

**saibabu**

**21bce9232**

**vitap**

## **assignment-2**

```
In [4]: import seaborn as sns  
import pandas as pd  
import matplotlib.pyplot as plt  
ds=sns.load_dataset("car_crashes")  
ds
```

Out[4]:

	total	speeding	alcohol	not_distracted	no_previous	ins_prem
<b>0</b>	18.8	7.332	5.640	18.048	15.040	78
<b>1</b>	18.1	7.421	4.525	16.290	17.014	105
<b>2</b>	18.6	6.510	5.208	15.624	17.856	89
<b>3</b>	22.4	4.032	5.824	21.056	21.280	82
<b>4</b>	12.0	4.200	3.360	10.920	10.680	87
<b>5</b>	13.6	5.032	3.808	10.744	12.920	83
<b>6</b>	10.8	4.968	3.888	9.396	8.856	106
<b>7</b>	16.2	6.156	4.860	14.094	16.038	113
<b>8</b>	5.9	2.006	1.593	5.900	5.900	127
<b>9</b>	17.9	3.759	5.191	16.468	16.826	116
<b>10</b>	15.6	2.964	3.900	14.820	14.508	91
<b>11</b>	17.5	9.450	7.175	14.350	15.225	86
<b>12</b>	15.3	5.508	4.437	13.005	14.994	64
<b>13</b>	12.8	4.608	4.352	12.032	12.288	80
<b>14</b>	14.5	3.625	4.205	13.775	13.775	71
<b>15</b>	15.7	2.669	3.925	15.229	13.659	64
<b>16</b>	17.8	4.806	4.272	13.706	15.130	78
<b>17</b>	21.4	4.066	4.922	16.692	16.264	87
<b>18</b>	20.5	7.175	6.765	14.965	20.090	128
<b>19</b>	15.1	5.738	4.530	13.137	12.684	66
<b>20</b>	12.5	4.250	4.000	8.875	12.375	104
<b>21</b>	8.2	1.886	2.870	7.134	6.560	101
<b>22</b>	14.1	3.384	3.948	13.395	10.857	111
<b>23</b>	9.6	2.208	2.784	8.448	8.448	77

	<b>total</b>	<b>speeding</b>	<b>alcohol</b>	<b>not_distracted</b>	<b>no_previous</b>	<b>ins_prem</b>
<b>24</b>	17.6	2.640	5.456	1.760	17.600	89
<b>25</b>	16.1	6.923	5.474	14.812	13.524	79
<b>26</b>	21.4	8.346	9.416	17.976	18.190	81
<b>27</b>	14.9	1.937	5.215	13.857	13.410	73
<b>28</b>	14.7	5.439	4.704	13.965	14.553	102
<b>29</b>	11.6	4.060	3.480	10.092	9.628	74
<b>30</b>	11.2	1.792	3.136	9.632	8.736	130
<b>31</b>	18.4	3.496	4.968	12.328	18.032	86
<b>32</b>	12.3	3.936	3.567	10.824	9.840	123
<b>33</b>	16.8	6.552	5.208	15.792	13.608	70
<b>34</b>	23.9	5.497	10.038	23.661	20.554	68
<b>35</b>	14.1	3.948	4.794	13.959	11.562	69
<b>36</b>	19.9	6.368	5.771	18.308	18.706	88
<b>37</b>	12.8	4.224	3.328	8.576	11.520	80
<b>38</b>	18.2	9.100	5.642	17.472	16.016	90
<b>39</b>	11.1	3.774	4.218	10.212	8.769	114
<b>40</b>	23.9	9.082	9.799	22.944	19.359	85
<b>41</b>	19.4	6.014	6.402	19.012	16.684	66
<b>42</b>	19.5	4.095	5.655	15.990	15.795	76
<b>43</b>	19.4	7.760	7.372	17.654	16.878	100
<b>44</b>	11.3	4.859	1.808	9.944	10.848	80
<b>45</b>	13.6	4.080	4.080	13.056	12.920	71
<b>46</b>	12.7	2.413	3.429	11.049	11.176	76
<b>47</b>	10.6	4.452	3.498	8.692	9.116	89

	total	speeding	alcohol	not_distracted	no_previous	ins_prem
<b>48</b>	23.8	8.092	6.664		23.086	20.706
<b>49</b>	13.8	4.968	4.554		5.382	11.592
<b>50</b>	17.4	7.308	5.568		14.094	15.660

## Inference: Load the data set of car\_crashes from sea born library

In [ ]:

```
[ ]:
```

In [ ]:

```
[ ]:
```

In [ ]:

```
[ ]:
```

In [ ]: `ds.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
```

# Inference: get the information of data set

In [ ]:

In [ ]:

In [ ]:

In [ ]: `ds.head()`

Out[ ]:

	total	speeding	alcohol	not_distracted	no_previous	ins_premi
0	18.8	7.332	5.640	18.048	15.040	784
1	18.1	7.421	4.525	16.290	17.014	1053
2	18.6	6.510	5.208	15.624	17.856	899
3	22.4	4.032	5.824	21.056	21.280	827
4	12.0	4.200	3.360	10.920	10.680	878



**Inference:** get the data set of first five rows of car\_crashes from sea born library

In [ ]:

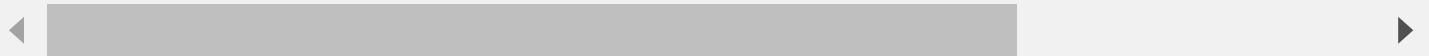
In [ ]:

In [ ]:

In [ ]: `ds.tail()`

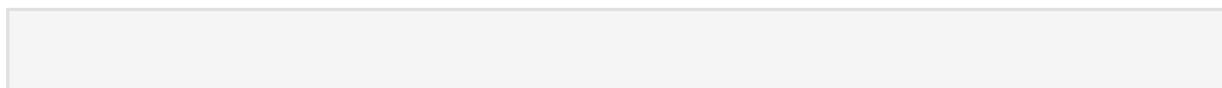
Out[ ]:

	total	speeding	alcohol	not_distracted	no_previous	ins_prem
46	12.7	2.413	3.429	11.049	11.176	76
47	10.6	4.452	3.498	8.692	9.116	89
48	23.8	8.092	6.664	23.086	20.706	99
49	13.8	4.968	4.554	5.382	11.592	67
50	17.4	7.308	5.568	14.094	15.660	79

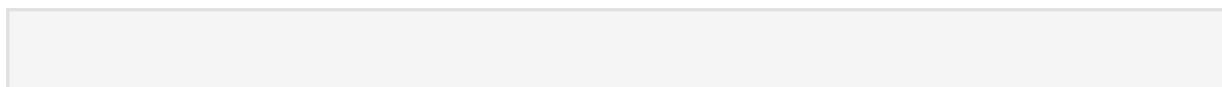


**Inference:** get the data set of lastt five rows of car\_crashes from sea born library

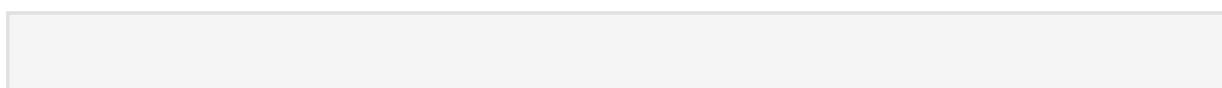
In [ ]:



In [ ]:

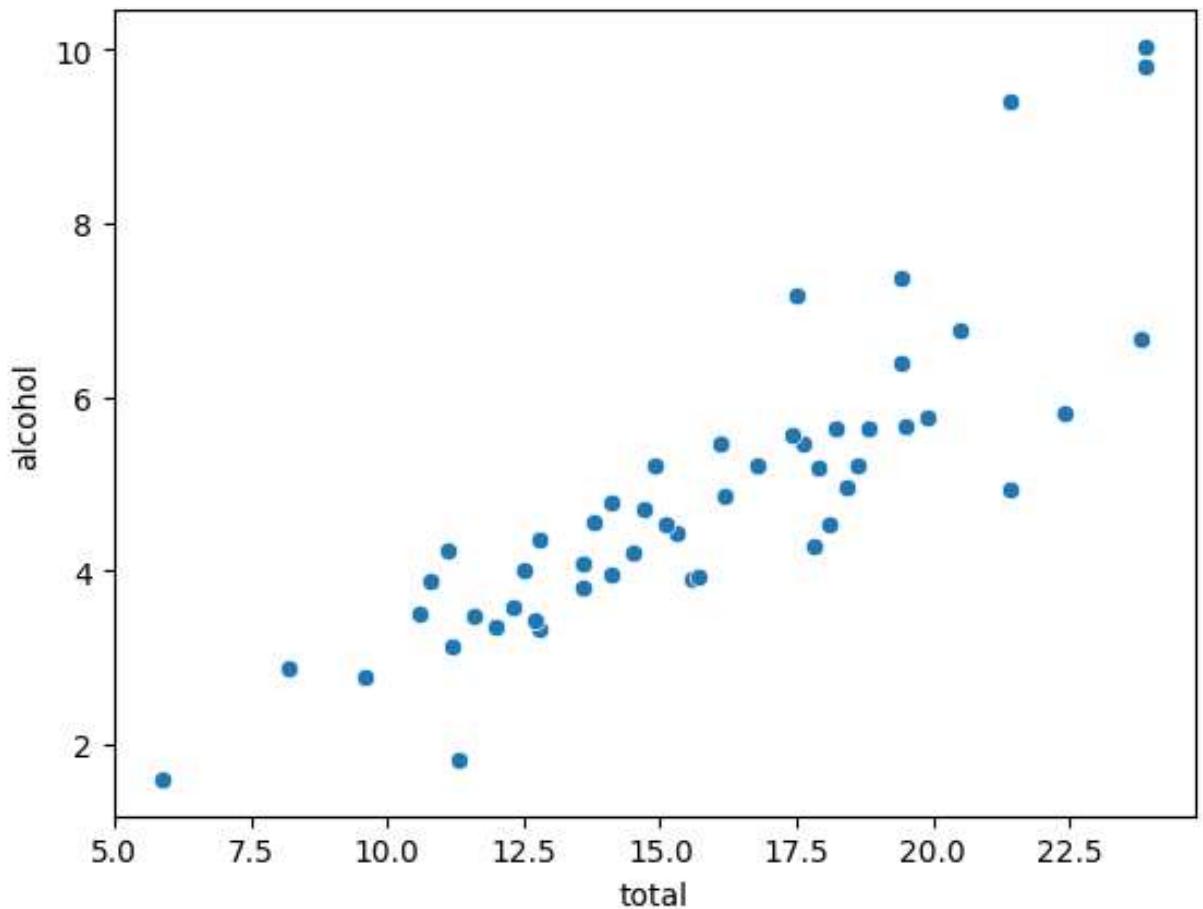


In [ ]:



In [ ]: sns.scatterplot(y="alcohol",x="total",data=ds)

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



**Inference:** from the plot we can say that as total increases alcohol is also increasing

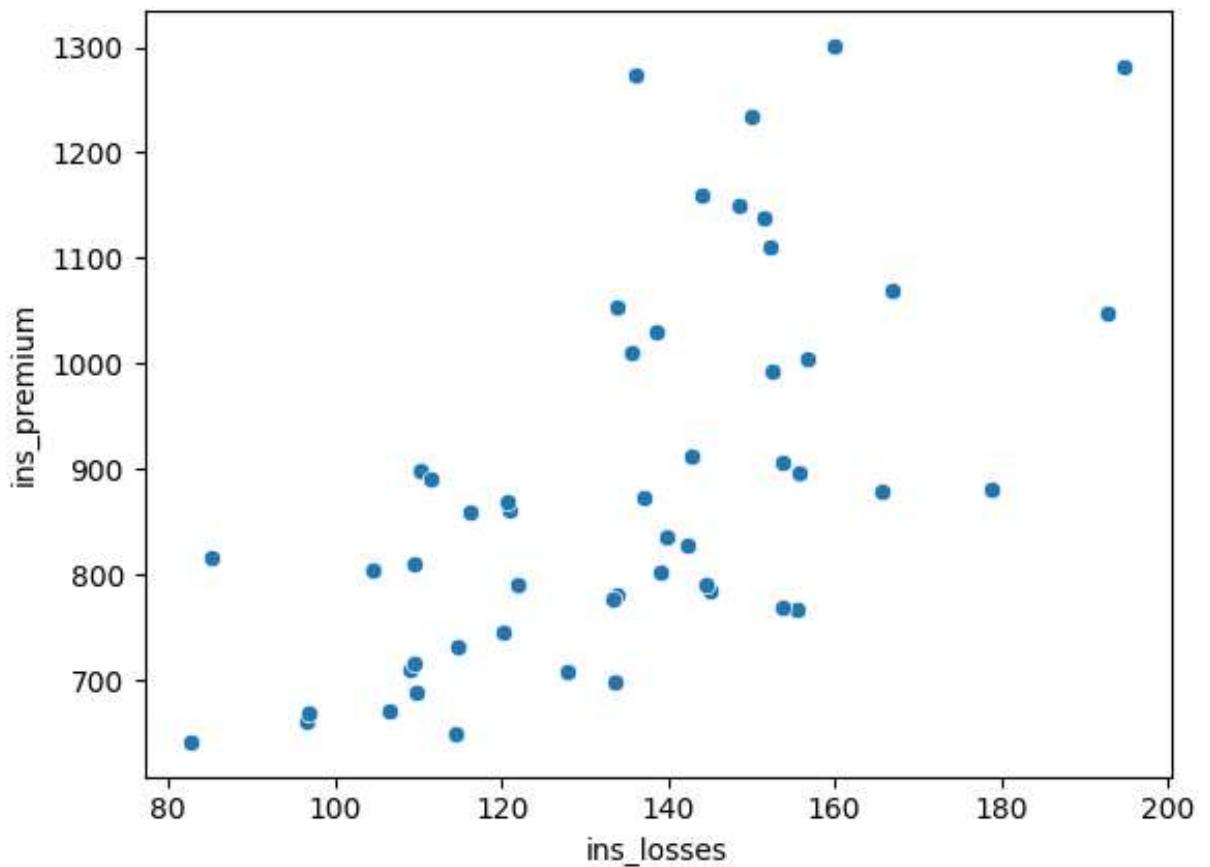
In [ ]:

In [ ]:

In [ ]:

In [ ]: sns.scatterplot(y="ins\_premium",x="ins\_losses",data=ds)

Out[ ]: <Axes: xlabel='ins\_losses', ylabel='ins\_premium'>



**Inference:**from the plot we can say that as ins\_losses increases ins\_premium is also increasing

In [ ]:

In [ ]:

In [ ]:

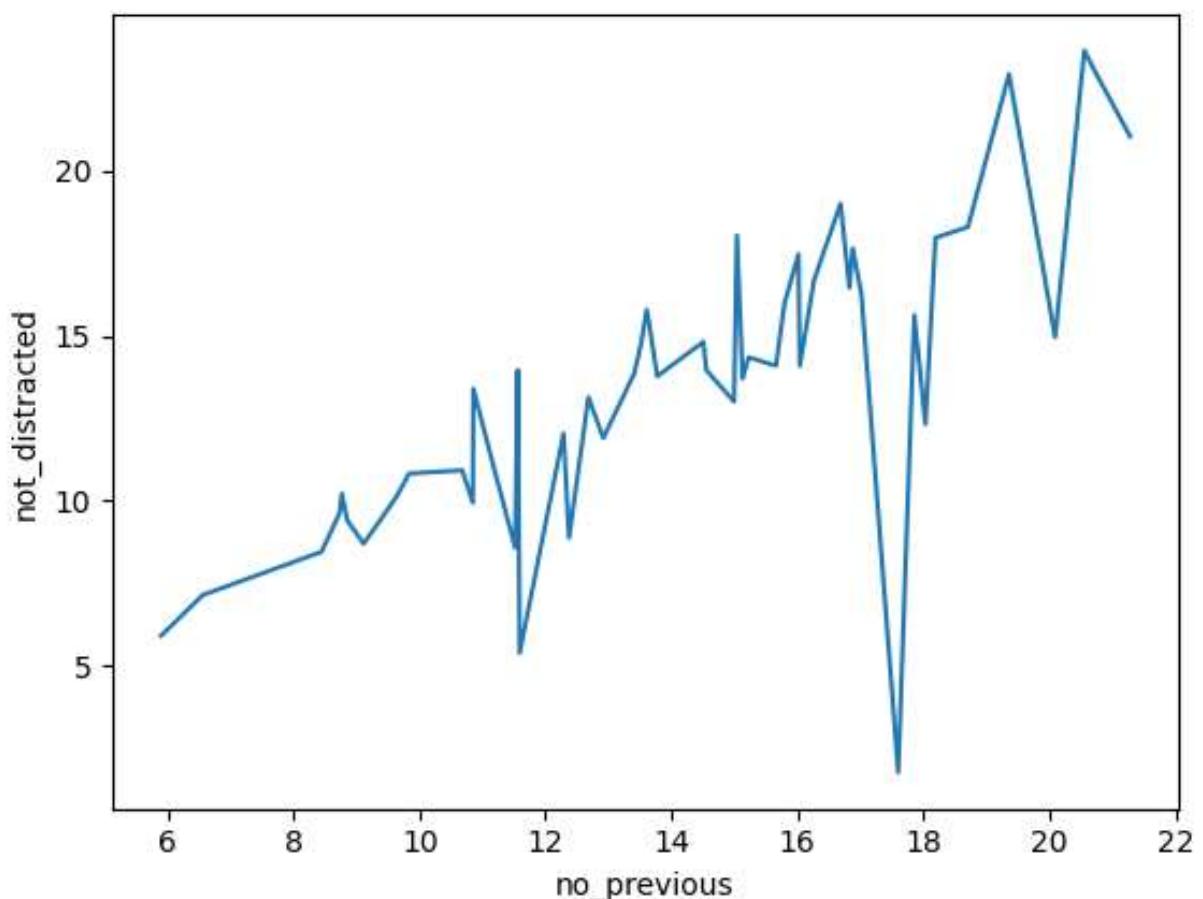
In [ ]: sns.lineplot(x="no\_previous",y="not\_distracted",data=ds,

```
<ipython-input-16-df72595b4fca>:1: FutureWarning:
```

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

```
    sns.lineplot(x="no_previous",y="not_distracted",data=d  
s, ci=None)
```

```
Out[ ]: <Axes: xlabel='no_previous', ylabel='not_distracted'>
```



**Inference:**from the plot we can say that as no\_previous increases not\_distrubuted is also increasing some point decreases and increase at a time

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]: sns.distplot(ds["total"])
```

```
<ipython-input-18-ecf0e2ee2d17>:1: UserWarning:
```

```
`distplot` is a deprecated function and will be removed  
in seaborn v0.14.0.
```

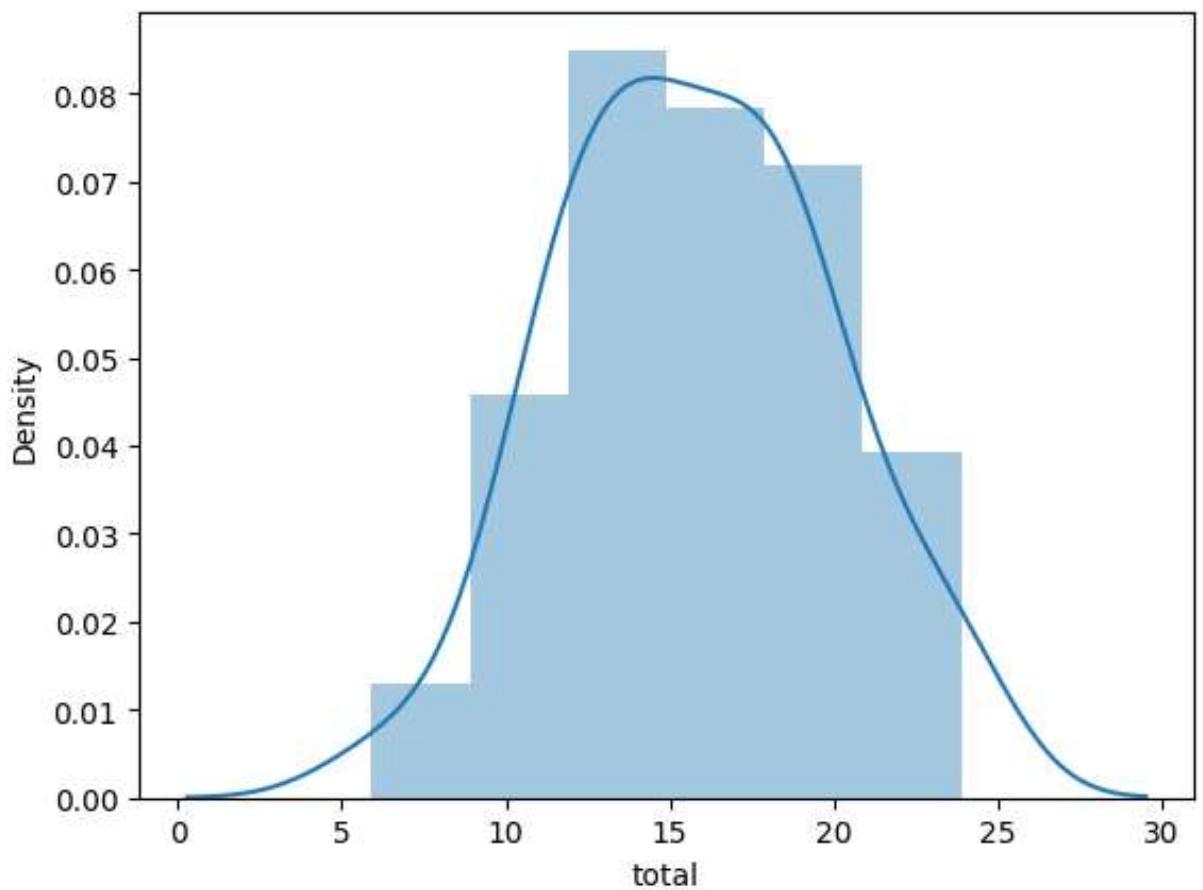
```
Please adapt your code to use either `displot` (a figure  
-level function with  
similar flexibility) or `histplot` (an axes-level functi  
on for histograms).
```

```
For a guide to updating your code to use the new functio  
ns, please see
```

```
https://gist.github.com/mwaskom/de44147ed2974457ad637275  
0bbe5751
```

```
sns.distplot(ds["total"])
```

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



**Inference:**from the plot we can say that as more density occurs in the range of 10-20 in total (xlabel)

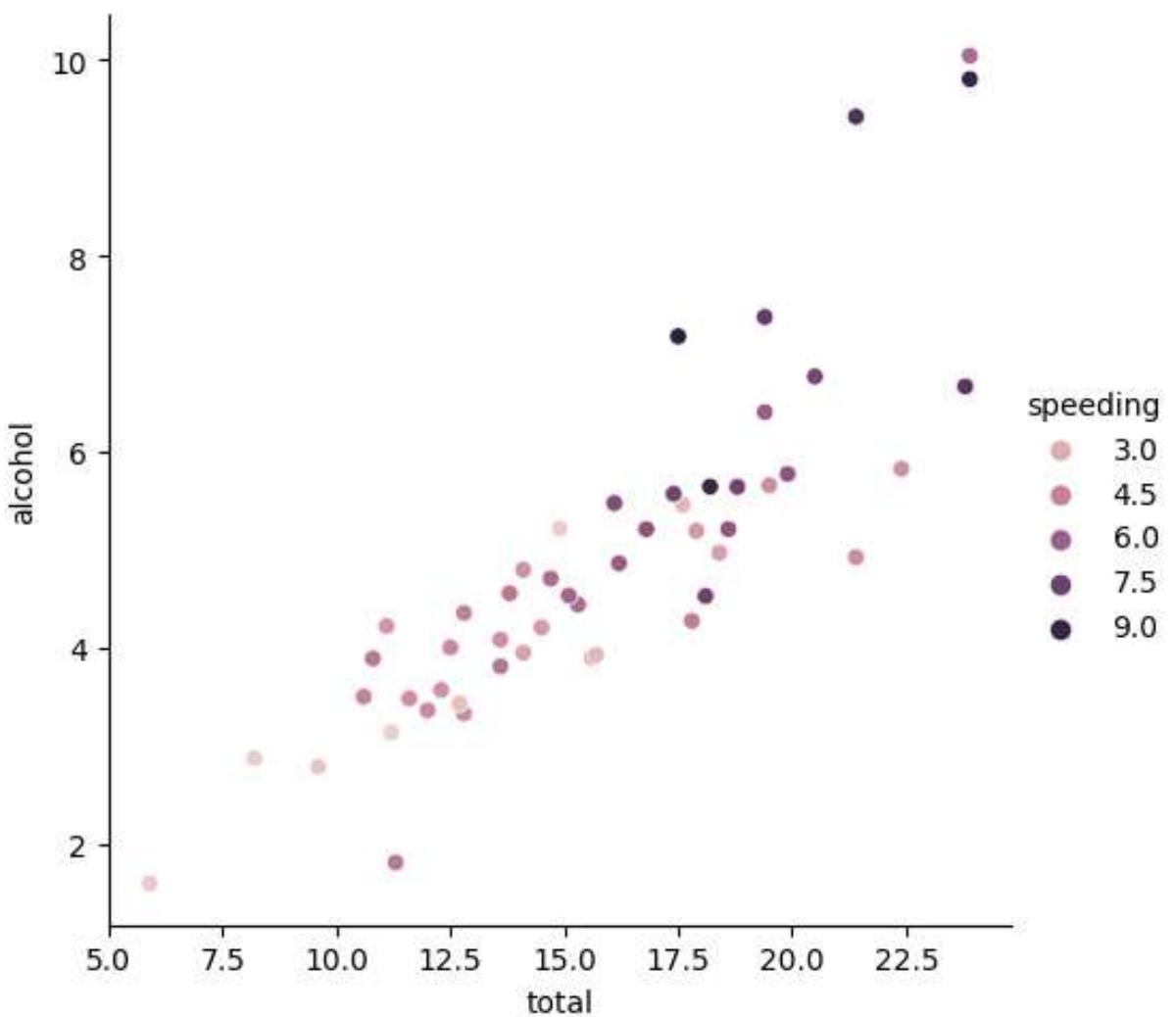
In [ ]:

In [ ]:

In [ ]:

In [ ]: `sns.relplot(x="total",y="alcohol",data=ds,hue="speeding")`

Out[ ]: <seaborn.axisgrid.FacetGrid at 0x7eaa08b735e0>



**Inference:**from the plot we can say that as total increases alcohol is also increasing with increasing in speeding

In [ ]:

In [ ]:

In [ ]:

In [ ]: `ds["alcohol"].value_counts()`

```
Out[ ]: 5.208      2
         5.640      1
         4.218      1
         4.704      1
         3.480      1
         3.136      1
         4.968      1
         3.567      1
         10.038     1
         4.794      1
         5.771      1
         3.328      1
         5.642      1
         9.799      1
         9.416      1
         6.402      1
         5.655      1
         7.372      1
         1.808      1
         4.080      1
         3.429      1
         3.498      1
         6.664      1
         4.554      1
         5.215      1
         5.474      1
         4.525      1
         5.456      1
         5.824      1
         3.360      1
         3.808      1
         3.888      1
         4.860      1
         1.593      1
         5.191      1
         3.900      1
         7.175      1
         4.437      1
         4.352      1
         4.205      1
         3.925      1
         4.272      1
         4.922      1
         6.765      1
```

```
4.530      1  
4.000      1  
2.870      1  
3.948      1  
2.784      1  
5.568      1  
Name: alcohol, dtype: int64
```

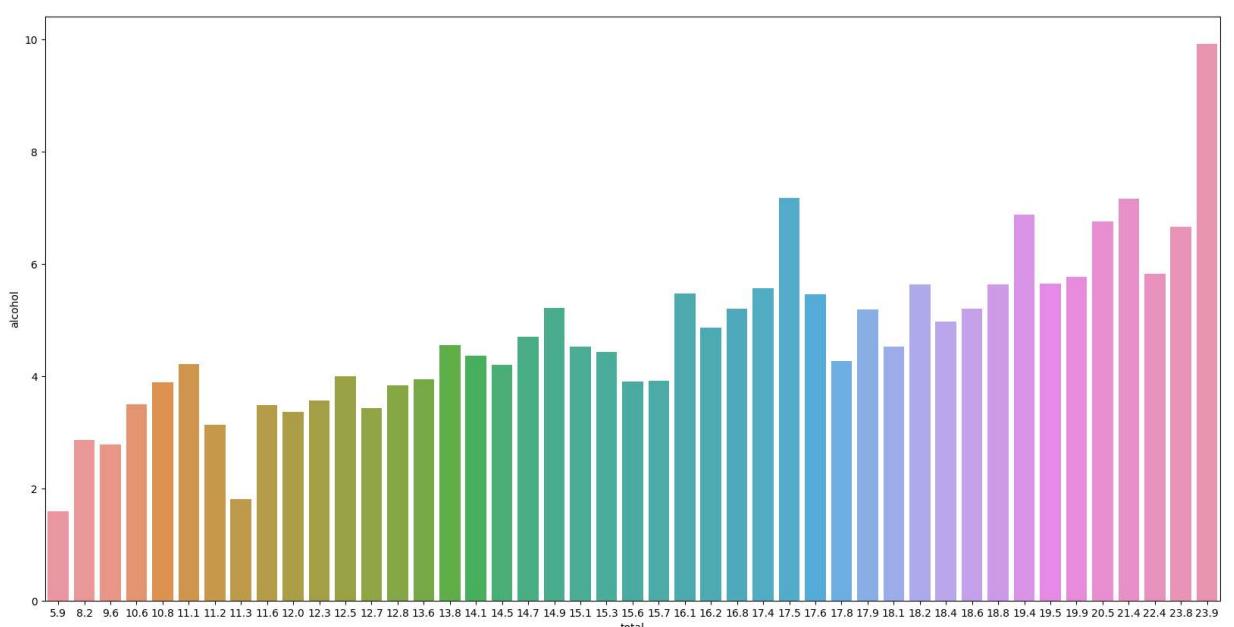
```
In [ ]: plt.subplots(figsize=(20,10))  
sns.barplot(data=ds,x="total",y="alcohol",ci=None)
```

```
<ipython-input-32-0389d71bee30>:2: FutureWarning:
```

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

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

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

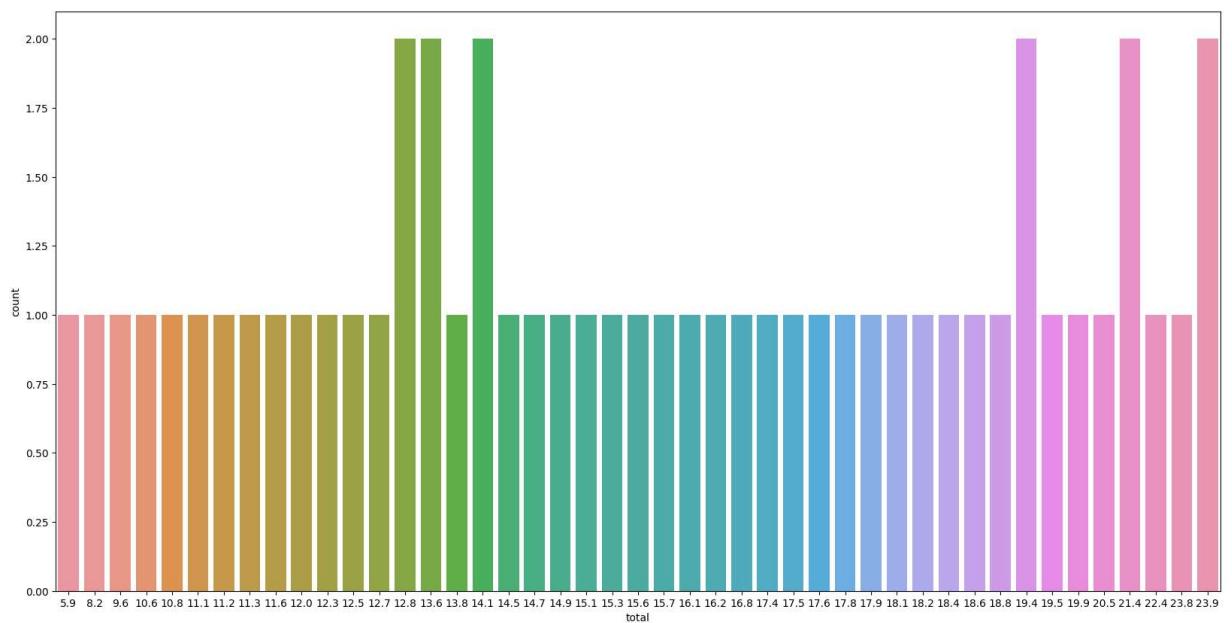


**Inference:**from the plot we can say that as total increases alcohol is also increasing

```
In [ ]:
```

```
In [ ]: plt.subplots(figsize=(20,10))  
sns.countplot(x="total", data=ds)
```

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



**Inference:**from the plot we can say that as total count of each occurrence is almost 1 time except 6 values occurred twice

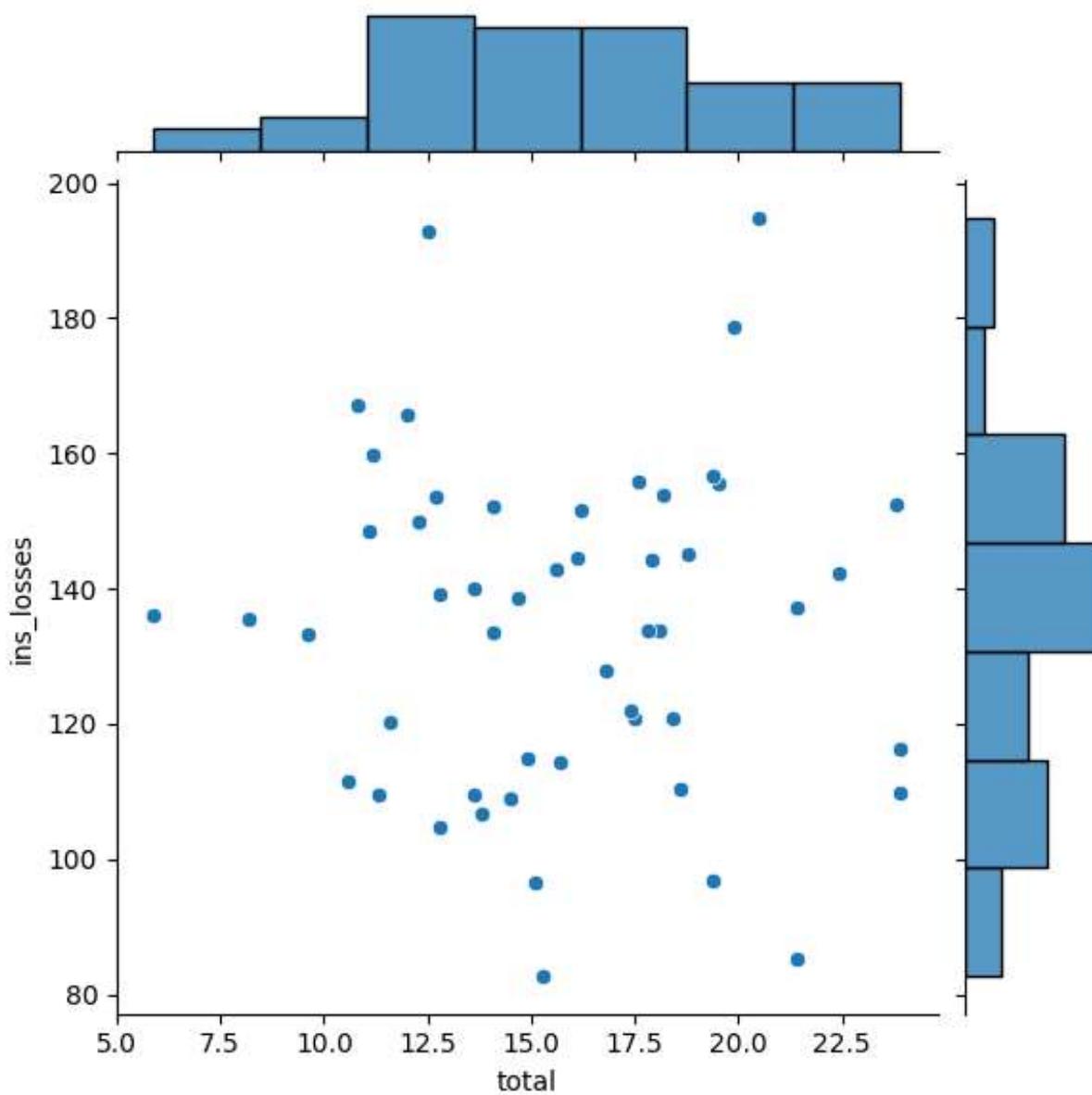
```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [3]: sns.jointplot(x="total",y="ins_losses",data=ds)
```

```
Out[3]: <seaborn.axisgrid.JointGrid at 0x7a85d19427a0>
```



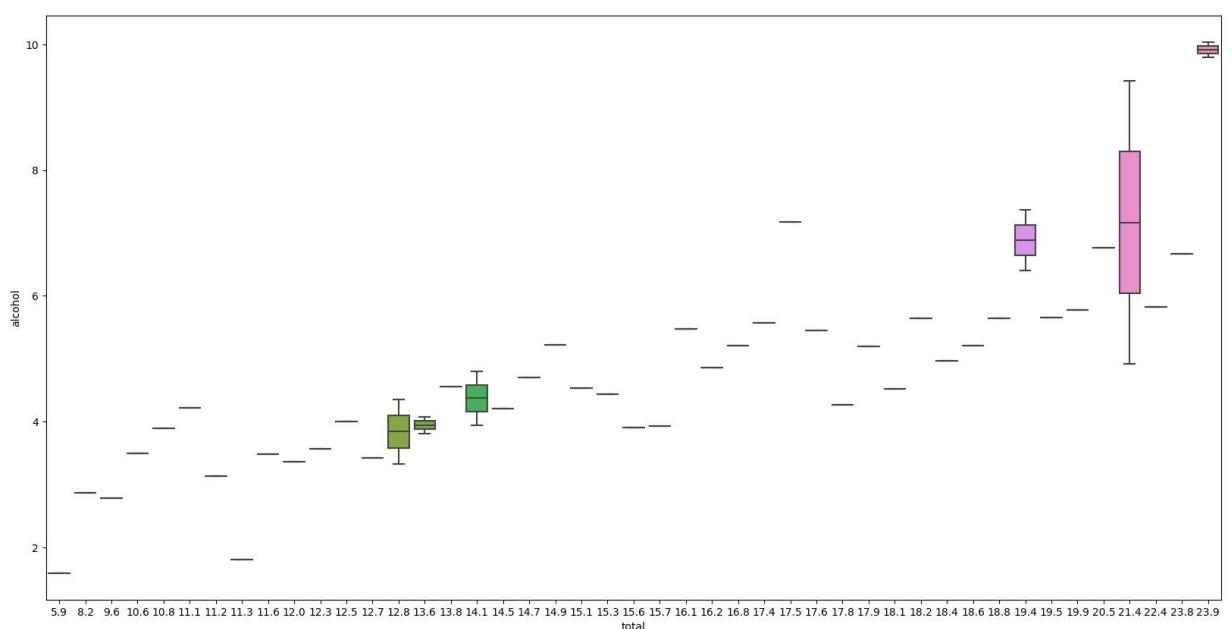
**Inference:**from the plot we can say that as ins\_losses increases total is also increasing get more frequency at middle values of both columns(total and ins\_losses)

In [ ]:

In [ ]:

In [ ]: `plt.subplots(figsize=(20,10))  
sns.boxplot(x="total",y="alcohol",data=ds)`

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



**Inference:**from the plot we can say that almost all alcohol

**medium values in the range3 to 6 ,maximum is 7,minimum is 1**

In [ ]:

In [ ]:

In [ ]: corr=ds.corr()  
corr

```
<ipython-input-50-5fb269adec>: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 columns or specify the value of numeric_only to silence this warning.  
corr=ds.corr()
```

Out[ ]:

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

**Inference:tabel represents correlation between every**

# attributes

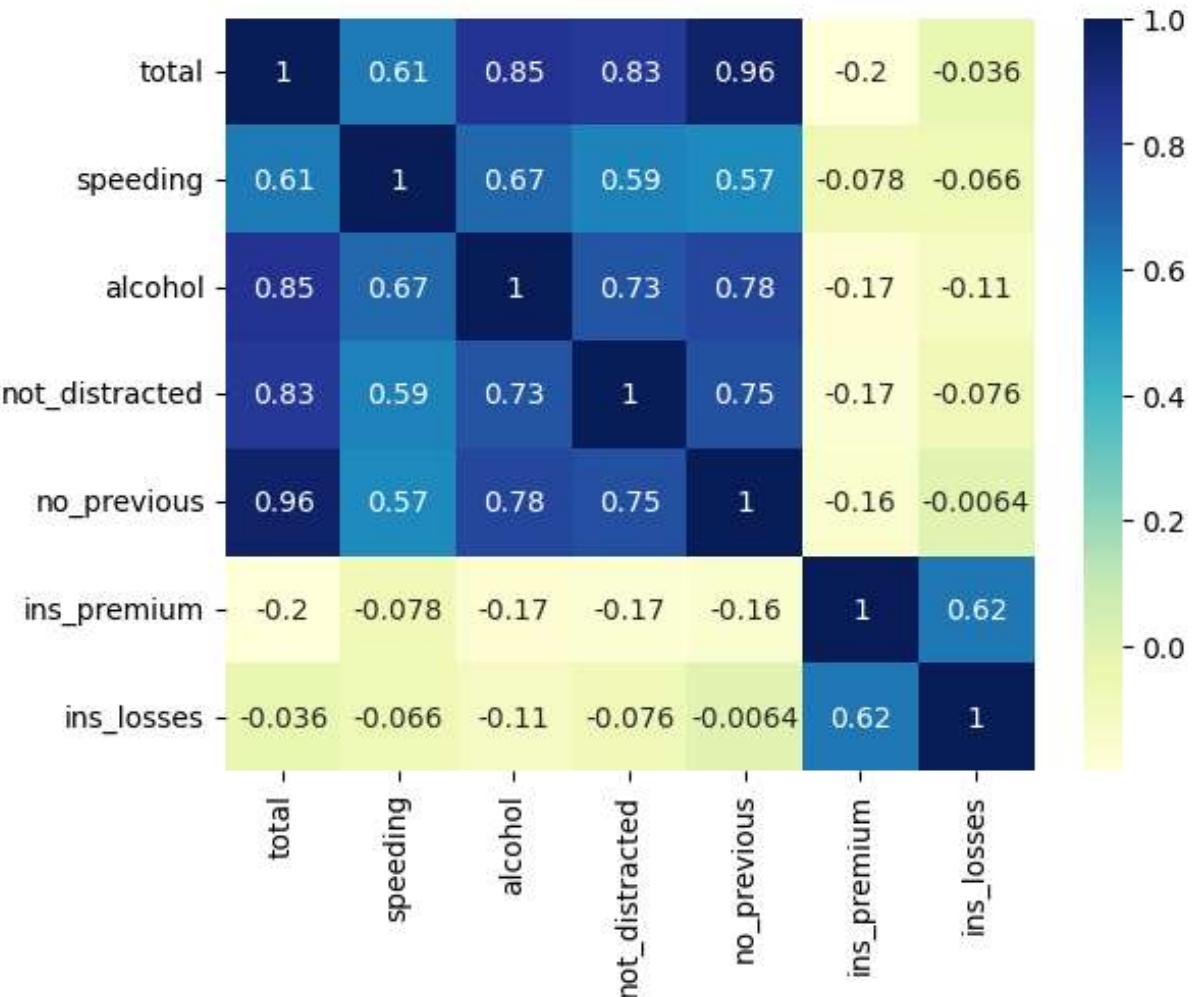
```
In [ ]: 
```

```
In [ ]: 
```

```
In [ ]: 
```

```
In [ ]: sns.heatmap(corr, annot=True, cmap="YlGnBu") 
```

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



**Inference:**from the plot we can say that as no\_previous and alcohol are more

**corelated(positive co  
relation)**