

Seaborn car_crashes dataset

```
In [1]: !pip install seaborn
```

```
Requirement already satisfied: seaborn in c:\users\admin\anaconda3\lib\site-packages (0.12.2)
Requirement already satisfied: numpy!=1.24.0,>=1.17 in c:\users\admin\anaconda3\lib\site-packages (from seaborn) (1.24.3)
Requirement already satisfied: pandas>=0.25 in c:\users\admin\anaconda3\lib\site-packages (from seaborn) (1.5.3)
Requirement already satisfied: matplotlib!=3.6.1,>=3.1 in c:\users\admin\anaconda3\lib\site-packages (from seaborn) (3.7.1)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\admin\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.0.5)
Requirement already satisfied: cycler>=0.10 in c:\users\admin\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\admin\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (4.25.0)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\admin\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\users\admin\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (23.0)
Requirement already satisfied: pillow>=6.2.0 in c:\users\admin\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in c:\users\admin\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (3.0.9)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\admin\anaconda3\lib\site-packages (from matplotlib!=3.6.1,>=3.1->seaborn) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\admin\anaconda3\lib\site-packages (from pandas>=0.25->seaborn) (2022.7)
Requirement already satisfied: six>=1.5 in c:\users\admin\anaconda3\lib\site-packages (from python-dateutil>=2.7->matplotlib!=3.6.1,>=3.1->seaborn) (1.16.0)
```

```
In [2]: import seaborn as sns
```

```
In [6]: 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', 'seice', 'taxis', 'tips', 'titanic']
```

```
In [7]: df=sns.load_dataset('car_crashes')
df
```

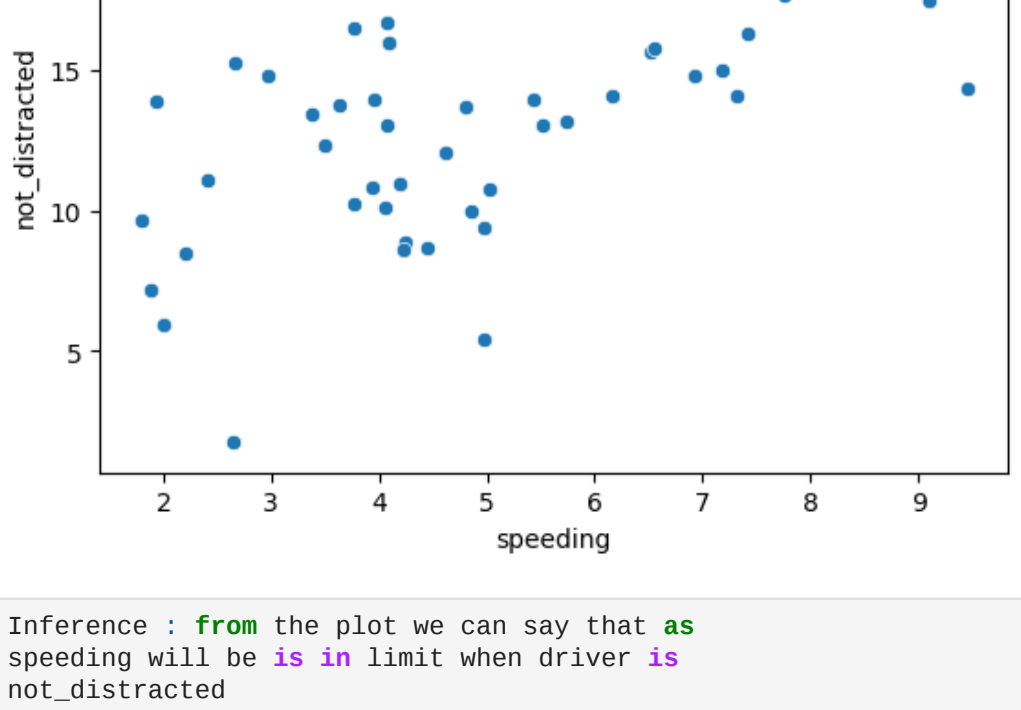
	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses
0	18.8	7.332	5.640	18.048	15.040	784.55	145.08
1	18.1	7.421	4.525	16.290	17.014	1053.48	133.93
2	18.6	6.510	5.208	15.624	17.856	899.47	110.35
3	22.4	4.032	5.824	21.056	21.280	827.34	142.39
4	12.0	4.200	3.360	10.920	10.680	878.41	165.63
5	13.6	5.032	3.808	10.744	12.920	835.50	139.91
6	10.8	4.968	3.888	9.396	8.856	1068.73	167.02
7	16.2	6.156	4.860	14.094	16.038	1137.87	151.48
8	5.9	2.006	1.593	5.900	5.900	1273.89	136.05
9	17.9	3.759	5.191	16.468	16.826	1160.13	144.18
10	15.6	2.964	3.900	14.820	14.508	913.15	142.80
11	17.5	9.450	7.175	14.350	15.225	861.18	120.92
12	15.3	5.508	4.437	13.005	14.994	641.96	82.75
13	12.8	4.608	4.352	12.032	12.288	803.11	139.15
14	14.5	3.625	4.205	13.775	13.775	710.46	108.92
15	15.7	2.669	3.925	15.229	13.659	649.06	114.47
16	17.8	4.806	4.272	13.706	15.130	780.45	133.80
17	21.4	4.066	4.922	16.692	16.264	872.51	137.13
18	20.5	7.175	6.765	14.965	20.090	1281.55	194.78
19	15.1	5.738	4.530	13.137	12.684	661.88	96.57
20	12.5	4.250	4.000	8.875	12.375	1048.78	192.70
21	8.2	1.886	2.870	7.134	6.560	1011.14	135.63
22	14.1	3.384	3.948	13.395	10.857	1110.61	152.26
23	9.6	2.208	2.784	8.448	8.448	777.18	133.35
24	17.6	2.640	5.456	1.760	17.600	896.07	155.77
25	16.1	6.923	5.474	14.812	13.524	790.32	144.45
26	21.4	8.346	9.416	17.976	18.190	816.21	85.15
27	14.9	1.937	5.215	13.857	13.410	732.28	114.82
28	14.7	5.439	4.704	13.965	14.553	1029.87	138.71
29	11.6	4.060	3.480	10.092	9.628	746.54	120.21
30	11.2	1.792	3.136	9.632	8.736	1301.52	159.85
31	18.4	3.496	4.968	12.328	18.032	869.85	120.75
32	12.3	3.936	3.567	10.824	9.840	1234.31	150.01
33	16.8	6.552	5.208	15.792	13.608	708.24	127.82
34	23.9	5.497	10.038	23.661	20.554	688.75	109.72
35	14.1	3.948	4.794	13.959	11.562	697.73	133.52
36	19.9	6.368	5.771	18.308	18.706	881.51	178.86
37	12.8	4.224	3.328	8.576	11.520	804.71	104.61
38	18.2	9.100	5.642	17.472	16.016	905.99	153.86
39	11.1	3.774	4.218	10.212	8.769	1148.99	148.58
40	23.9	9.082	9.799	22.944	19.359	858.97	116.29
41	19.4	6.014	6.402	19.012	16.684	669.91	96.87
42	19.5	4.095	5.655	15.990	15.795	767.91	155.57
43	19.4	7.760	7.372	17.654	16.878	1004.75	156.83
44	11.3	4.859	1.808	9.944	10.848	809.38	109.48
45	13.6	4.080	4.080	13.056	12.920	716.20	109.61
46	12.7	2.413	3.429	11.049	11.176	768.95	153.72
47	10.6	4.452	3.498	8.692	9.116	890.03	111.62
48	23.8	8.092	6.664	23.086	20.706	992.61	152.56
49	13.8	4.968	4.554	5.382	11.592	670.31	106.62
50	17.4	7.308	5.568	14.094	15.660	791.14	122.04

```
In [8]: 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 [9]: sns.scatterplot(x='speeding',y='alcohol',data=df)
```

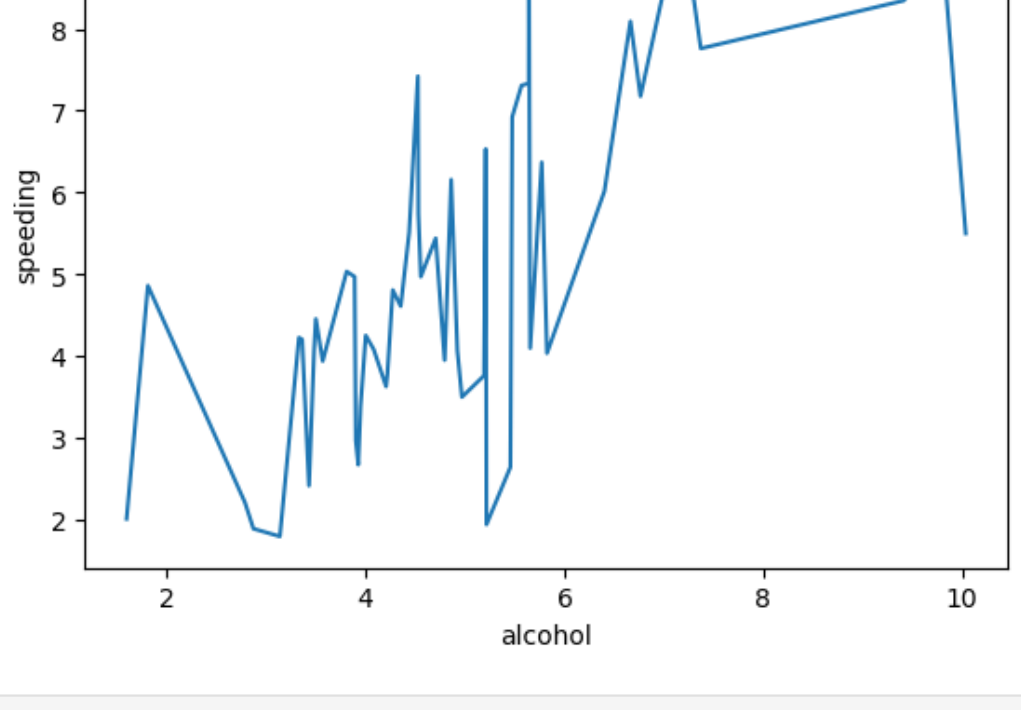
```
Out[9]: <Axes: xlabel='speeding', ylabel='alcohol'>
```



Inference : from the plot we can say that as alcohol **is** directly proportional to speeding **as** speed **is** increases after alcohol **is** consumed

```
In [16]: sns.scatterplot(x='speeding',y='not_distracted',data=df)
```

```
Out[16]: <Axes: xlabel='speeding', ylabel='not_distracted'>
```

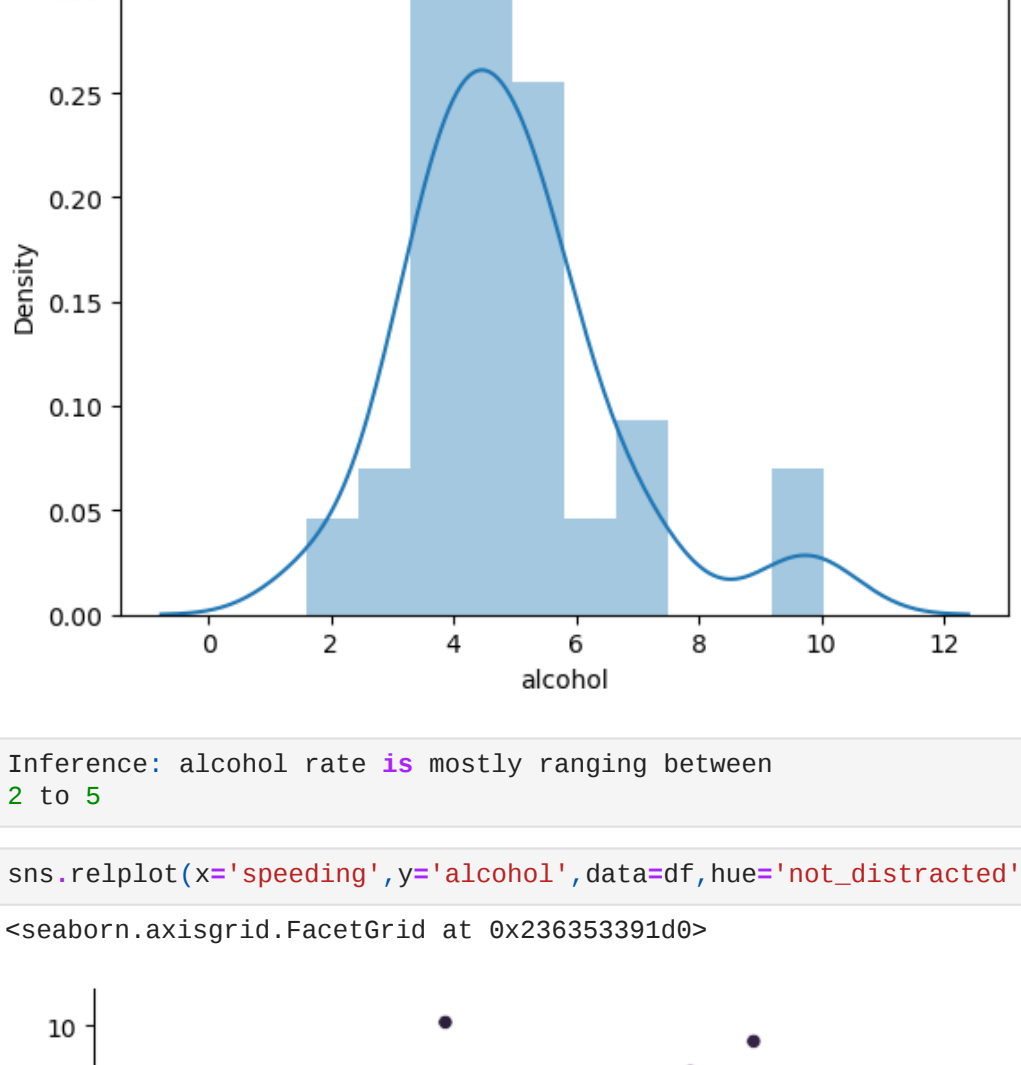


Inference : from the plot we can say that as speeding will be **is in** limit when driver **is** not_distracted

```
In [21]: sns.lineplot(x='alcohol',y='speeding',data=df,ci=None)
```

C:\Users\Admin\AppData\Local\Temp\ipykernel_12808\972903478.py:1: FutureWarning: The 'ci' parameter is deprecated. Use 'errorbar=None' for the same effect.

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



Inference: with **inc** in alcohol rate speeding also **increasing**

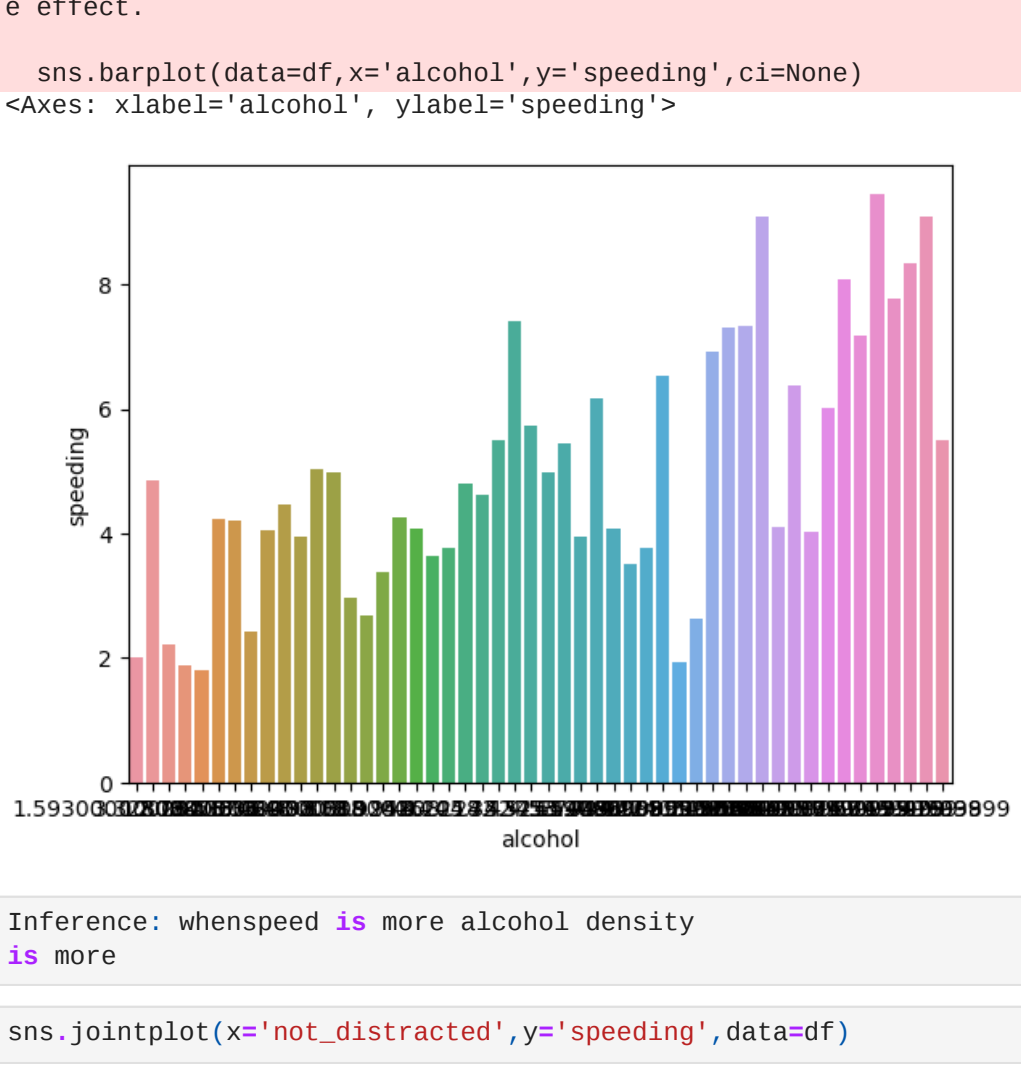
```
In [22]: sns.distplot(df['alcohol'])
```

C:\Users\Admin\AppData\Local\Temp\ipykernel_12808\2398698211.py:1: FutureWarning: 'distplot' is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either 'distplot' (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/de4147ed297445d74ad6372750bbe5751>

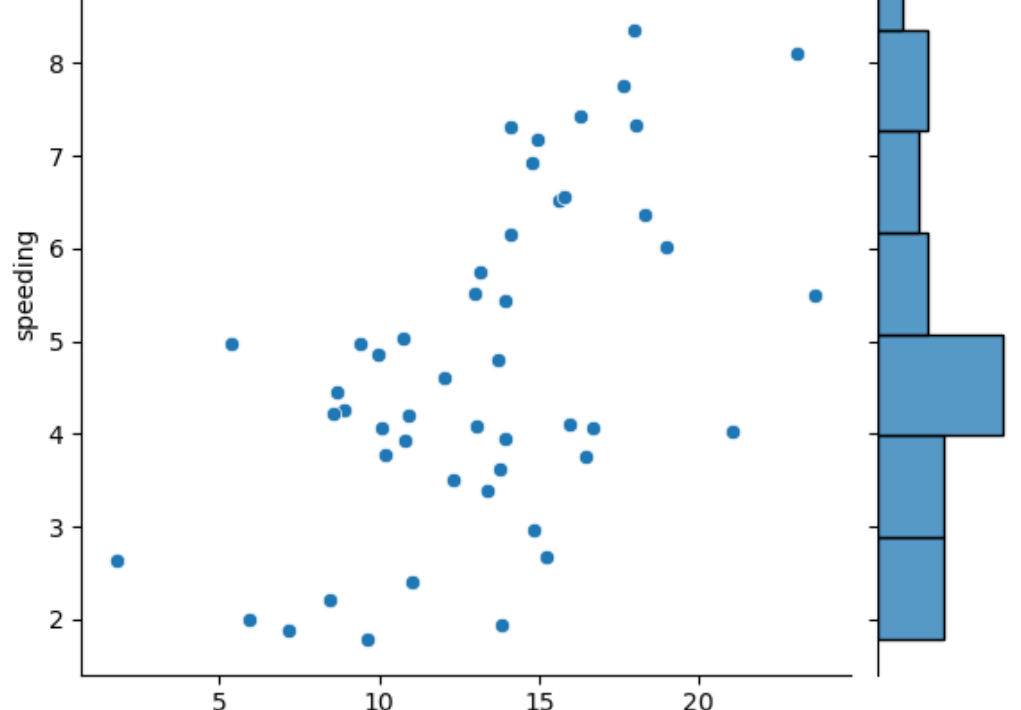
```
Out[22]: <Axes: xlabel='alcohol', ylabel='Density'>
```



Inference: alcohol rate **is** mostly ranging **2** to **5**

```
In [41]: sns.relplot(x='speeding',y='alcohol',data=df,hue='not_distracted')
```

```
Out[41]: <seaborn.axisgrid.FacetGrid at 0x23635391d0>
```

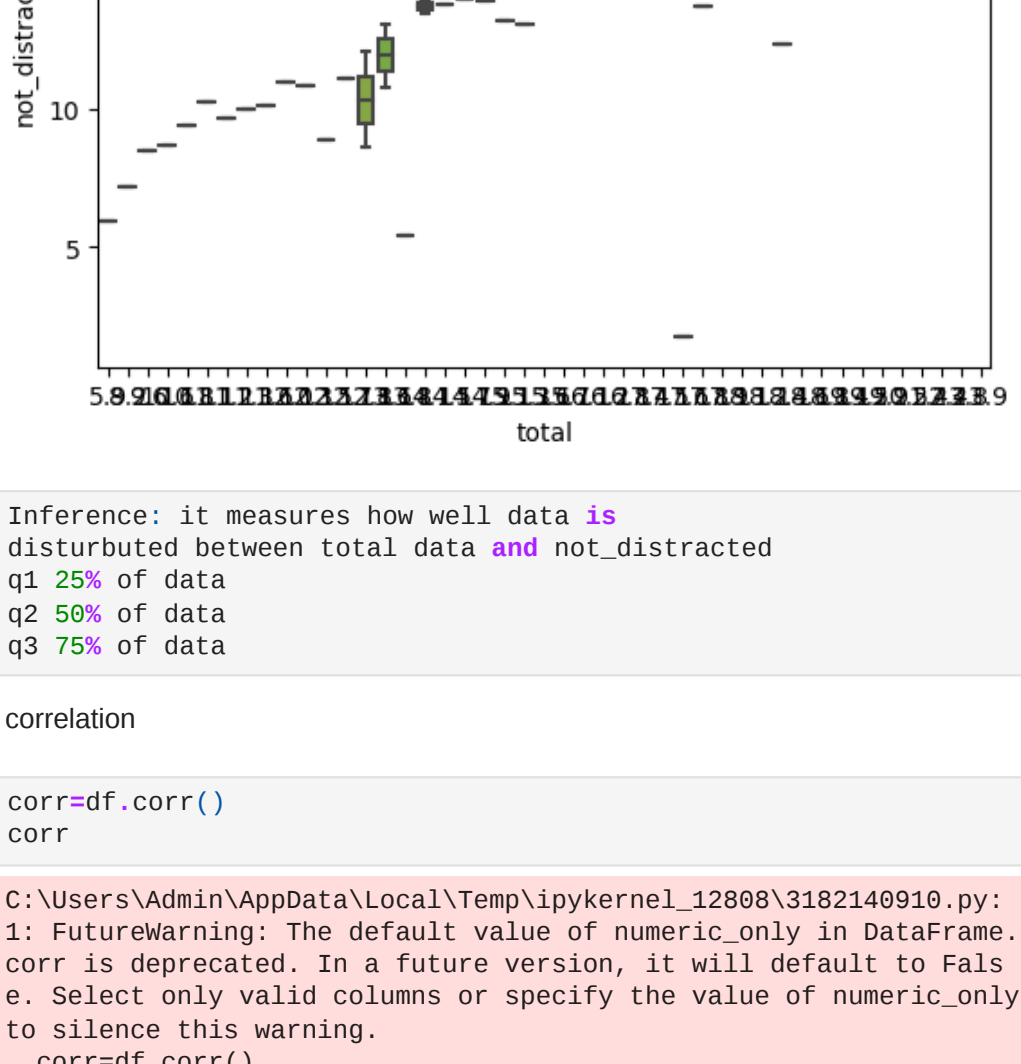


Inference: when drivers alcohol **and** speed are more he was more distracted **else** less distracted

```
In [66]: import matplotlib.pyplot as plt
fig=plt.figure()
car=fig.add_axes([0.1,0.1,0.8,0.8])
sns.barplot(data=df,x='alcohol',y='speeding',ci=None)
```

C:\Users\Admin\AppData\Local\Temp\ipykernel_12808\1537573723.py:4: FutureWarning: The 'ci' parameter is deprecated. Use 'errorbar=None' for the same effect.

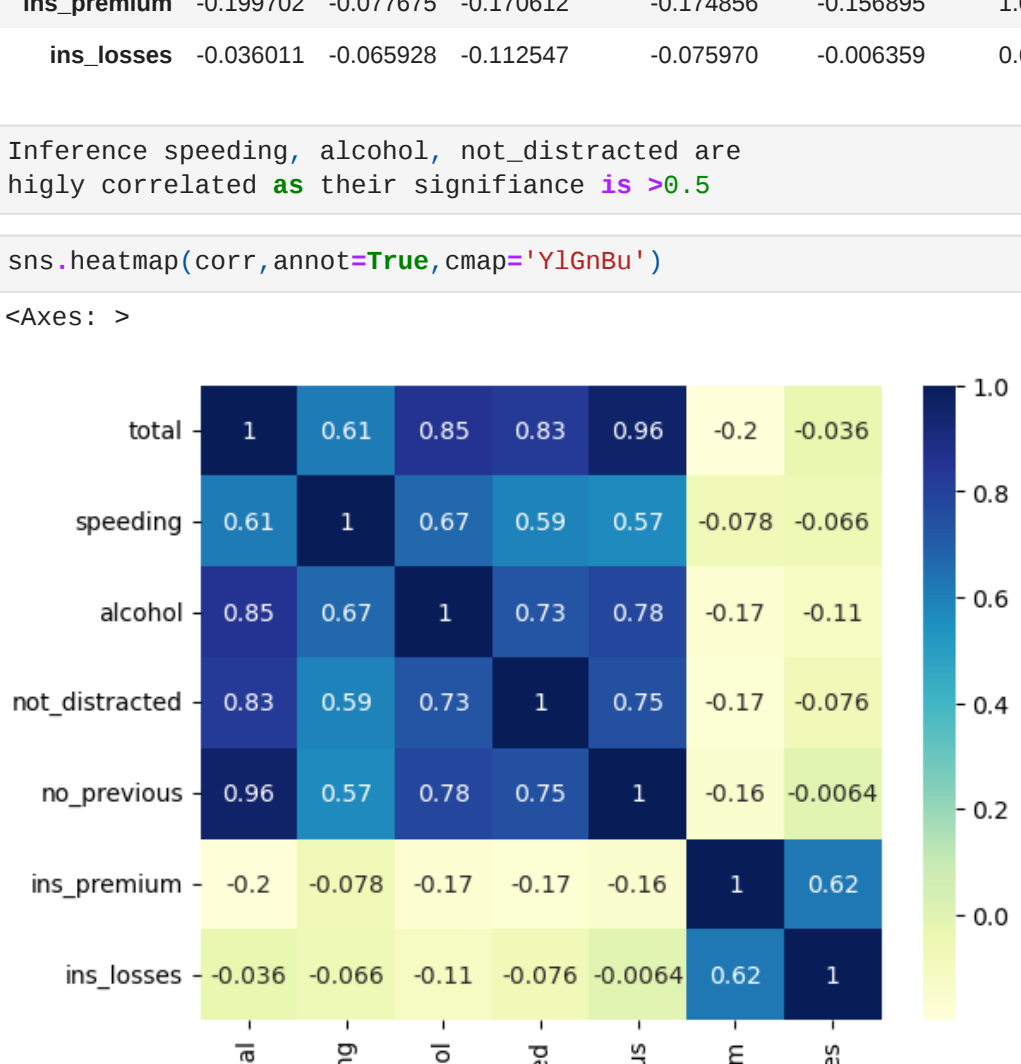
```
Out[66]: <Axes: xlabel='alcohol', ylabel='speeding'>
```



Inference: whenspeed **is** more alcohol density **is** more

```
In [68]: sns.jointplot(x='not_distracted',y='speeding',data=df)
```

```
Out[68]: <seaborn.axisgrid.JointGrid at 0x23653755f10>
```



Inference: speed **is** more when person got distracted most

```
In [77]: sns.boxplot(x='total',y='not_distracted',data=df)
```

```
Out[77]: <Axes: xlabel='total', ylabel='not_distracted'>
```



Inference: it measures how well data **is** distributed between total data **and** not_distracted

q1 25% of data

q2 50% of data

q3 75% of data

correlation

```
In [78]: corr=df.corr()
corr
```

C:\Users\Admin\AppData\Local\Temp\ipykernel_12808\3182140910.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 columns or specify the value of numeric_only to silence this warning.

	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	0.620000
ins_losses	-0.036011	-0.065928	-0.112547	-0.075970	-0.006359	0.620000	1.000000

Inference speeding, alcohol, not_distracted are highly correlated **as** their significance **is** >0.5

```
In [84]: sns.heatmap(corr,annot=True,cmap='YlGnBu')
```

```
Out[84]: <Axes: >
```


Inference: Litter shade of same color represents high correlation than darker ones

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```