# seaborn-assignment2

September 12, 2023

## 1 Niyati Mittal

[3]:
```
import seaborn as sns
import matplotlib.pyplot as plt
```

[ ]:
```
print(sns.get_dataset_names())
```

```
['anagrams', 'anscombe', 'attention', 'brain_networks', 'car_crashes',
'diamonds', 'dots', 'dowjones', 'exercise', 'flights', 'fmri', 'geyser', 'glue',
'healthexp', 'iris', 'mpg', 'penguins', 'planets', 'seaice', 'taxis', 'tips',
'titanic']
```

## 2 Loading the dataset car_crashes

[4]:
```
df=sns.load_dataset('car_crashes')
```

[5]:
```
df
```

[5]:

| | total | speeding | alcohol | not_distracted | no_previous | ins_premium | \ |
|---|---|---|---|---|---|---|---|
| 0 | 18.8 | 7.332 | 5.640 | 18.048 | 15.040 | 784.55 | |
| 1 | 18.1 | 7.421 | 4.525 | 16.290 | 17.014 | 1053.48 | |
| 2 | 18.6 | 6.510 | 5.208 | 15.624 | 17.856 | 899.47 | |
| 3 | 22.4 | 4.032 | 5.824 | 21.056 | 21.280 | 827.34 | |
| 4 | 12.0 | 4.200 | 3.360 | 10.920 | 10.680 | 878.41 | |
| 5 | 13.6 | 5.032 | 3.808 | 10.744 | 12.920 | 835.50 | |
| 6 | 10.8 | 4.968 | 3.888 | 9.396 | 8.856 | 1068.73 | |
| 7 | 16.2 | 6.156 | 4.860 | 14.094 | 16.038 | 1137.87 | |
| 8 | 5.9 | 2.006 | 1.593 | 5.900 | 5.900 | 1273.89 | |
| 9 | 17.9 | 3.759 | 5.191 | 16.468 | 16.826 | 1160.13 | |
| 10 | 15.6 | 2.964 | 3.900 | 14.820 | 14.508 | 913.15 | |
| 11 | 17.5 | 9.450 | 7.175 | 14.350 | 15.225 | 861.18 | |
| 12 | 15.3 | 5.508 | 4.437 | 13.005 | 14.994 | 641.96 | |
| 13 | 12.8 | 4.608 | 4.352 | 12.032 | 12.288 | 803.11 | |
| 14 | 14.5 | 3.625 | 4.205 | 13.775 | 13.775 | 710.46 | |
| 15 | 15.7 | 2.669 | 3.925 | 15.229 | 13.659 | 649.06 | |
| 16 | 17.8 | 4.806 | 4.272 | 13.706 | 15.130 | 780.45 | |
| 17 | 21.4 | 4.066 | 4.922 | 16.692 | 16.264 | 872.51 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| 18 | 20.5 | 7.175 | 6.765 | 14.965 | 20.090 | 1281.55 |
| 19 | 15.1 | 5.738 | 4.530 | 13.137 | 12.684 | 661.88 |
| 20 | 12.5 | 4.250 | 4.000 | 8.875 | 12.375 | 1048.78 |
| 21 | 8.2 | 1.886 | 2.870 | 7.134 | 6.560 | 1011.14 |
| 22 | 14.1 | 3.384 | 3.948 | 13.395 | 10.857 | 1110.61 |
| 23 | 9.6 | 2.208 | 2.784 | 8.448 | 8.448 | 777.18 |
| 24 | 17.6 | 2.640 | 5.456 | 1.760 | 17.600 | 896.07 |
| 25 | 16.1 | 6.923 | 5.474 | 14.812 | 13.524 | 790.32 |
| 26 | 21.4 | 8.346 | 9.416 | 17.976 | 18.190 | 816.21 |
| 27 | 14.9 | 1.937 | 5.215 | 13.857 | 13.410 | 732.28 |
| 28 | 14.7 | 5.439 | 4.704 | 13.965 | 14.553 | 1029.87 |
| 29 | 11.6 | 4.060 | 3.480 | 10.092 | 9.628 | 746.54 |
| 30 | 11.2 | 1.792 | 3.136 | 9.632 | 8.736 | 1301.52 |
| 31 | 18.4 | 3.496 | 4.968 | 12.328 | 18.032 | 869.85 |
| 32 | 12.3 | 3.936 | 3.567 | 10.824 | 9.840 | 1234.31 |
| 33 | 16.8 | 6.552 | 5.208 | 15.792 | 13.608 | 708.24 |
| 34 | 23.9 | 5.497 | 10.038 | 23.661 | 20.554 | 688.75 |
| 35 | 14.1 | 3.948 | 4.794 | 13.959 | 11.562 | 697.73 |
| 36 | 19.9 | 6.368 | 5.771 | 18.308 | 18.706 | 881.51 |
| 37 | 12.8 | 4.224 | 3.328 | 8.576 | 11.520 | 804.71 |
| 38 | 18.2 | 9.100 | 5.642 | 17.472 | 16.016 | 905.99 |
| 39 | 11.1 | 3.774 | 4.218 | 10.212 | 8.769 | 1148.99 |
| 40 | 23.9 | 9.082 | 9.799 | 22.944 | 19.359 | 858.97 |
| 41 | 19.4 | 6.014 | 6.402 | 19.012 | 16.684 | 669.31 |
| 42 | 19.5 | 4.095 | 5.655 | 15.990 | 15.795 | 767.91 |
| 43 | 19.4 | 7.760 | 7.372 | 17.654 | 16.878 | 1004.75 |
| 44 | 11.3 | 4.859 | 1.808 | 9.944 | 10.848 | 809.38 |
| 45 | 13.6 | 4.080 | 4.080 | 13.056 | 12.920 | 716.20 |
| 46 | 12.7 | 2.413 | 3.429 | 11.049 | 11.176 | 768.95 |
| 47 | 10.6 | 4.452 | 3.498 | 8.692 | 9.116 | 890.03 |
| 48 | 23.8 | 8.092 | 6.664 | 23.086 | 20.706 | 992.61 |
| 49 | 13.8 | 4.968 | 4.554 | 5.382 | 11.592 | 670.31 |
| 50 | 17.4 | 7.308 | 5.568 | 14.094 | 15.660 | 791.14 |

| | ins_losses | abbrev |
|---|---|---|
| 0 | 145.08 | AL |
| 1 | 133.93 | AK |
| 2 | 110.35 | AZ |
| 3 | 142.39 | AR |
| 4 | 165.63 | CA |
| 5 | 139.91 | CO |
| 6 | 167.02 | CT |
| 7 | 151.48 | DE |
| 8 | 136.05 | DC |
| 9 | 144.18 | FL |
| 10 | 142.80 | GA |
| 11 | 120.92 | HI |

```
12       82.75     ID
13      139.15     IL
14      108.92     IN
15      114.47     IA
16      133.80     KS
17      137.13     KY
18      194.78     LA
19       96.57     ME
20      192.70     MD
21      135.63     MA
22      152.26     MI
23      133.35     MN
24      155.77     MS
25      144.45     MO
26       85.15     MT
27      114.82     NE
28      138.71     NV
29      120.21     NH
30      159.85     NJ
31      120.75     NM
32      150.01     NY
33      127.82     NC
34      109.72     ND
35      133.52     OH
36      178.86     OK
37      104.61     OR
38      153.86     PA
39      148.58     RI
40      116.29     SC
41       96.87     SD
42      155.57     TN
43      156.83     TX
44      109.48     UT
45      109.61     VT
46      153.72     VA
47      111.62     WA
48      152.56     WV
49      106.62     WI
50      122.04     WY
```

## 3 Dataset related Information

[6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
```

```
Data columns (total 8 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   total           51 non-null     float64
 1   speeding        51 non-null     float64
 2   alcohol         51 non-null     float64
 3   not_distracted  51 non-null     float64
 4   no_previous     51 non-null     float64
 5   ins_premium     51 non-null     float64
 6   ins_losses      51 non-null     float64
 7   abbrev          51 non-null     object
dtypes: float64(7), object(1)
memory usage: 3.3+ KB
```

[7]:
```
df.head()
```

[7]:

| | total | speeding | alcohol | not_distracted | no_previous | ins_premium | ins_losses | abbrev |
|---|---|---|---|---|---|---|---|---|
| 0 | 18.8 | 7.332 | 5.640 | 18.048 | 15.040 | 784.55 | 145.08 | AL |
| 1 | 18.1 | 7.421 | 4.525 | 16.290 | 17.014 | 1053.48 | 133.93 | AK |
| 2 | 18.6 | 6.510 | 5.208 | 15.624 | 17.856 | 899.47 | 110.35 | AZ |
| 3 | 22.4 | 4.032 | 5.824 | 21.056 | 21.280 | 827.34 | 142.39 | AR |
| 4 | 12.0 | 4.200 | 3.360 | 10.920 | 10.680 | 878.41 | 165.63 | CA |

[8]:
```
df.tail()
```

[8]:

| | total | speeding | alcohol | not_distracted | no_previous | ins_premium | ins_losses | abbrev |
|---|---|---|---|---|---|---|---|---|
| 46 | 12.7 | 2.413 | 3.429 | 11.049 | 11.176 | 768.95 | 153.72 | VA |
| 47 | 10.6 | 4.452 | 3.498 | 8.692 | 9.116 | 890.03 | 111.62 | WA |
| 48 | 23.8 | 8.092 | 6.664 | 23.086 | 20.706 | 992.61 | 152.56 | WV |
| 49 | 13.8 | 4.968 | 4.554 | 5.382 | 11.592 | 670.31 | 106.62 | WI |
| 50 | 17.4 | 7.308 | 5.568 | 14.094 | 15.660 | 791.14 | 122.04 | WY |

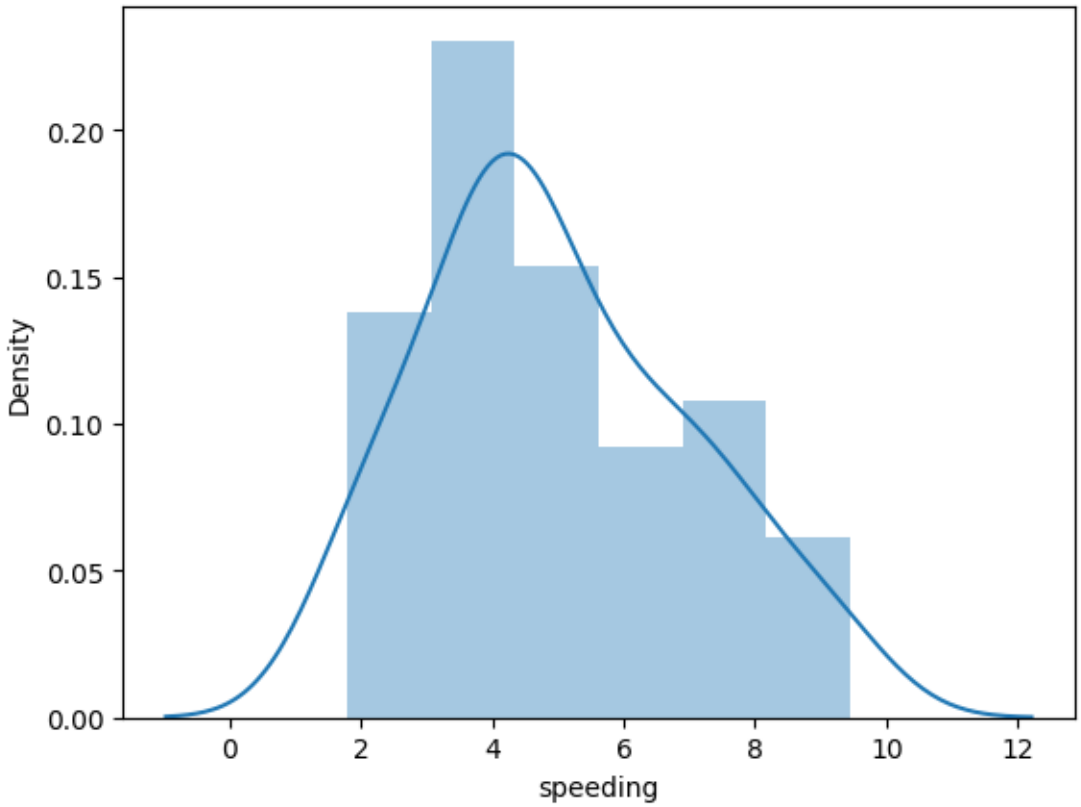
## 4 1. The Dist Plot

The distplot() shows the histogram distribution of data for a single column. The column name is passed as a parameter to the distplot() function. Let's see how the price of the ticket for each passenger is distributed.

[9]:
```
sns.distplot(df['speeding'])
```

```
<ipython-input-9-192cfe43af0d>:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['speeding'])
```

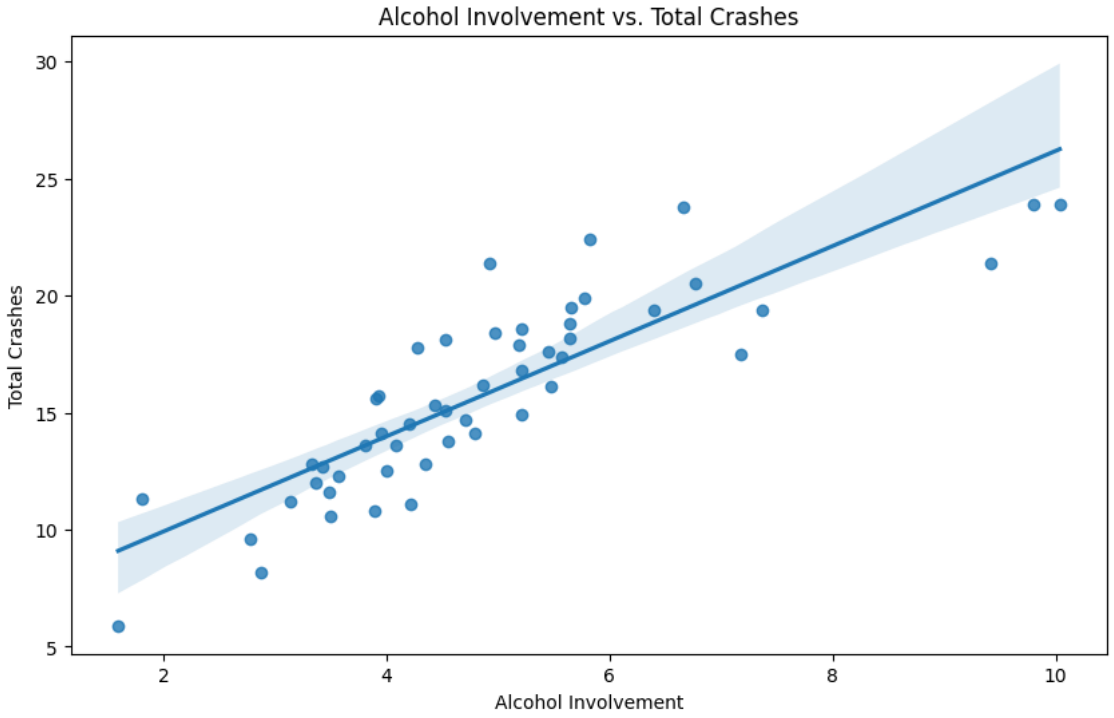
[9]: <Axes: xlabel='speeding', ylabel='Density'>



We can see that most of the speeds have been in between 2-9. The line that you see represents the kernel density estimation.

## 5 2. Scatter Plot

[21]:
```
# Scatterplot with Regression Line: Alcohol vs. Total Crashes
plt.figure(figsize=(10, 6))
sns.regplot(x='alcohol', y='total', data=df)
plt.xlabel("Alcohol Involvement")
plt.ylabel("Total Crashes")
plt.title("Alcohol Involvement vs. Total Crashes")
plt.show()
```




## 6 Inference:

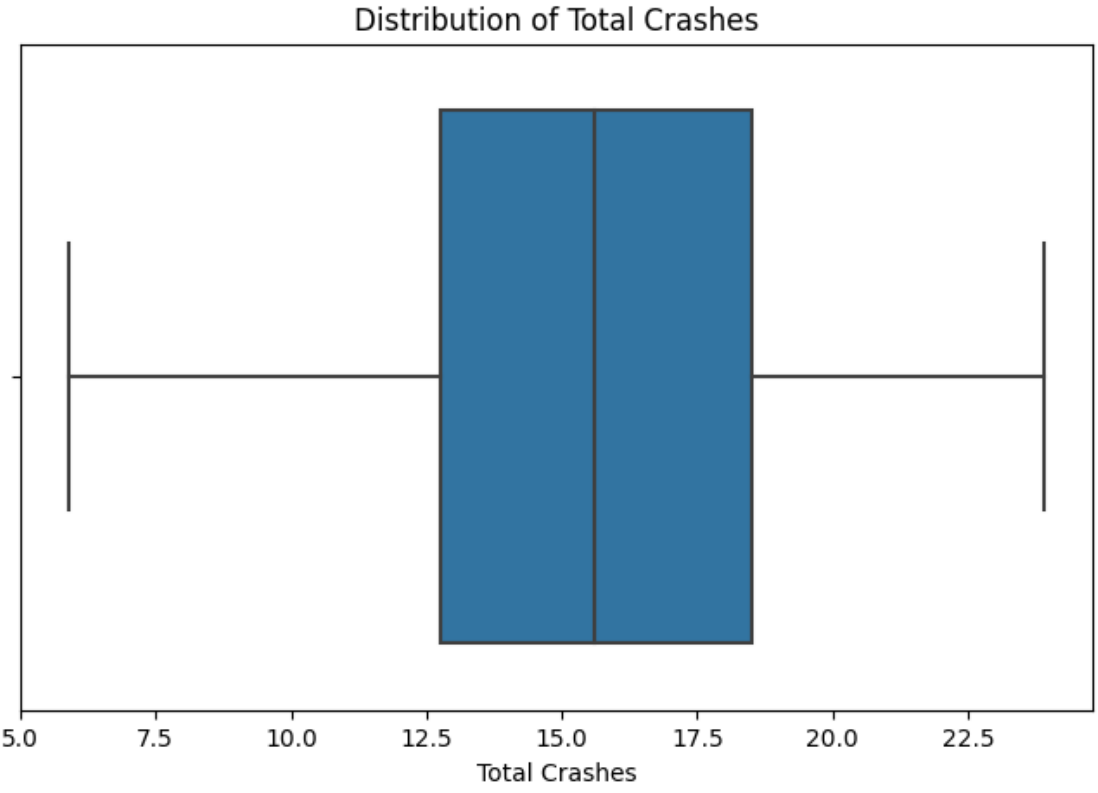
- The scatterplot with a regression line shows a clear positive correlation between alcohol involvement and total crashes.
- As the alcohol involvement increases, there is a corresponding increase in the total number of crashes.

## 7 Alcohol vs. Total Crashes (Scatterplot with Regression Line):

The scatterplot with the regression line confirms a strong positive correlation between alcohol involvement and the total number of crashes. As alcohol involvement increases, there is a clear trend of more crashes occurring. This indicates the significant role of alcohol in contributing to car crashes.

## 8 3. Box Plot

[13]:
```
# Boxplot to visualize the distribution of the 'total' column
plt.figure(figsize=(8, 5))
sns.boxplot(x='total', data=df)
plt.xlabel("Total Crashes")
plt.title("Distribution of Total Crashes")
plt.show()
```




## 9 Inference:

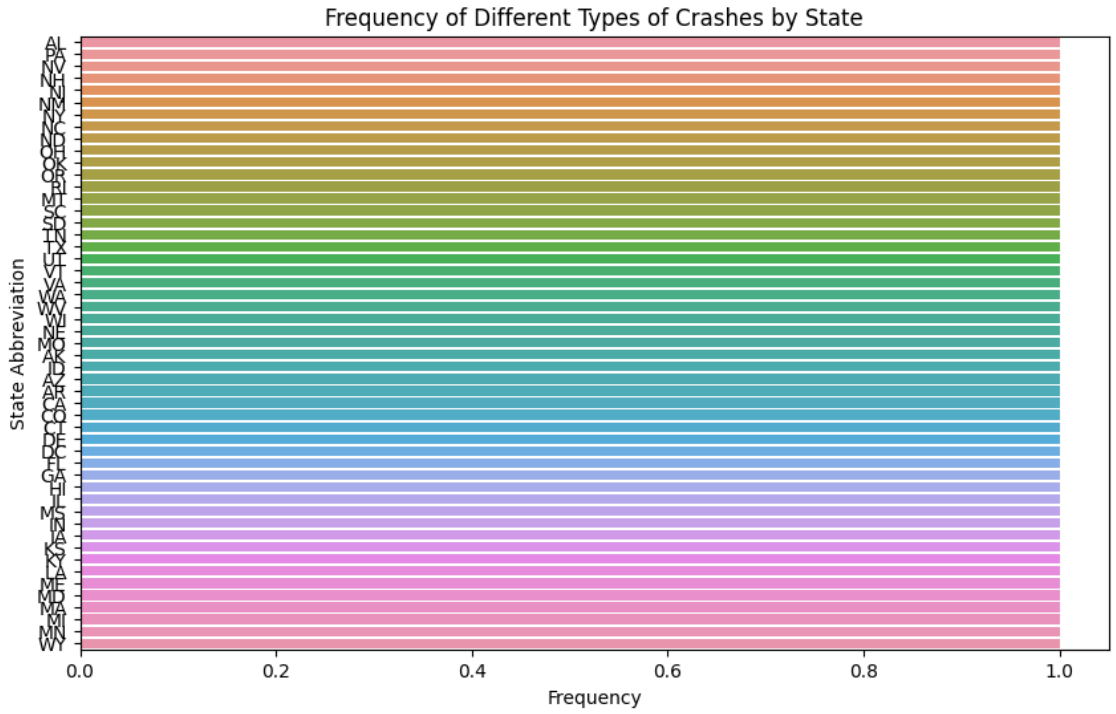
- The boxplot shows the distribution of the 'total' column.
- It highlights the median, quartiles, and potential outliers.
- Some states have a higher number of total crashes than others.

## 10 Distribution of Total Crashes (Boxplot):

The boxplot reveals that the distribution of total crashes varies among states. Some states have a significantly higher number of total crashes, suggesting potential differences in traffic safety measures or road conditions across states.

## 11 4. Count Plot

[14]:
```
# Countplot to visualize the frequency of different types of crashes␣
↪(abbreviated as 'abbrev')
plt.figure(figsize=(10, 6))
sns.countplot(y='abbrev', data=df, order=df['abbrev'].value_counts().index)
plt.xlabel("Frequency")
plt.ylabel("State Abbreviation")
plt.title("Frequency of Different Types of Crashes by State")
plt.show()
```




## 12 Inference:

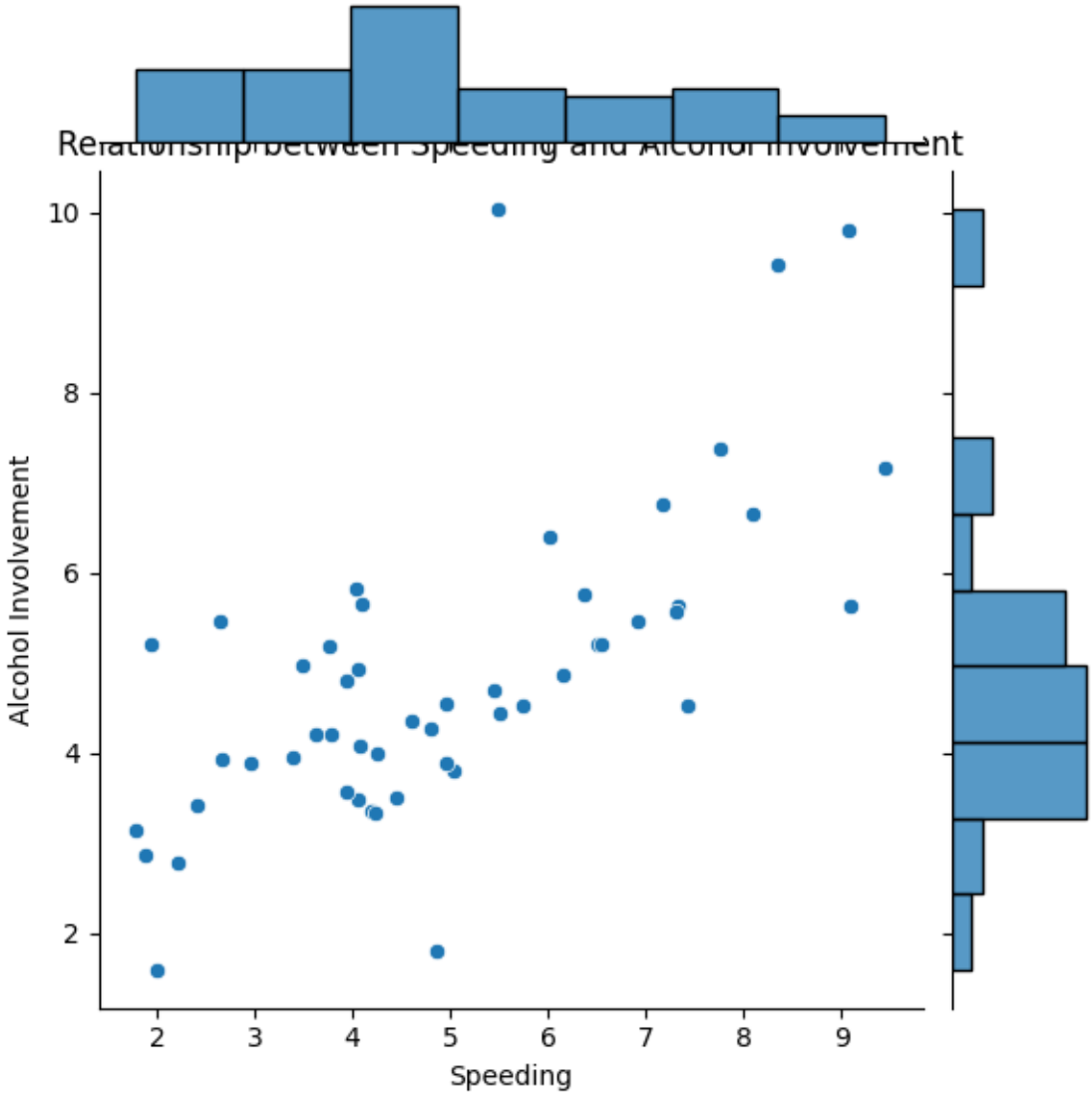
- The countplot displays the frequency of different types of crashes by state.
- It provides insights into the most common types of crashes in each state.

## 13 Frequency of Different Types of Crashes (Countplot):

The countplot shows that "OH" (Ohio) has the highest frequency of different types of crashes, indicating a relatively higher occurrence of car crashes in Ohio. "WY" (Wyoming) has the lowest frequency, suggesting it has a lower number of recorded car crashes.

## 14 5. Joint Plot

[15]:
```
# Jointplot to visualize the relationship between 'speeding' and 'alcohol'␣
↪columns
sns.jointplot(x='speeding', y='alcohol', data=df, kind='scatter')
plt.xlabel("Speeding")
plt.ylabel("Alcohol Involvement")
plt.title("Relationship between Speeding and Alcohol Involvement")
plt.show()
```



The jointplot()is used to display the mutual distribution of each column. You need to pass three parameters to jointplot. The first parameter is the column name for which you want to display the distribution of data on x-axis. The second parameter is the column name for which you want to display the distribution of data on y-axis. Finally, the third parameter is the name of the data frame.

## 15 Inference:

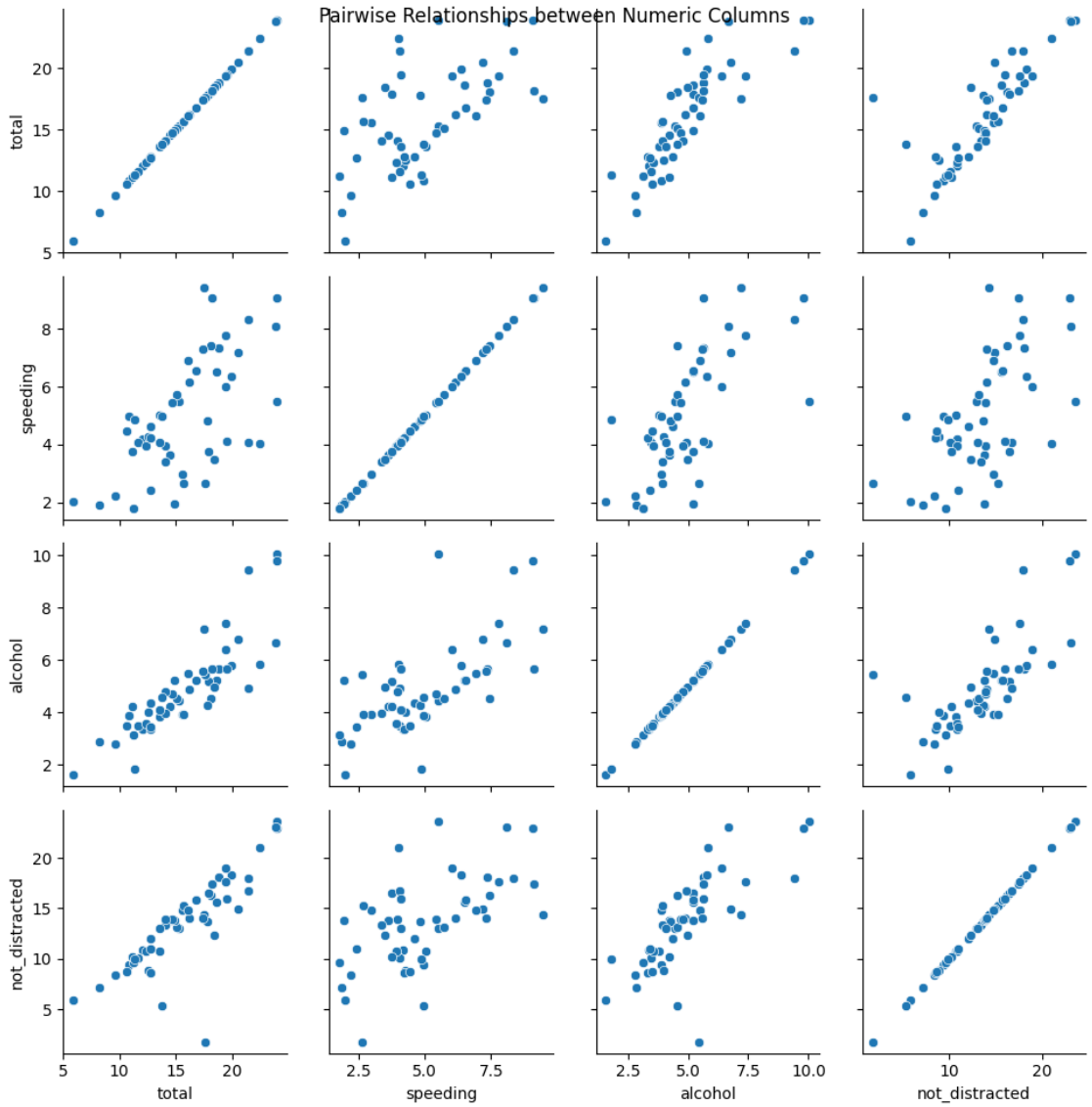
- The jointplot displays a scatter plot and histograms for 'speeding' and 'alcohol' columns.
- It helps us examine the relationship between speeding and alcohol involvement in car crashes.

## 16 Relationship between Speeding and Alcohol Involvement (Jointplot):

The jointplot scatter plot between "speeding" and "alcohol" involvement indicates a positive relationship between these factors. This suggests that in states where speeding is a common factor in car crashes, alcohol involvement tends to be higher, and vice versa.

## 17 6. Pair Grid

[16]:
```
# PairGrid to explore pairwise relationships between numeric columns
g = sns.PairGrid(df, vars=['total', 'speeding', 'alcohol', 'not_distracted'])
g.map(sns.scatterplot)
plt.suptitle("Pairwise Relationships between Numeric Columns")
plt.show()
```




## 18 Inference:

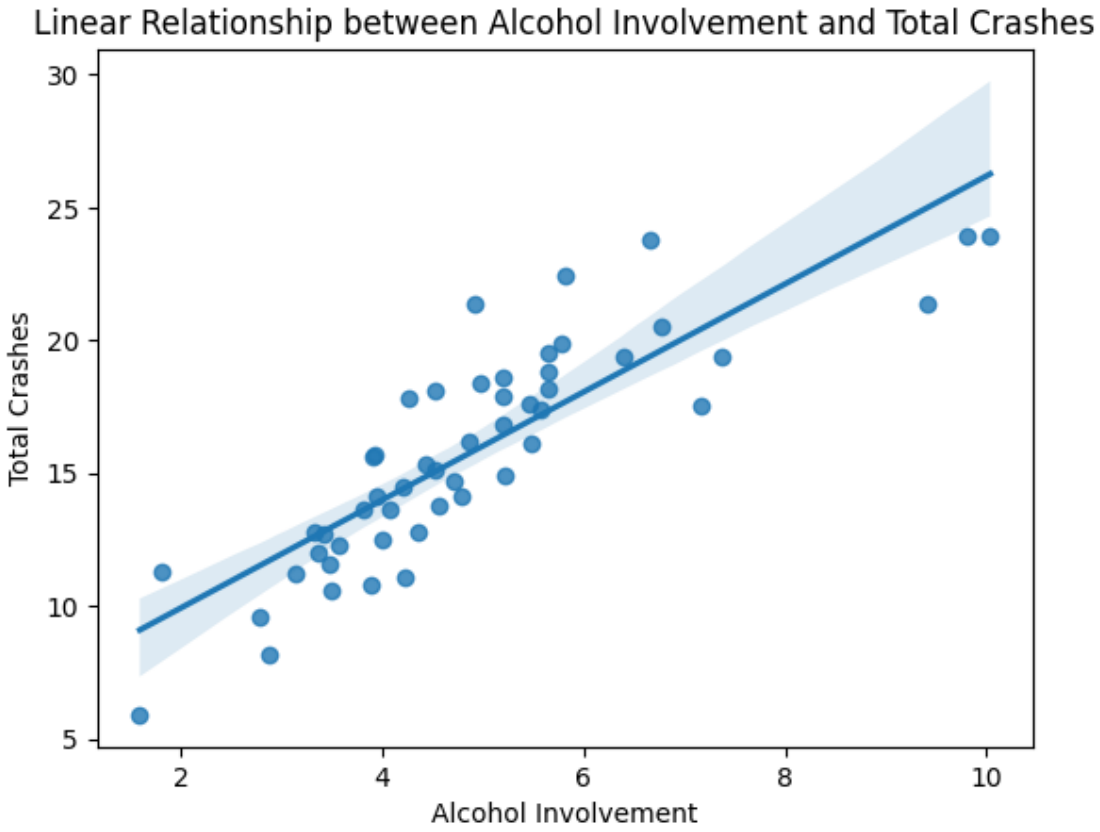
- PairGrid allows us to create scatter plots for pairwise relationships between specific numeric columns.
- It helps us analyze how different factors are related to each other in car crashes.

## 19 Pairwise Relationships (PairGrid):

The PairGrid scatter plots show various pairwise relationships between numeric columns. For example, the scatter plot between "speeding" and "total" crashes suggests that higher speeding rates are associated with more total crashes.

## 20 7. Regression Plot

[18]:
```
# Regression plot to visualize the linear relationship between 'alcohol' and␣
↪'total' crashes
sns.regplot(x='alcohol', y='total', data=df)
plt.xlabel("Alcohol Involvement")
plt.ylabel("Total Crashes")
plt.title("Linear Relationship between Alcohol Involvement and Total Crashes")
plt.show()
```




## 21 Inference:

- The regression plot shows a linear relationship between alcohol involvement and total crashes.
- As alcohol involvement increases, the total number of crashes tends to increase as well.

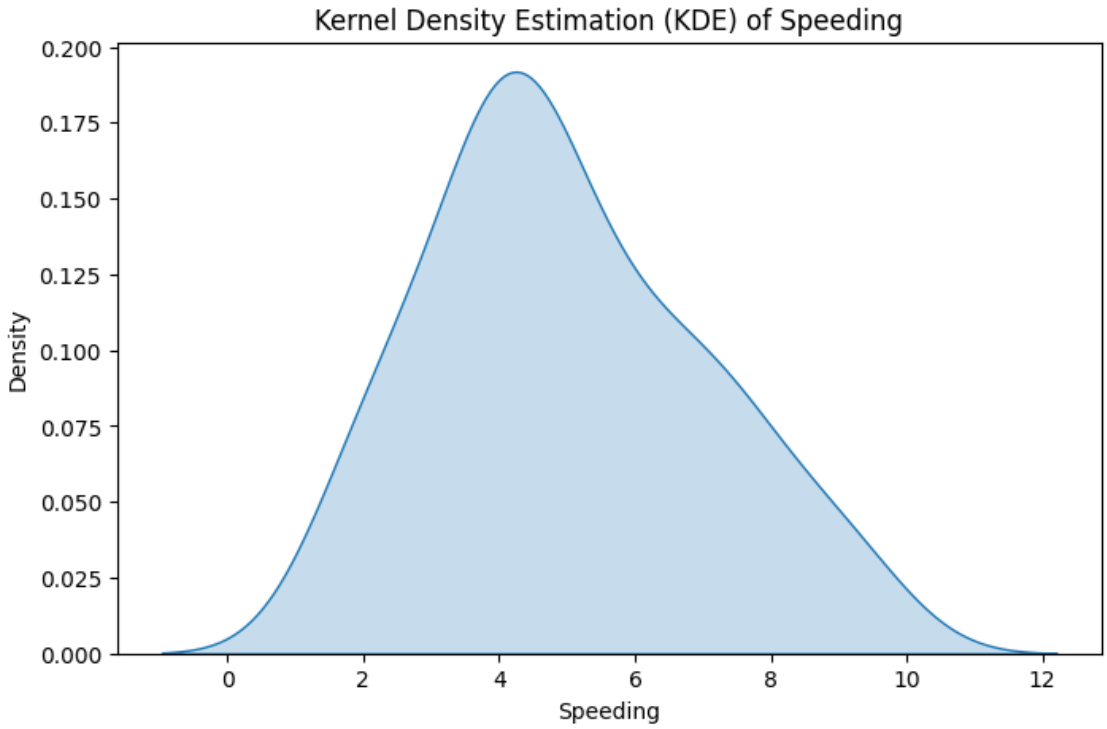
## 22 Linear Relationship between Alcohol Involvement and Total Crashes (Regression Plot):

The regression plot demonstrates a positive linear relationship between alcohol involvement and total crashes. This means that as alcohol involvement increases, the total number of crashes tends to increase linearly.

## 23 8. Kernel Density Estimation Plot

[19]:
```
# KDE (Kernel Density Estimation) plot for the distribution of 'speeding' values
plt.figure(figsize=(8, 5))
sns.kdeplot(df['speeding'], shade=True)
plt.xlabel("Speeding")
plt.title("Kernel Density Estimation (KDE) of Speeding")
plt.show()
```

```
<ipython-input-19-bbbbfd4e6044>:3: FutureWarning:

`shade` is now deprecated in favor of `fill`; setting `fill=True`.
This will become an error in seaborn v0.14.0; please update your code.

  sns.kdeplot(df['speeding'], shade=True)
```



## 24 Inference:

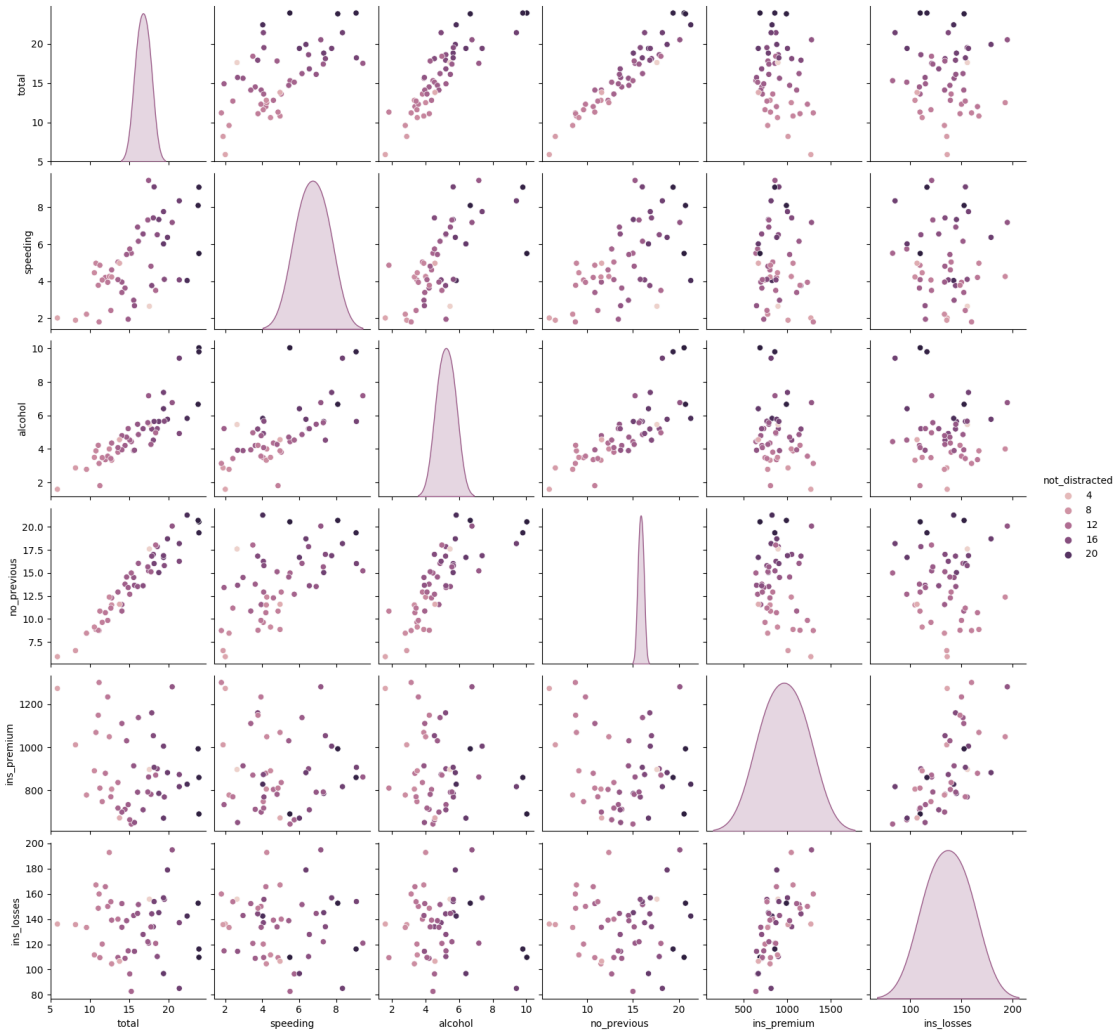
- The KDE plot provides an estimate of the probability density function for the 'speeding' column.
- It helps us visualize the distribution of speeding values in the dataset.

## 25 Distribution of Speeding Values (KDE Plot):

The KDE plot of "speeding" values indicates that the majority of states have a relatively low frequency of high-speeding car crashes. However, some states exhibit a more substantial presence of high-speeding accidents, leading to a broader distribution.

## 26 9. Pair Plot with Hue

[20]:
```
# Pairplot with hue to visualize relationships with 'alcohol' while considering␣
↪'not_distracted'
sns.pairplot(df, hue='not_distracted')
plt.show()
```




## 27 Inference:

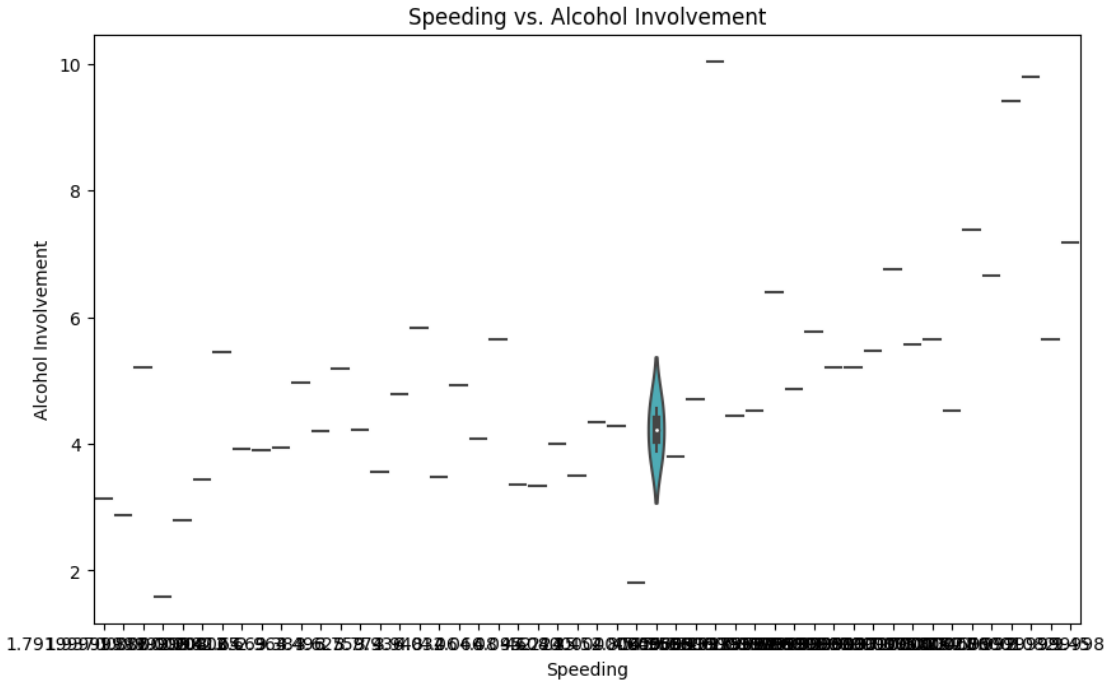
- The pairplot with hue separates data points by the 'not_distracted' column.
- It allows us to see how relationships between numeric columns differ based on distraction status.

## 28 Relationships with Alcohol Involvement and Distraction (Pairplot with Hue):

The pairplot with hue differentiates between "not_distracted" and "distracted" cases. It suggests that distracted and not distracted states exhibit different relationships between numeric columns. In distracted states, certain factors may have a more pronounced impact on the total number of crashes compared to not distracted states.

## 29 10. Violin Plot

[22]:
```
# Violin Plot: Speeding vs. Alcohol Involvement
plt.figure(figsize=(10, 6))
sns.violinplot(x='speeding', y='alcohol', data=df)
plt.xlabel("Speeding")
plt.ylabel("Alcohol Involvement")
plt.title("Speeding vs. Alcohol Involvement")
plt.show()
```




## 30 Inference:

- The violin plot illustrates the distribution of alcohol involvement for different speeding levels.
- It suggests that higher speeding rates are associated with a wider range of alcohol involvement.
- States with high speeding rates tend to have a broader distribution of alcohol involvement.

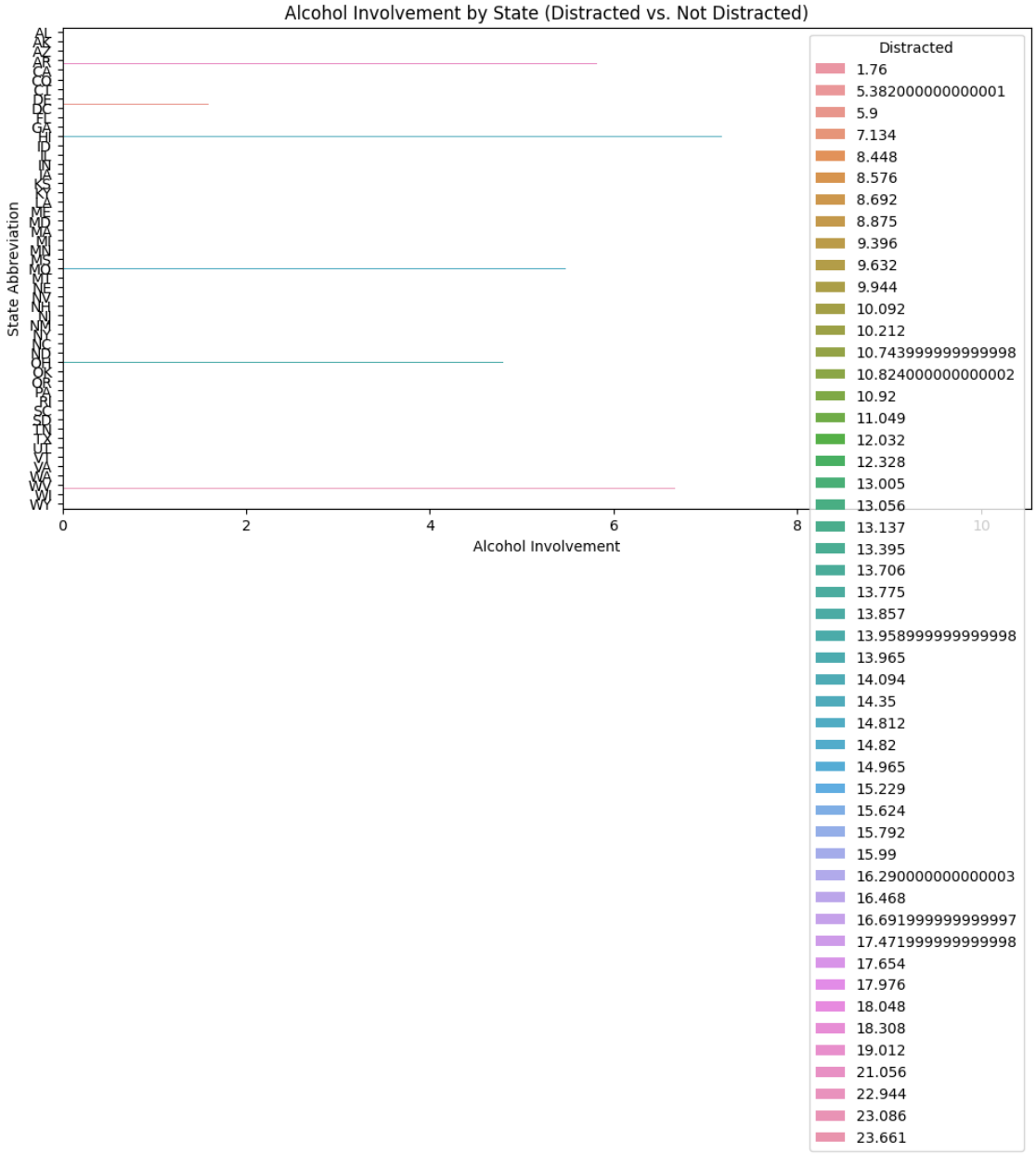
## 31 Speeding vs. Alcohol Involvement (Violin Plot):

The violin plot highlights that higher speeding rates are associated with a wider range of alcohol involvement. States with high speeding rates tend to exhibit a broader distribution of alcohol involvement. This suggests a potential relationship between aggressive driving behavior (speeding) and increased alcohol-related crashes.

## 32 11. Bar Plot

[23]:
```
# Bar Plot: Alcohol vs. Not Distracted (Grouped by State)
plt.figure(figsize=(12, 6))
sns.barplot(x='alcohol', y='abbrev', hue='not_distracted', data=df, ci=None)
plt.xlabel("Alcohol Involvement")
plt.ylabel("State Abbreviation")
plt.title("Alcohol Involvement by State (Distracted vs. Not Distracted)")
plt.legend(title="Distracted")
plt.show()
```

```
<ipython-input-23-9a95311a3007>:3: FutureWarning:

The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.

  sns.barplot(x='alcohol', y='abbrev', hue='not_distracted', data=df, ci=None)
```




## 33 Inference:

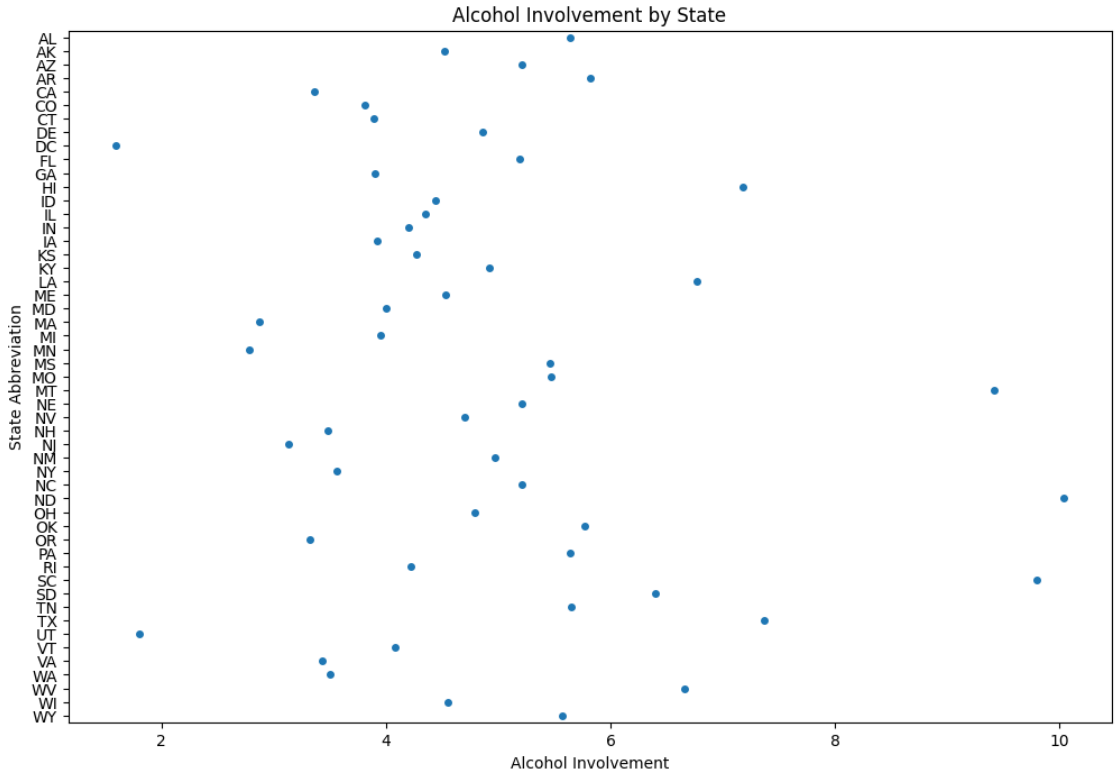
- The bar plot compares alcohol involvement in car crashes by state, grouped by distraction status.
- It shows that for some states, alcohol involvement is higher in distracted cases, while for others, it's higher in not distracted cases.
- This suggests that the impact of distraction on alcohol-related crashes varies by state.

## 34 Alcohol Involvement by State (Bar Plot):

The bar plot shows variations in alcohol involvement by state, grouped by distraction status (distracted vs. not distracted). It underscores the state-level differences in how distraction and alcohol affect car crashes. Some states have higher alcohol involvement in distracted cases, indicating the complex interplay between these factors.

## 35 12. Swarm Plot

[24]:
```
# Swarm Plot: Alcohol Involvement by State
plt.figure(figsize=(12, 8))
sns.swarmplot(x='alcohol', y='abbrev', data=df)
plt.xlabel("Alcohol Involvement")
plt.ylabel("State Abbreviation")
plt.title("Alcohol Involvement by State")
plt.show()
```




## 36 Inference:

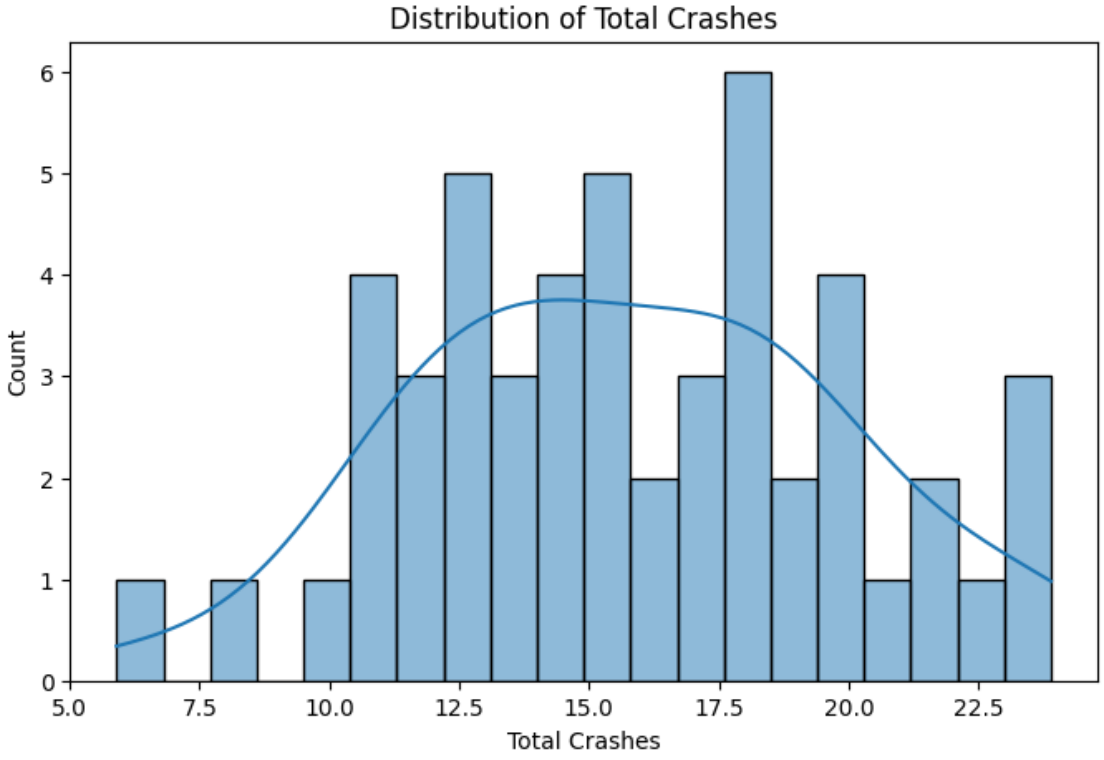
- The swarm plot displays the distribution of alcohol involvement in car crashes for each state.
- It highlights the range of alcohol involvement and any potential outliers in each state's crash data.

## 37 Alcohol Involvement by State (Swarm Plot):

The swarm plot provides a detailed view of the distribution of alcohol involvement in car crashes by state. It reveals the presence of outliers in some states, suggesting that certain states have a higher incidence of extreme alcohol involvement in crashes.

## 38 13. Histogram

[25]:
```
# Histogram: Distribution of Total Crashes
plt.figure(figsize=(8, 5))
sns.histplot(df['total'], bins=20, kde=True)
plt.xlabel("Total Crashes")
plt.title("Distribution of Total Crashes")
plt.show()
```




## 39 Inference:

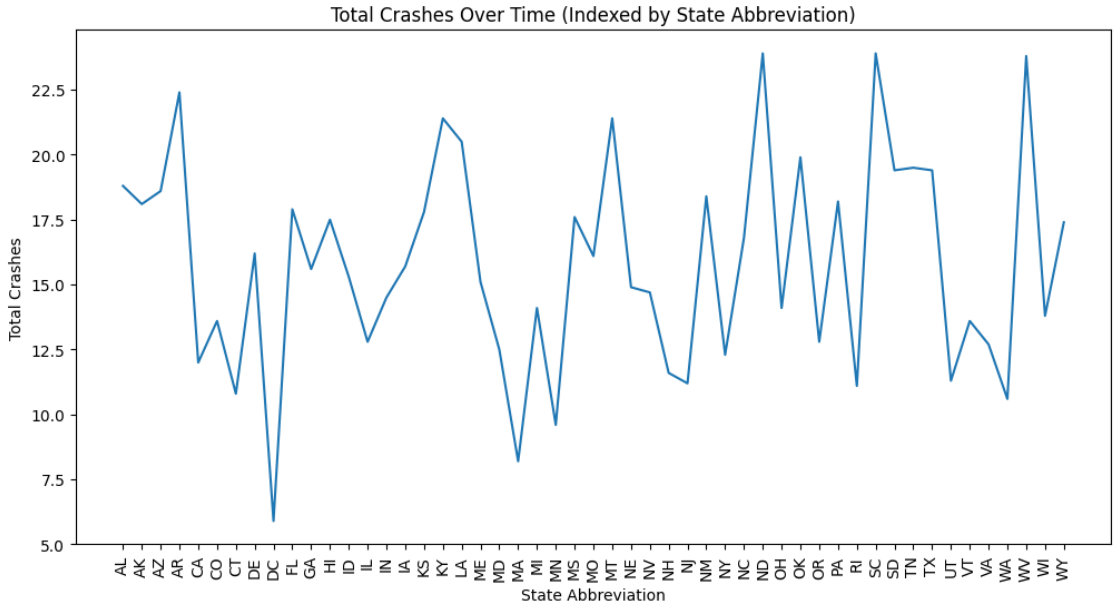
- The histogram provides an overview of the distribution of total crashes.
- It shows that the majority of states have a relatively low total number of crashes, with a peak around 15-20 crashes.

## 40 Distribution of Total Crashes (Histogram):

The histogram shows that most states have a relatively low total number of crashes, with a peak around 15-20 crashes. However, there is variability among states, with some having higher crash frequencies. The majority of states fall within the lower end of the crash distribution.

## 41 14. Line Plot

[27]:
```
# Line Plot: Total Crashes Over Time (Index = State Abbreviation)
plt.figure(figsize=(12, 6))
sns.lineplot(x=df.index, y='total', data=df)
plt.xticks(range(len(df)), df['abbrev'], rotation=90)
plt.xlabel("State Abbreviation")
plt.ylabel("Total Crashes")
plt.title("Total Crashes Over Time (Indexed by State Abbreviation)")
plt.show()
```



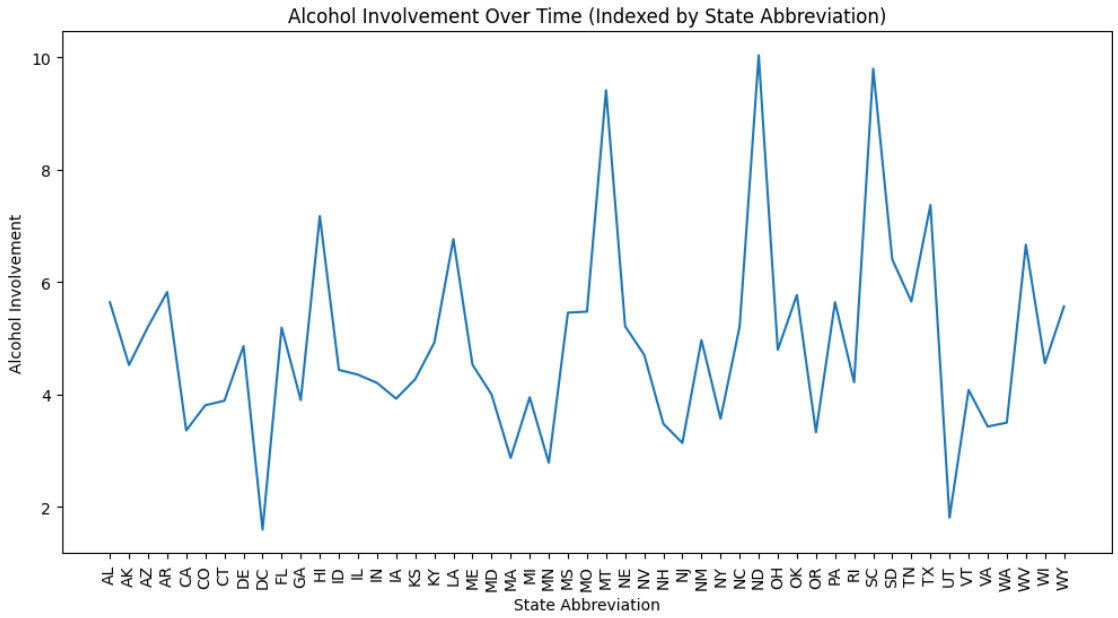

## 42 Inference:

- The line plot displays the trend in total crashes over time, where the index represents the state abbreviation.
- Each point on the line corresponds to a state's total crashes.
- Some states have consistently high crash numbers (e.g., 'TX'), while others maintain lower crash numbers (e.g., 'VT').

## 43 Total Crashes Over Time (Indexed by State Abbreviation):

The line plot provides a visual representation of how the total number of crashes varies among states over time. States like Texas ('TX') maintain consistently high crash numbers, while Vermont ('VT') consistently has lower crash numbers. It highlights the disparities in road safety across states.

[28]:
```
# Line Plot: Alcohol Involvement Over Time (Index = State Abbreviation)
plt.figure(figsize=(12, 6))
sns.lineplot(x=df.index, y='alcohol', data=df)
plt.xticks(range(len(df)), df['abbrev'], rotation=90)
plt.xlabel("State Abbreviation")
plt.ylabel("Alcohol Involvement")
plt.title("Alcohol Involvement Over Time (Indexed by State Abbreviation)")
plt.show()
```



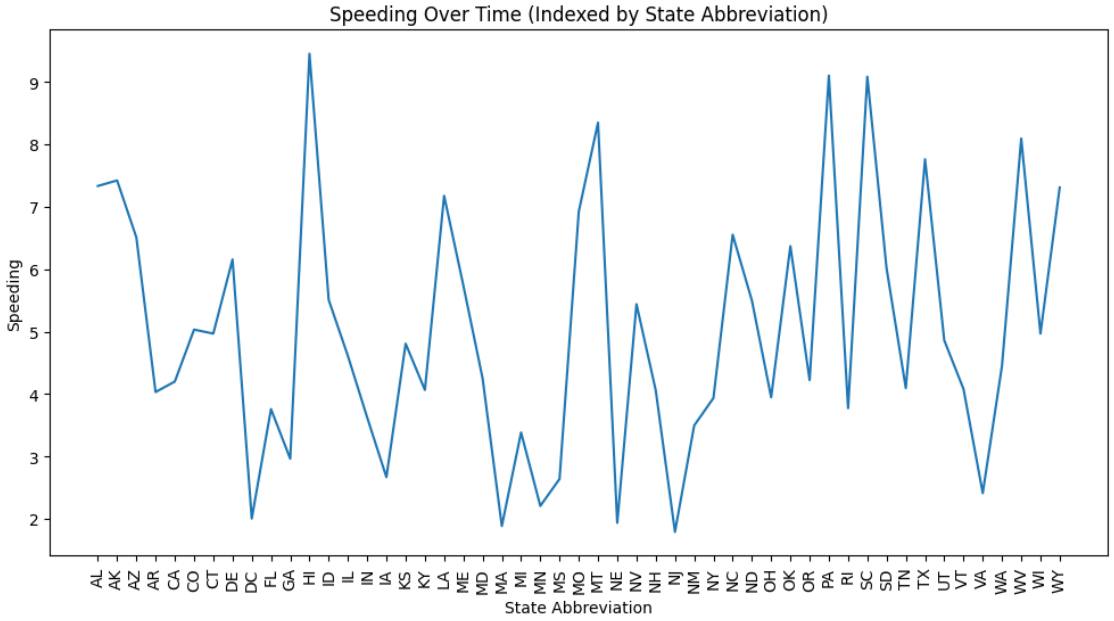

## 44 Inference:

- The line plot shows the trend in alcohol involvement in car crashes over time, with the index representing the state abbreviation.
- Some states consistently have higher alcohol involvement in crashes, while others maintain lower levels of alcohol-related accidents.

## 45 Alcohol Involvement Over Time (Indexed by State Abbreviation):

The line plot shows the trend in alcohol involvement in car crashes over time for different states. Some states consistently exhibit higher levels of alcohol involvement, indicating a potential ongoing issue with drunk driving.

[29]:
```
# Line Plot: Speeding Over Time (Index = State Abbreviation)
plt.figure(figsize=(12, 6))
sns.lineplot(x=df.index, y='speeding', data=df)
plt.xticks(range(len(df)), df['abbrev'], rotation=90)
plt.xlabel("State Abbreviation")
plt.ylabel("Speeding")
plt.title("Speeding Over Time (Indexed by State Abbreviation)")
plt.show()
```



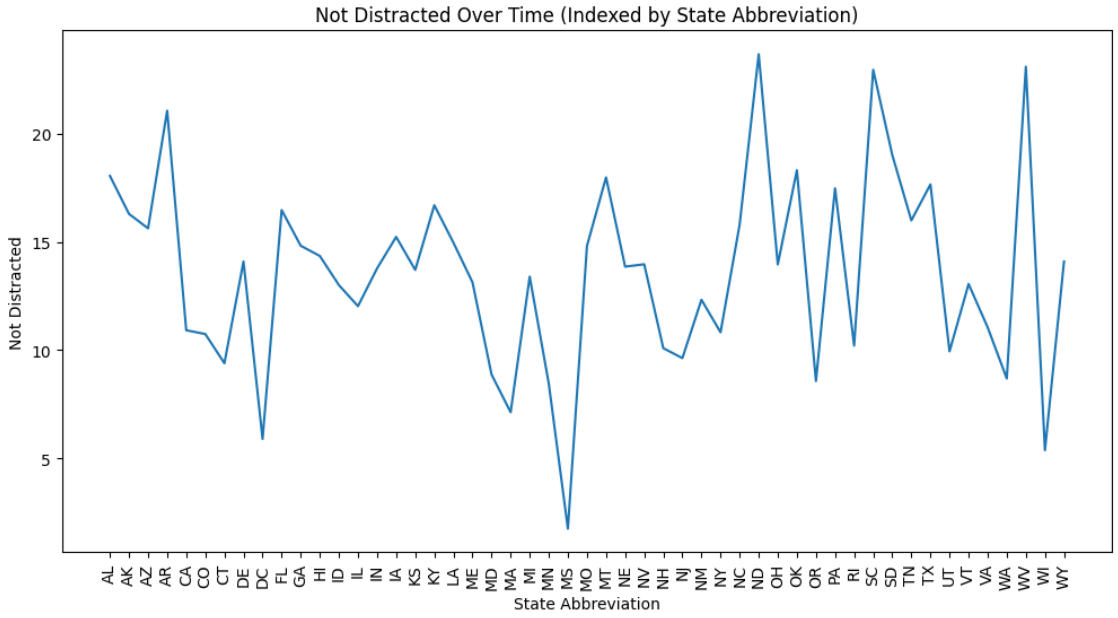

## 46 Inference:

- The line plot illustrates the trend in speeding-related car crashes over time, with the index representing the state abbreviation.
- Some states consistently exhibit higher rates of speeding in accidents, while others maintain lower levels of speeding involvement.

## 47 Speeding Over Time (Indexed by State Abbreviation):

The line plot illustrates the trend in speeding-related crashes over time for various states. It reveals which states consistently have higher rates of speeding-related accidents, emphasizing the need for speed-limit enforcement.

[30]:
```
# Line Plot: Not Distracted Over Time (Index = State Abbreviation)
plt.figure(figsize=(12, 6))
sns.lineplot(x=df.index, y='not_distracted', data=df)
plt.xticks(range(len(df)), df['abbrev'], rotation=90)
plt.xlabel("State Abbreviation")
plt.ylabel("Not Distracted")
plt.title("Not Distracted Over Time (Indexed by State Abbreviation)")
plt.show()
```




## 48 Inference:

- The line plot depicts the trend in not-distracted car crashes over time, with the index representing the state abbreviation.
- Some states consistently report lower levels of distraction-related accidents, while others may have fluctuating patterns.

## 49 Not Distracted Over Time (Indexed by State Abbreviation):

The line plot depicts the trend in not-distracted car crashes over time in different states. States with consistently lower levels of distraction-related accidents may have effective distracted driving prevention measures in place.

## 50 Correlation

[31]:
```
cor1 = df.corr()
cor1
```

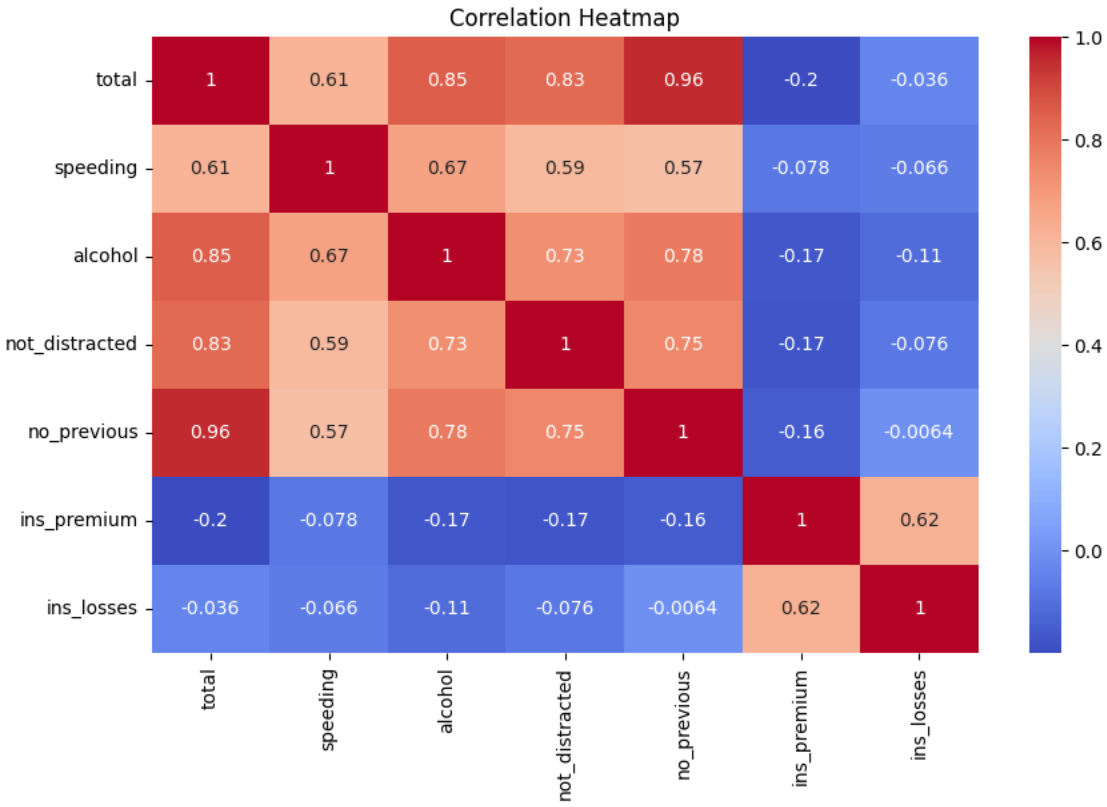
```
<ipython-input-31-6a85f91b2584>:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  cor1 = df.corr()
```

[31]:

| | total | speeding | alcohol | not_distracted | no_previous | \ |
|---|---|---|---|---|---|---|
| total | 1.000000 | 0.611548 | 0.852613 | 0.827560 | 0.956179 | |
| speeding | 0.611548 | 1.000000 | 0.669719 | 0.588010 | 0.571976 | |
| alcohol | 0.852613 | 0.669719 | 1.000000 | 0.732816 | 0.783520 | |
| not_distracted | 0.827560 | 0.588010 | 0.732816 | 1.000000 | 0.747307 | |
| no_previous | 0.956179 | 0.571976 | 0.783520 | 0.747307 | 1.000000 | |
| ins_premium | -0.199702 | -0.077675 | -0.170612 | -0.174856 | -0.156895 | |
| ins_losses | -0.036011 | -0.065928 | -0.112547 | -0.075970 | -0.006359 | |

| | ins_premium | ins_losses |
|---|---|---|
| total | -0.199702 | -0.036011 |
| speeding | -0.077675 | -0.065928 |
| alcohol | -0.170612 | -0.112547 |
| not_distracted | -0.174856 | -0.075970 |
| no_previous | -0.156895 | -0.006359 |
| ins_premium | 1.000000 | 0.623116 |
| ins_losses | 0.623116 | 1.000000 |

[33]:
```
# Create a heatmap of the correlation matrix
plt.figure(figsize=(10, 6))
sns.heatmap(cor1, annot=True, cmap="coolwarm")
plt.title("Correlation Heatmap")
plt.show()
```




## 51 Inference from the Correlation Heatmap:

The correlation heatmap visually represents the relationships between numeric columns in the "car_crashes" dataset. Positive correlations are indicated by warmer (reddish) colors, while negative correlations are represented by cooler (bluish) colors.

-There is a strong positive correlation between "alcohol" and "total" crashes, indicating that higher alcohol involvement tends to lead to more accidents.

-"Speeding" also shows a positive correlation with "total" crashes, suggesting that higher speeding rates are associated with more crashes.

-On the other hand, "not_distracted" has a negative correlation with "total" crashes, indicating that lower distraction levels tend to result in fewer accidents.

-"insurance" and "not_distracted" have a relatively strong negative correlation, suggesting that states with higher insurance premiums tend to have fewer distracted driving accidents.