

Project Milestone-1

Agasthya Katkam

For this milestone, I focused on preparing the data and conducting an initial exploratory data analysis to gain a clearer understanding of the dataset's structure and content. Here are some reflections on this stage of the project:

1. **Data Preparation and Cleaning:** Handling missing values and ensuring consistent formatting were essential steps. A clean dataset enables smooth visualization and analysis processes later. For instance, I chose to drop rows with missing values, but depending on the missing data pattern, it might be beneficial to consider imputation.

```
In [1]: # Import necessary Libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [2]: insurance_data = pd.read_csv("Automobile_Insurance_Complaint_Rankings_Beginning_2009.csv")
```

```
In [3]: insurance_data.head()
```

```
Out[3]:
```

	NAIC	Company Name	Ratio	Upheld Complaints	Question of Fact Complaints	Not Upheld Complaints	Total Complaints	Premiums Written (in Millions)	Rank	Filing Year
0	35882	GEICO General Insurance Company	0.0	0	2	10	12	2716.297785	1	2022
1	15024	Preferred Mutual Insurance Company	0.0	0	1	10	11	55.354694	2	2022
2	23787	Nationwide Mutual Insurance Company	0.0	0	2	3	5	41.949818	3	2022
3	42919	Progressive Northwestern Insurance Company	0.0	0	3	3	6	39.621536	4	2022
4	40169	Farmers Casualty Insurance Company	0.0	0	0	6	6	37.100223	5	2022

2. **Exploratory Data Analysis (EDA):** Conducting EDA provided initial insights, such as identifying which companies receive the most complaints and how complaint volumes change over time. Visualizations like the line plot for trends over time revealed potential patterns that could be explored in greater depth in subsequent milestones.

```
In [5]: # Data Cleaning
# Check for missing values
print("\nMissing values per column:")
print(insurance_data.isnull().sum())
```

```
Missing values per column:
NAIC                                0
Company Name                        0
Ratio                              0
Upheld Complaints                   0
Question of Fact Complaints         0
Not Upheld Complaints               0
Total Complaints                    0
Premiums Written (in Millions)     0
Rank                                0
Filing Year                         0
dtype: int64
```

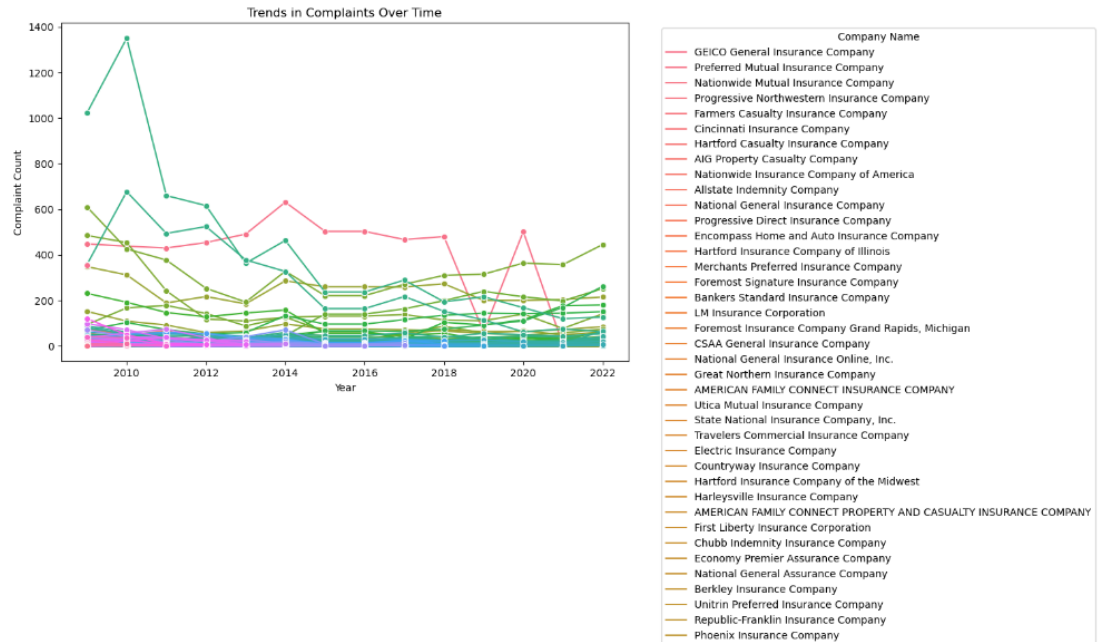
```
In [6]: # Drop rows or fill missing values if any
insurance_data = insurance_data.dropna() # Alternatively, use data.fillna() if you'd like to fill in missing data
```

```
In [8]: if 'Year' in insurance_data.columns:
        data['Year'] = insurance_data['Year'].astype(int)
```

```
In [10]: # Exploratory Data Analysis (EDA)
# Generate summary statistics
print("\nSummary Statistics:")
print(insurance_data.describe())
```

```
Summary Statistics:
      NAIC      Ratio  Upheld Complaints \
count  2209.000000  2209.000000    2209.000000
mean   23769.832051    30.321138         2.993210
std     9572.953302   1087.406803        11.073595
min    10022.000000     0.000000         0.000000
25%    15210.000000     0.000000         0.000000
50%    23450.000000     0.010000         1.000000
75%    29939.000000     0.080000         2.000000
max     43974.000000  48780.487800        265.000000
```

```
In [17]: # Visualize complaints over years
if 'Filing Year' in insurance_data.columns:
    plt.figure(figsize=(10, 6))
    sns.lineplot(data=insurance_data, x='Filing Year', y='Total Complaints', hue='Company Name', marker="o")
    plt.title('Trends in Complaints Over Time')
    plt.xlabel('Year')
    plt.ylabel('Complaint Count')
    plt.legend(title='Company Name', bbox_to_anchor=(1.05, 1), loc='upper left')
    plt.show()
```



3. **Challenges:** Challenge was in deciding the best approach to handle missing data without biasing the analysis and we did not face this challenge as there are no missing values.
4. **Next Steps:** Based on these insights, I plan to move on to creating interactive visualizations to allow stakeholders to explore the data dynamically, which will help in identifying more granular patterns within specific complaint types and companies.