NAME: SHAIK MOHID BABU

REG. NO: 21BCE9569

Branch: Computer Science and Engineering with Specialization in Artificial Intelligence and Machine Learning

Campus: VIT-AP

Email: mohidbabu.21bce9569@vitapstudent.ac.in

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:

```
data=sns.load_dataset('car_crashes')
```

In [3]:

```
data.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
```

In [4]:

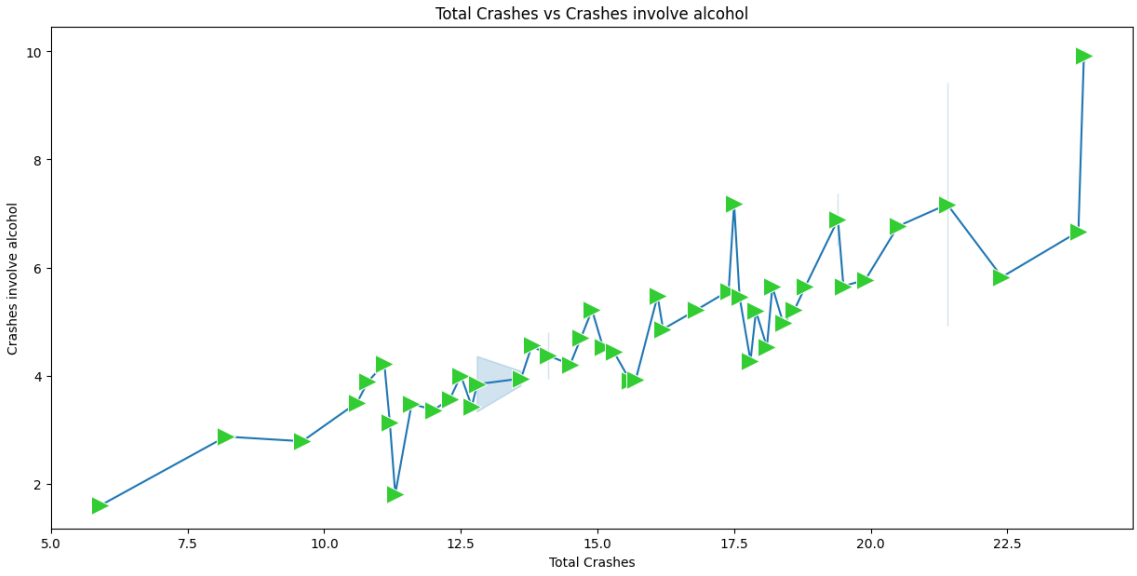
```
data.head()
```

Out[4]:

	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

In [5]:

```
plt.figure(figsize=(15,7))
sns.lineplot(x="total",y="alcohol",data=data,marker='>',markerfacecolor='limegreen',mark
plt.xlabel("Total Crashes")
plt.ylabel("Crashes involve alcohol")
plt.title("Total Crashes vs Crashes involve alcohol")
plt.show()
```



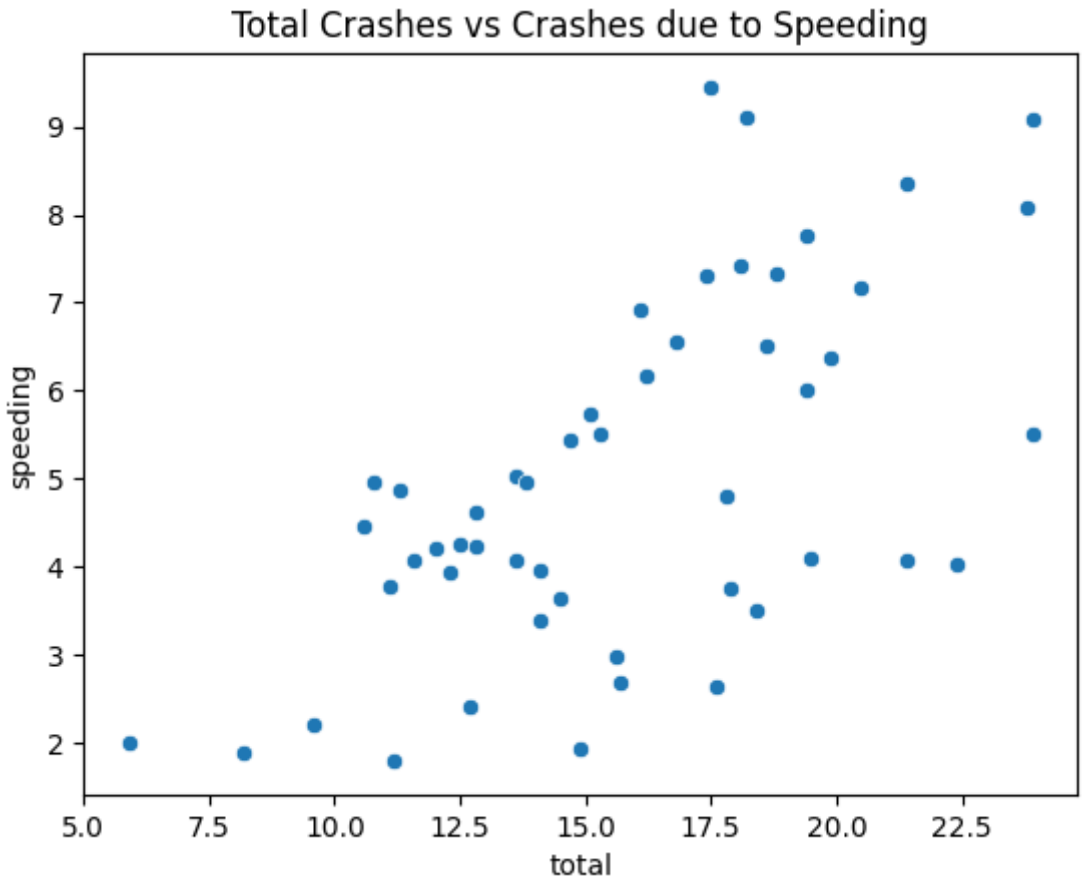

```
Inference:
A line plot shows that as alcohol consumption rises, total crashes increase.
```

In [6]:

```
sns.scatterplot(x='total',y='speeding',data=data)
plt.title("Total Crashes vs Crashes due to Speeding")
```

Out[6]:

```
Text(0.5, 1.0, 'Total Crashes vs Crashes due to Speeding')
```



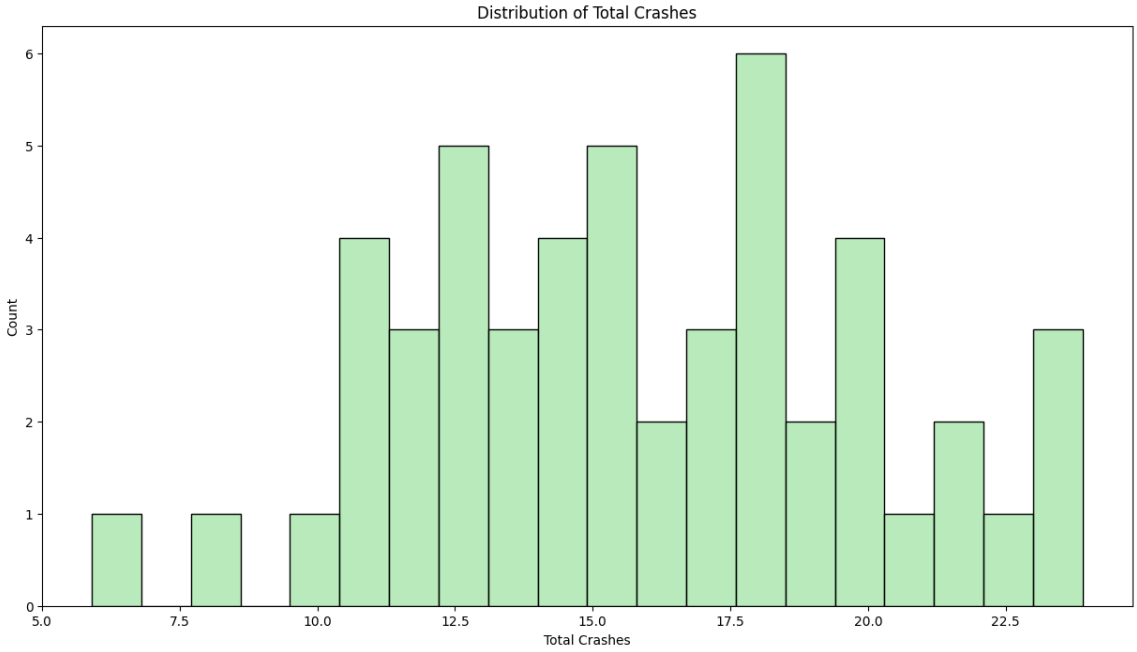
Inference: Scatter plot compares the total crashes with the increases due to speeding

In [7]:

```
plt.figure(figsize=(15,8))
sns.histplot(x="total",data=data,color='#A0E3A5',bins=20)
plt.title("Distribution of Total Crashes")
plt.xlabel("Total Crashes")
```

Out[7]:

```
Text(0.5, 0, 'Total Crashes')
```

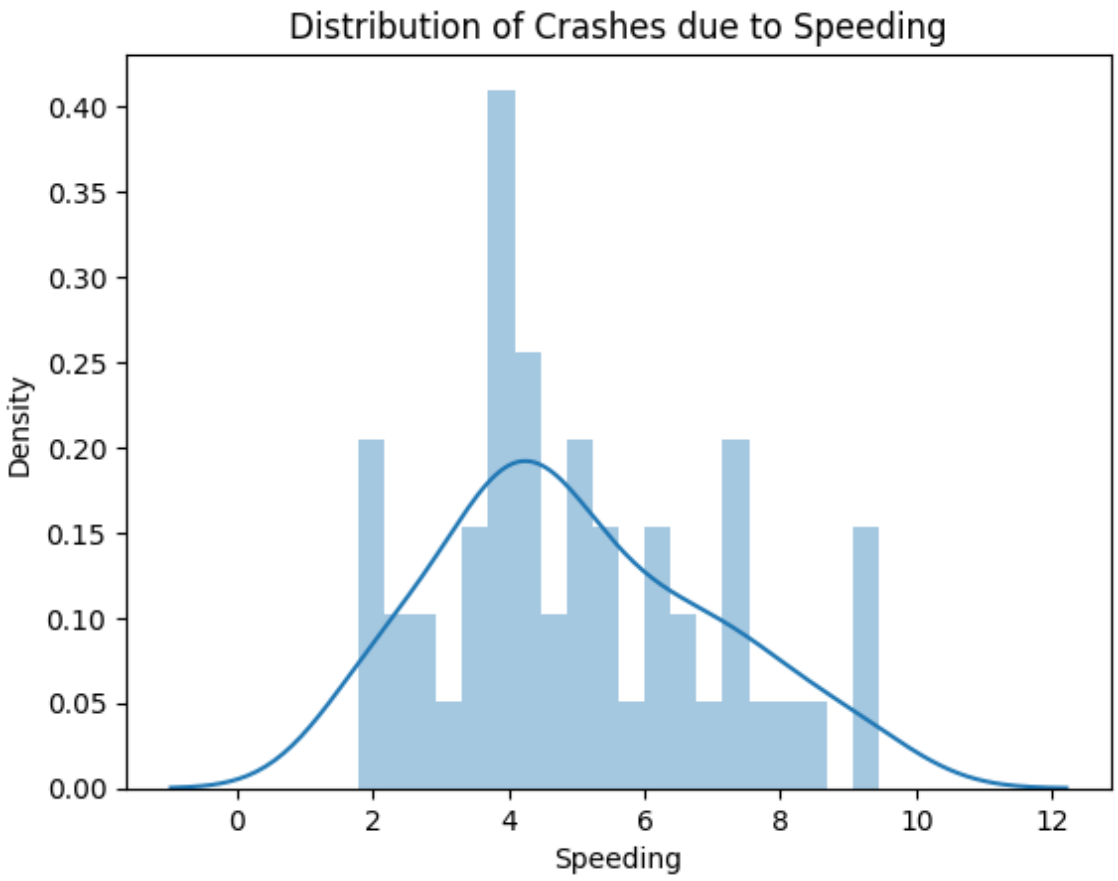


Inference:The plot reveals the distribution of total crashes, aiding in identifying common crash rate ranges.

In [8]:

```
sns.distplot(data['speeding'],bins=20)
plt.xlabel("Speeding")
plt.ylabel('Density')
plt.title("Distribution of Crashes due to Speeding")
plt.show()
```

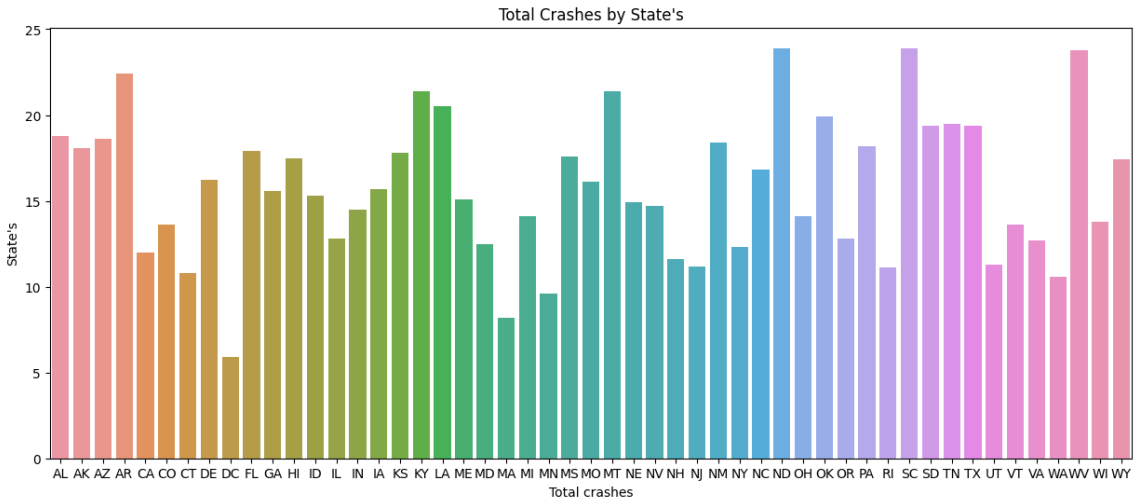
```
<ipython-input-8-12e29da558d7>:1: UserWarning:

`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 wi
th
similar flexibility) or `histplot` (an axes-level function for histogram
s).

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

  sns.distplot(data['speeding'],bins=20)
```



Inference: The distplot displayed above illustrates the density levels of speeding-related crashes within the dataset.
The highest density, which is 0.38, corresponds to a situation where there were 4 speeding-related crashes.

In [9]:

```
plt.figure(figsize=(15,6))
sns.barplot(x="abbrev",y="total",data=data,)
plt.xlabel("Total crashes")
plt.ylabel("State's")
plt.title("Total Crashes by State's")
plt.show()
```



Inference: The barplot shows the total number of accidents by state.

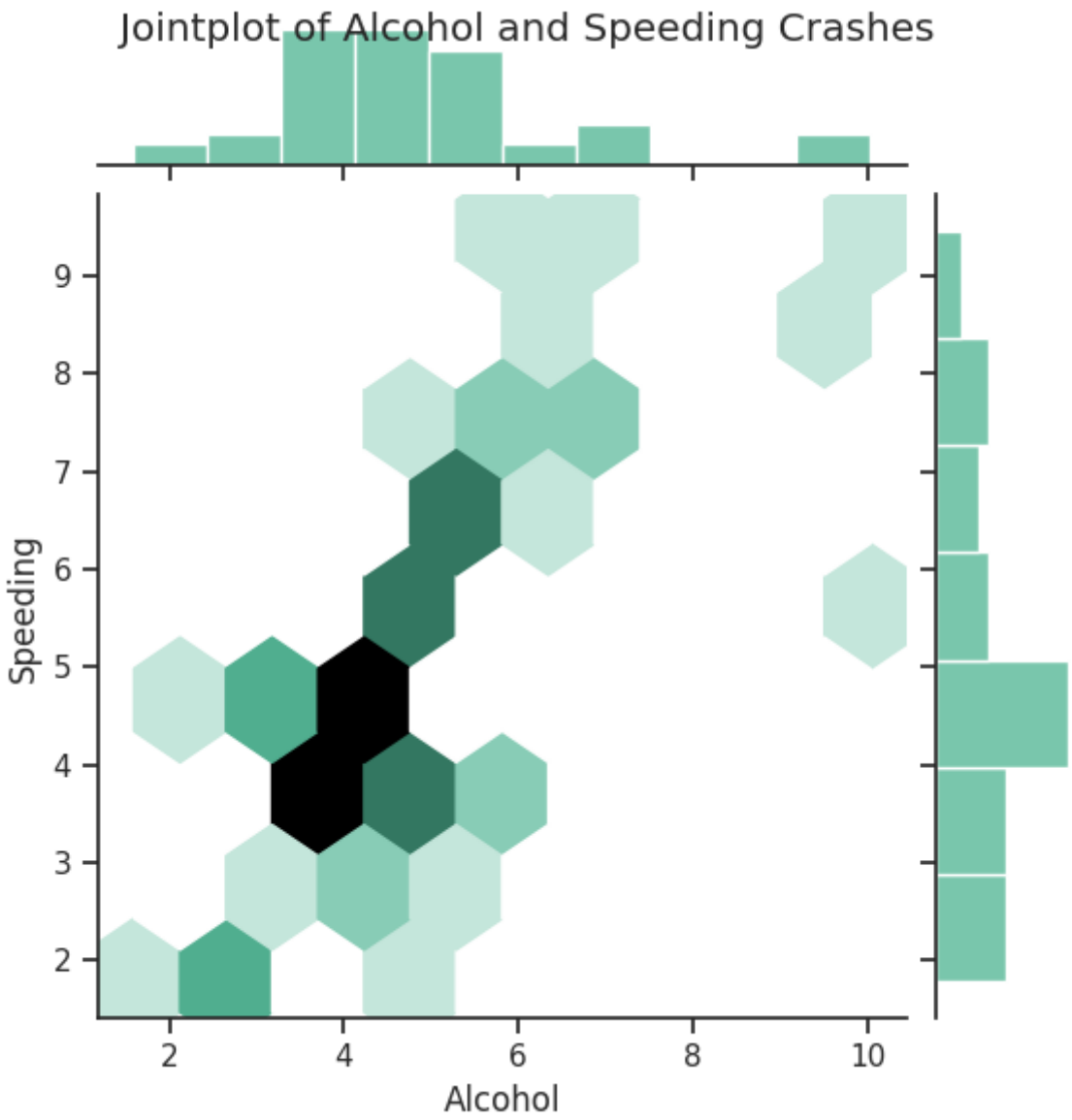
In [10]:

```
plt.figure(figsize=(15,8))
sns.set_theme(style="ticks")
sns.jointplot(x="alcohol",y="speeding",data=data,kind="hex", color="#4CB391")
plt.xlabel("Alcohol")
plt.ylabel("Speeding")
plt.suptitle("Jointplot of Alcohol and Speeding Crashes")
```

Out[10]:

```
Text(0.5, 0.98, 'Jointplot of Alcohol and Speeding Crashes')
```

```
<Figure size 1500x800 with 0 Axes>
```

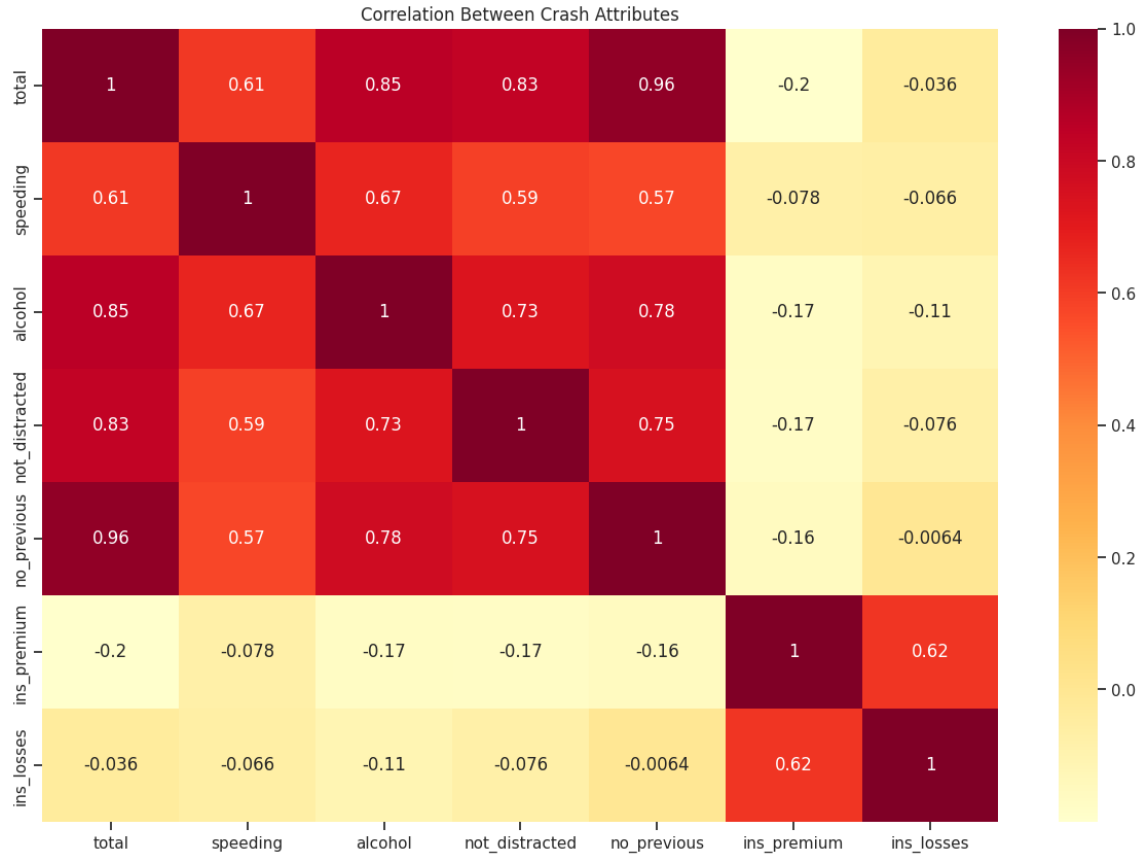



```
Inference: The jointplot provides a visual representation of the relationship between
alcohol-related accidents and speeding-related accidents.
It suggests that there is a tendency for speeding-related accidents to increase as the
number of alcohol-related accidents increases.
```

In [11]:

```
cor = data.corr()
plt.figure(figsize=(15, 10))
sns.heatmap(cor, annot=True, cmap="YlOrRd")
plt.title("Correlation Between Crash Attributes")
plt.show()
```

```
<ipython-input-11-ca45b563cfdd>:1: FutureWarning: The default value of num
eric_only in DataFrame.corr is deprecated. In a future version, it will de
fault to False. Select only valid columns or specify the value of numeric_
only to silence this warning.
  cor = data.corr()
```



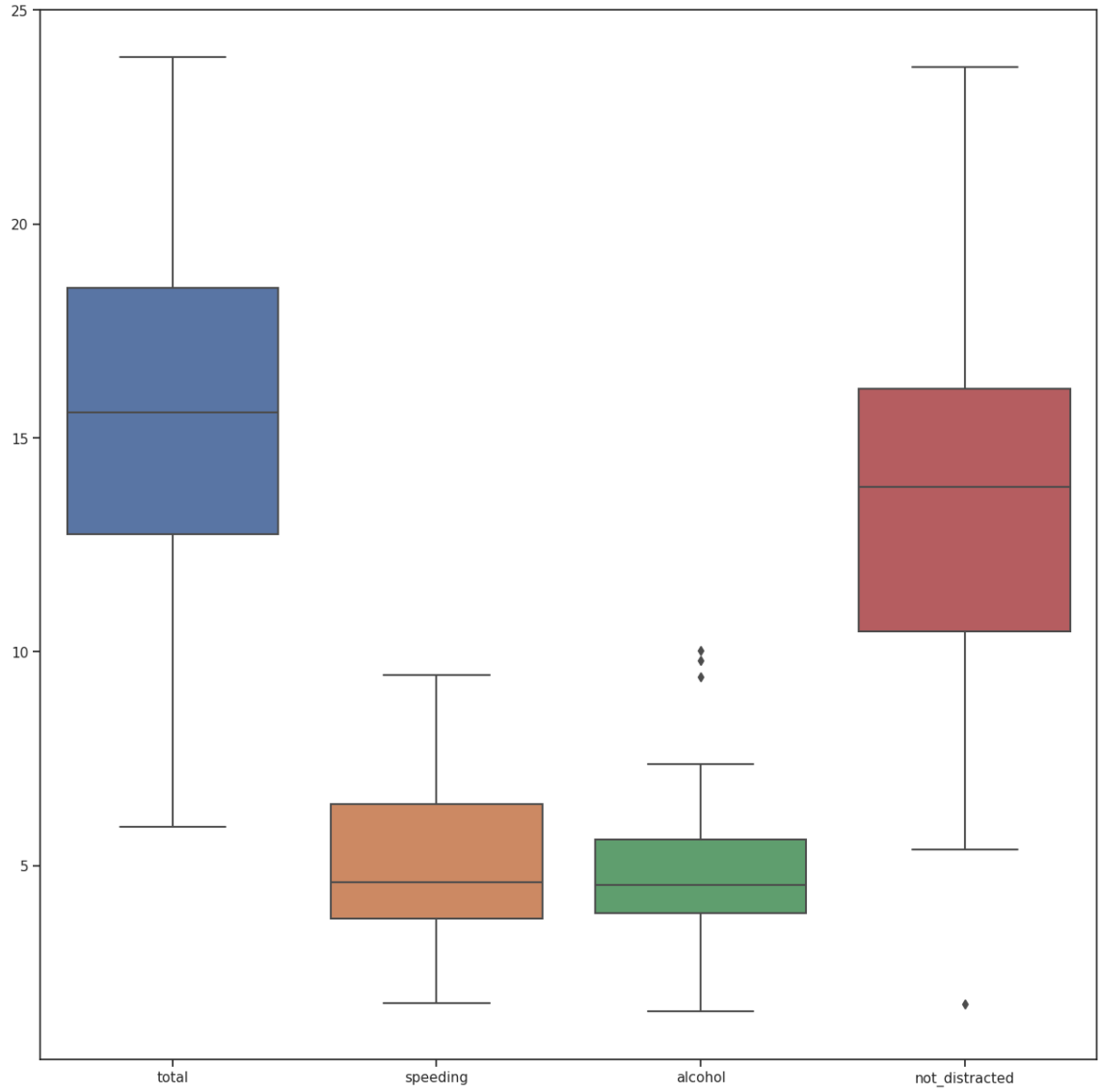
Inference: The heatmap, which displays the correlation between attributes related to crashes, reveals a positive correlation among certain attributes.

In [12]:

```
plt.figure(figsize=(15,15))
sns.boxplot(data.iloc[:,0:4])
```

Out[12]:

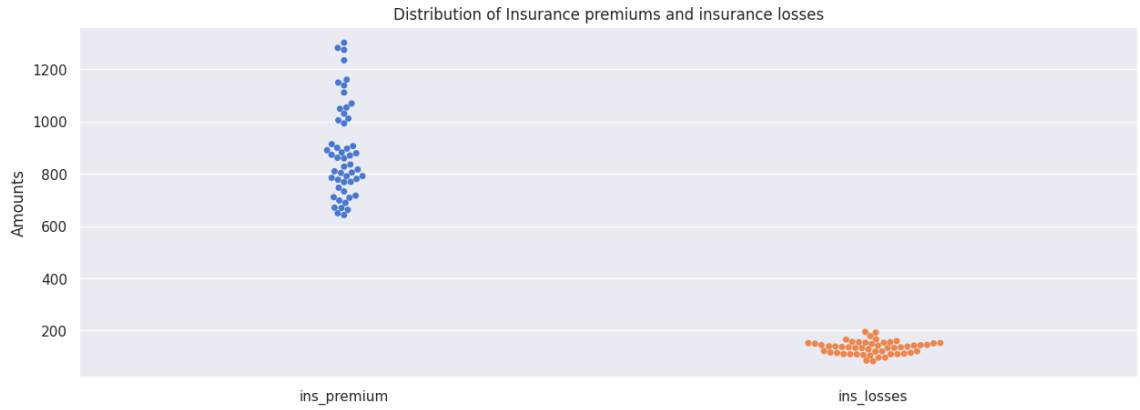
<Axes: >



Inference: The boxplot of the dataset indicates the presence of outliers for the number of crashes related to alcohol and crashes that are not distracted.

In [13]:

```
plt.figure(figsize=(15, 5))
sns.set_theme(style="darkgrid", palette="muted")
sns.swarmplot(data.iloc[:,5:7],)
plt.title("Distribution of Insurance premiums and insurance losses")
plt.ylabel("Amounts")
plt.show()
```



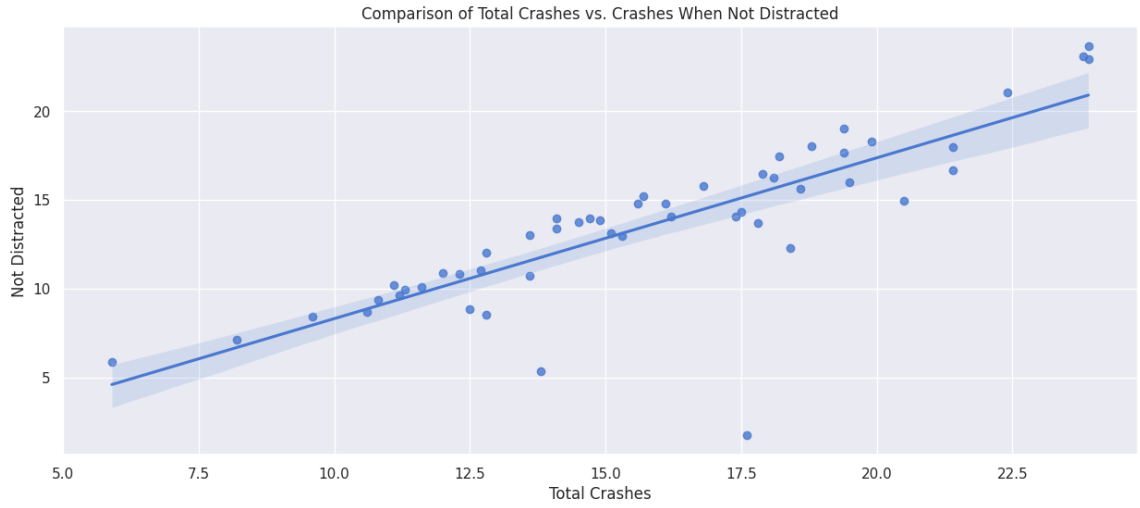
Inference: Swarm plot reveals ins_premium and ins_losses distribution: losses range 0-200, premiums 600-1200, emphasizing premiums greatly exceed losses in the dataset.

In [14]:

```
plt.figure(figsize=(15,6))
sns.regplot(x="total",y="not_distracted",data=data)
plt.xlabel("Total Crashes")
plt.ylabel("Not Distracted")
plt.title("Comparison of Total Crashes vs. Crashes When Not Distracted")
```

Out[14]:

```
Text(0.5, 1.0, 'Comparison of Total Crashes vs. Crashes When Not Distracte
d')
```



Inference: The inference is that there is a positive correlation between the total number of crashes and the number of crashes that occur when the driver is not distracted.

```
This means that as the number of crashes when not distracted increases, the total
number of crashes also increases.
```