```python
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```python
import pandas as pd

# Define the file paths for the CSV files to be loaded
filepaths = ["/content/drive/MyDrive/Data visualozation/Crime/Auto_theft.csv",
"/content/drive/MyDrive/Data visualozation/Crime/Complaints_against_police.csv",
"/content/drive/MyDrive/Data visualozation/Crime/Murders.csv",
"/content/drive/MyDrive/Data visualozation/Crime/Property_stolen_and_recovered.csv",
"/content/drive/MyDrive/Data visualozation/Crime/Rape_Victims.csv"]

# Define the column names to be selected from each CSV file
column_names = [["Area_Name", "Year", "Auto_Theft_Stolen"],
                ["CPA_Cases_Registered"],
                ["Victims_Total"],
                ["Cases_Property_Stolen"],
                ["Rape_Cases_Reported"]]

# Load the CSV files and select the desired columns
dfs = []
for filepath, cols in zip(filepaths, column_names):
    df = pd.read_csv(filepath, usecols=cols)
    dfs.append(df)

# Merge the selected columns into a single DataFrame
merged_df = pd.concat(dfs, axis=1)

# Define the file path and name for the new CSV file
output_filepath = "/content/drive/MyDrive/Data visualozation/Crime/Murged_data/output1.csv"

# Save the merged DataFrame to the new CSV file
merged_df.to_csv(output_filepath, index=False)
```

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

# Load the CSV file into a pandas DataFrame
df = pd.read_csv('/content/drive/MyDrive/Data visualozation/Murged_data/output1.csv')

# Group the data by year and calculate the mean of each numerical column
grouped = df.groupby('Year').mean()

# Plot each numerical column in the same plot
grouped.plot(kind='line', subplots=True, figsize=(10, 8))

# Set the plot title and labels
plt.suptitle('Numerical data by year')
plt.xlabel('Year')
plt.ylabel('Value')

# Show the plot
plt.show()
```
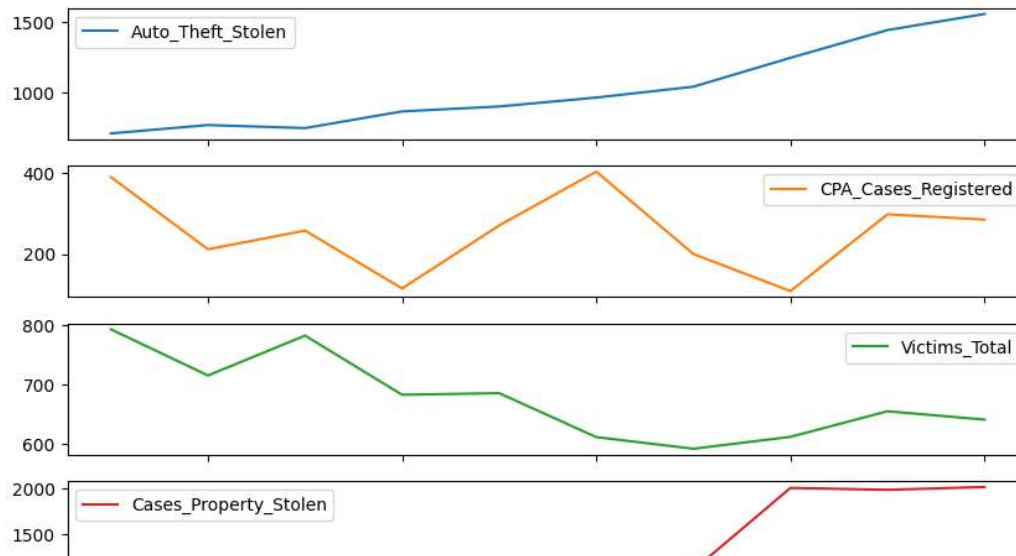
```
<ipython-input-5-1450d9fa08e2>:8: FutureWarning: The default value of numeric_only in DataFrameGroupBy.
  grouped = df.groupby('Year').mean()
```


Numerical data by year

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

# Load the CSV file into a pandas DataFrame
df = pd.read_csv('/content/drive/MyDrive/Data visualozation/Murged_data/output1.csv')

# Group the data by year and calculate the mean of each numerical column
grouped = df.groupby('Year').mean()

# Get a list of the numerical column names
num_cols = grouped.columns

# Create a bar plot with five bars for each year
fig, ax = plt.subplots(figsize=(15, 10))
bar_width = 0.15
for i, col in enumerate(num_cols):
    ax.bar(grouped.index + (i * bar_width), grouped[col], width=bar_width, label=col)

# Set the plot title and labels
ax.set_title('Numerical data by year')
ax.set_xlabel('Year')
ax.set_ylabel('Value')
ax.set_xticks(grouped.index + (2 * bar_width))
ax.set_xticklabels(grouped.index)
ax.legend()

# Show the plot
plt.show()
```
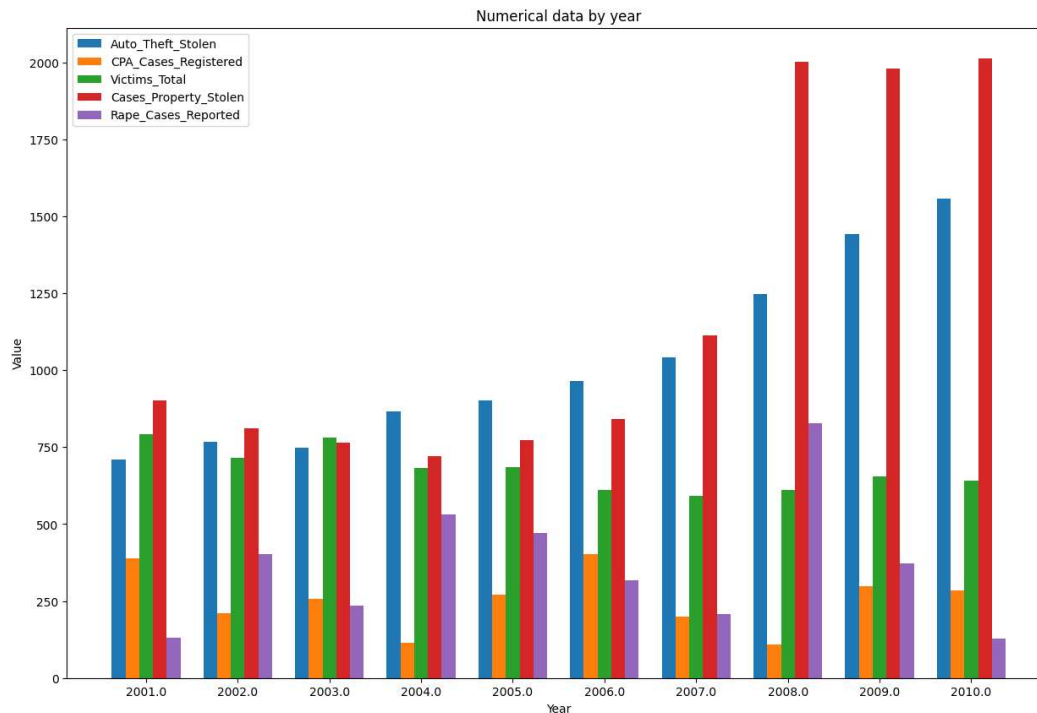
```
<ipython-input-6-356bf4e63b0c>:8: FutureWarning: The default value of numeric_only in DataFrameGroupBy.
  grouped = df.groupby('Year').mean()
```



```python
import pandas as pd

# Define the file paths for the CSV files to be loaded
filepaths = ["/content/drive/MyDrive/Data visualozation/Crime/Auto_theft.csv",
"/content/drive/MyDrive/Data visualozation/Crime/Complaints_against_police.csv",
"/content/drive/MyDrive/Data visualozation/Crime/Murders.csv",
"/content/drive/MyDrive/Data visualozation/Crime/Property_stolen_and_recovered.csv",
"/content/drive/MyDrive/Data visualozation/Crime/Rape_Victims.csv"]

# Define the column names to be selected from each CSV file
column_names = [["Area_Name", "Auto_Theft_Stolen"],
                ["CPA_Cases_Registered"],
                ["Victims_Total"],
                ["Cases_Property_Stolen"],
                ["Rape_Cases_Reported"]]

# Load the CSV files and select the desired columns
dfs = []
for filepath, cols in zip(filepaths, column_names):
    df = pd.read_csv(filepath, usecols=cols)
    dfs.append(df)

# Merge the selected columns into a single DataFrame
merged_df = pd.concat(dfs, axis=1)

# Define the file path and name for the new CSV file
output_filepath = "/content/drive/MyDrive/Data visualozation/Murged_data/output2.csv"

# Save the merged DataFrame to the new CSV file
merged_df.to_csv(output_filepath, index=False)


import pandas as pd
import matplotlib.pyplot as plt

# Load the CSV file into a pandas DataFrame
df = pd.read_csv('/content/drive/MyDrive/Data visualozation/Murged_data/output2.csv')

# Group the data by state and calculate the mean of each numerical column
grouped = df.groupby('Area_Name').mean()
```

```
grouped = df.groupby('Area_Name').mean()

# Get a list of the numerical column names
num_cols = grouped.columns

# Create a bar plot with five bars for each state
fig, ax = plt.subplots(figsize=(20, 12))
bar_width = 0.15
for i, col in enumerate(num_cols):
    ax.bar([x + (i * bar_width) for x in range(len(grouped))], grouped[col], width=bar_width, label=col)

# Set the plot title and labels
ax.set_title('Numerical data by state')
ax.set_xlabel('State')
ax.set_ylabel('Value')
ax.set_xticks([x + (2 * bar_width) for x in range(len(grouped))])
ax.set_xticklabels(['{}'.format(x) for x in grouped.index.tolist()],rotation=90)
ax.legend()

# Show the plot
plt.show()
```
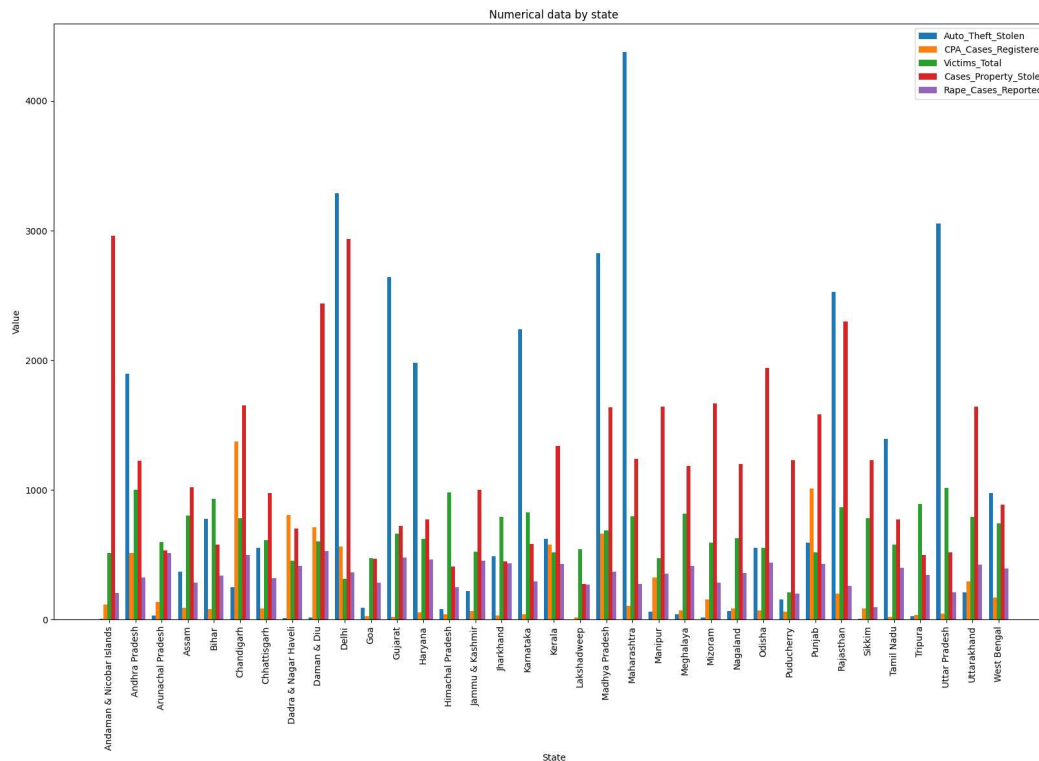
```python
import pandas as pd
import matplotlib.pyplot as plt

# Load the CSV file into a pandas DataFrame
df = pd.read_csv('/content/drive/MyDrive/Data visualozation/Murged_data/output2.csv')

# Get the sum of all numerical columns
sums = df.select_dtypes(include='number').sum()

# Create a pie chart with a larger size
fig, ax = plt.subplots(figsize=(8, 8))
ax.pie(sums, labels=sums.index, autopct='%1.1f%%')
ax.set_title('Sum of Numerical Columns')

plt.show()
```

Sum of Numerical Columns
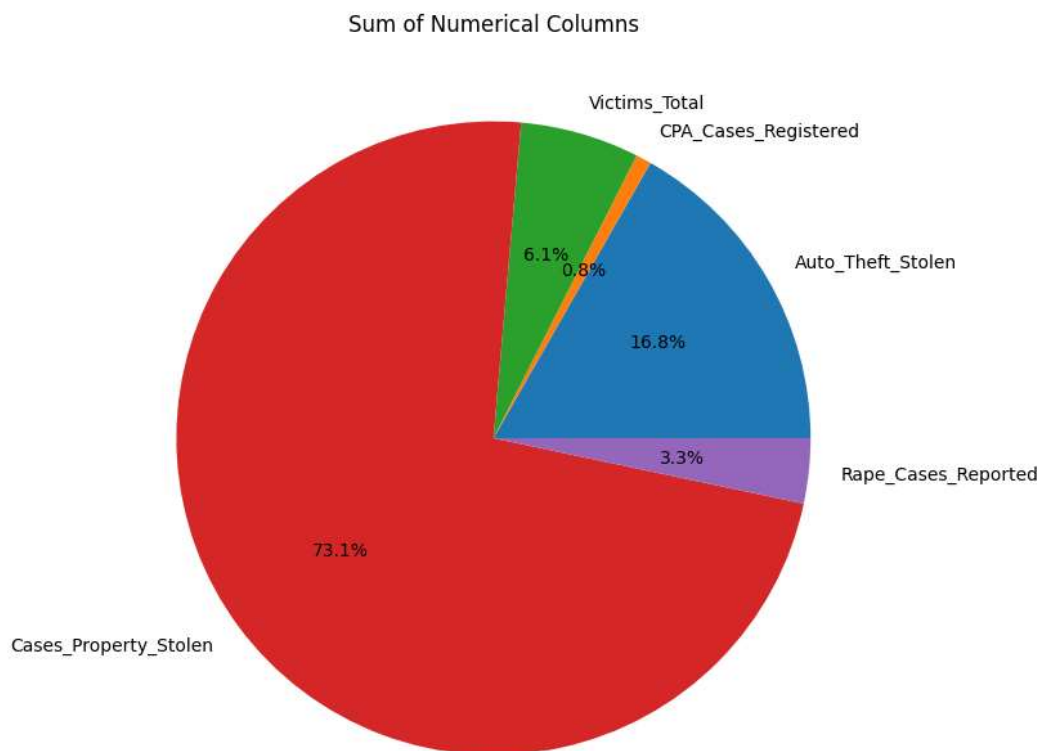
```python
#["E:\IIT_Guwhati\Semester_02\CS - 595 - Data visualization lab\Project\Crime\Auto_theft.csv",
#"E:\IIT_Guwhati\Semester_02\CS - 595 - Data visualization lab\Project\Crime\Complaints_against_police.csv",
#"E:\IIT_Guwhati\Semester_02\CS - 595 - Data visualization lab\Project\Crime\Murders.csv",
#"E:\IIT_Guwhati\Semester_02\CS - 595 - Data visualization lab\Project\Crime\Property_stolen_and_recovered.csv",
#"E:\IIT_Guwhati\Semester_02\CS - 595 - Data visualization lab\Project\Crime\Rape_Victims.csv"]


#col1 = "Auto_Theft_Stolen" # replace with actual column name from file1.csv
#col2 = "CPA_Cases_Registered" # replace with actual column name from file2.csv
#col3 = "Victims_Total" # replace with actual column name from file3.csv
#col4 = "Cases_Property_Stolen" # replace with actual column name from file4.csv
#col5 = "Rape_Cases_Reported" # replace with actual column name from file5.csv


#merged_df = pd.merge(df1, df2, on='Year')
#merged_df = pd.merge(merged_df, df3, on='Year')
#merged_df = pd.merge(merged_df, df4, on='Year')
#merged_df = pd.merge(merged_df, df5, on='Year')


#df1 = pd.read_csv('file1.csv')[['year', 'crime_type1']]
#df2 = pd.read_csv('file2.csv')[['year', 'crime_type2']]
#df3 = pd.read_csv('file3.csv')[['year', 'crime_type3']]
#df4 = pd.read_csv('file4.csv')[['year', 'crime_type4']]
#df5 = pd.read_csv('file5.csv')[['year', 'crime_type5']]


#df1 = pd.read_csv("E:\IIT_Guwhati\Semester_02\CS_595_Data_visualization_lab\Project\Crime\Auto_theft.csv")[['Year', 'Auto_Theft_Stolen']]
#df2 = pd.read_csv("E:\IIT_Guwhati\Semester_02\CS_595_Data_visualization_lab\Project\Crime\Complaints_against_police.csv")[['Year', 'CPA_Case
#df3 = pd.read_csv("E:\IIT_Guwhati\Semester_02\CS_595_Data_visualization_lab\Project\Crime\Murders.csv")[['Year', 'Victims_Total']]
#df4 = pd.read_csv("E:\IIT_Guwhati\Semester_02\CS_595_Data_visualization_lab\Project\Crime\Property_stolen_and_recovered.csv")[['Year', 'Case
#f5 = pd.read_csv("E:\IIT_Guwhati\Semester_02\CS_595_Data_visualization_lab\Project\Crime\Rape_Victims.csv")[['Year', 'Rape_Cases_Reported']]


#merged_df = merged_df.set_index('Year')


#merged_df = merged_df.transpose()


#merged_df.plot(kind='bar', figsize=(10, 5))
#plt.title('Crimes by year and type')
#plt.xlabel('Crime type')
#plt.ylabel('Number of crimes')
#plt.show()
```

Colab paid products - Cancel contracts here

✓ 0s   completed at 10:56 PM   ● ✕