

Assignment_2

September 14, 2023

1 Assignment-2

Anand Misra

21BAI1105

1.1 Importing the required packages

```
[1]: import pandas as pd
import seaborn as sns
from matplotlib import pyplot as plt
```

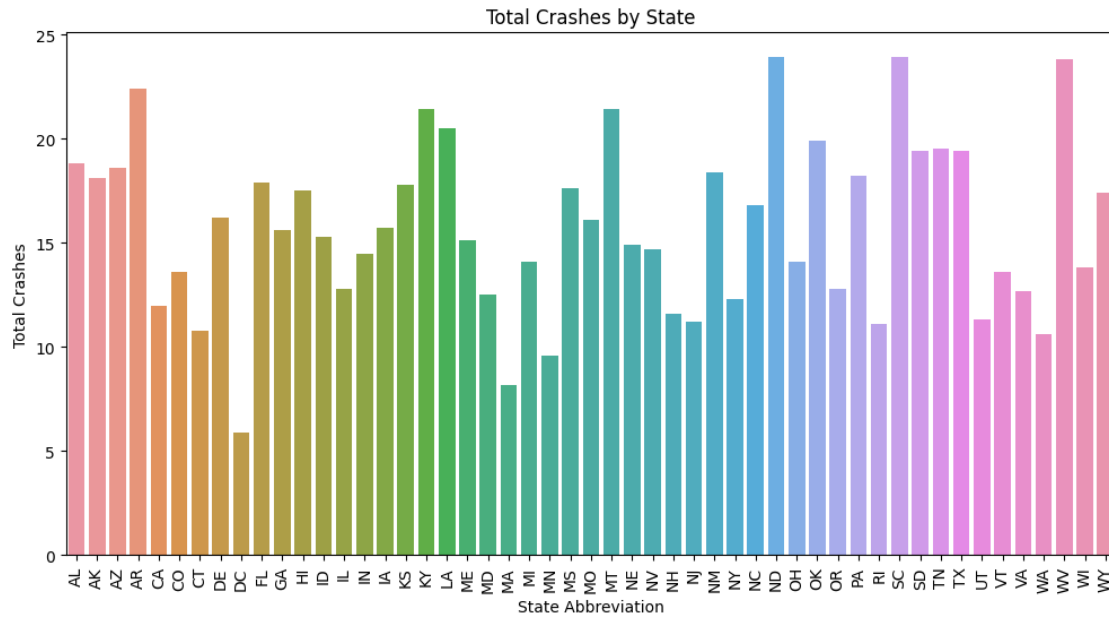
1.2 Loading the dataset

```
[3]: df = sns.load_dataset('car_crashes')
```

1.3 Performing data visualization with their inferences

1.3.1 1. Bar plot: Total crashes by state

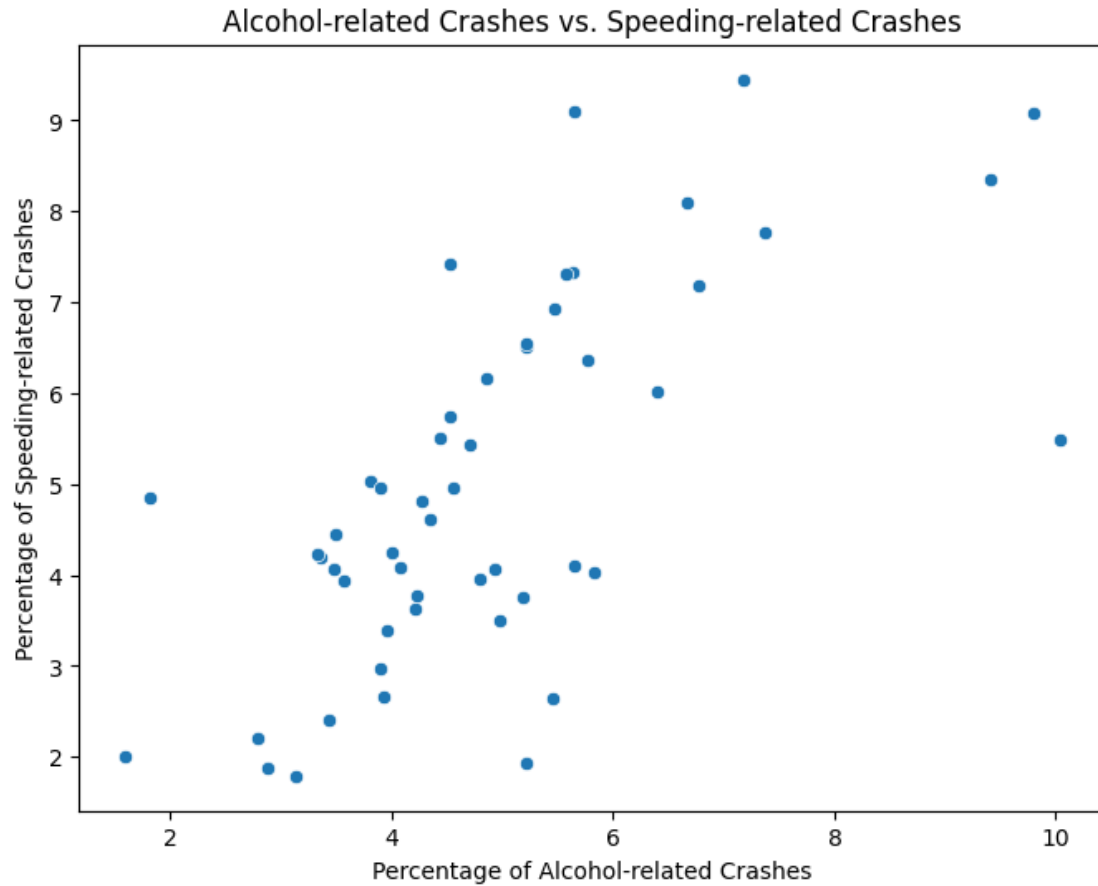
```
[4]: plt.figure(figsize=(12, 6))
sns.barplot(x='abbrev', y='total', data=df)
plt.title('Total Crashes by State')
plt.xlabel('State Abbreviation')
plt.ylabel('Total Crashes')
plt.xticks(rotation=90)
plt.show()
```



Inference: The total number of crashes varies by state, with some states having notably higher crash rates than others.

1.3.2 2. Scatter plot: Alcohol-related crashes vs. Speeding-related crashes

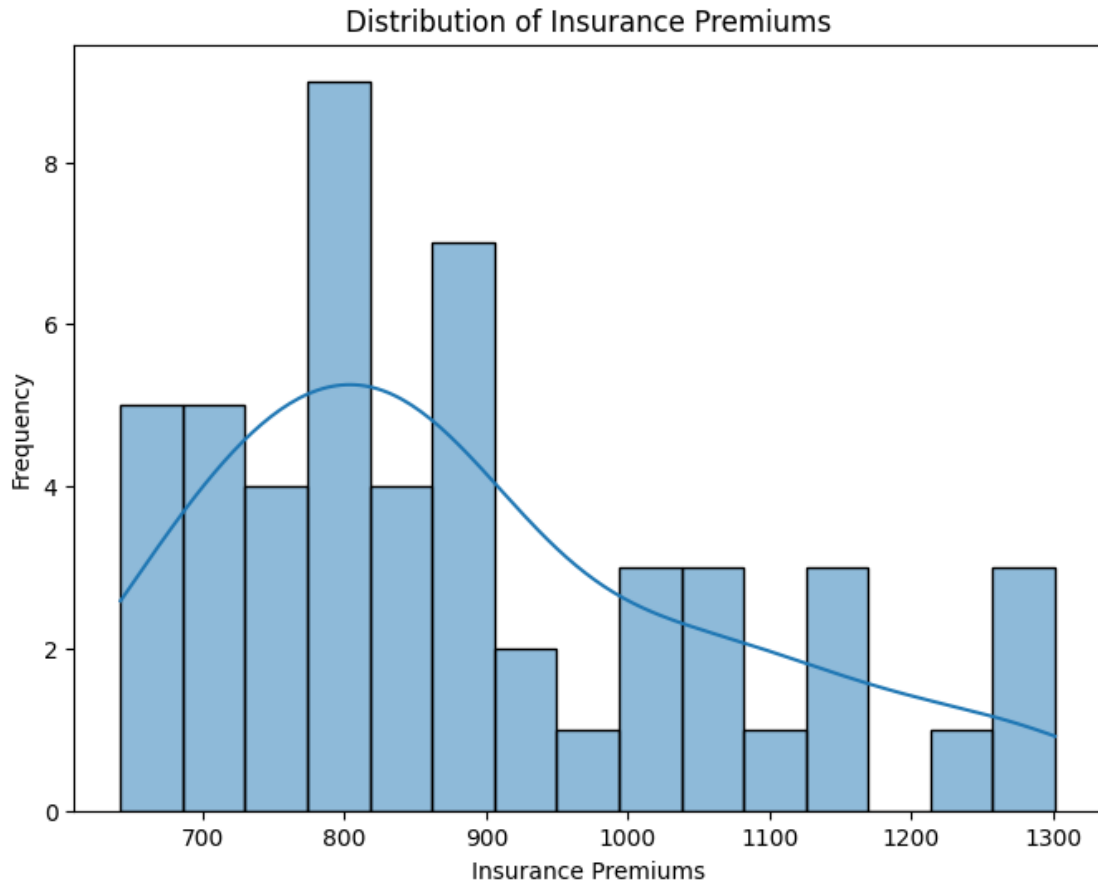
```
[5]: plt.figure(figsize=(8, 6))
sns.scatterplot(x='alcohol', y='speeding', data=df)
plt.title('Alcohol-related Crashes vs. Speeding-related Crashes')
plt.xlabel('Percentage of Alcohol-related Crashes')
plt.ylabel('Percentage of Speeding-related Crashes')
plt.show()
```



Inference: There is a positive correlation between alcohol-related crashes and speeding-related crashes.

1.3.3 3. Histogram: Distribution of insurance premiums

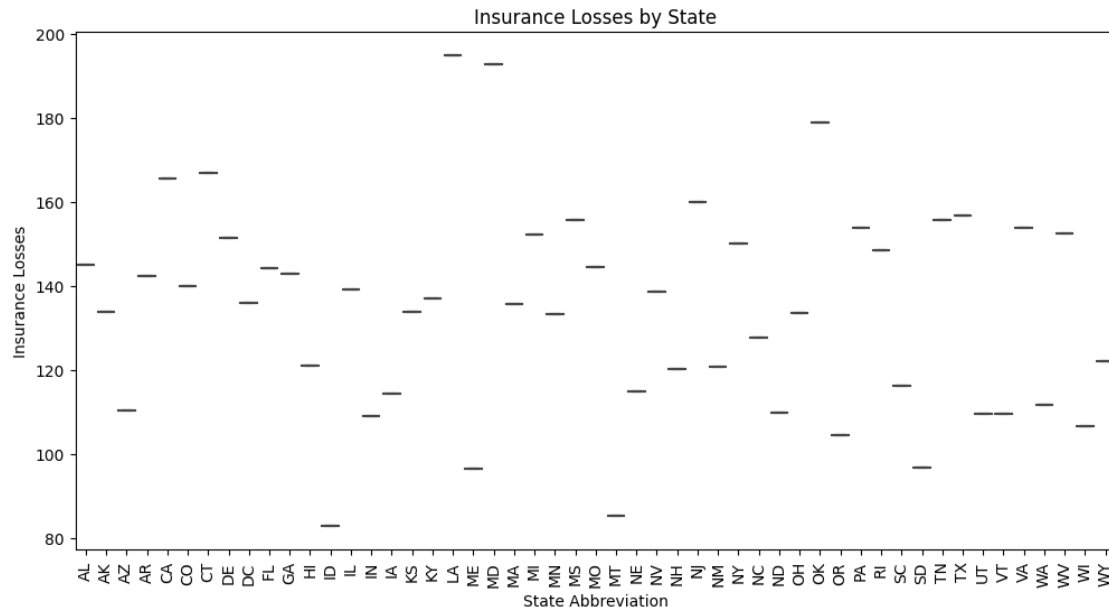
```
[8]: plt.figure(figsize=(8, 6))
sns.histplot(df['ins_premium'], bins=15, kde=True)
plt.title('Distribution of Insurance Premiums')
plt.xlabel('Insurance Premiums')
plt.ylabel('Frequency')
plt.show()
```



Inference: The distribution of insurance premiums appears to be right-skewed, with most states having lower premiums.

1.3.4 4. Box plot: Insurance losses by state

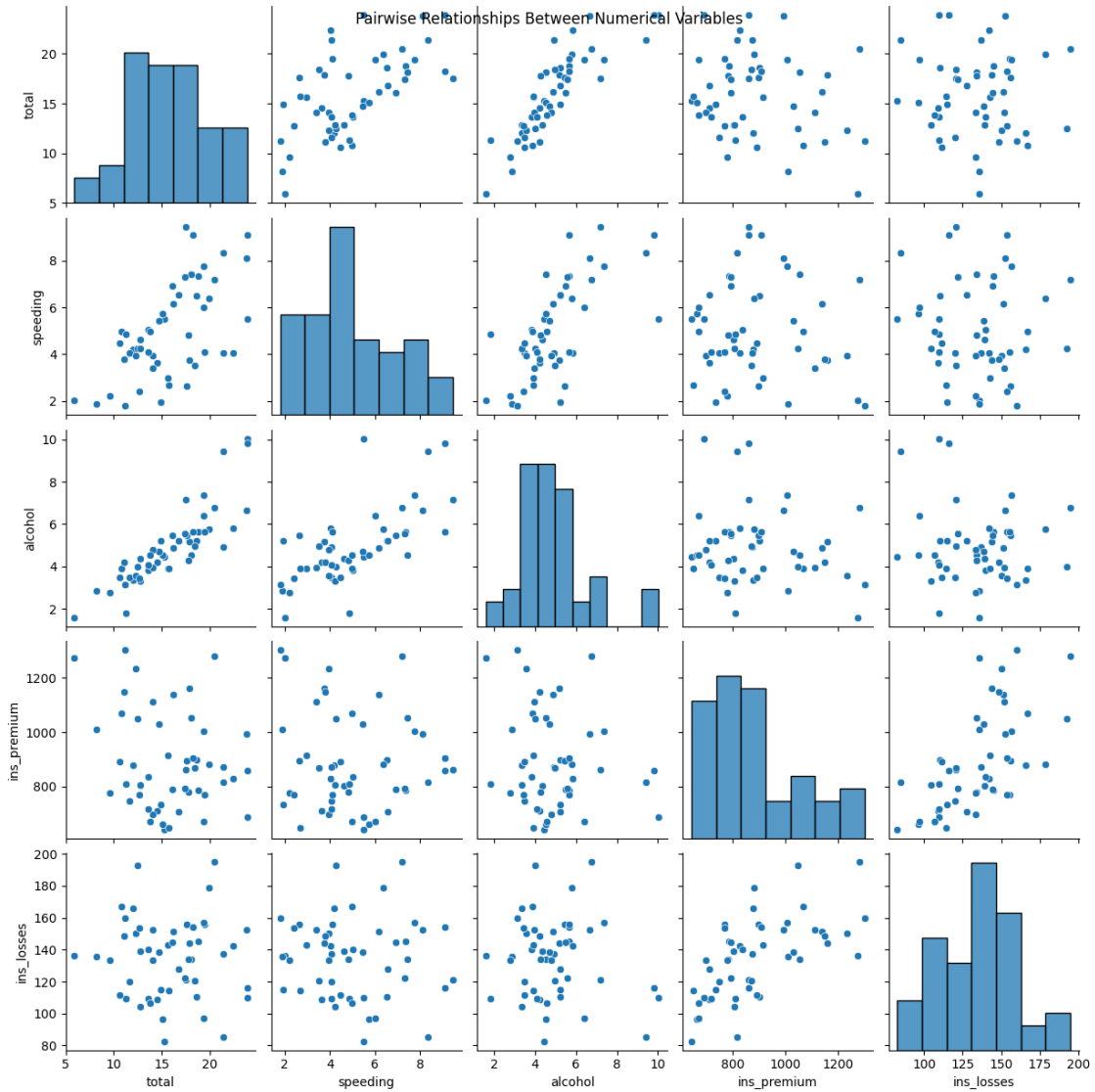
```
[9]: plt.figure(figsize=(12, 6))
sns.boxplot(x='abbrev', y='ins_losses', data=df)
plt.title('Insurance Losses by State')
plt.xlabel('State Abbreviation')
plt.ylabel('Insurance Losses')
plt.xticks(rotation=90)
plt.show()
```



Inference: There is significant variability in insurance losses by state, with some states having higher losses than others.

1.3.5 5. Pair plot: Pairwise relationships between numerical variables

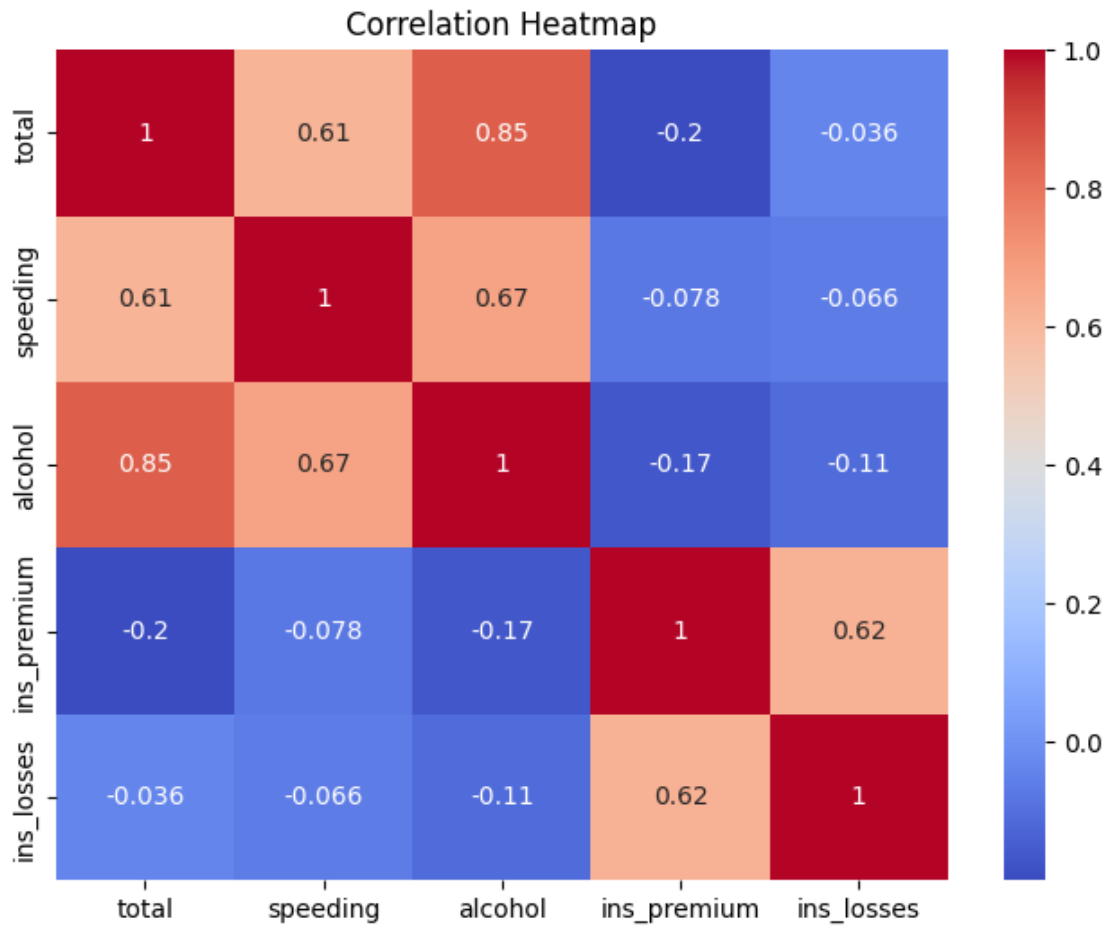
```
[10]: sns.pairplot(df[['total', 'speeding', 'alcohol', 'ins_premium', 'ins_losses']])
plt.suptitle('Pairwise Relationships Between Numerical Variables')
plt.show()
```



Inference: The pair plot shows the relationships and distributions between numerical variables.

1.3.6 6. Heatmap: Correlation between numerical variables

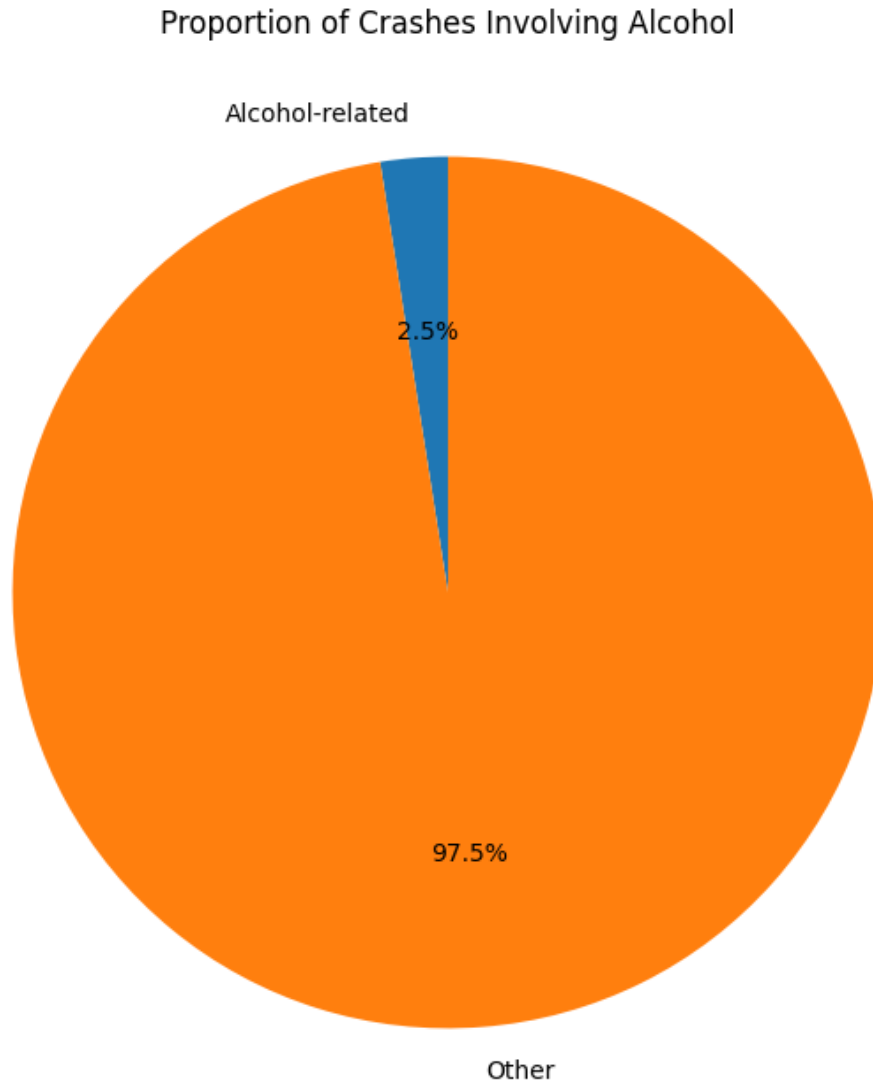
```
[11]: plt.figure(figsize=(8, 6))
sns.heatmap(df[['total', 'speeding', 'alcohol', 'ins_premium', 'ins_losses']].
            corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```



Inference: The heatmap shows the correlation between numerical variables; some variables are positively correlated, while others are negatively correlated.

1.3.7 7. Pie chart: Proportion of crashes involving alcohol

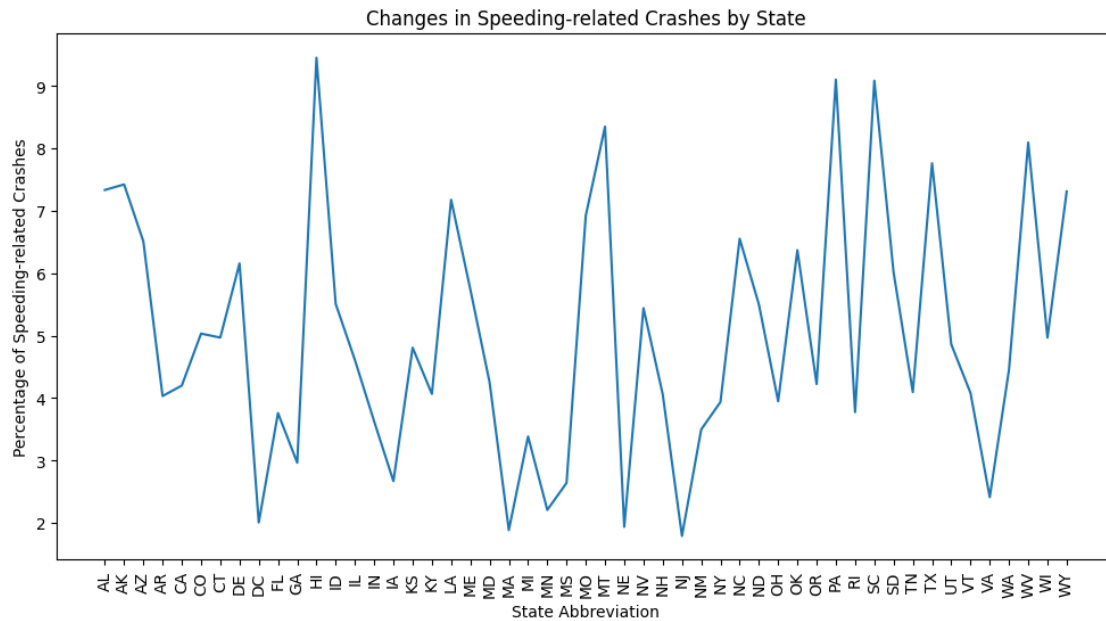
```
[12]: plt.figure(figsize=(8, 8))
alcohol_proportions = df['alcohol'].sum() / 100
labels = ['Alcohol-related', 'Other']
sizes = [alcohol_proportions, 100 - alcohol_proportions]
plt.pie(sizes, labels=labels, autopct='%1.1f%%', startangle=90)
plt.title('Proportion of Crashes Involving Alcohol')
plt.show()
```



Inference: A significant proportion of crashes involve alcohol.

1.3.8 8. Line plot: Changes in speeding-related crashes over states

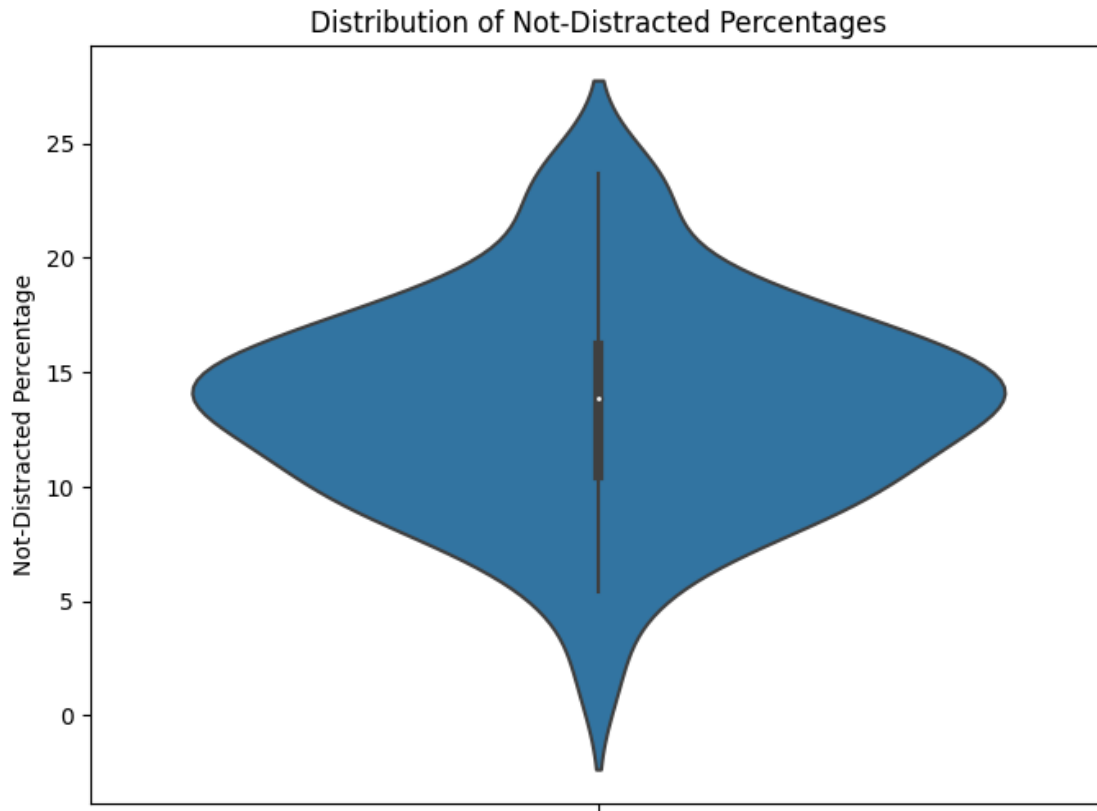
```
[13]: plt.figure(figsize=(12, 6))
sns.lineplot(x='abbrev', y='speeding', data=df)
plt.title('Changes in Speeding-related Crashes by State')
plt.xlabel('State Abbreviation')
plt.ylabel('Percentage of Speeding-related Crashes')
plt.xticks(rotation=90)
plt.show()
```

Inference: There are variations in the percentage of speeding-related crashes across different states.

1.3.9 9. Violin plot: Distribution of not-distracted percentages

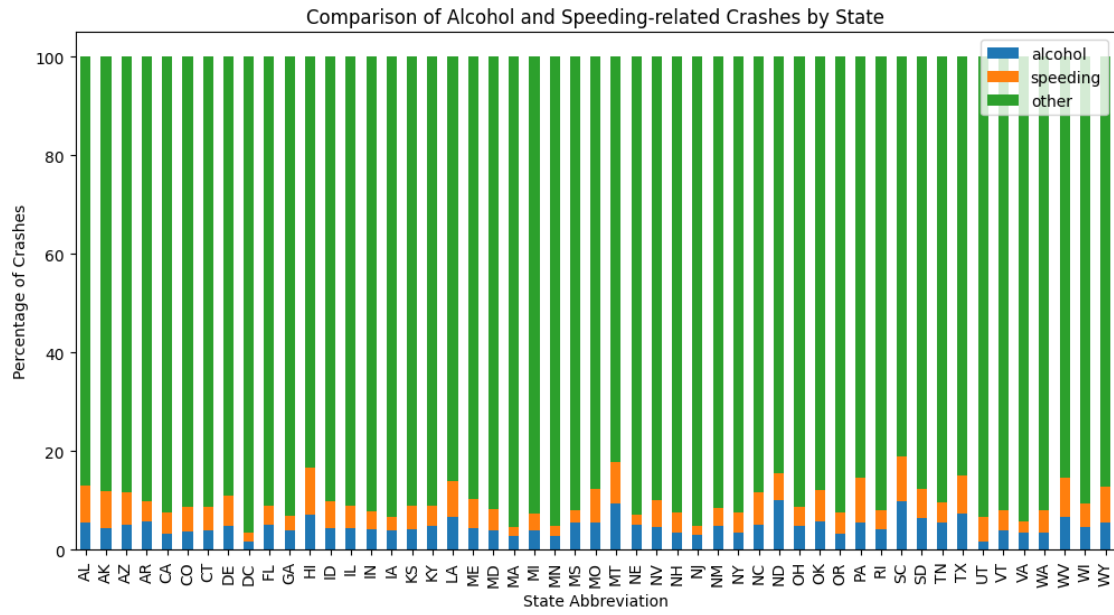
```
[14]: plt.figure(figsize=(8, 6))
sns.violinplot(y='not_distracted', data=df)
plt.title('Distribution of Not-Distracted Percentages')
plt.ylabel('Not-Distracted Percentage')
plt.show()
```



Inference: The distribution of not-distracted percentages is relatively symmetric, with some outliers.

1.3.10 10. Stacked bar plot: Comparison of alcohol and speeding-related crashes by state

```
[15]: df['other'] = 100 - df['alcohol'] - df['speeding']
df[['alcohol', 'speeding', 'other']].set_index(df['abbrev']).plot(kind='bar',
    ↪stacked=True, figsize=(12, 6))
plt.title('Comparison of Alcohol and Speeding-related Crashes by State')
plt.xlabel('State Abbreviation')
plt.ylabel('Percentage of Crashes')
plt.xticks(rotation=90)
plt.show()
```



Inference: This plot visually compares the percentage of alcohol-related, speeding-related, and other crashes by state.

[]: