Assignment -2

DATA VISUALIZATION

Name: Dhivyadharsini.S. K

Reg no:21BCE0776

Branch: B. Tech computer science and engineering

Campus: VIT Vellore

Question:

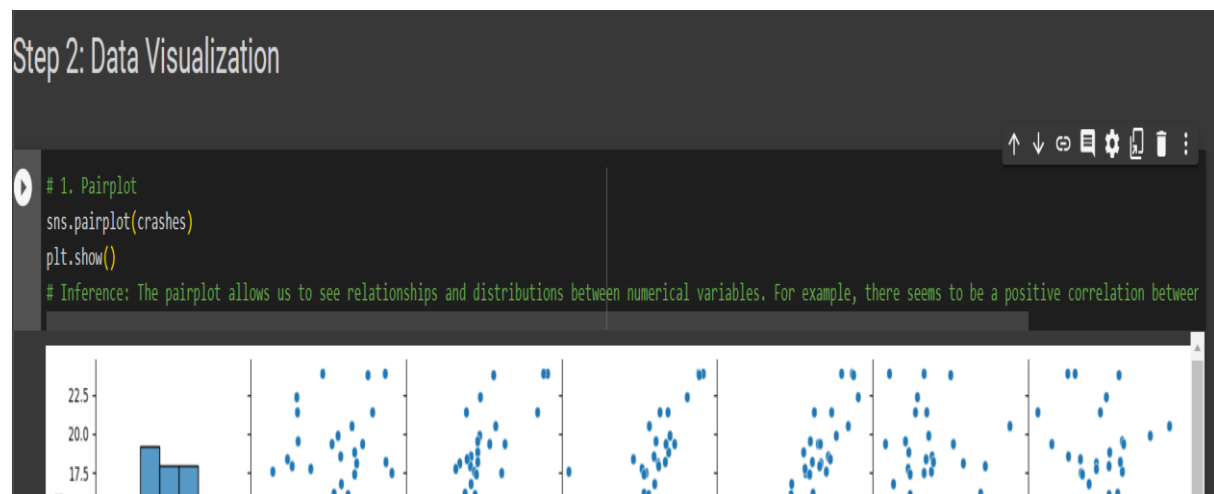**1.take a car crashes dataset from seaborn library**

2.load the dataset

```
[1]  # Import necessary libraries
     import seaborn as sns
     import matplotlib.pyplot as plt

[2]  # Step 1: Load the car crashes dataset from Seaborn
     crashes = sns.load_dataset('car_crashes')
```
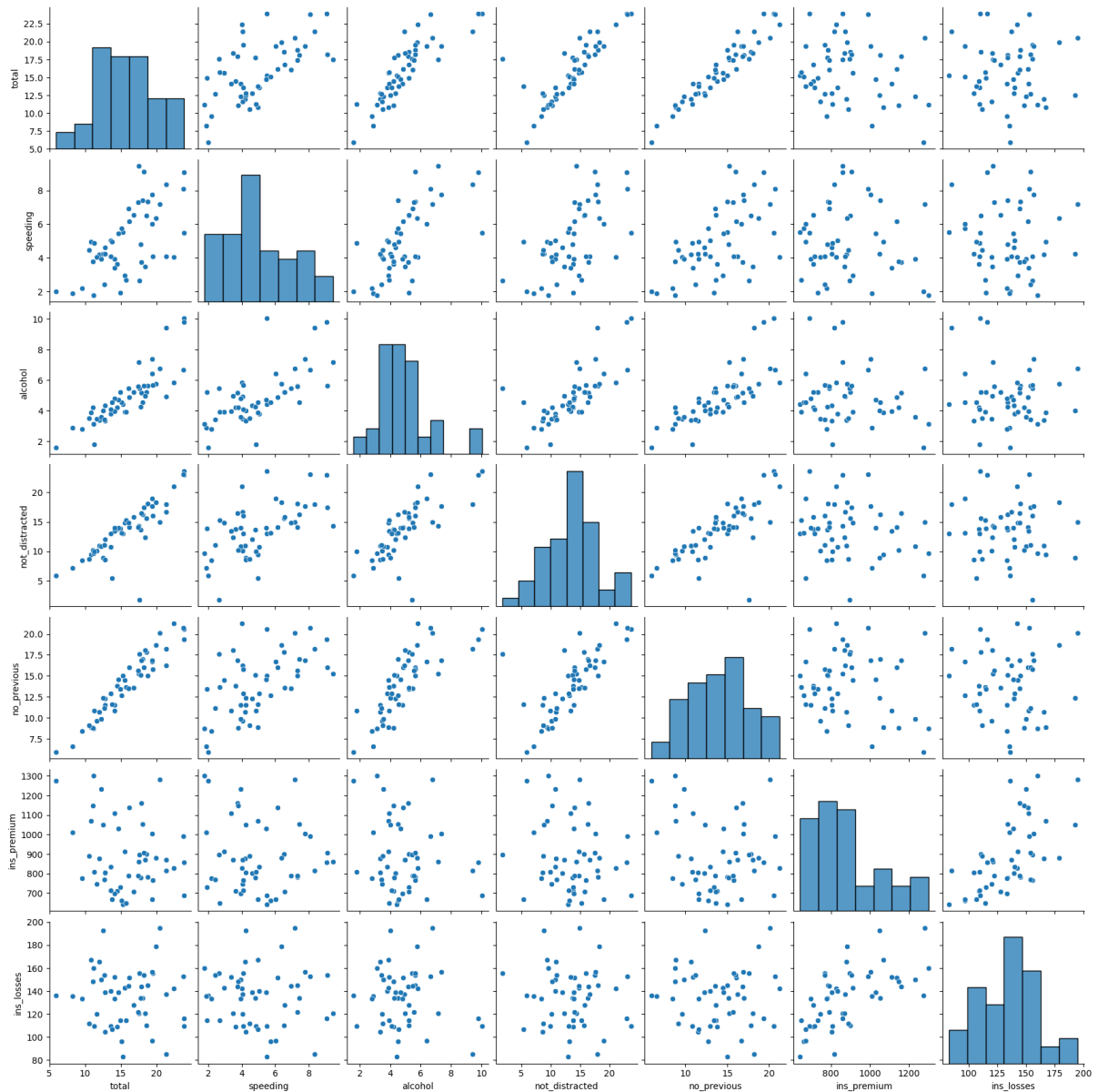
3.data visualiation

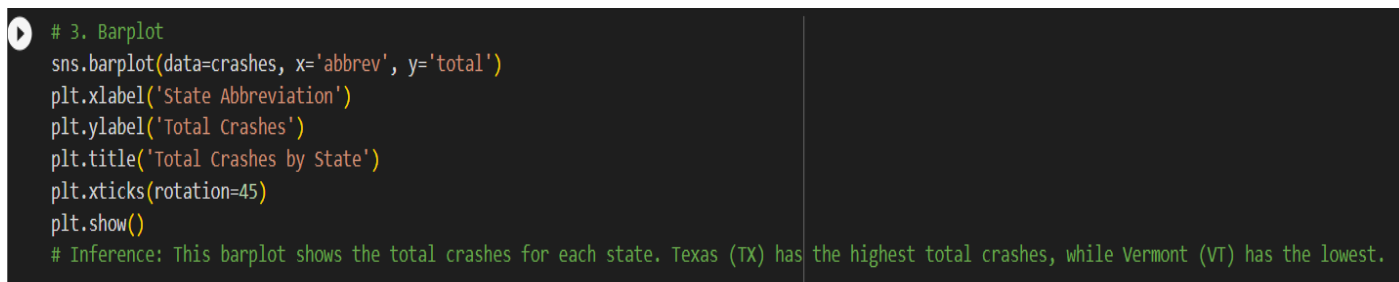**4.inference is must for each every graphy**


```
# 1. Pairplot
sns.pairplot(crashes)
plt.show()
# Inference: The pairplot allows us to see relationships and distributions between numerical variables. For example, there seems to be a positive correlation betweer
```
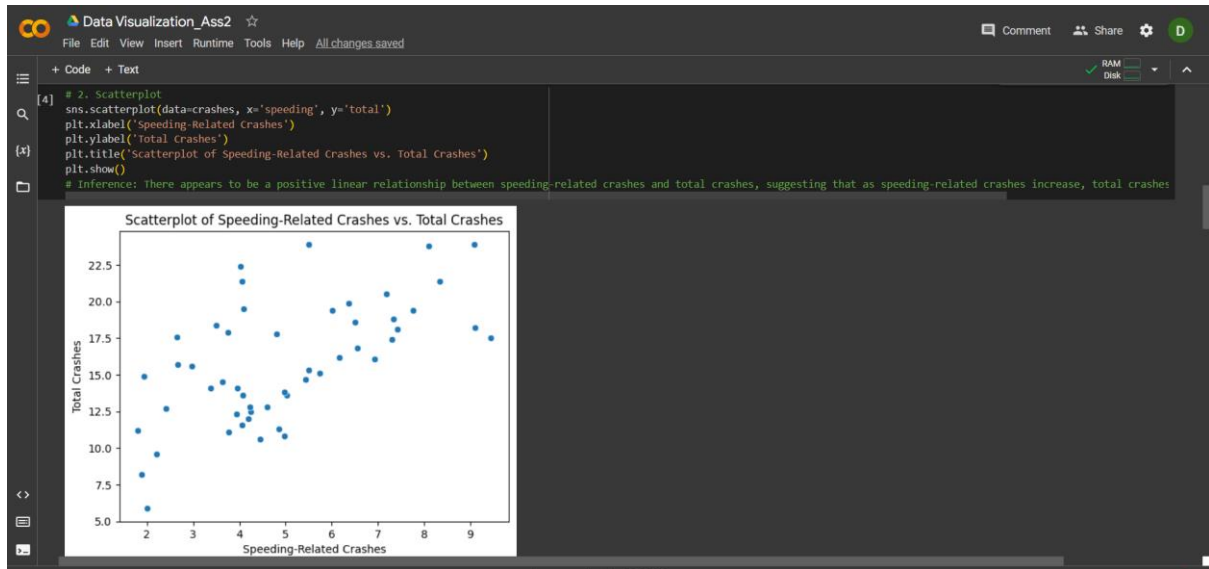
Assignment -2

DATA VISUALIZATION

Assignment -2

DATA VISUALIZATION


```
# 2. Scatterplot
sns.scatterplot(data=crashes, x='speeding', y='total')
plt.xlabel('Speeding-Related Crashes')
plt.ylabel('Total Crashes')
plt.title('Scatterplot of Speeding-Related Crashes vs. Total Crashes')
plt.show()
# Inference: There appears to be a positive linear relationship between speeding-related crashes and total crashes, suggesting that as speeding-related crashes increase, total crashes
```



```
# 3. Barplot
sns.barplot(data=crashes, x='abbrev', y='total')
plt.xlabel('State Abbreviation')
plt.ylabel('Total Crashes')
plt.title('Total Crashes by State')
plt.xticks(rotation=45)
plt.show()
# Inference: This barplot shows the total crashes for each state. Texas (TX) has the highest total crashes, while Vermont (VT) has the lowest.
```
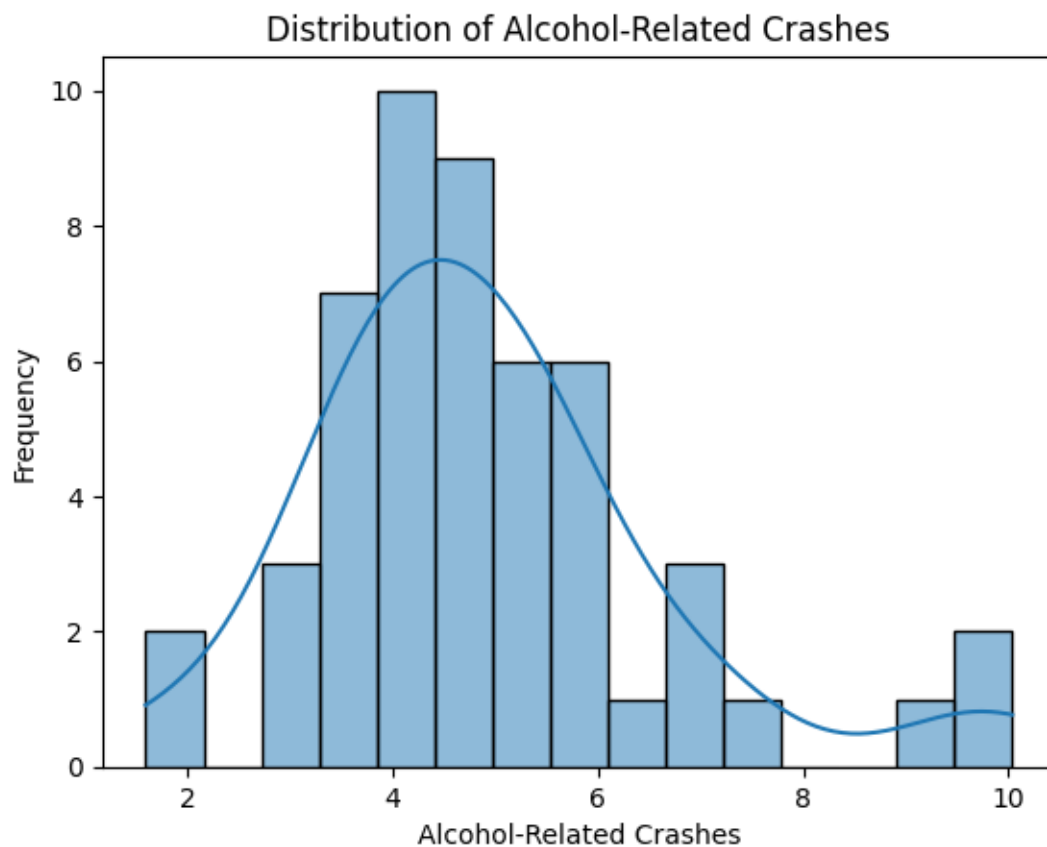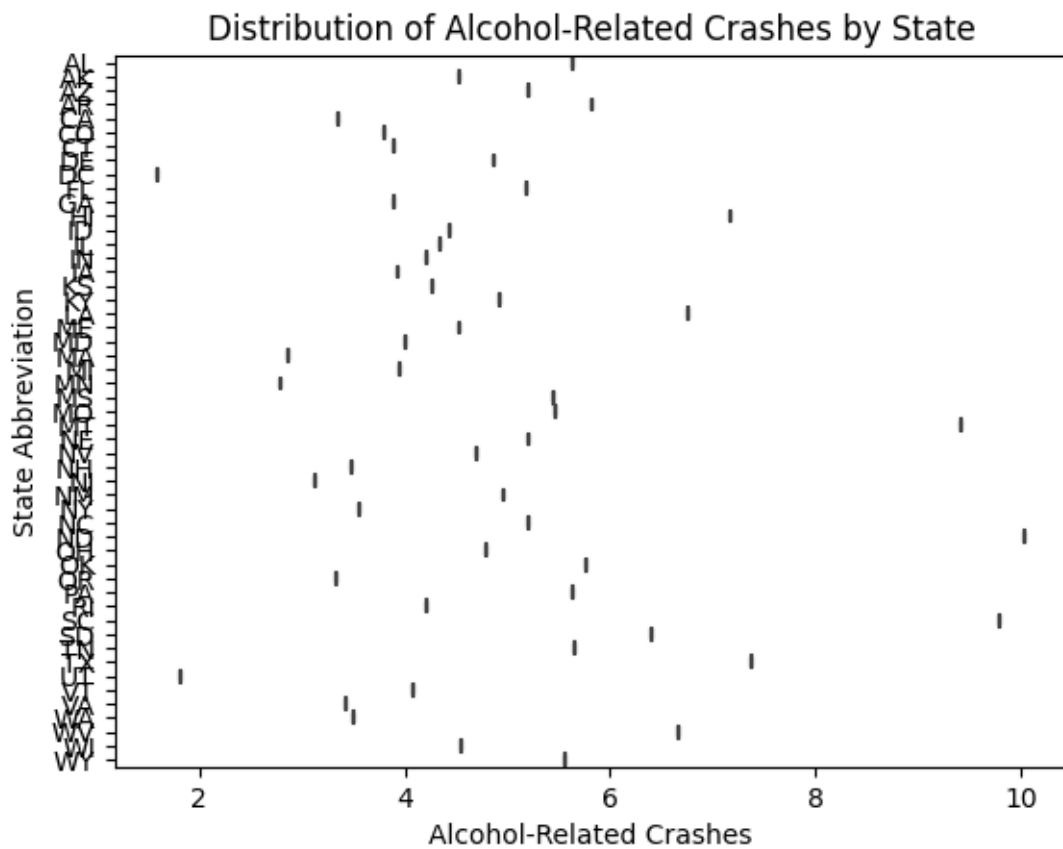
Assignment -2

DATA VISUALIZATION

```python
# 4. Histogram
sns.histplot(crashes['alcohol'], bins=15, kde=True)
plt.xlabel('Alcohol-Related Crashes')
plt.ylabel('Frequency')
plt.title('Distribution of Alcohol-Related Crashes')
plt.show()
# Inference: The histogram illustrates the distribution of alcohol-related crashes. It
```


Distribution of Alcohol-Related Crashes

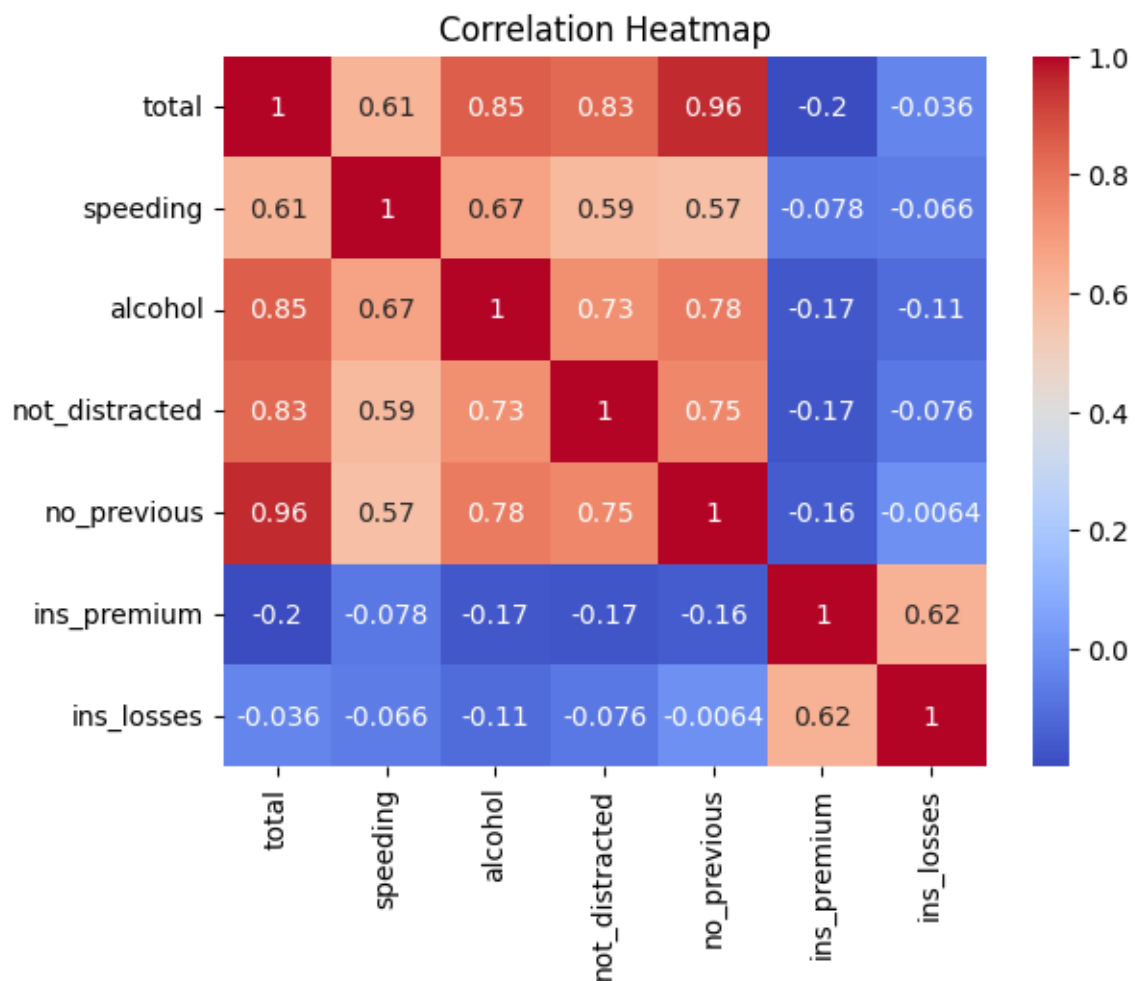# Assignment -2

DATA VISUALIZATION

```
# 5. Boxplot
sns.boxplot(data=crashes, x='alcohol', y='abbrev')
plt.xlabel('Alcohol-Related Crashes')
plt.ylabel('State Abbreviation')
plt.title('Distribution of Alcohol-Related Crashes by State')
plt.show()
# Inference: The boxplot shows the distribution of alcohol-related crashes by state.
```


Distribution of Alcohol-Related Crashes by State

```
# 6. Heatmap (Correlation)
correlation_matrix = crashes.corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
# Inference: The heatmap displays the correlation between numerical variables.
```


Correlation Heatmap

Assignment -2

DATA VISUALIZATION

```
# 7. Violinplot
sns.violinplot(data=crashes, x='ins_premium', y='abbrev', inner='quartile')
plt.xlabel('Insurance Premiums')
plt.ylabel('State Abbreviation')
plt.title('Distribution of Insurance Premiums by State')
plt.show()
# Inference: The violinplot shows the distribution of insurance premiums by state, including quartile summaries.
```


Distribution of Insurance Premiums by State
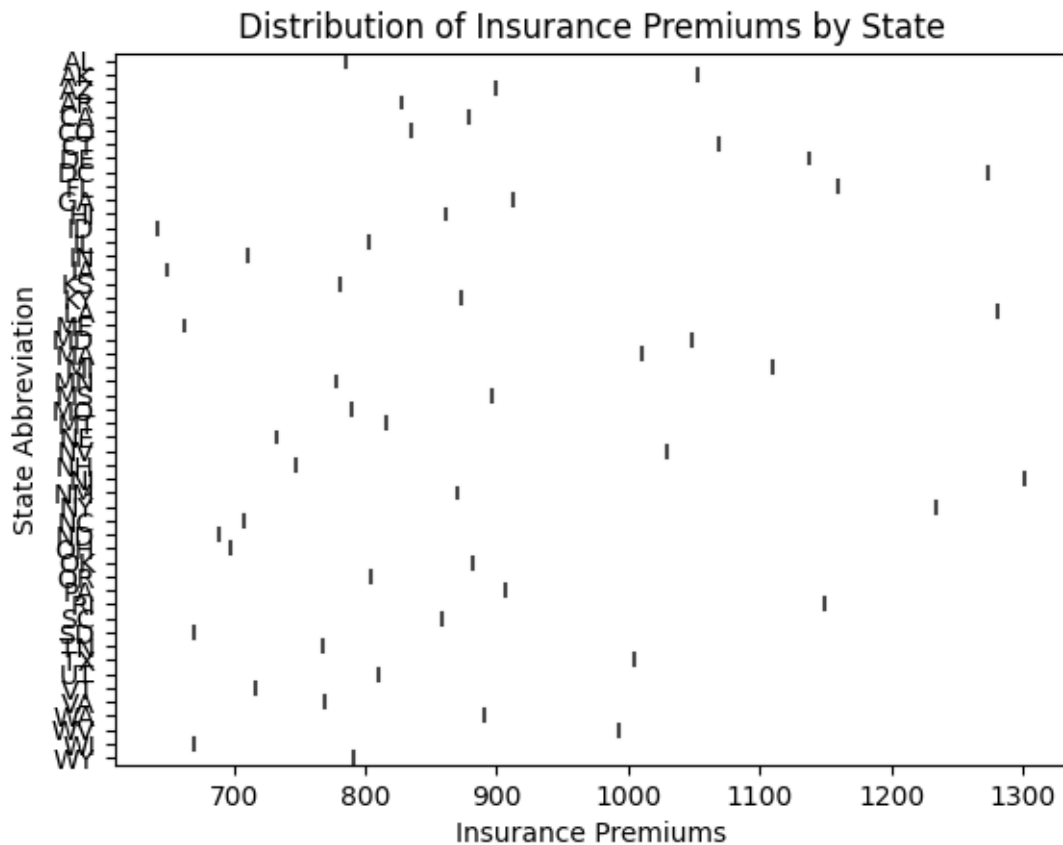
DATA VISUALIZATION

```
# 8. Countplot based on 'not_distracted' variable
sns.countplot(data=crashes, x='not_distracted')
plt.xlabel('Not Distracted')
plt.ylabel('Count')
plt.title('Count of Not Distracted vs. Distracted Drivers')
plt.show()
```


Count of Not Distracted vs. Distracted Drivers