Assmnt Week-2

```
Assignment 8 th september
1.Take car crashes dataset from seaborn library
2.load the dataset
3.data visualiation
4.Inference is must for each and every graph
5.Submit it by wednesday in pdf format
```

In [30]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:
```
sns.get_dataset_names()
```

Out[2]:
```
['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 [3]:
```
df = sns.load_dataset('car_crashes')
```

In [27]:
```
df.head(5)
```

Out[27]:

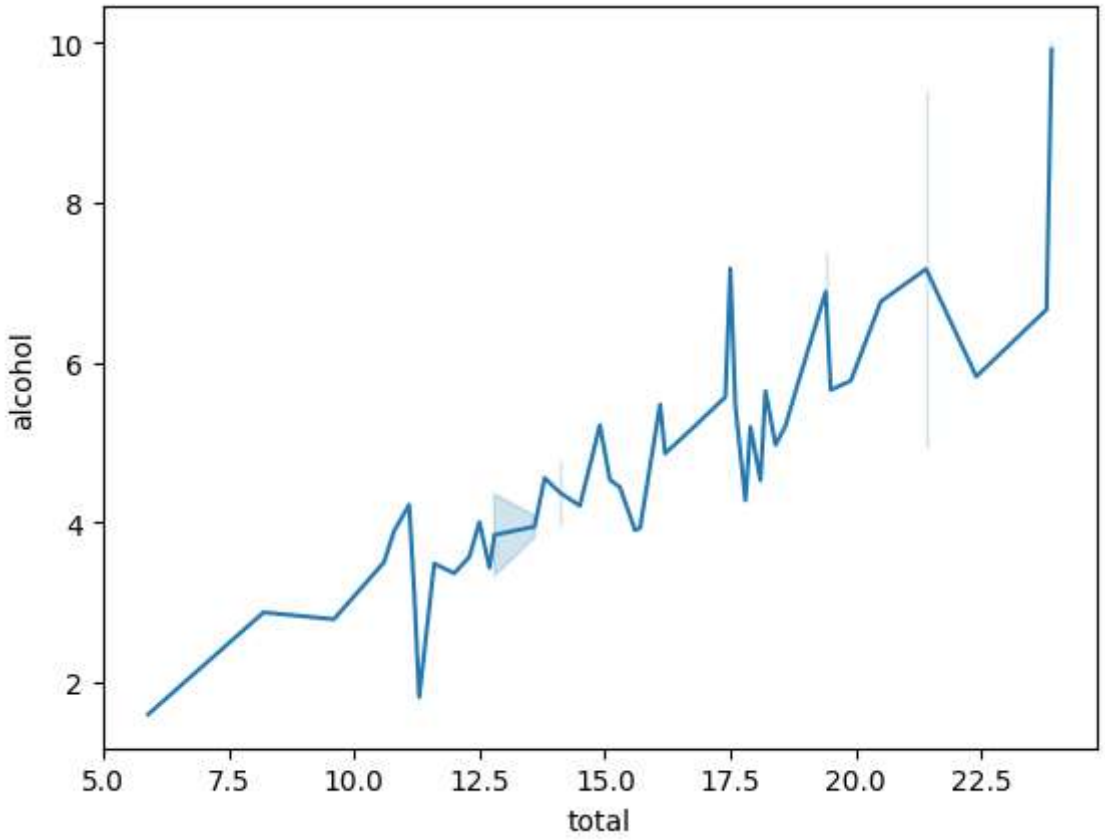
	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses	abbrev
0	18.8	7.332	5.640	18.048	15.040	784.55	145.08	AL
1	18.1	7.421	4.525	16.290	17.014	1053.48	133.93	AK
2	18.6	6.510	5.208	15.624	17.856	899.47	110.35	AZ
3	22.4	4.032	5.824	21.056	21.280	827.34	142.39	AR
4	12.0	4.200	3.360	10.920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

In [5]:
```
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
```

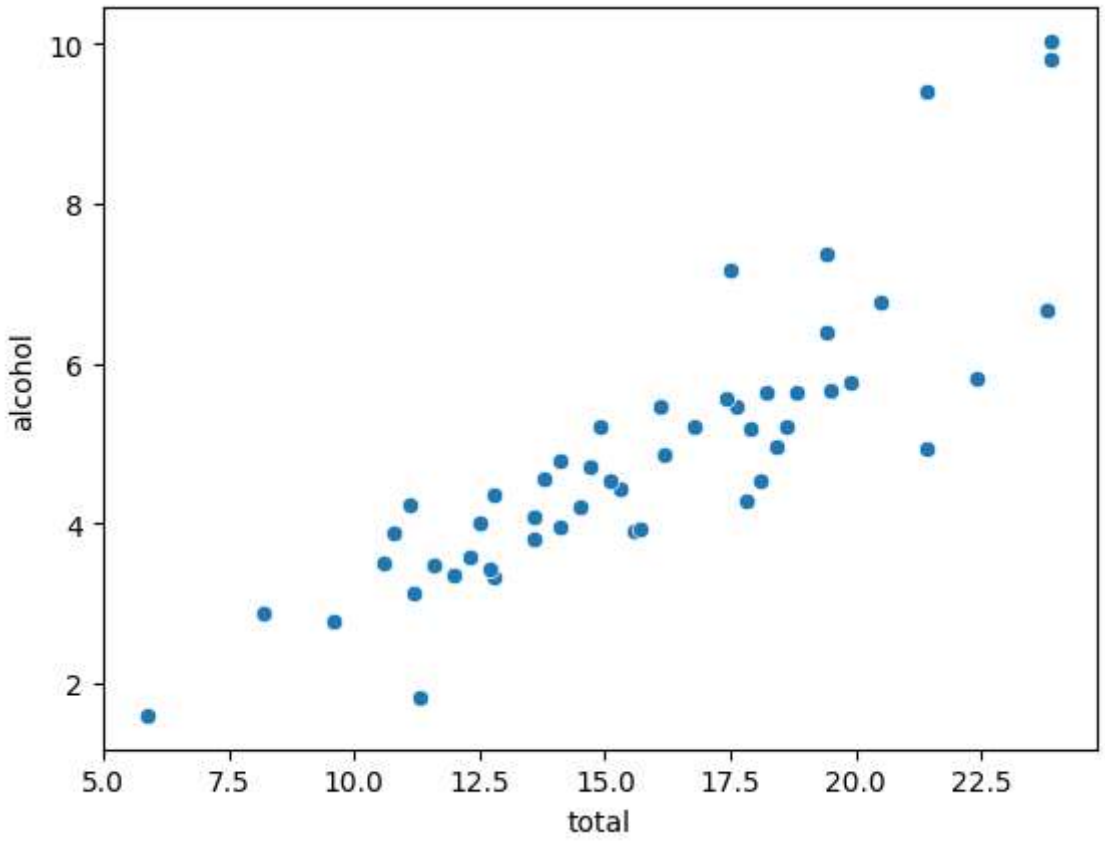
In [13]:
```
sns.lineplot(x = 'total', y = 'alcohol', data = df)
# Inference: With this the consuption of alcohol has a direct impact on No of accidents
```

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



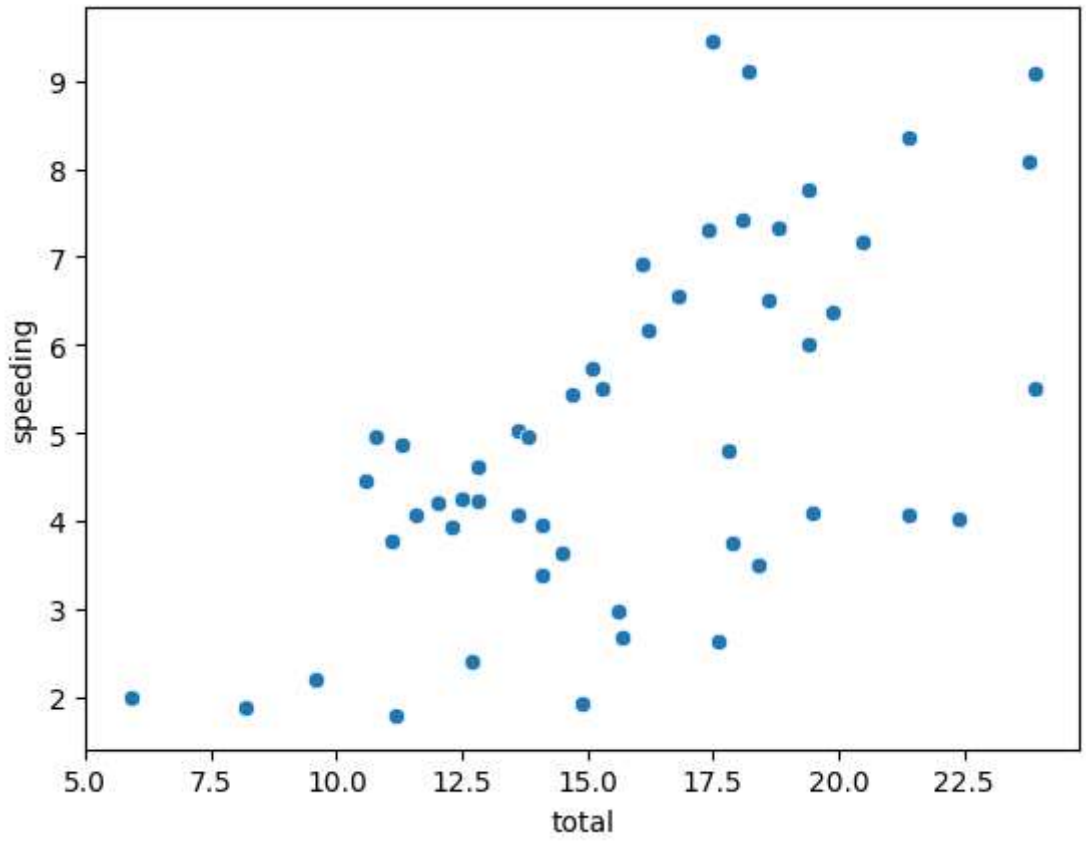
In [10]:
```
sns.scatterplot(x = 'total', y = 'alcohol', data = df)
# Inference: With this the consuption of alcohol has a direct impact on No of accidents
```

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



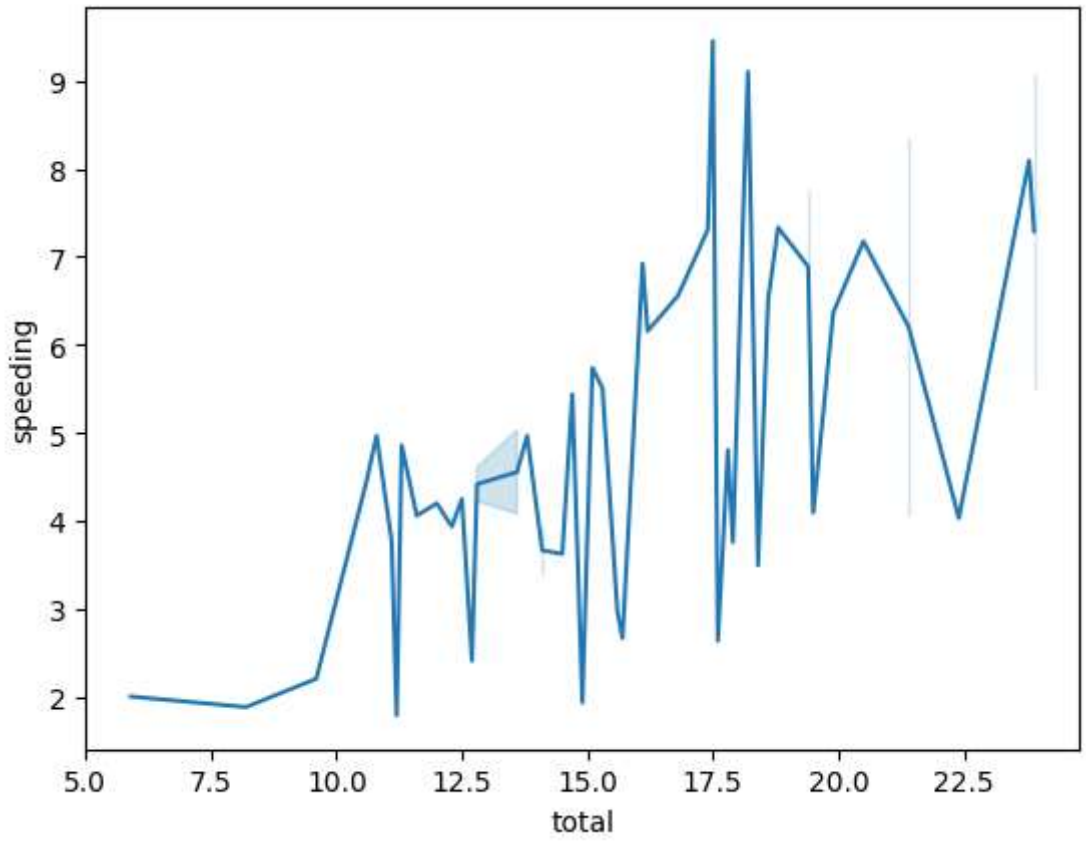
In [8]:
```
sns.scatterplot(x = 'total', y = 'speeding', data = df)
# Inference: With the increse in speed the number of accidents also increses
```

Out[8]: <Axes: xlabel='total', ylabel='speeding'>



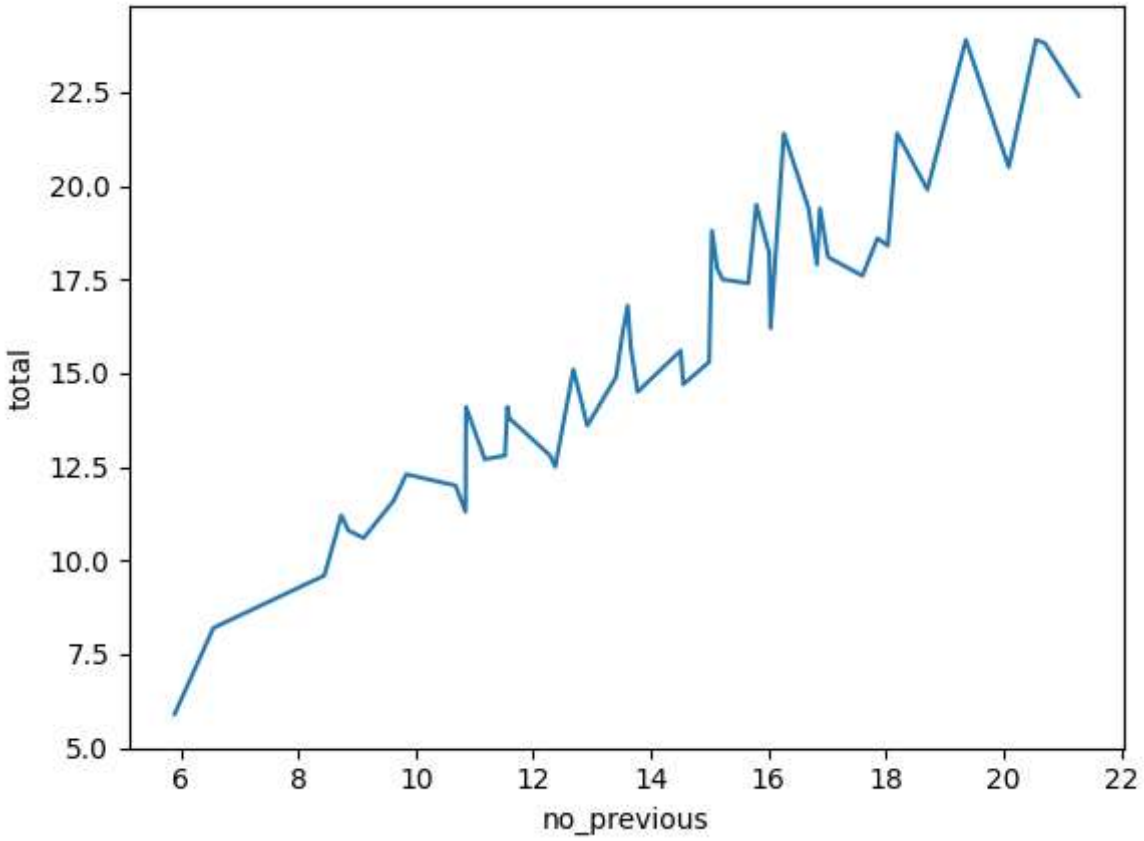
In [26]:
```
sns.lineplot(x = 'total', y = 'speeding', data = df)
# Inference: With the increse in speed the number of accidents also increses
```

Out[26]: <Axes: xlabel='total', ylabel='speeding'>



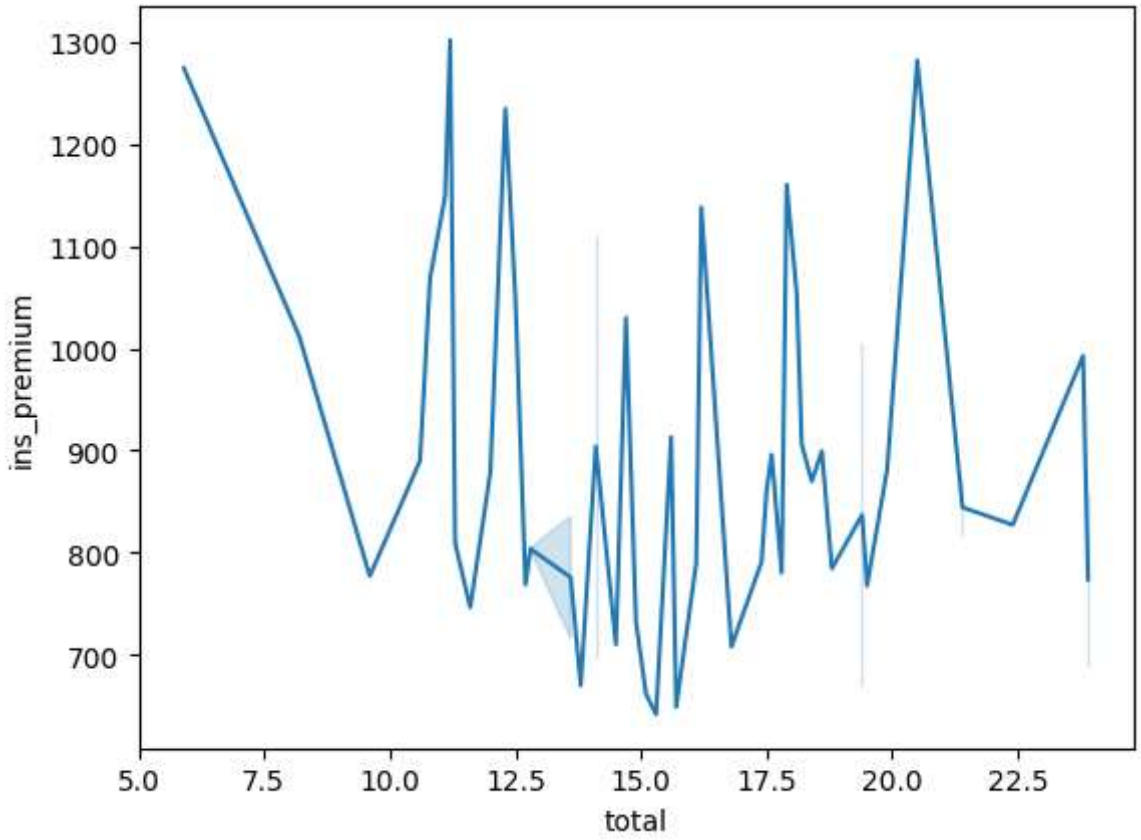
In [39]:
```
sns.lineplot(y = 'total', x = 'no_previous', data = df)
# Inference: The total is directly propotional to the no_previous
```

Out[39]: <Axes: xlabel='no_previous', ylabel='total'>



In [36]:
```
sns.lineplot(x = 'total', y = 'ins_premium', data = df)
# Inference: The Total Num of accidents has no such impact with the ins_premium
```

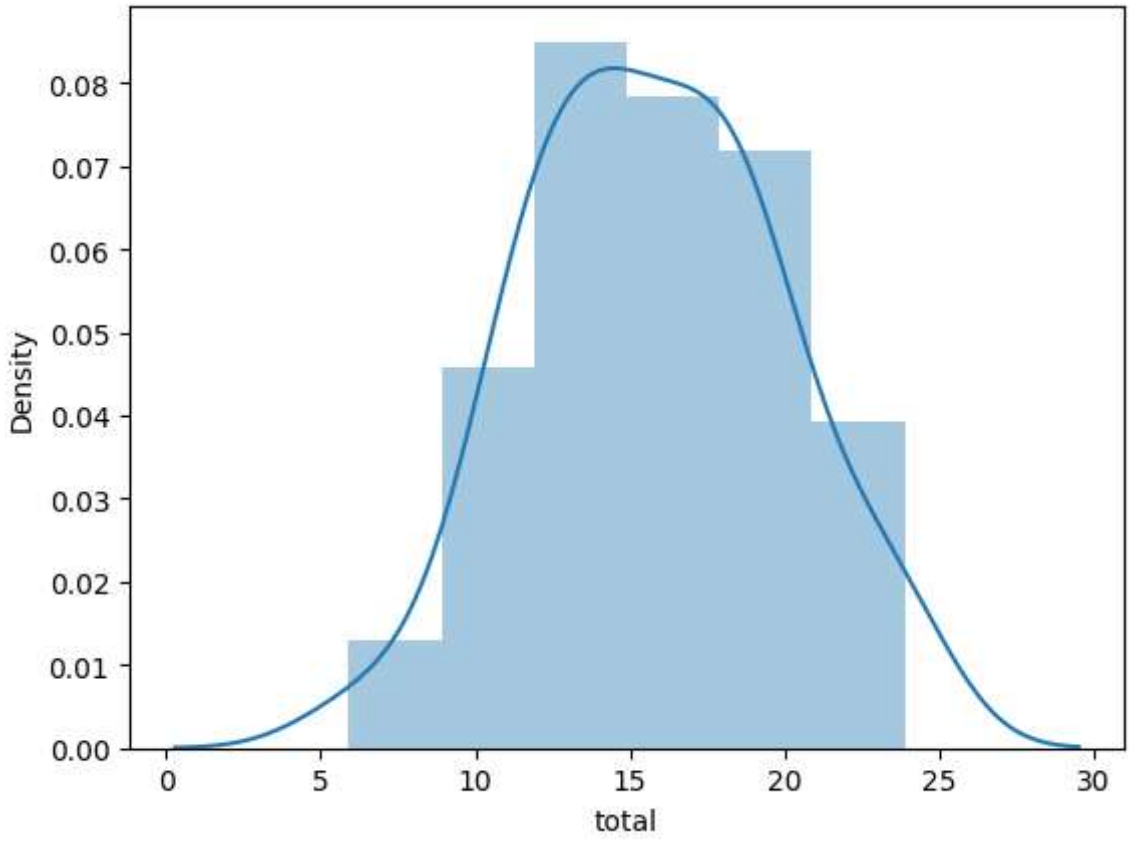
Out[36]: <Axes: xlabel='total', ylabel='ins_premium'>



In [18]:
```
sns.distplot(df['total'])
```

```
<ipython-input-18-2ba73417f012>: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(df['total'])
```

Out[18]:
```
<Axes: xlabel='total', ylabel='Density'>
```

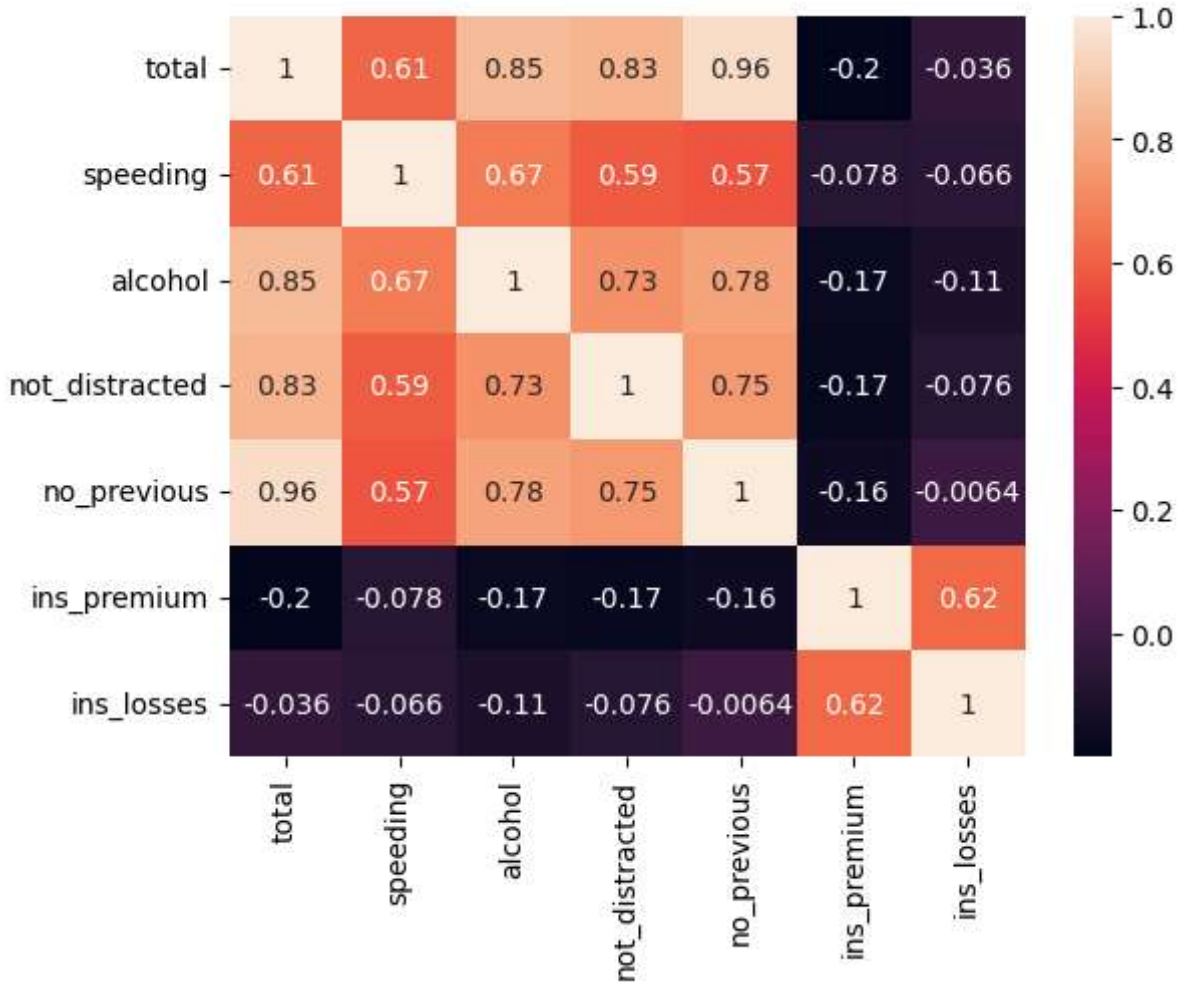


In [28]:
```
corr = df.corr()
```

```
<ipython-input-28-45893e33df67>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is de
precated. In a future version, it will default to False. Select only valid columns or specify the value of
numeric_only to silence this warning.
  corr = df.corr()
```

In [34]:
```
sns.heatmap(corr, annot=True)
```

Out[34]: <Axes: >



Inference

- The total number of accidents is mostly correlated with the no_previous
- The total number of accidents is highly dependent or correlated with the alcohol
- The total number of accidents is not correlated with the ins_premium and ins_losses
- ins_premium and ins_losses are correlated with each other