MUNI ASWANTH PRASAD A - 21BCE8854

Data Visualisation Techniques Exercise

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

In [11]:
```
sns.get_dataset_names()
```

Out[11]:
```
['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 [4]:
```
data = sns.load_dataset('car_crashes')
```

In [5]:
```
data.head()
```

Out[5]:

	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 [12]:
```
data['not_distracted'].value_counts()
```

Out[12]:
```
14.094    2
18.048    1
17.472    1
13.965    1
10.092    1
9.632     1
12.328    1
10.824    1
15.792    1
23.661    1
13.959    1
18.308    1
8.576     1
10.212    1
17.976    1
22.944    1
19.012    1
15.990    1
17.654    1
9.944     1
13.056    1
11.049    1
8.692     1
23.086    1
13.857    1
14.812    1
16.290    1
1.760     1
15.624    1
21.056    1
10.920    1
10.744    1
9.396     1
5.900     1
16.468    1
14.820    1
14.350    1
13.005    1
12.032    1
13.775    1
15.229    1
13.706    1
16.692    1
14.965    1
13.137    1
8.875     1
7.134     1
13.395    1
8.448     1
5.382     1
Name: not_distracted, dtype: int64
```

In [6]:
```
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
```

Data Visualisation Techniques

In [16]:
```
print(np.array(data).shape)
print(len(data.columns))
print(len(data.index))
```

```
(51, 8)
8
51
```

In [17]:
```
#Checking for Missing Values
np.sum(data.isnull())
```

Out[17]:
```
total             0
speeding          0
alcohol           0
not_distracted    0
no_previous       0
ins_premium       0
ins_losses        0
abbrev            0
dtype: int64
```

In [20]:
```
data.corr()
```

```
C:\Users\ayyam\AppData\Local\Temp\ipykernel_31576\2627137660.py:1: FutureWarning: The default value of numeric_only in DataFram
e.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.
  data.corr()
```

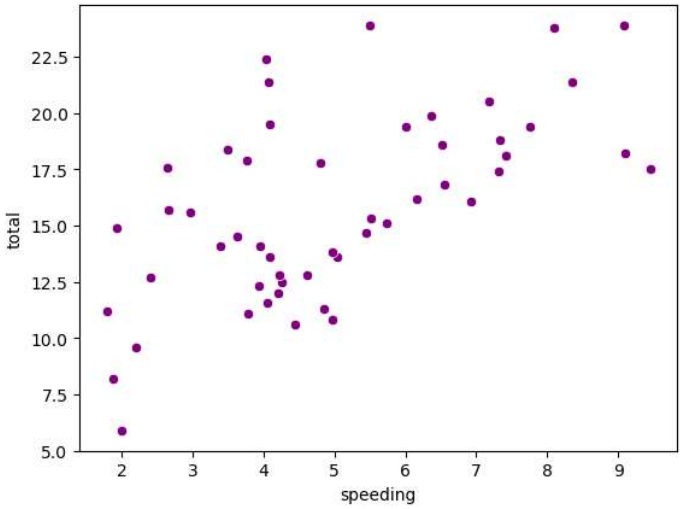
Out[20]:

	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses
total	1.000000	0.611548	0.852613	0.827560	0.956179	-0.199702	-0.036011
speeding	0.611548	1.000000	0.669719	0.588010	0.571976	-0.077675	-0.065928
alcohol	0.852613	0.669719	1.000000	0.732816	0.783520	-0.170612	-0.112547
not_distracted	0.827560	0.588010	0.732816	1.000000	0.747307	-0.174856	-0.075970
no_previous	0.956179	0.571976	0.783520	0.747307	1.000000	-0.156895	-0.006359
ins_premium	-0.199702	-0.077675	-0.170612	-0.174856	-0.156895	1.000000	0.623116
ins_losses	-0.036011	-0.065928	-0.112547	-0.075970	-0.006359	0.623116	1.000000

SCATRTERPLOT

In [21]:
```
sns.scatterplot(y="total",x="speeding",data=data, color="purple")
```

Out[21]:
```
<Axes: xlabel='speeding', ylabel='total'>
```

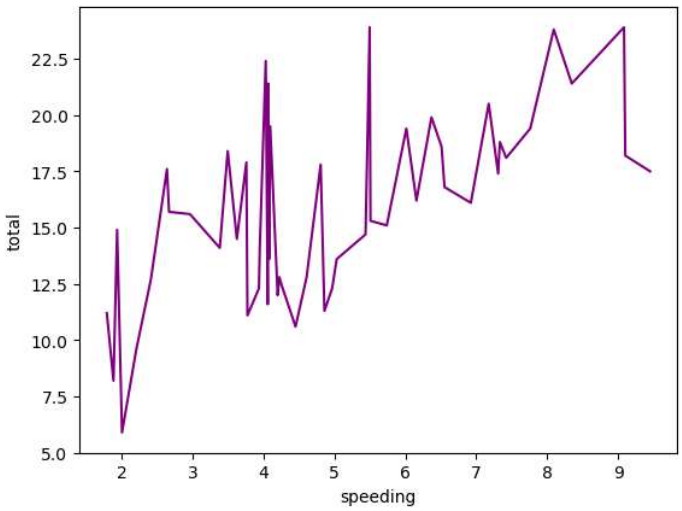


From the above Scatterplot the speed increaces and the level of alcohol increases the number of car crashes also increased

LINEPLOT

In [23]:
```
sns.lineplot(x="speeding",y="total",data=data,errorbar=None,color="purple")
```

Out[23]:
```
<Axes: xlabel='speeding', ylabel='total'>
```

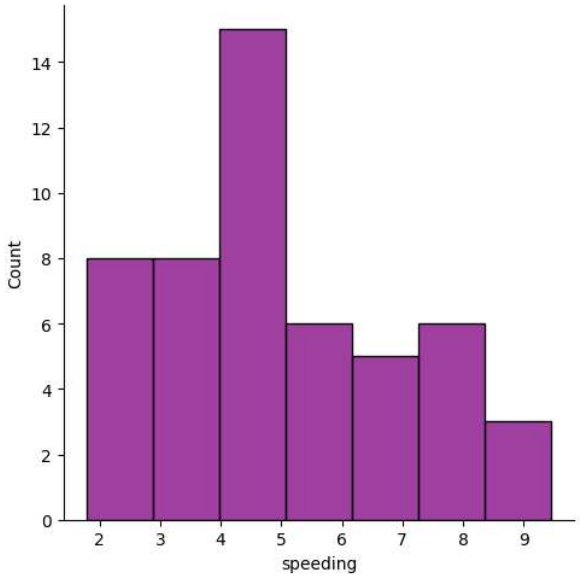


From the above lineplot as the speed increaces the number of car crashes also increased

DISPLOT

In [25]:
```
sns.displot(data['speeding'], color ='purple')
```

Out[25]:
```
<seaborn.axisgrid.FacetGrid at 0x192c640fed0>
```



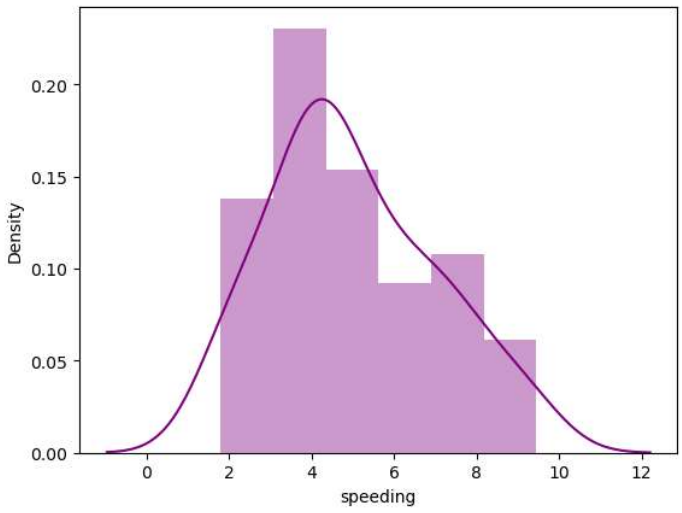
From the above Displot the speed increaces the number of car crashes also increased

Distribution Plot

In [28]:
```
sns.distplot(data['speeding'], color='purple')
```

```
C:\Users\ayyam\AppData\Local\Temp\ipykernel_31576\2105138372.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(data['speeding'], color='purple')
```

Out[28]:
```
<Axes: xlabel='speeding', ylabel='Density'>
```



In [35]:
```
print("mean")
data['speeding'].mean()
```

```
mean
```

Out[35]:
```
4.998196078431373
```

In [36]:
```
print("median")
data['speeding'].median()
```

```
median
```

Out[36]:
```
4.6080000000000005
```

In [37]:
```
print("mode")
data['speeding'].mode()
```

mode

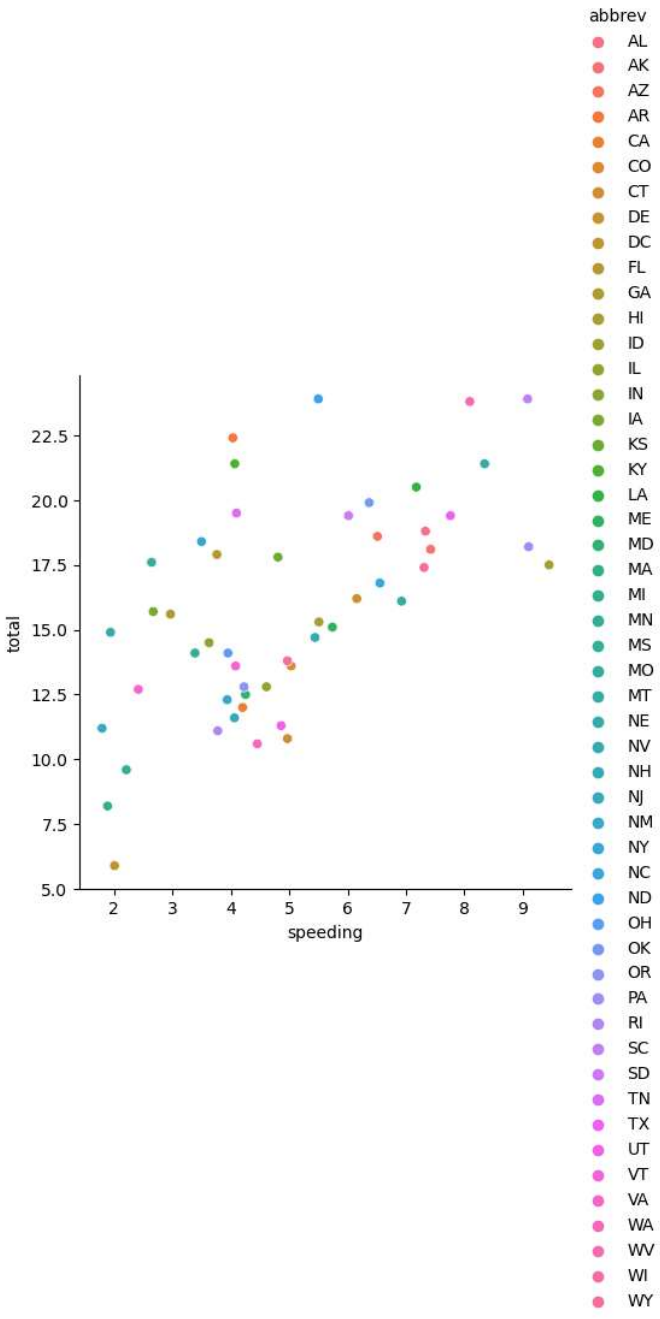
Out[37]:
```
0    4.968
Name: speeding, dtype: float64
```

From the above Distribution plot the distribution is almost symmetric (skewed symmetric)

RELPLOT

In [43]:
```
sns.relplot(x='speeding',y='total',data=data,hue='abbrev')
```

Out[43]:
```
<seaborn.axisgrid.FacetGrid at 0x192c64d2010>
```

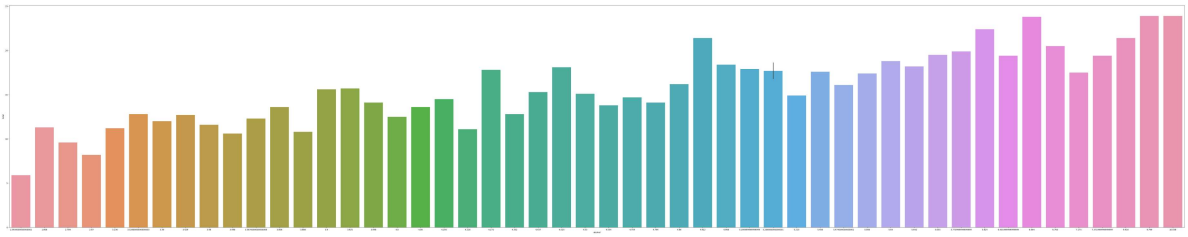


From the above Relplot, it is bases on states

BARPLOT

In [45]:
```
plt.subplots(figsize=(106,20))
sns.barplot(x='alcohol',y='total',data=data)
```

Out[45]:
```
<Axes: xlabel='alcohol', ylabel='total'>
```

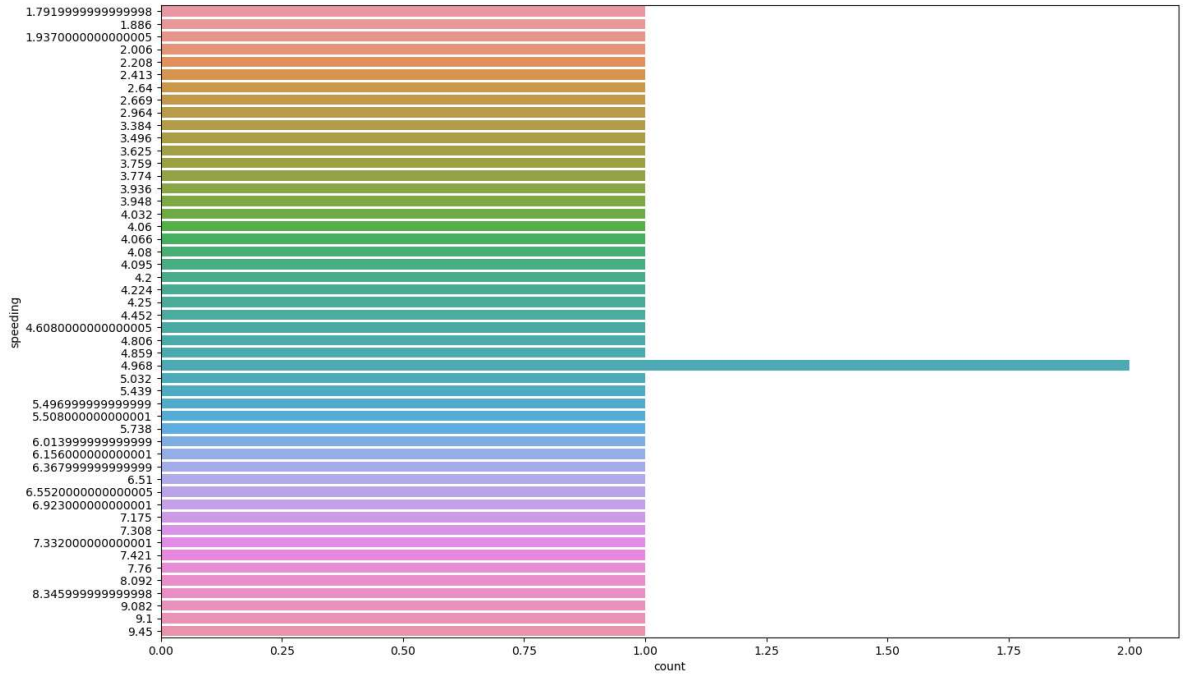


from the above Bargarph or Barplot it is observed that as the level of alcohol inccreased the total number of acidents also increased

COUNTPLOT

In [47]:
```
plt.subplots(figsize=(16,10))
sns.countplot(y='speeding',data=data,orient='h')
```

Out[47]:
```
<Axes: xlabel='count', ylabel='speeding'>
```



from the above Countplot there are more number of cases recorded when speed is at 4.968 units

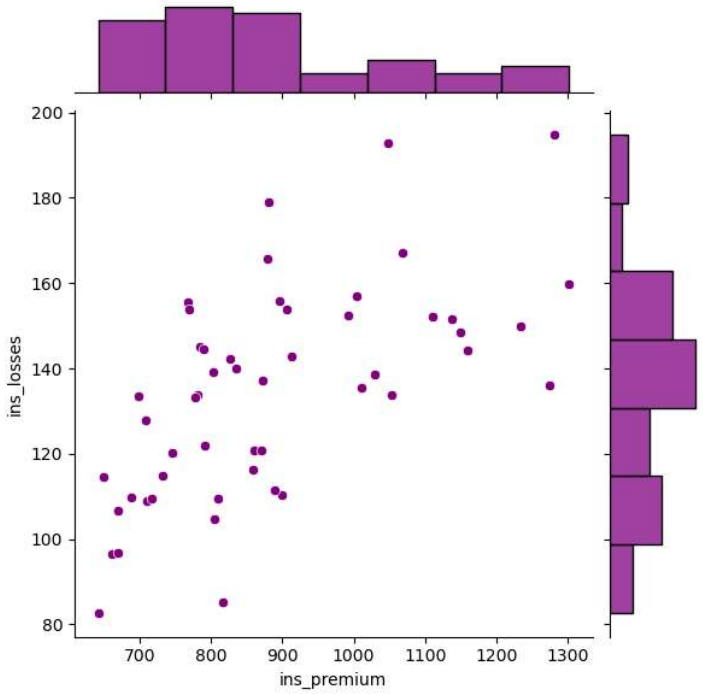
In [48]:
```
data['speeding'].mode()
```

Out[48]:
```
0    4.968
Name: speeding, dtype: float64
```

JOINTPLOT

In [50]:
```
sns.jointplot(x='ins_premium',y='ins_losses',data=data,color='purple')
```

Out[50]:
```
<seaborn.axisgrid.JointGrid at 0x192cbe48210>
```



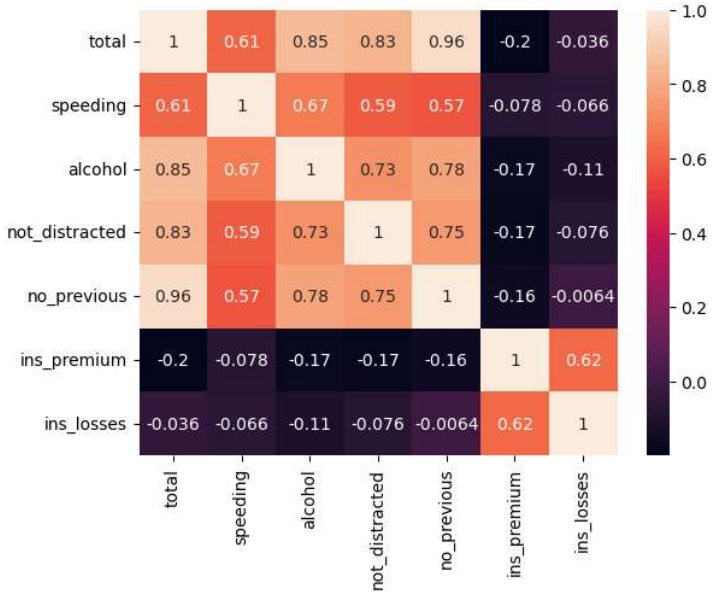
from the above Jointplot though the insurance premium was taken the insurance losses is high

HEATMAP (CORRELATION MATRIX)

In [55]:
```
sns.heatmap(data.corr(),annot=True,)
```

```
C:\Users\ayyam\AppData\Local\Temp\ipykernel_31576\1682716602.py:1: FutureWarning: The default value of numeric_only in DataFram
e.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.
  sns.heatmap(data.corr(),annot=True,)
```

Out[55]:
```
<Axes: >
```

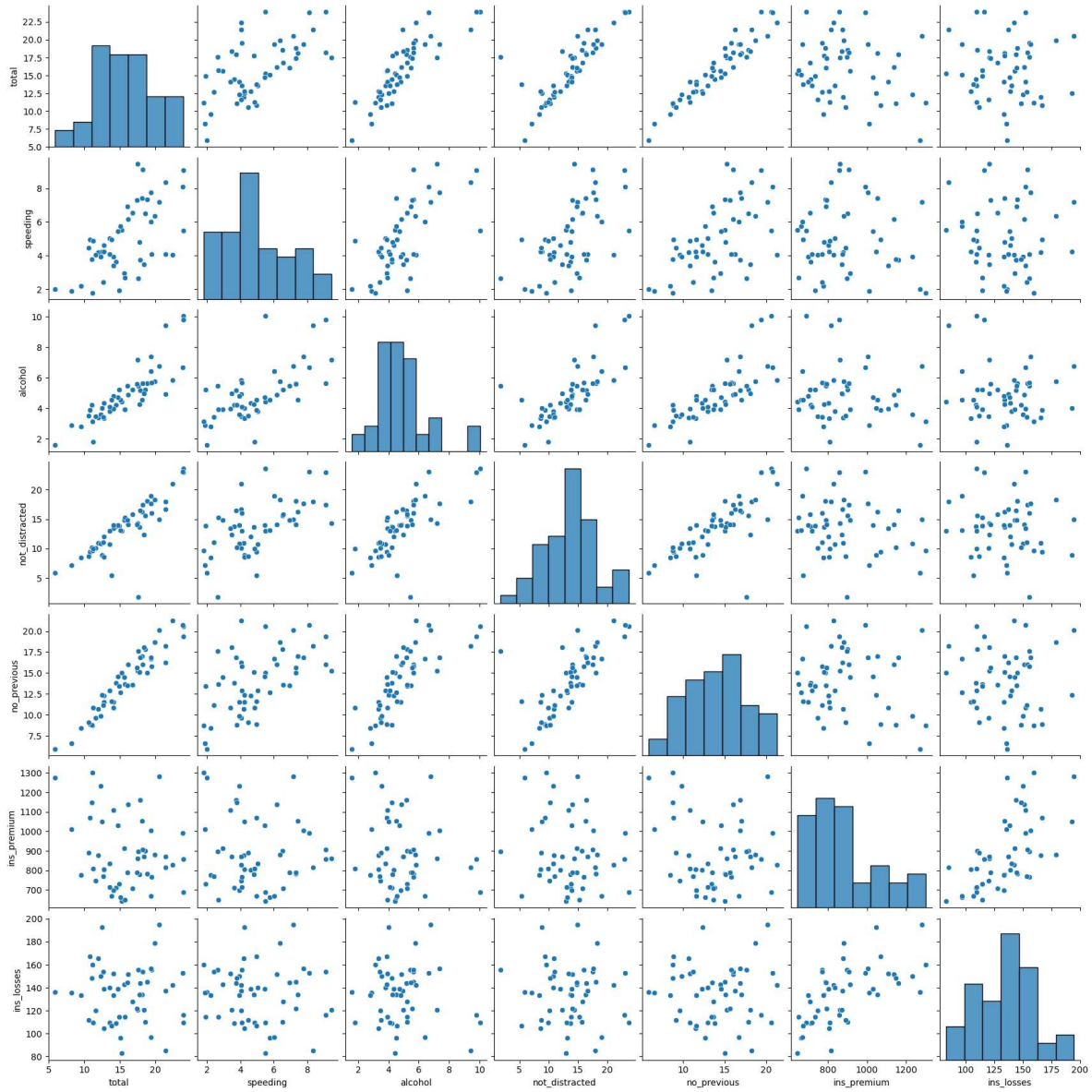


from the above Heatmap ,it is observed that the correlation between each and every attribute.

PAIRPLOT

In [64]:
```
sns.pairplot(data)
```

Out[64]:
```
<seaborn.axisgrid.PairGrid at 0x192db1572d0>
```

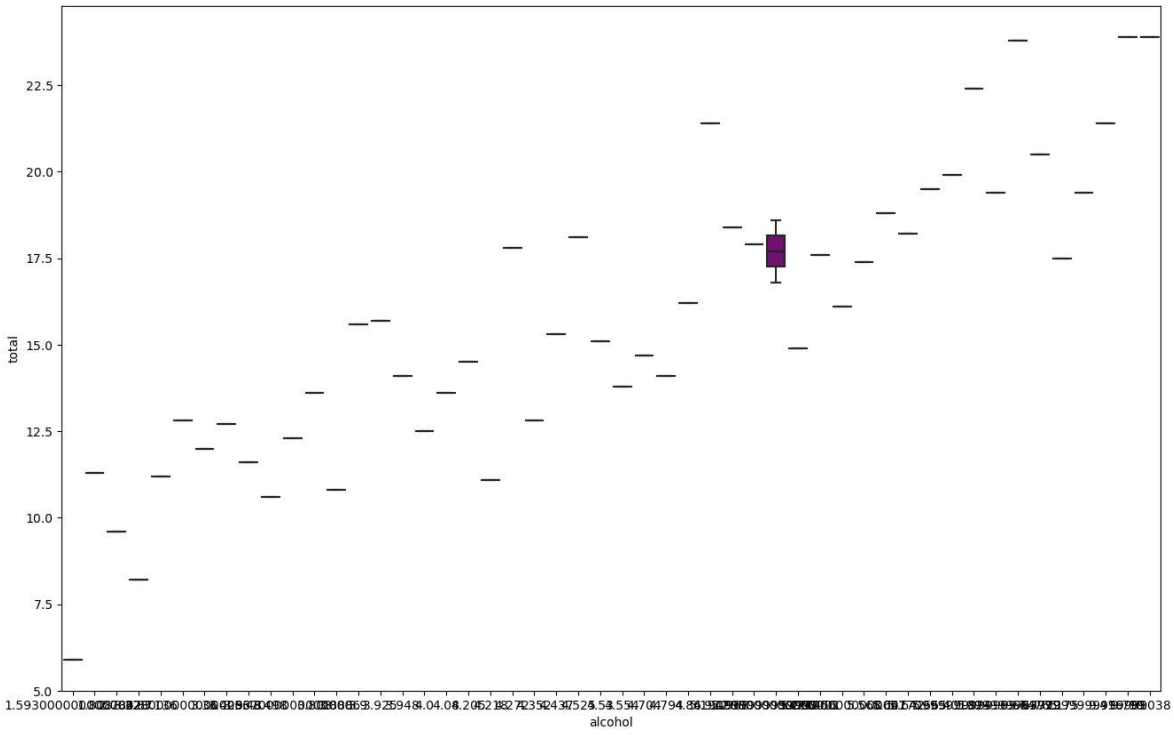


from the above Pairplot we can observe the trend between the attributes

BOXPLOT

In [67]:
```
plt.subplots(figsize=(16,10))
sns.boxplot(x='alcohol',y='total',data=data, color='purple')
```

Out[67]: <Axes: xlabel='alcohol', ylabel='total'>



In [ ]: