

ASSIGNMENT-2

NAME:S.manikanta

REG NO:21BCE7217

In [1]:

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

In [2]:

```
dataset=pd.read_csv("car_crashes.csv")
dataset
```

Out[2]:

	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
5	13.6	5.032	3.808	10.744	12.920	835.50	139.91	CO
6	10.8	4.968	3.888	9.396	8.856	1068.73	167.02	CT
7	16.2	6.156	4.860	14.094	16.038	1137.87	151.48	DE
8	5.9	2.006	1.593	5.900	5.900	1273.89	136.05	DC
9	17.9	3.759	5.191	16.468	16.826	1160.13	144.18	FL

In [3]:

```
dataset.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]:

```
dataset.head(8)
```

Out[4]:

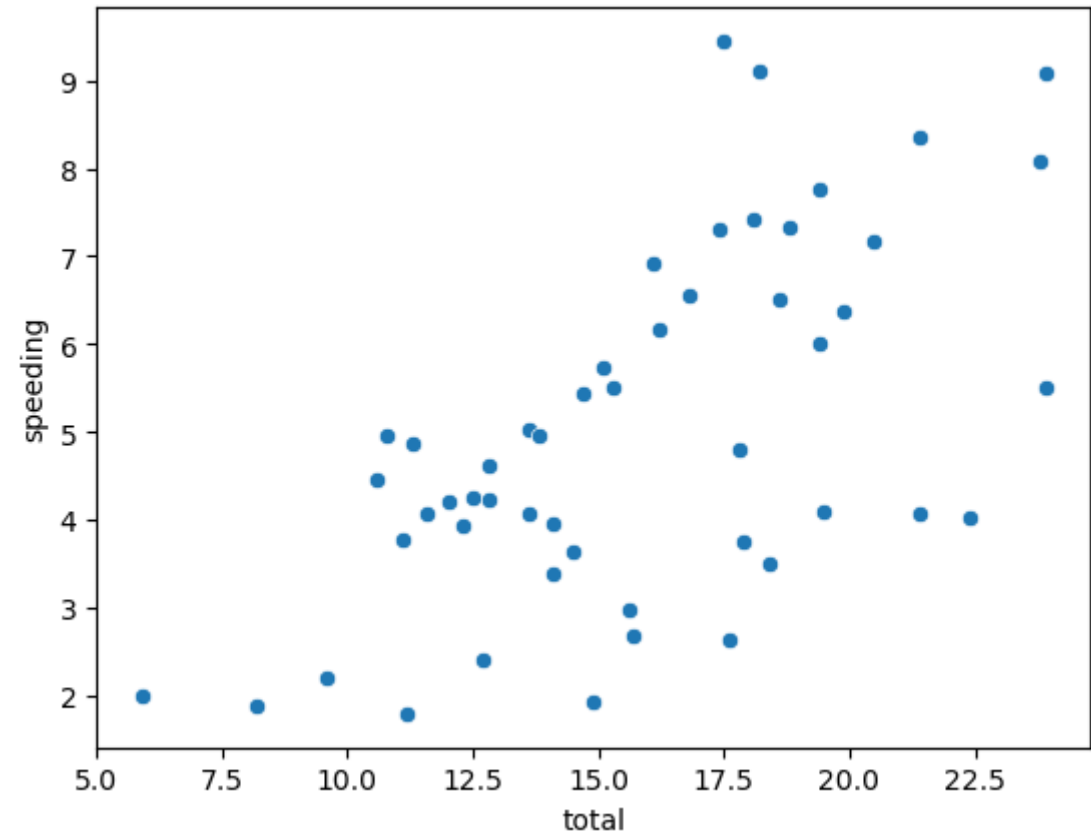
	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses	abbrev
0	18.8	7.332	5.640	18.048	15.040	784.55	145.08	AL
1	18.1	7.421	4.525	16.290	17.014	1053.48	133.93	AK
2	18.6	6.510	5.208	15.624	17.856	899.47	110.35	AZ
3	22.4	4.032	5.824	21.056	21.280	827.34	142.39	AR
4	12.0	4.200	3.360	10.920	10.680	878.41	165.63	CA
5	13.6	5.032	3.808	10.744	12.920	835.50	139.91	CO
6	10.8	4.968	3.888	9.396	8.856	1068.73	167.02	CT
7	16.2	6.156	4.860	14.094	16.038	1137.87	151.48	DE

In [5]:

```
sns.scatterplot(x="total",y="speeding",data=dataset)
```

Out[5]:

<Axes: xlabel='total', ylabel='speeding'>

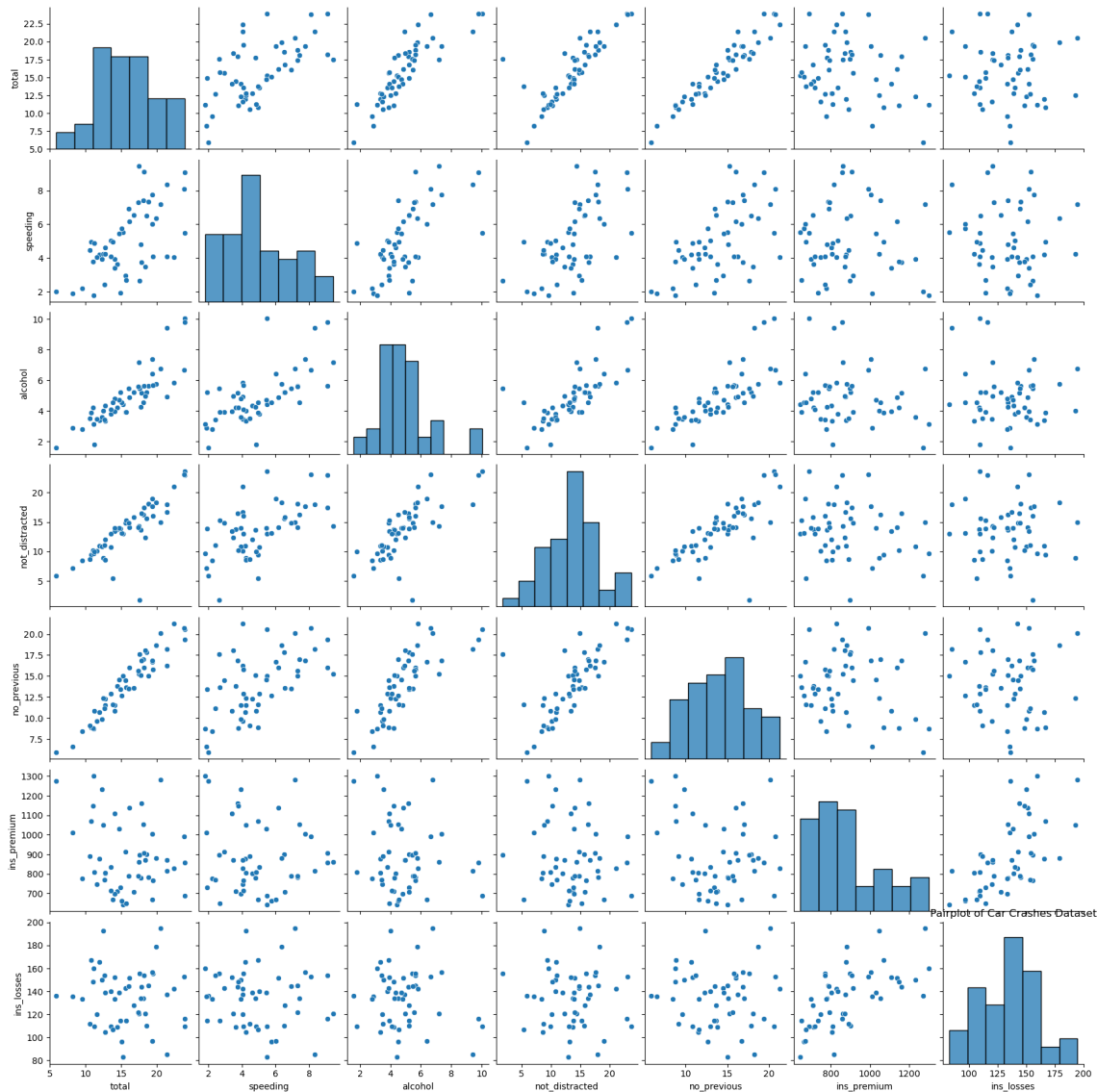


In [6]:

```
# Inference: from the plot we can say that as the total increases speeding is decreases
```

In [7]:

```
sns.pairplot(dataset)  
plt.title("Pairplot of Car Crashes Dataset")  
plt.show()
```



In [8]:

```
# Inference: The pairplot provides a quick overview of the relationships between numeric
```

In [9]:

```
sns.distplot(dataset["total"], bins=20, kde=True)
plt.title("Histogram of Total Number of Accidents")
plt.xlabel("Total Accidents")
plt.ylabel("Frequency")
plt.show()
```

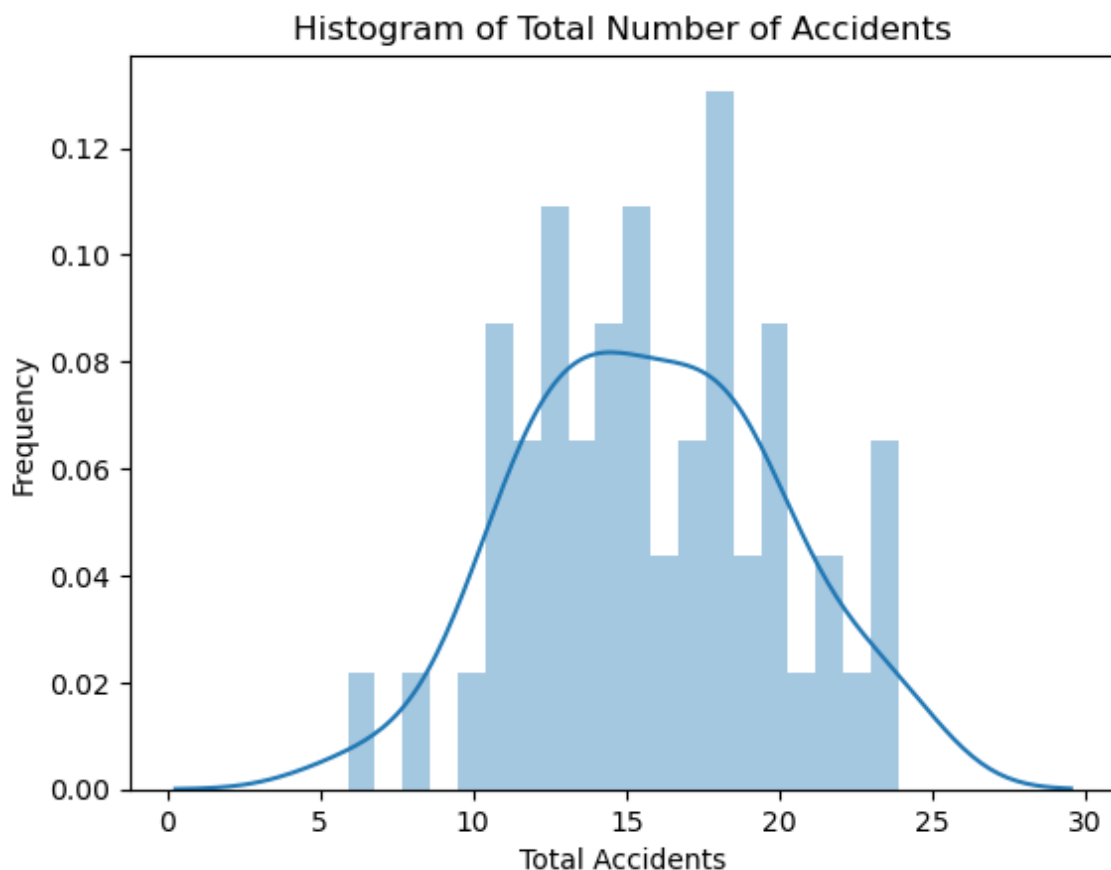
C:\Users\seepana manikanta\AppData\Local\Temp\ipykernel_9816\819728707.py:
1: UserWarning:

`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> (<https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>)

```
sns.distplot(dataset["total"], bins=20, kde=True)
```

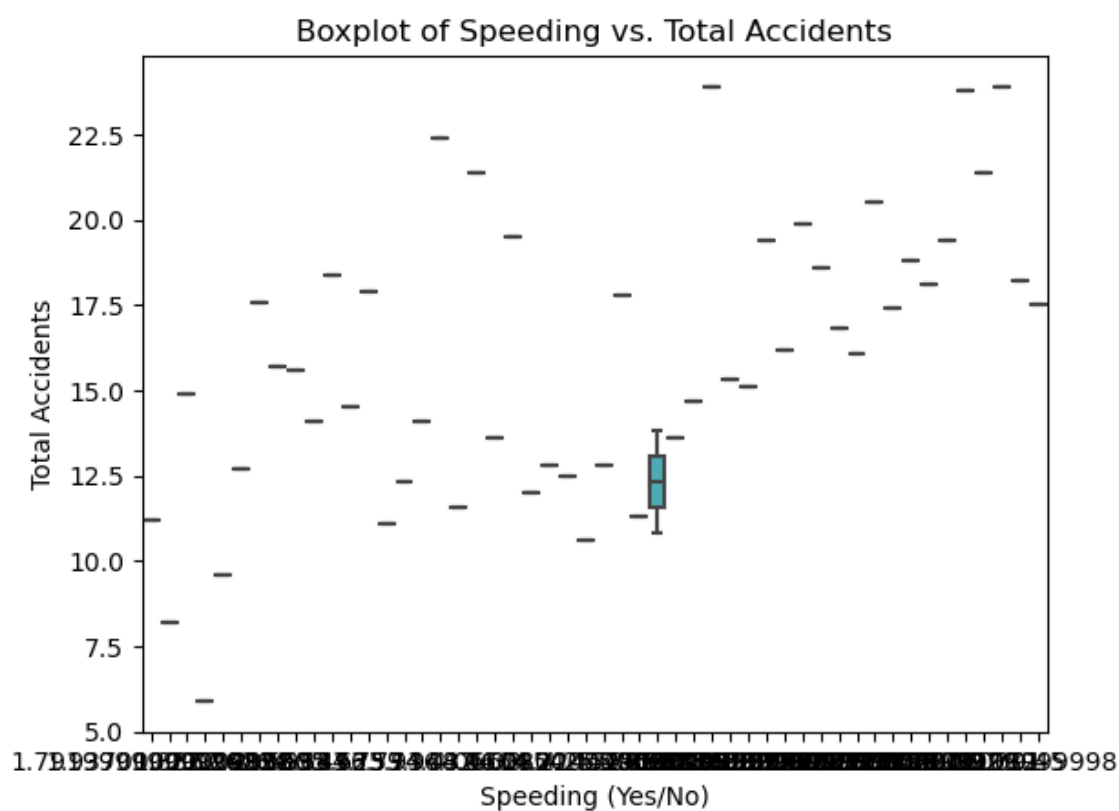


In [10]:

```
# Inference: The histogram shows the distribution of total accidents. Most states have a
```

In [11]:

```
sns.boxplot(x="speeding", y="total", data=dataset)
plt.title("Boxplot of Speeding vs. Total Accidents")
plt.xlabel("Speeding (Yes/No)")
plt.ylabel("Total Accidents")
plt.show()
```



In [12]:

```
# Inference: The boxplot illustrates the relationship between speeding (yes/no) and the
```

In [13]:

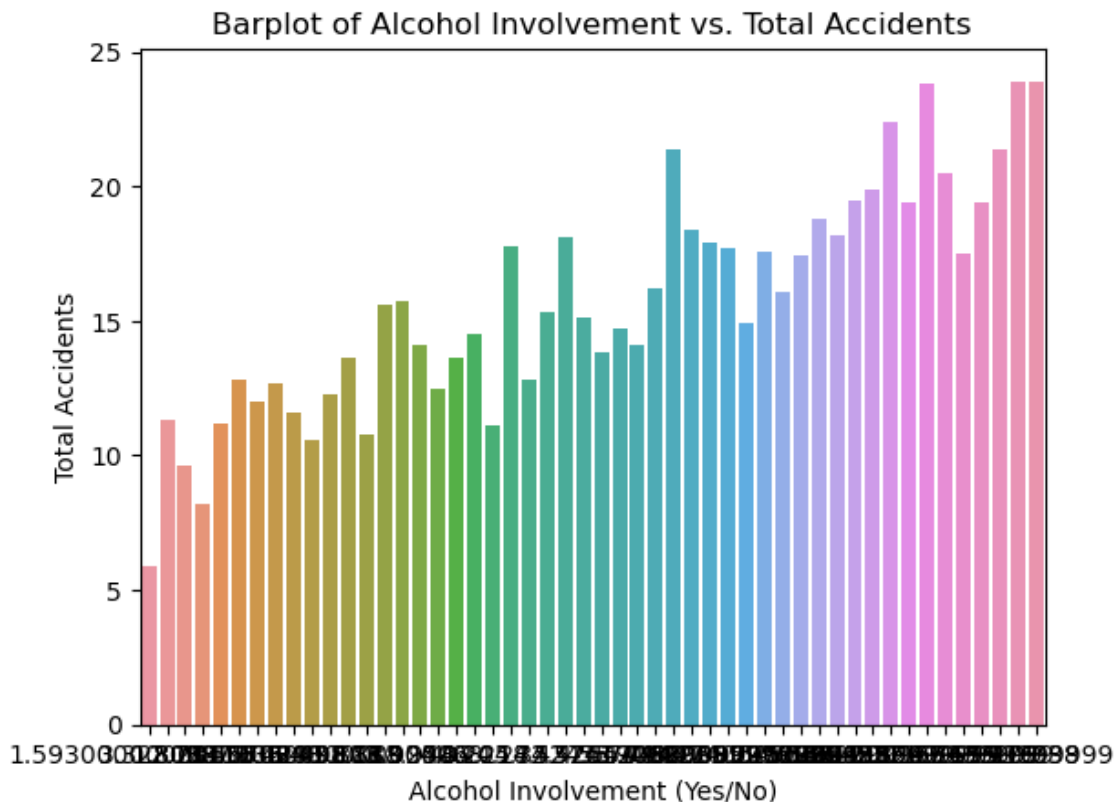
```
sns.barplot(x="alcohol", y="total", data=dataset, ci=None)
plt.title("Barplot of Alcohol Involvement vs. Total Accidents")
plt.xlabel("Alcohol Involvement (Yes/No)")
plt.ylabel("Total Accidents")
plt.show()
```

C:\Users\seepana manikanta\AppData\Local\Temp\ipykernel_9816\13257844.py:

1: FutureWarning:

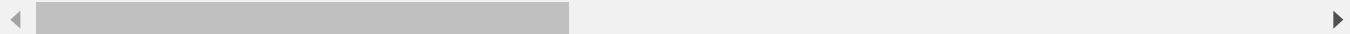
The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.

```
sns.barplot(x="alcohol", y="total", data=dataset, ci=None)
```



In [14]:

```
# Inference: The barplot compares the total number of accidents for states with and with
```

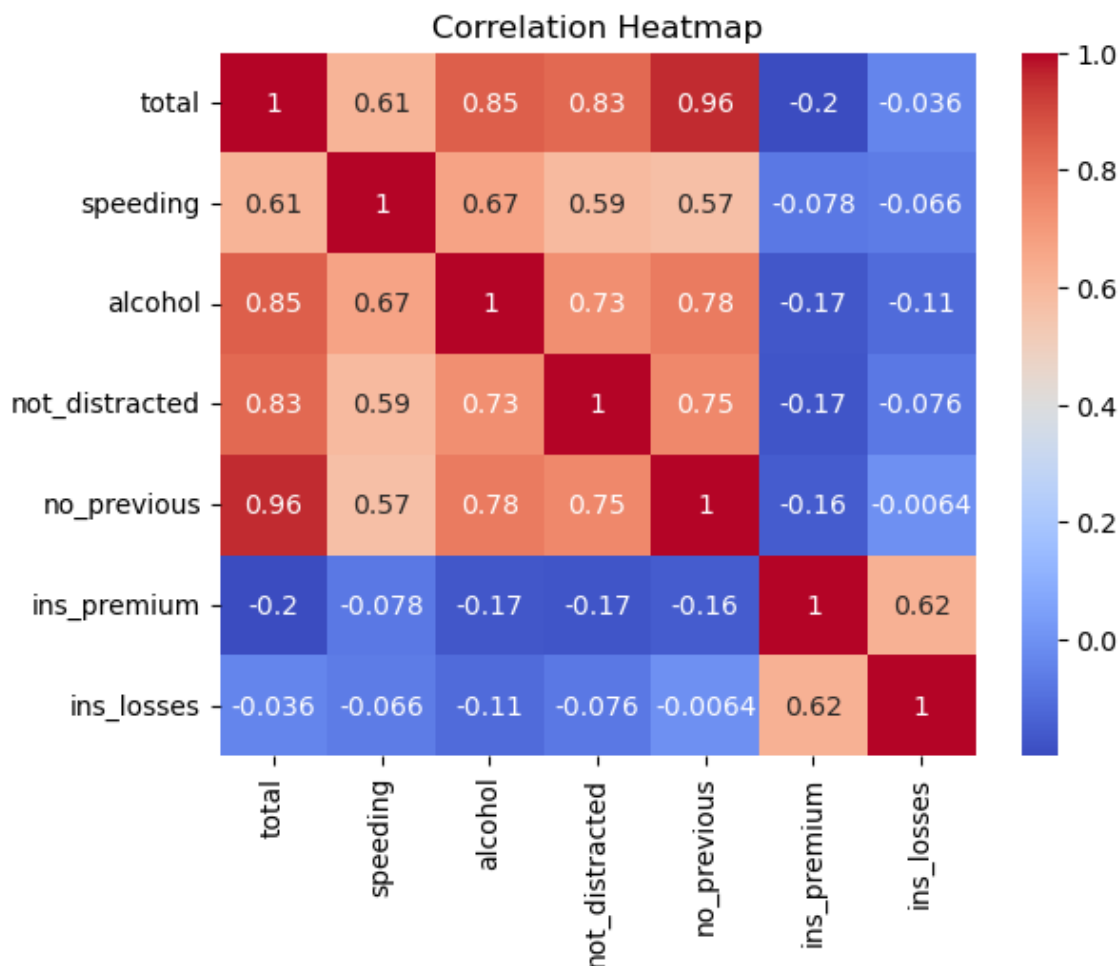


In [15]:

```
correlation_matrix = dataset.corr()
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm")
plt.title("Correlation Heatmap")
plt.show()
```

C:\Users\seepana manikanta\AppData\Local\Temp\ipykernel_9816\1427295036.py: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 id columns or specify the value of numeric_only to silence this warning.

```
correlation_matrix = dataset.corr()
```



In [16]:

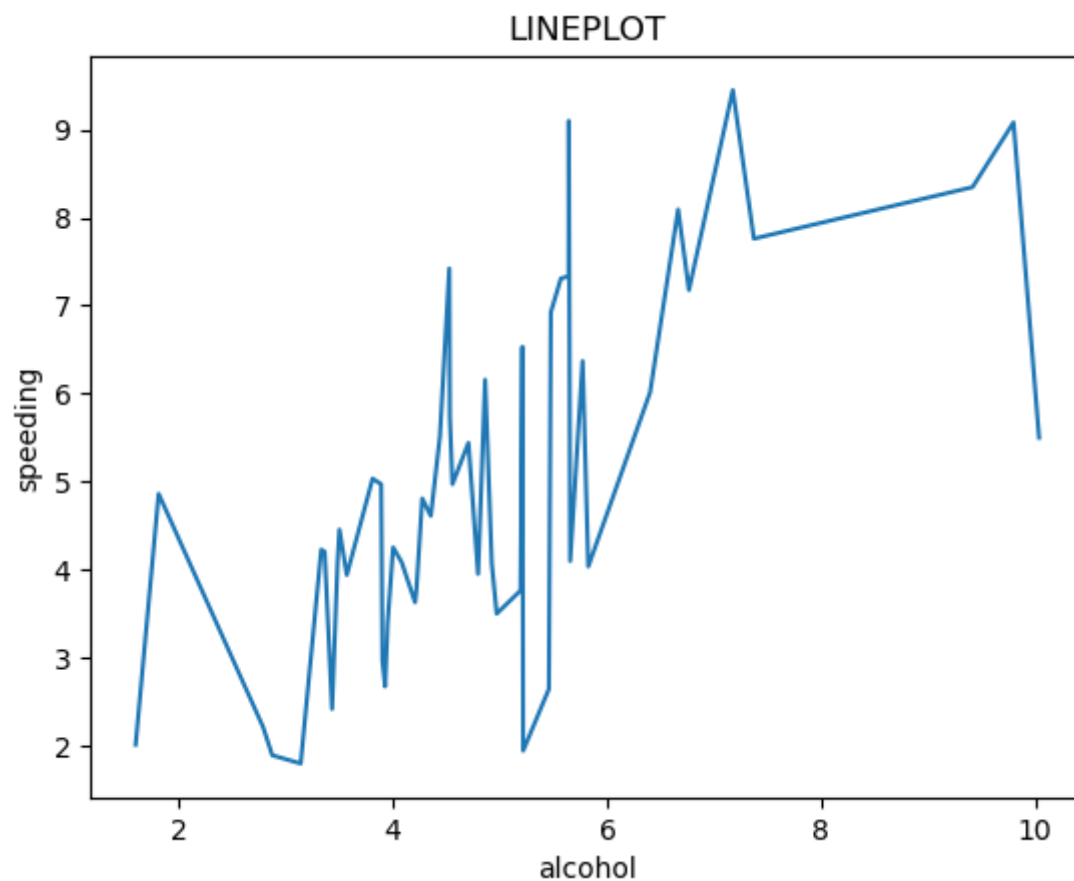
```
# Inference: The heatmap displays the correlation between numeric variables in the datas
```

In [17]:

```
sns.lineplot(x="alcohol",y="speeding",data=dataset)  
plt.title("LINEPLOT")
```

Out[17]:

Text(0.5, 1.0, 'LINEPLOT')



In [18]:

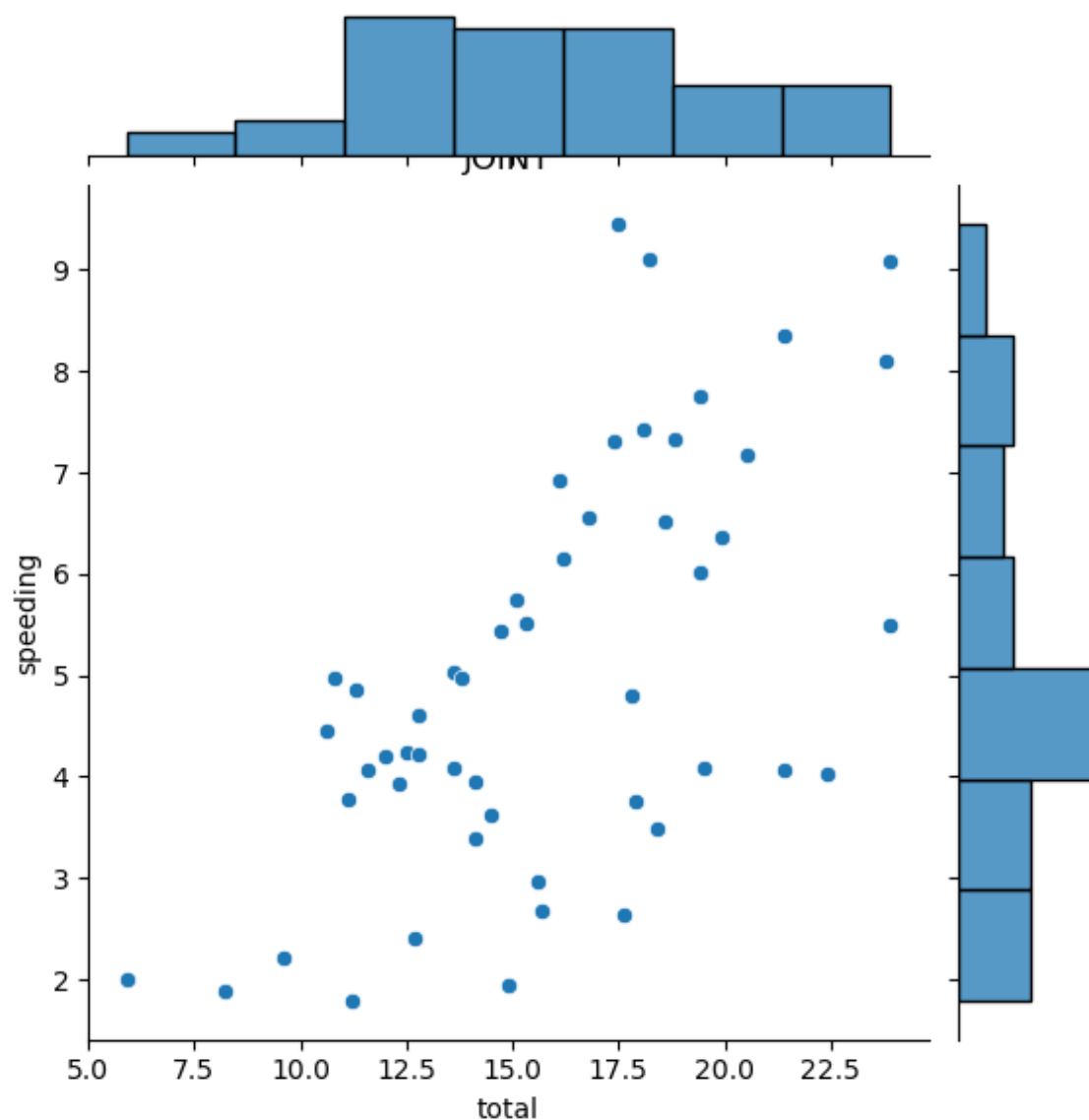
```
# Inference: The line plot comparing "Alcohol" and "Speeding" incidents in car crashes s
```


In [19]:

```
sns.jointplot(x="total",y="speeding",data=dataset)  
plt.title("JOINT")
```

Out[19]:

Text(0.5, 1.0, 'JOINT')

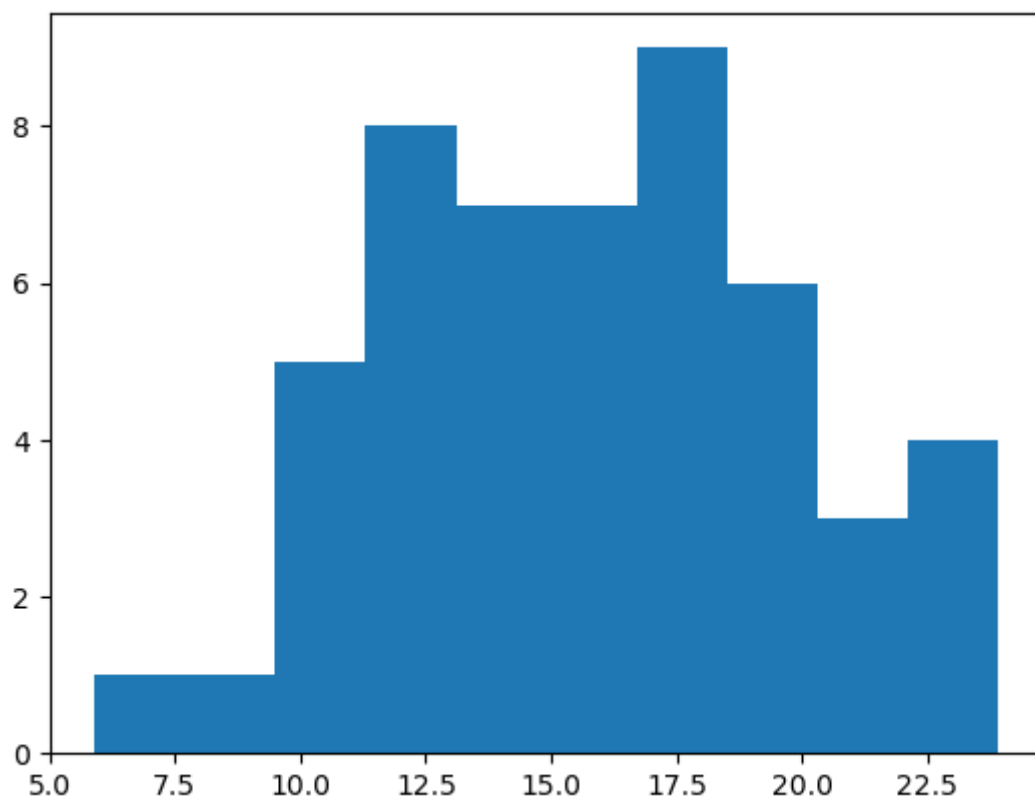


In [20]:

```
# INFERENCE :States with a higher rate of "Speeding" incidents tend to have a wider rang
```

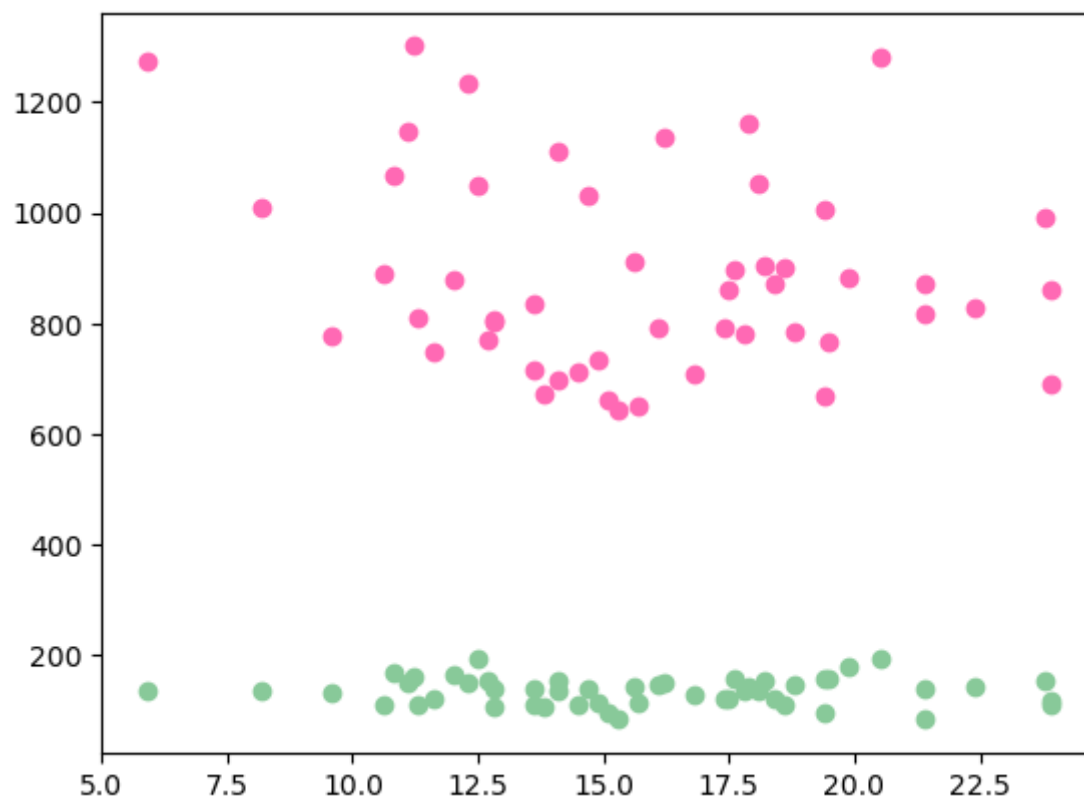
In [21]:

```
plt.hist(dataset['total'])  
plt.show()
```



In [22]:

```
plt.scatter(dataset['total'],dataset['ins_premium'],color='hotpink')  
plt.scatter(dataset['total'],dataset['ins_losses'],color='#88c999')  
plt.show()
```

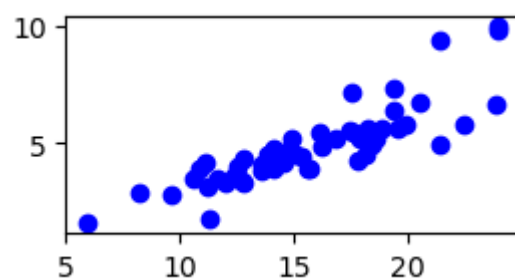


In [23]:

```
plt.subplot(3,2,2)  
plt.scatter(dataset['total'],dataset['alcohol'],color='blue')
```

Out[23]:

<matplotlib.collections.PathCollection at 0x125084d4fd0>



In []:

```
# INFERENCE :
```