Name: Chirag

Email: chirag.2021@vitstudent.ac.in

Branch: IT

Campus: VIT Vellore

In [1]:
```
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import warnings
```

In [2]:
```
# Ignore FutureWarnings
warnings.simplefilter(action='ignore', category=FutureWarning)
```

In [3]:
```
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']
```

In [5]:
```
df = pd.read_csv('car_crashes.csv')
```

In [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
```

Inferences

- We have No Null values, and all the columns are Floating Number

In [7]:
```
df.describe()
```

Out[7]:

	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses
count	51.000000	51.000000	51.000000	51.000000	51.000000	51.000000	51.000000
mean	15.790196	4.998196	4.886784	13.573176	14.004882	886.957647	134.493137
std	4.122002	2.017747	1.729133	4.508977	3.764672	178.296285	24.835922
min	5.900000	1.792000	1.593000	1.760000	5.900000	641.960000	82.750000
25%	12.750000	3.766500	3.894000	10.478000	11.348000	768.430000	114.645000
50%	15.600000	4.608000	4.554000	13.857000	13.775000	858.970000	136.050000
75%	18.500000	6.439000	5.604000	16.140000	16.755000	1007.945000	151.870000
max	23.900000	9.450000	10.038000	23.661000	21.280000	1301.520000	194.780000

In [8]:
```
df.head()
```

Out[8]:

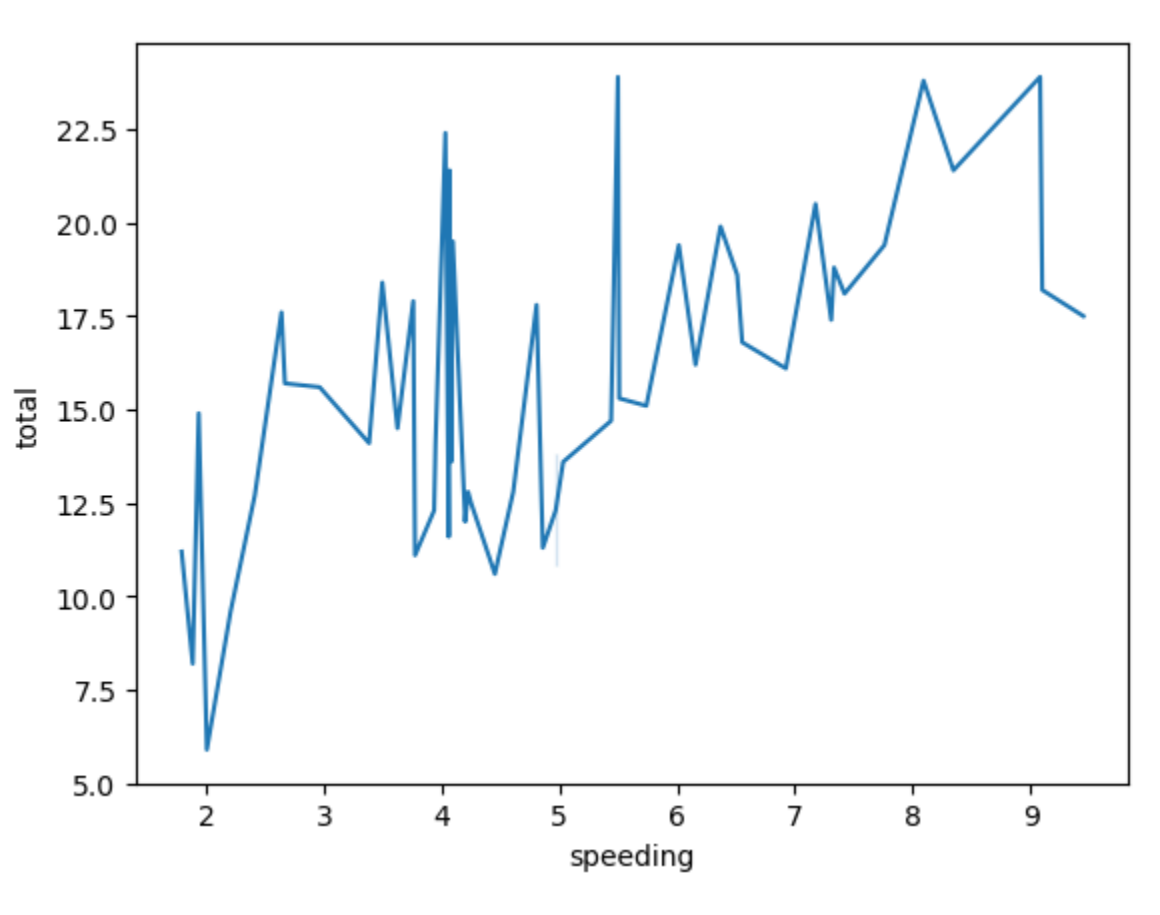
	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses	abbrev
0	18.8	7.332	5.640	18.048	15.040	784.55	145.08	AL
1	18.1	7.421	4.525	16.290	17.014	1053.48	133.93	AK
2	18.6	6.510	5.208	15.624	17.856	899.47	110.35	AZ
3	22.4	4.032	5.824	21.056	21.280	827.34	142.39	AR
4	12.0	4.200	3.360	10.920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

Inference

- Since we have all the of the data as Numerical vs Numerical we may use Scatter Plot

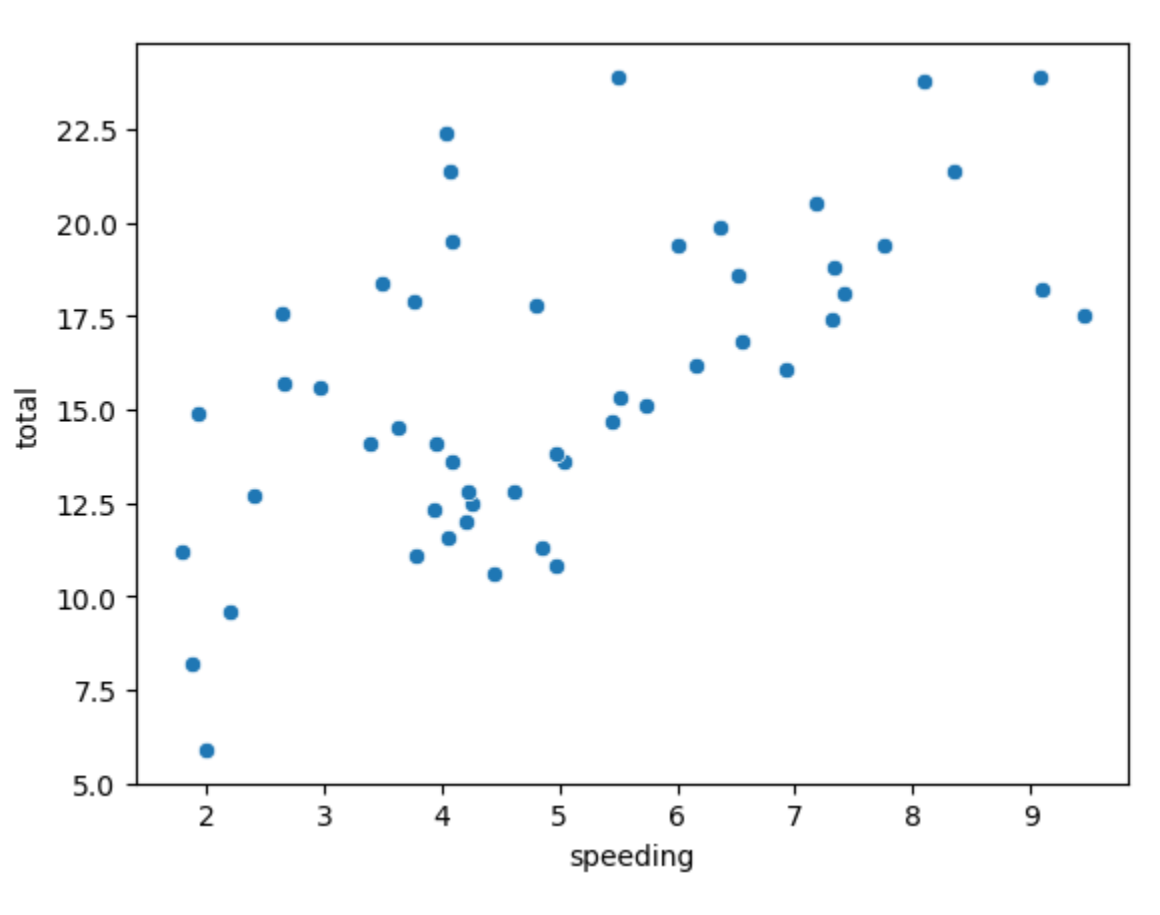
In [9]:
```
sns.lineplot(y='total',x='speeding',data=df)
```

Out[9]: `<Axes: xlabel='speeding', ylabel='total'>`



In [10]:
```
sns.scatterplot(y='total',x='speeding',data=df)
```

Out[10]: `<Axes: xlabel='speeding', ylabel='total'>`

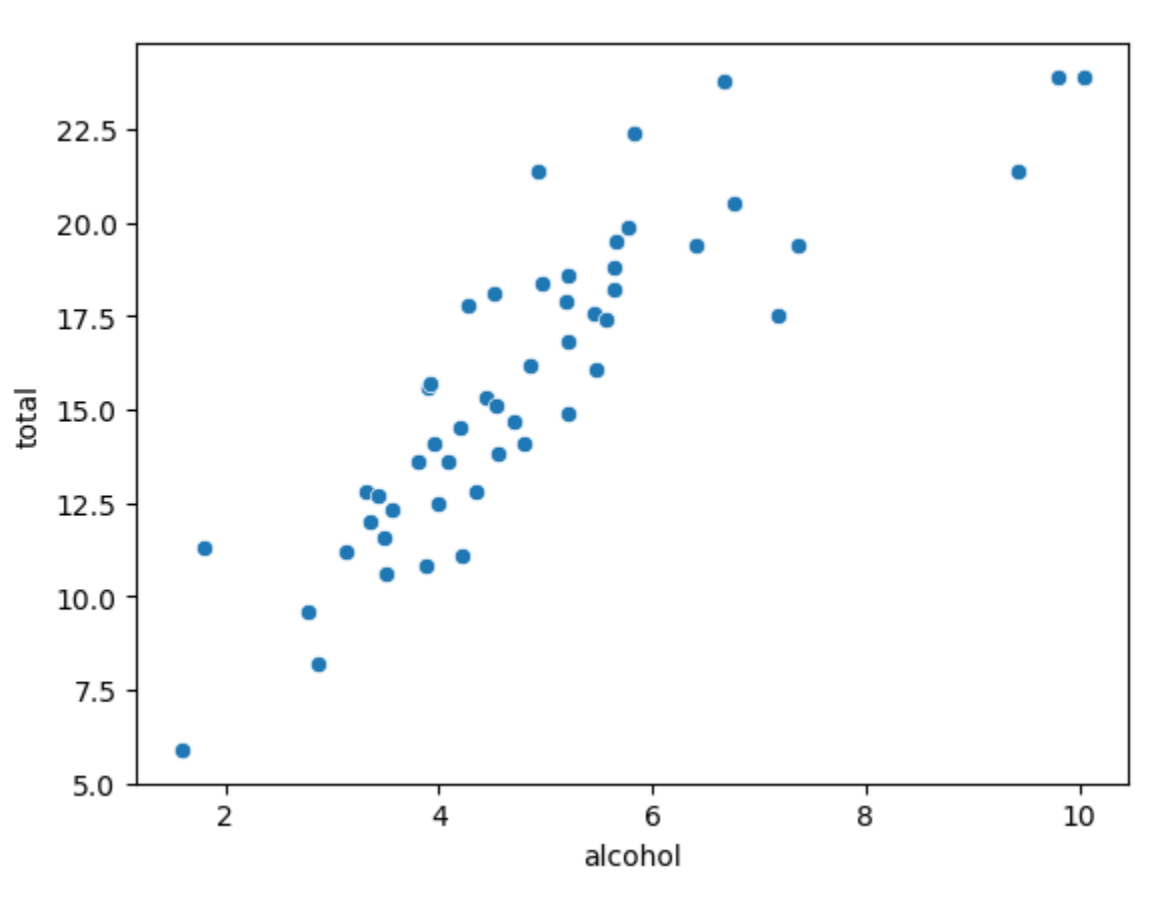


Inferences From Above Plot:

- Here we Can see as the Speeding is increased the total number of Crashes is also increasing

In [11]:
```
sns.scatterplot(y='total',x='alcohol',data=df)
```

Out[11]: `<Axes: xlabel='alcohol', ylabel='total'>`

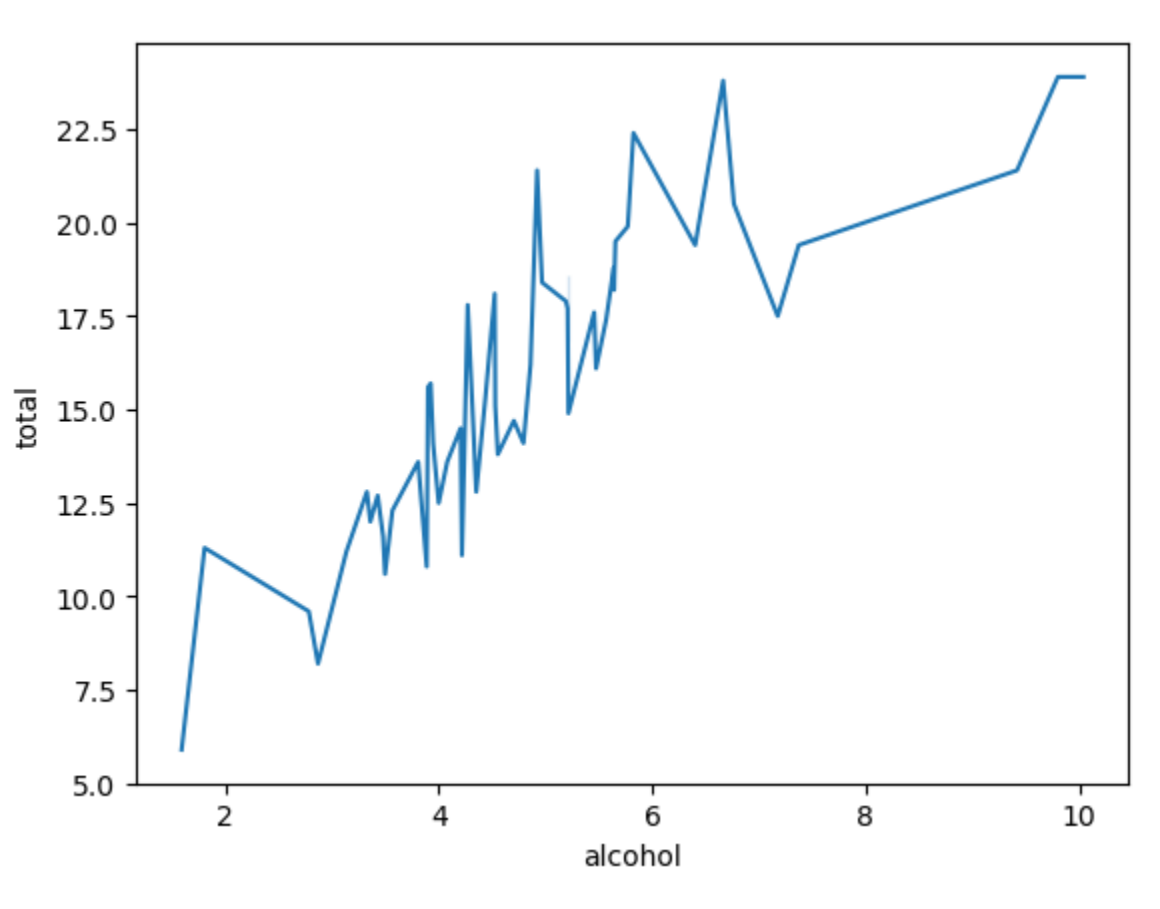


Inferences From Above Plot:(+ve Correlation)

Here we Can see as the Speeding is increased the total number of Crashes is also increasing

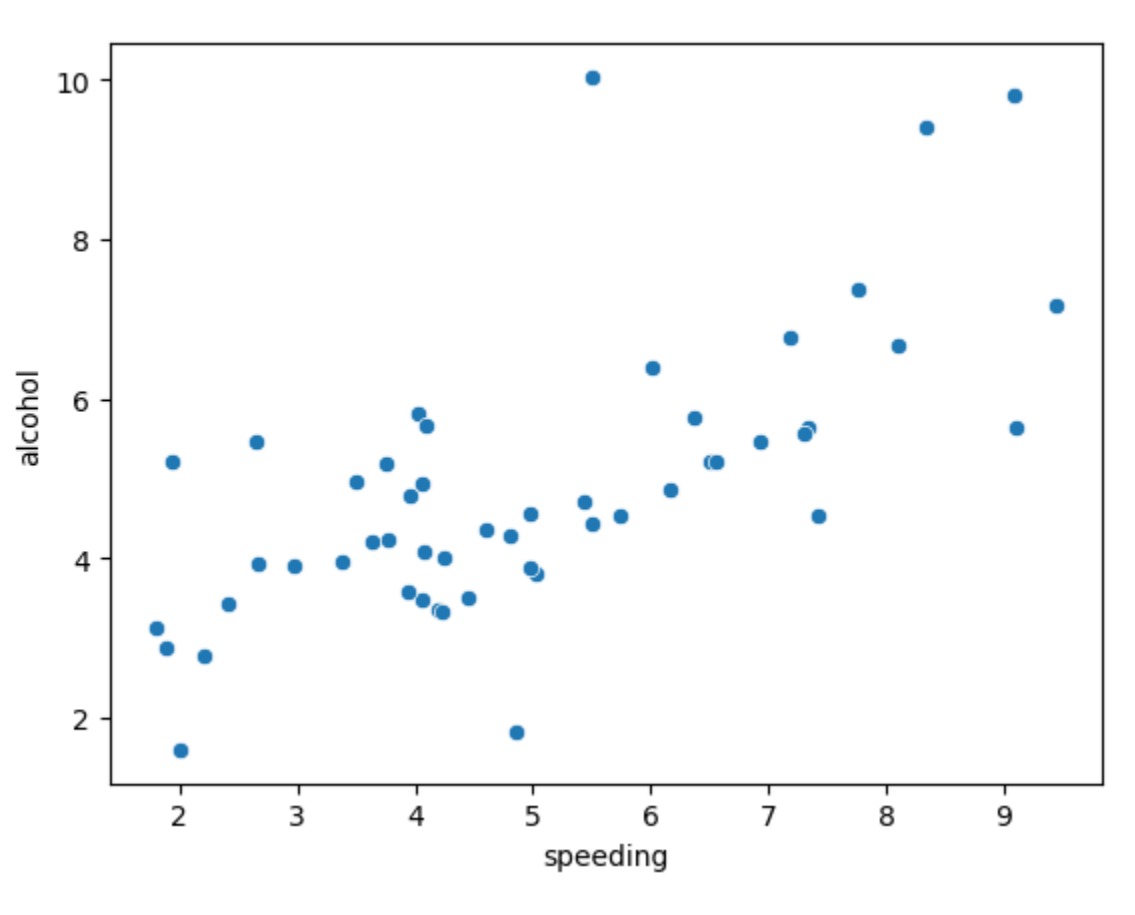
In [12]:
```
sns.lineplot(y='total',x='alcohol',data=df)
```

Out[12]: `<Axes: xlabel='alcohol', ylabel='total'>`



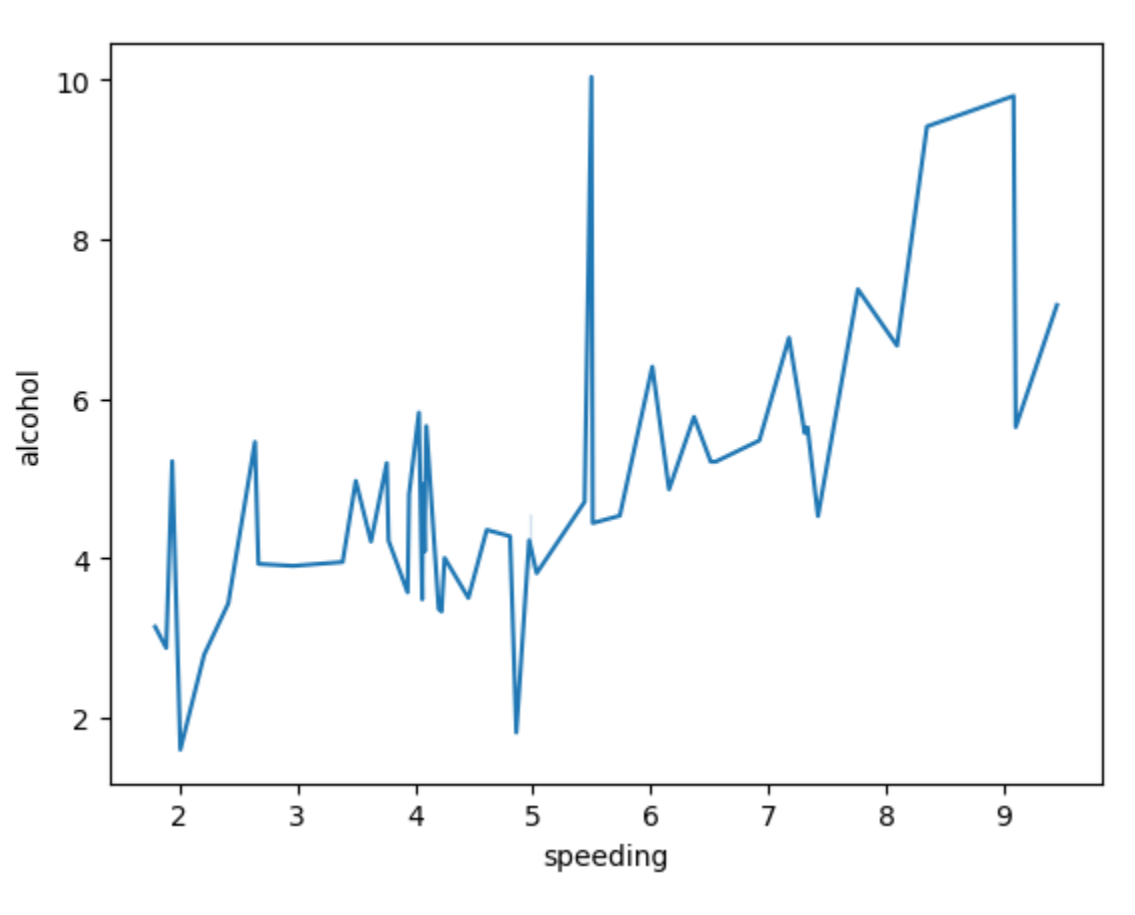
In [13]:
```
sns.scatterplot(y='alcohol',x='speeding',data=df)
```

Out[13]: `<Axes: xlabel='speeding', ylabel='alcohol'>`



In [15]:
```
sns.lineplot(y='alcohol',x='speeding',data=df)
```

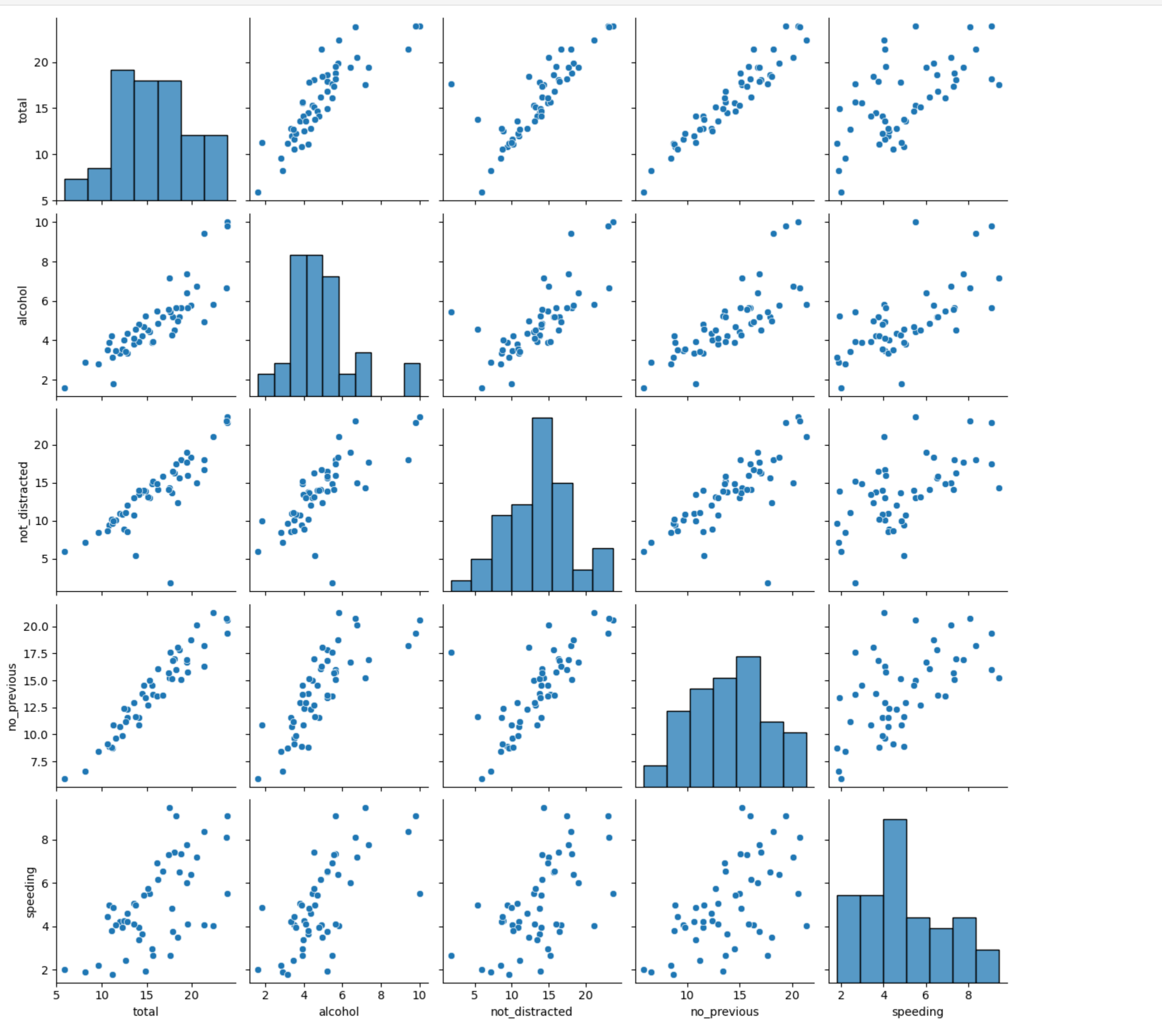
Out[15]: `<Axes: xlabel='speeding', ylabel='alcohol'>`



Inferences From Above Graphs:

- here we see that generally, as we drink more alchohol more chances of speeding

In [16]:
```
sns.pairplot(df[['total','alcohol','not_distracted','no_previous','speeding']])
plt.show()
```



Inferences

- From the Above Plots we can see that:

In [ ]:
```
-Realations According to the Total:(Positively related)
1:- total vs Alcohol
2:- total vs not_distracted
3:- total vs no_previous
This Symbolises that the number of acident increasing as the Y-axis values increases;
9/13/23, 12:31 PM Assignment2.ipynb - Colaboratory
https://colab.research.google.com/drive/1RrJdLAjcrr2K5c9XknXKhcbXEsPa5ukH#scrollTo=SZNnt74sV2NS&printMode=true 6/9
-Relations According to the Speeding:
1:- speeding vs alcohol
This shows that speeding is done by the ones who drank more alcohol.
```

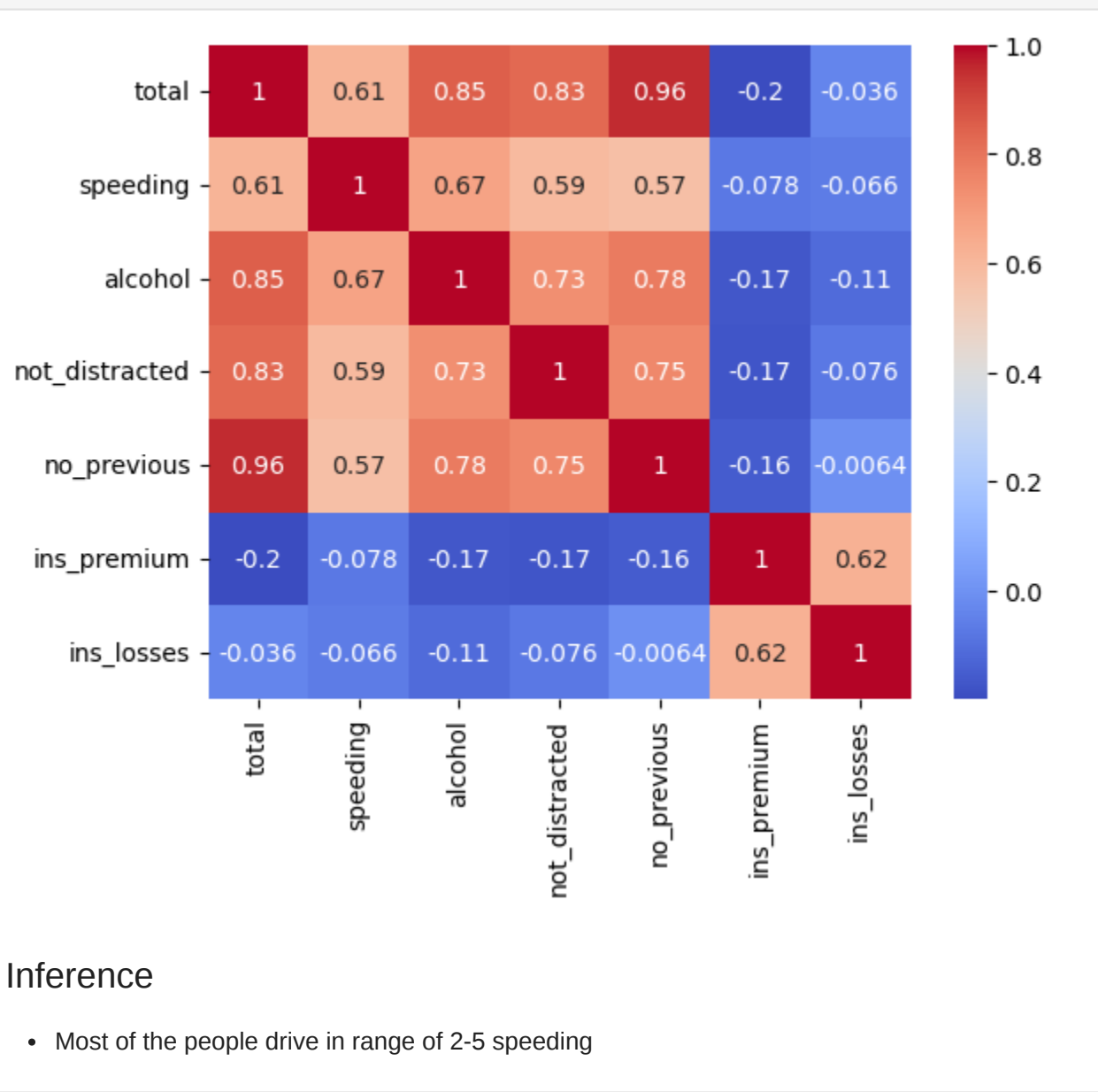
-Relations According to the Alcohol: 1:- Alcohol vs no_previous 2:- Alcholo vs no_distracted From the Graphs we can see a relationship that persons who drank the most are less likely to be distracted. The persons who drinks more had no_previous crashes

---

In [ ]:
```
-Relations According to the not_distracted:
1:- not_distracted vs no_previous
This says that persons who are not distracted , have had no previous accident.
```

Realtionship between other plot are hard to draw because of does not follow any Good pattern.

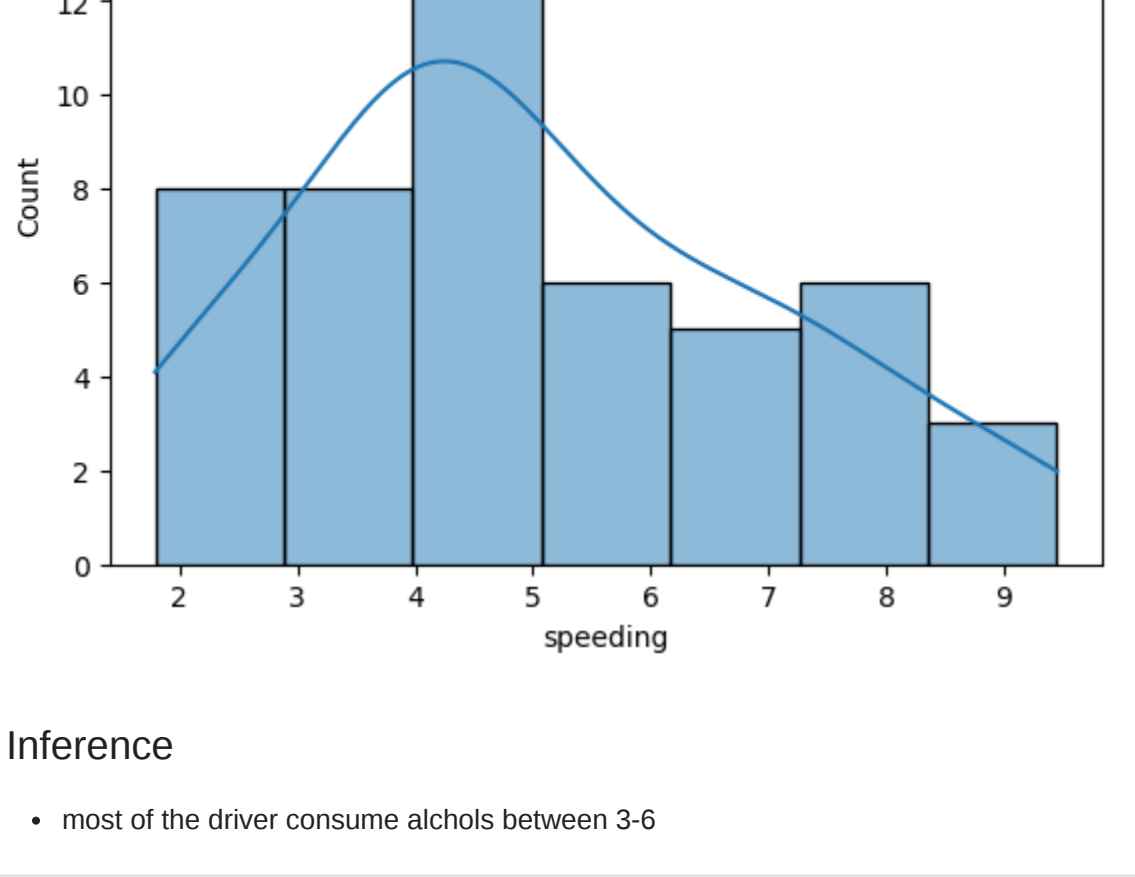
In [17]:
```
correlation_matrix = df.corr(numeric_only=True)
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.show()
```



Inference

- Most of the people drive in range of 2-5 speeding

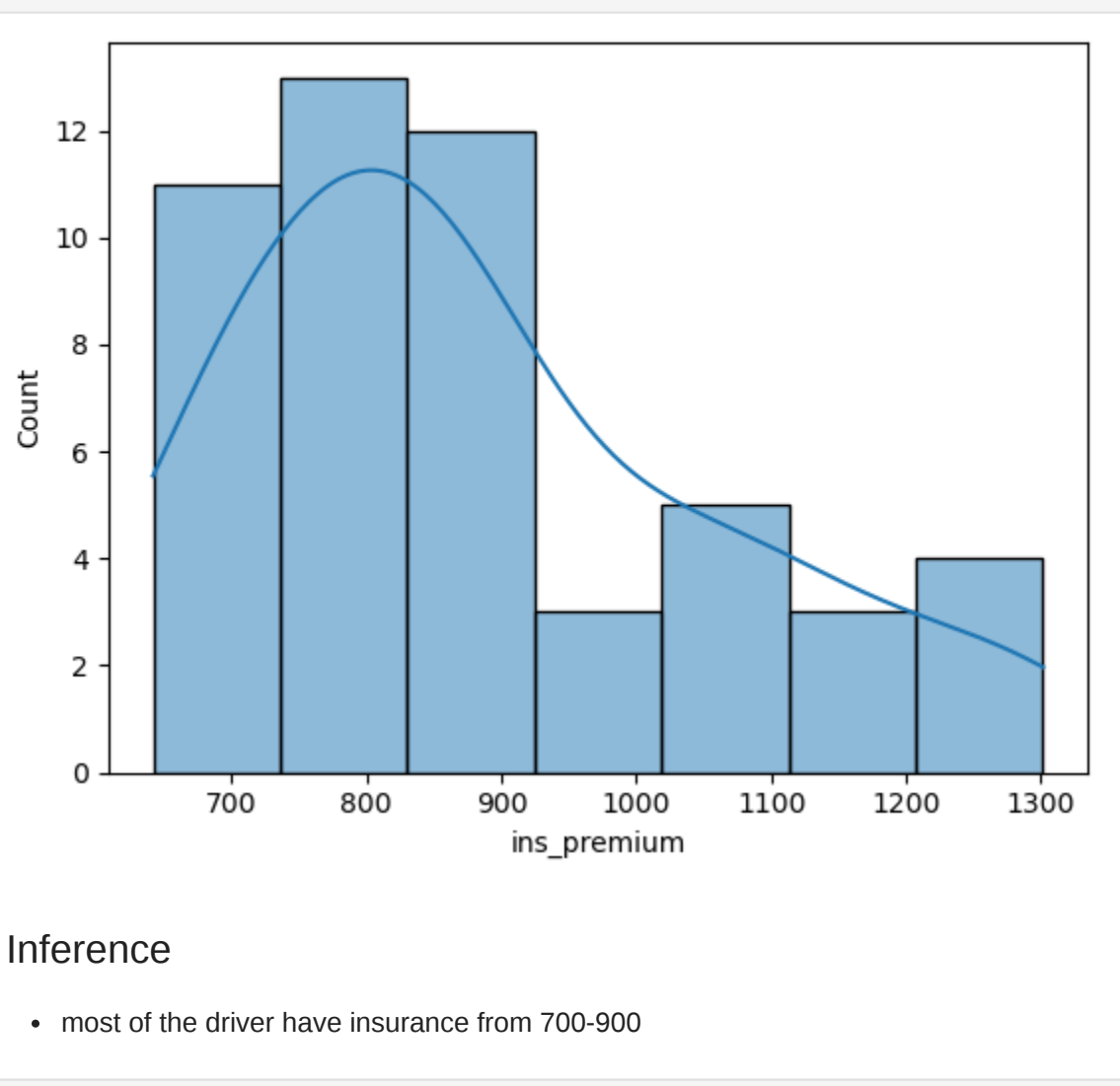
In [18]:
```
sns.histplot(df['speeding'], kde=True)
plt.show()
```



Inference

- most of the driver consume alchols between 3-6

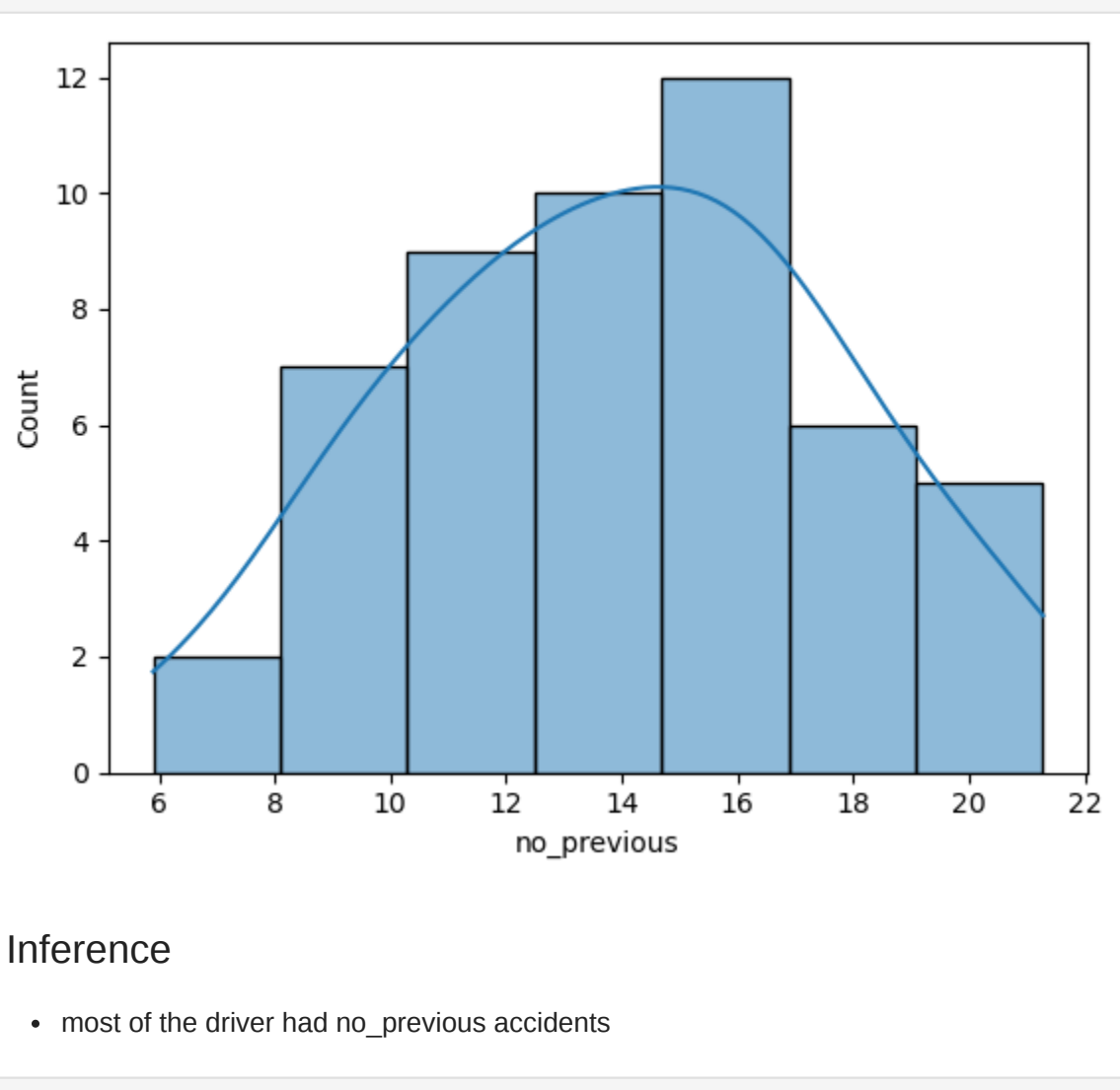
In [19]:
```
sns.histplot(df['ins_premium'], kde=True)
plt.show()
```



Inference

- most of the driver have insurance from 700-900

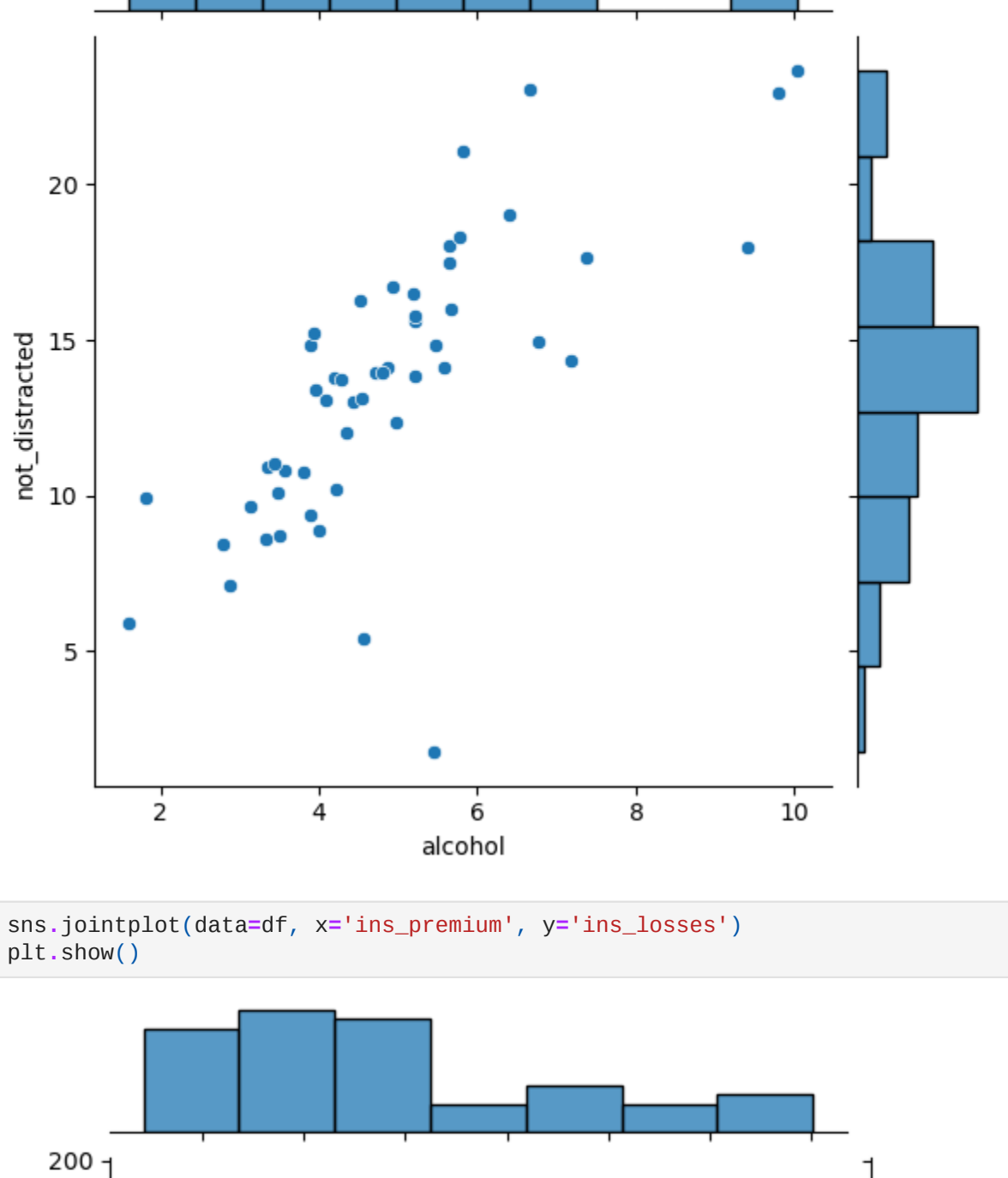
In [20]:
```
sns.histplot(df['no_previous'], kde=True)
plt.show()
```



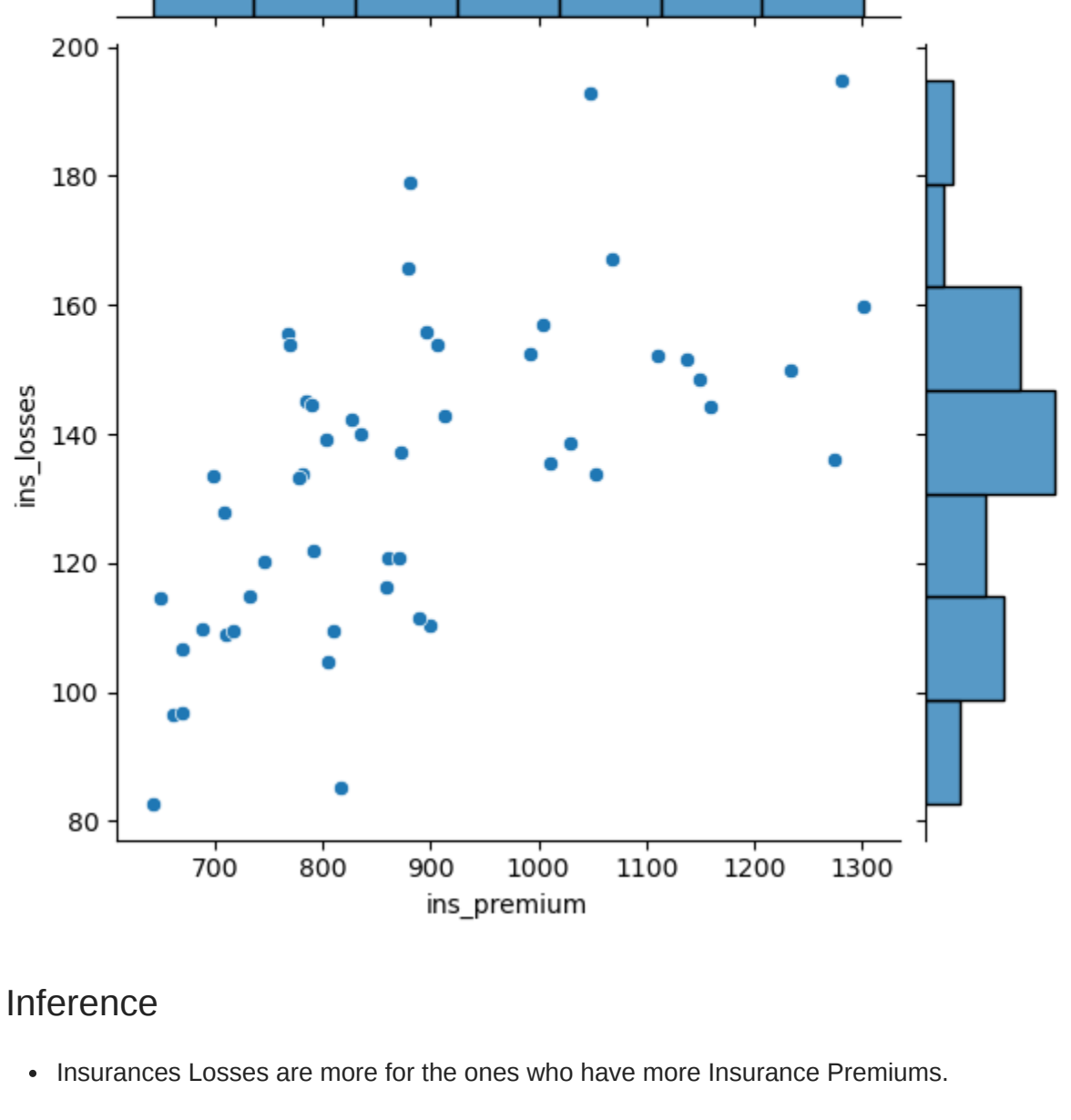
Inference

- most of the driver had no_previous accidents

In [21]:
```
sns.jointplot(data=df, x='alcohol', y='not_distracted')
plt.show()
```



In [22]:
```
sns.jointplot(data=df, x='ins_premium', y='ins_losses')
plt.show()
```



Inference

- Insurances Losses are more for the ones who have more Insurance Premiums.