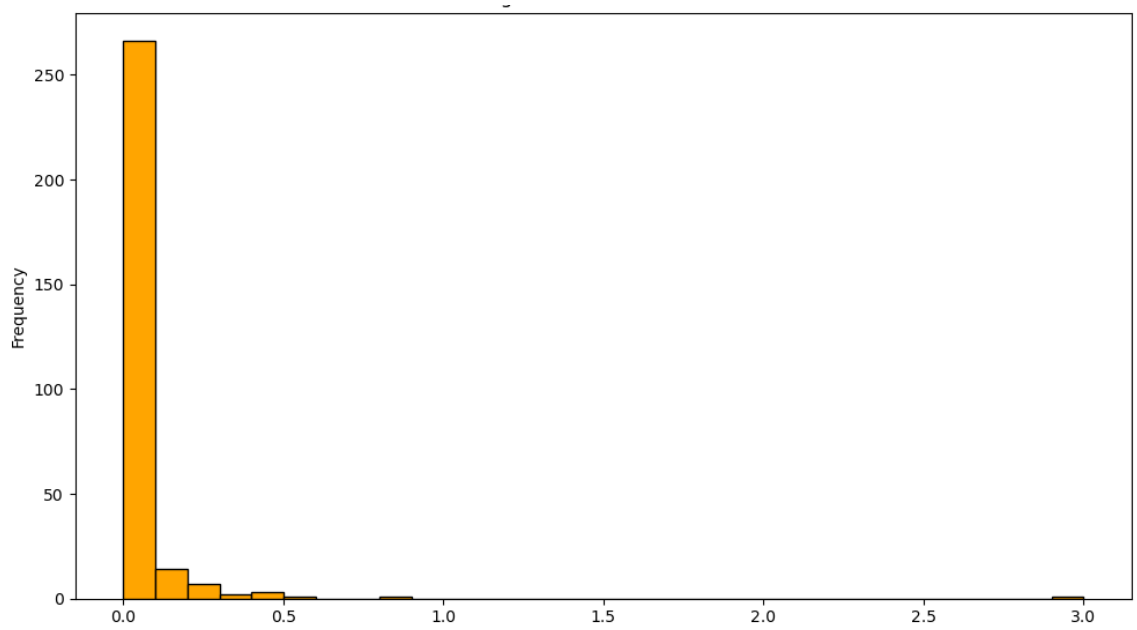


2. Histograms

```
[15]: #histograms plot
plt.figure(figsize=(10,6))
plt.hist(df['Records'], bins=30, color='orange', edgecolor='black')
#naming the lablels
plt.title("Histogram of Records Breached")
plt.xlabel("Records")
plt.ylabel("Frequency")
#displaying
plt.tight_layout()
plt.show
```

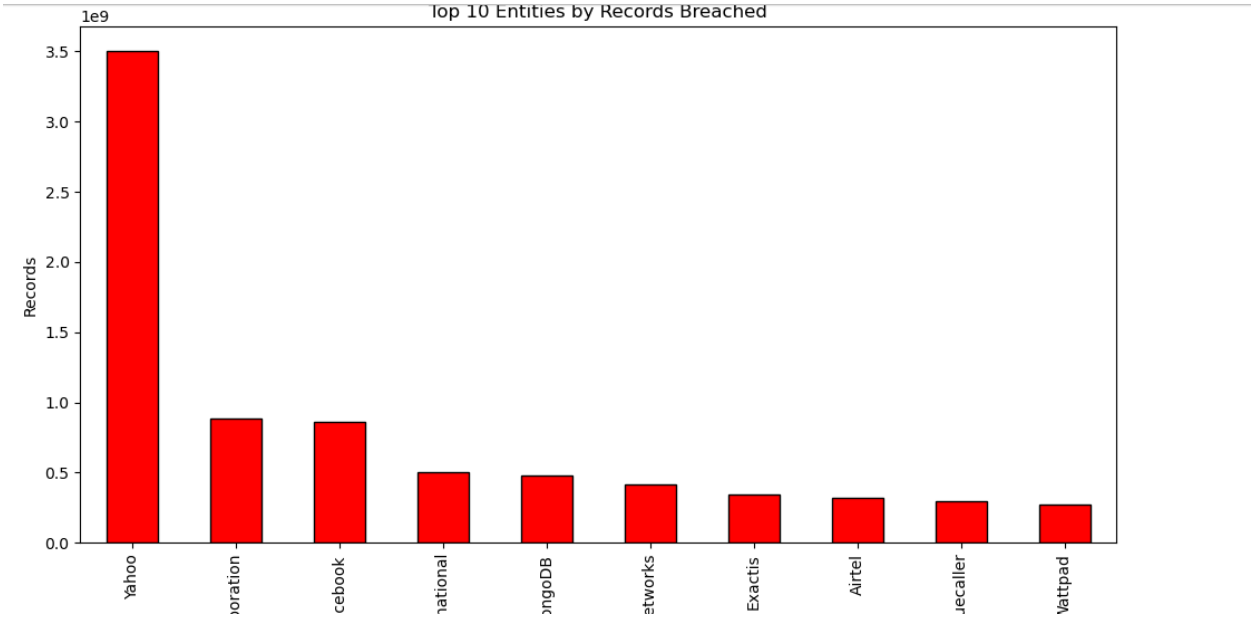
```
[15]: <function matplotlib.pyplot.show(close=None, block=None)>
```

Histogram of Records Breached



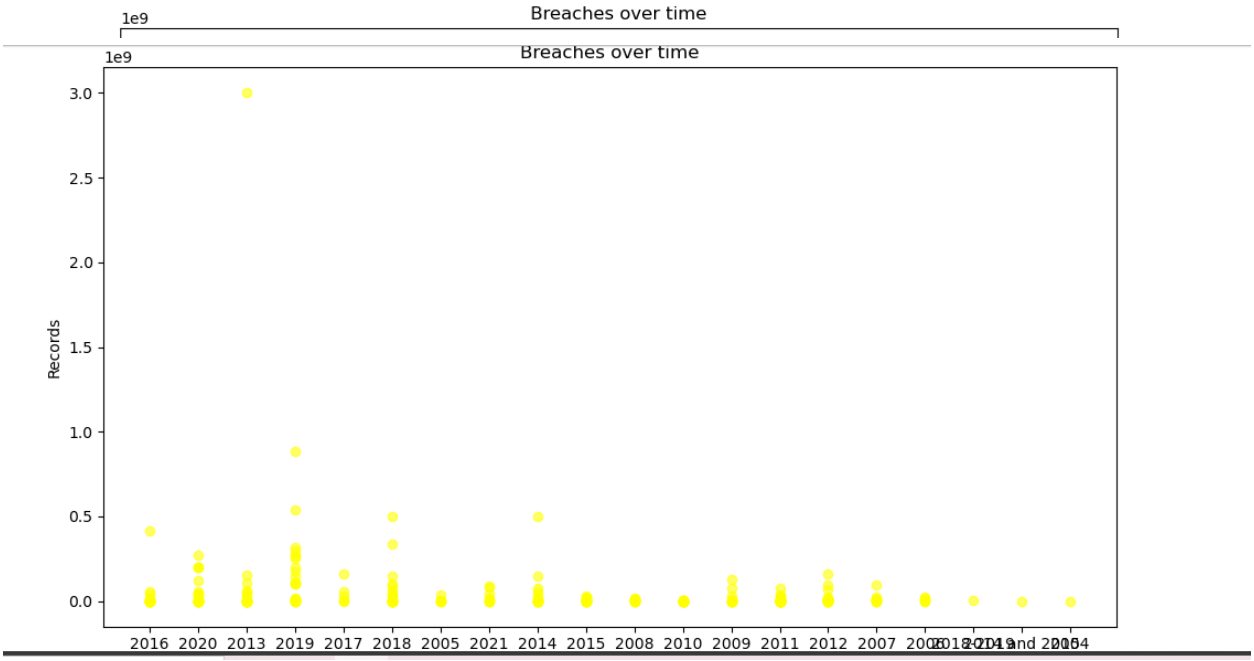
3. Bar Graph

```
[18]: #bar graph
top_entities = df.groupby('Entity')['Records'].sum().sort_values(ascending=False).head(10)
plt.figure(figsize=(12,6))
top_entities.plot(kind='bar', color='red', edgecolor='black')
#naming the labels
plt.title("Top 10 Entities by Records Breached")
plt.xlabel("entity")
plt.ylabel("Records")
plt.xtick(rotation=45, ha='right')#rotate x-axis and aligning the lebels to right

plt.tigh_layout()
plt.show()
```



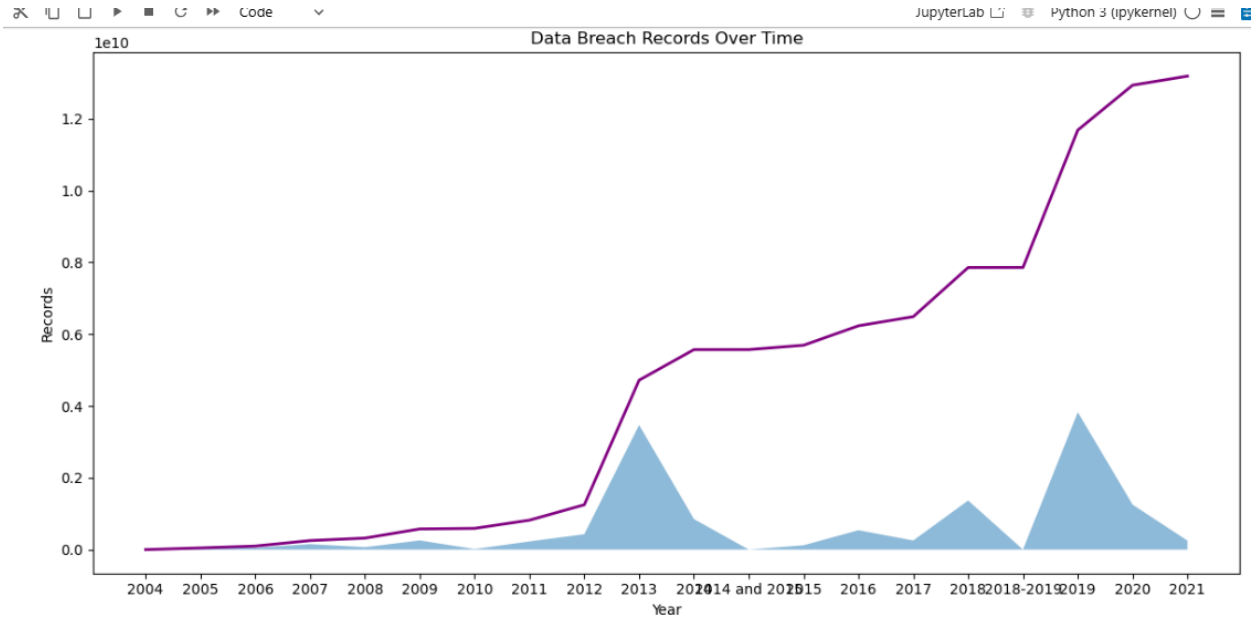
4. Scatter Plot

[21]:
```
#scatter plot
plt.figure(figsize=(10, 6))
plt.scatter(df['Year'], df['Records'], color='yellow', alpha=0.6)

#naming the lebels
plt.title("Breaches over time")
plt.xlabel("Year")
plt.ylabel("Records")
#diplay
plt.tight_layout()
plt.show()
```



5. Area Plot

[24]:
```
#area plot
#calculate cumulative breach size over time and total records
records_per_year = df.groupby('Year')['Records'].sum().sort_index()
cumulative_records = records_per_year.cumsum()
#plotting
plt.figure(figsize=(12,6))
plt.fill_between(records_per_year.index, records_per_year.values, alpha=0.5, label='Total Records (per year)')
plt.plot(cumulative_records.index, cumulative_records.values, color='purple', label='Cumulative Records', linewidth=2)

# naming Labels
plt.title("Data Breach Records Over Time")
plt.xlabel("Year")
plt.ylabel("Records")
# Display the plot
plt.tight_layout()
plt.show()
```



6. Hexagonal bin Plot / hexbin

```
36]: #hexbin plot / hexagonal Bin
     #plotting
     #convert any string in years to numeric value
     df['Year'] = pd.to_numeric(df['Year'], errors='coerce')
     plt.figure(figsize=(10, 6))
     plt.hexbin(df['Year'], df['Records'], gridsize=30, cmap='Blues', bins='log')
     # naming labels
     plt.title("Breaches Records over time")
     plt.xlabel("year")
     plt.ylabel("Records")

     #display
     plt.tight_layout()
     plt.show()
```