document

September 13, 2023

# **0.0.1 Name : G.VISHNU VARDHAN 0.0.2 Reg no : 21BCE8760 0.0.3 Mobile no : 8106217741 0.0.4 Campus : VIT - AP**

[2]:

**import**

**seaborn**

**as**

**sns**

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']

[3]:

df

=

sns

.

load\_dataset(

'

car\_crashes

'

)

df

[3]: total speeding alcohol not\_distracted no\_previous ins\_premium \

0 18.8 7.332 5.640 18.048 15.040 784.55 1 18.1 7.421 4.525 16.290 17.014 1053.48

1. 18.6 6.510 5.208 15.624 17.856 899.47
2. 22.4 4.032 5.824 21.056 21.280 827.34
3. 12.0 4.200 3.360 10.920 10.680 878.41
4. 13.6 5.032 3.808 10.744 12.920 835.50 6 10.8 4.968 3.888 9.396 8.856 1068.73 7 16.2 6.156 4.860 14.094 16.038 1137.87
5. 5.9 2.006 1.593 5.900 5.900 1273.89
6. 17.9 3.759 5.191 16.468 16.826 1160.13
7. 15.6 2.964 3.900 14.820 14.508 913.15
8. 17.5 9.450 7.175 14.350 15.225 861.18
9. 15.3 5.508 4.437 13.005 14.994 641.96
10. 12.8 4.608 4.352 12.032 12.288 803.11
11. 14.5 3.625 4.205 13.775 13.775 710.46
12. 15.7 2.669 3.925 15.229 13.659 649.06
13. 17.8 4.806 4.272 13.706 15.130 780.45
14. 21.4 4.066 4.922 16.692 16.264 872.51 18 20.5 7.175 6.765 14.965 20.090 1281.55 19 15.1 5.738 4.530 13.137 12.684 661.88
15. 12.5 4.250 4.000 8.875 12.375 1048.78
16. 8.2 1.886 2.870 7.134 6.560 1011.14
17. 14.1 3.384 3.948 13.395 10.857 1110.61
18. 9.6 2.208 2.784 8.448 8.448 777.18
19. 17.6 2.640 5.456 1.760 17.600 896.07
20. 16.1 6.923 5.474 14.812 13.524 790.32
21. 21.4 8.346 9.416 17.976 18.190 816.21
22. 14.9 1.937 5.215 13.857 13.410 732.28 28 14.7 5.439 4.704 13.965 14.553 1029.87
23. 11.6 4.060 3.480 10.092 9.628 746.54
24. 11.2 1.792 3.136 9.632 8.736 1301.52
25. 18.4 3.496 4.968 12.328 18.032 869.85
26. 12.3 3.936 3.567 10.824 9.840 1234.31
27. 16.8 6.552 5.208 15.792 13.608 708.24
28. 23.9 5.497 10.038 23.661 20.554 688.75
29. 14.1 3.948 4.794 13.959 11.562 697.73
30. 19.9 6.368 5.771 18.308 18.706 881.51
31. 12.8 4.224 3.328 8.576 11.520 804.71
32. 18.2 9.100 5.642 17.472 16.016 905.99
33. 11.1 3.774 4.218 10.212 8.769 1148.99
34. 23.9 9.082 9.799 22.944 19.359 858.97
35. 19.4 6.014 6.402 19.012 16.684 669.31
36. 19.5 4.095 5.655 15.990 15.795 767.91 43 19.4 7.760 7.372 17.654 16.878 1004.75
37. 11.3 4.859 1.808 9.944 10.848 809.38
38. 13.6 4.080 4.080 13.056 12.920 716.20
39. 12.7 2.413 3.429 11.049 11.176 768.95
40. 10.6 4.452 3.498 8.692 9.116 890.03 48 23.8 8.092 6.664 23.086 20.706 992.61
41. 13.8 4.968 4.554 5.382 11.592 670.31
42. 17.4 7.308 5.568 14.094 15.660 791.14

ins\_losses abbrev

1. 145.08 AL
2. 133.93 AK
3. 110.35 AZ
4. 142.39 AR
5. 165.63 CA
6. 139.91 CO
7. 167.02 CT
8. 151.48 DE
9. 136.05 DC
10. 144.18 FL
11. 142.80 GA
12. 120.92 HI
13. 82.75 ID

|  |  |  |
| --- | --- | --- |
| 13 | 139.15 | IL |
| 14 | 108.92 | IN |
| 15 | 114.47 | IA |
| 16 | 133.80 | KS |
| 17 | 137.13 | KY |
| 18 | 194.78 | LA |
| 19 | 96.57 | ME |
| 20 | 192.70 | MD |
| 21 | 135.63 | MA |
| 22 | 152.26 | MI |
| 23 | 133.35 | MN |
| 24 | 155.77 | MS |
| 25 | 144.45 | MO |
| 26 | 85.15 | MT |
| 27 | 114.82 | NE |
| 28 | 138.71 | NV |
| 29 | 120.21 | NH |
| 30 | 159.85 | NJ |
| 31 | 120.75 | NM |
| 32 | 150.01 | NY |
| 33 | 127.82 | NC |
| 34 | 109.72 | ND |
| 35 | 133.52 | OH |
| 36 | 178.86 | OK |
| 37 | 104.61 | OR |
| 38 | 153.86 | PA |
| 39 | 148.58 | RI |
| 40 | 116.29 | SC |
| 41 | 96.87 | SD |
| 42 | 155.57 | TN |
| 43 | 156.83 | TX |
| 44 | 109.48 | UT |
| 45 | 109.61 | VT |
| 46 | 153.72 | VA |
| 47 | 111.62 | WA |
| 48 | 152.56 | WV |
| 49 | 106.62 | WI |
| 50 | 122.04 | WY |

[10]:

x

=

df

.

head(

5

)

x

[10]: total speeding alcohol not\_distracted no\_previous ins\_premium \

0 18.8 7.332 5.640 18.048 15.040 784.55 1 18.1 7.421 4.525 16.290 17.014 1053.48

* 1. 18.6 6.510 5.208 15.624 17.856 899.47
  2. 22.4 4.032 5.824 21.056 21.280 827.34
  3. 12.0 4.200 3.360 10.920 10.680 878.41

ins\_losses abbrev

|  |  |  |
| --- | --- | --- |
| 0 | 145.08 | AL |
| 1 | 133.93 | AK |
| 2 | 110.35 | AZ |
| 3 | 142.39 | AR |
| 4 | 165.63 | CA |

[5]:

df

.

info()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 51 entries, 0 to 50 Data columns (total 8 columns):

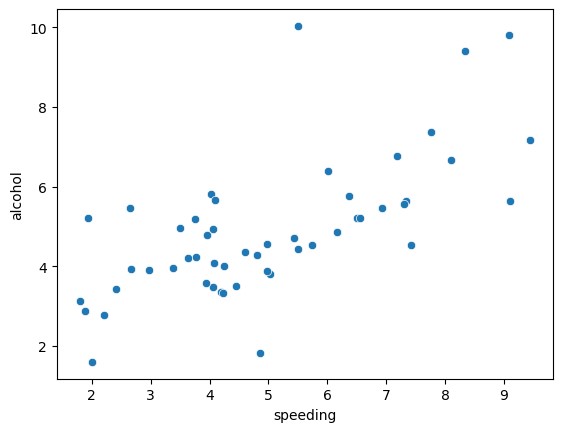
|  |  |
| --- | --- |
| # Column Non-Null Count Dtype  --- ------ -------------- ----- | |
| 0 total 51 non-null | float64 |
| 1 speeding 51 non-null | float64 |
| 2 alcohol 51 non-null | float64 |
| 3 not\_distracted 51 non-null | float64 |
| 4 no\_previous 51 non-null | float64 |
| 5 ins\_premium 51 non-null | float64 |
| 6 ins\_losses 51 non-null | float64 |
| 7 abbrev 51 non-null  dtypes: float64(7), object(1) memory usage: 3.3+ KB | object |

[7]: sns.scatterplot(x="speeding",y="alcohol",data=df)

*# inference*

*#most of the drivers who drunk more alcohol have droven with more speed*

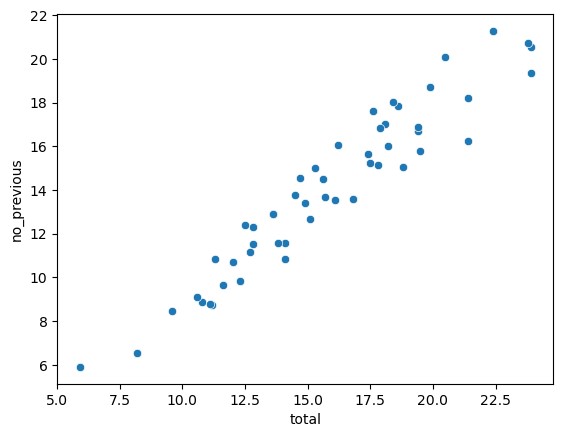
[7]: <Axes: xlabel='speeding', ylabel='alcohol'>



[46]: sns.scatterplot(x="total",y="no\_previous",data=df)

*# the given plot shows thge relation between no previous and total*

[46]: <Axes: xlabel='total', ylabel='no\_previous'>

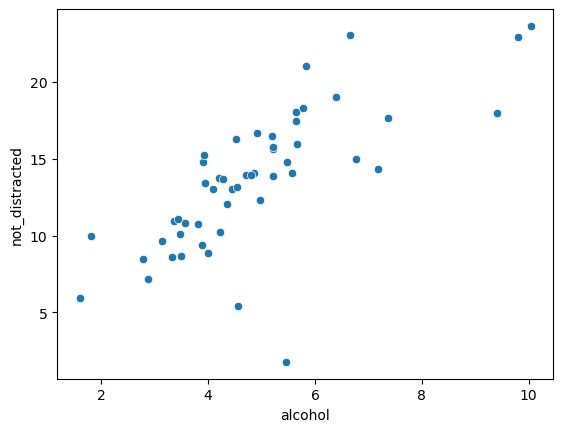


[8]: sns.scatterplot(x="alcohol",y="not\_distracted",data=df)

*#inference*

*# people who drunk less alochol they are less not\_distracted*

[8]: <Axes: xlabel='alcohol', ylabel='not\_distracted'>

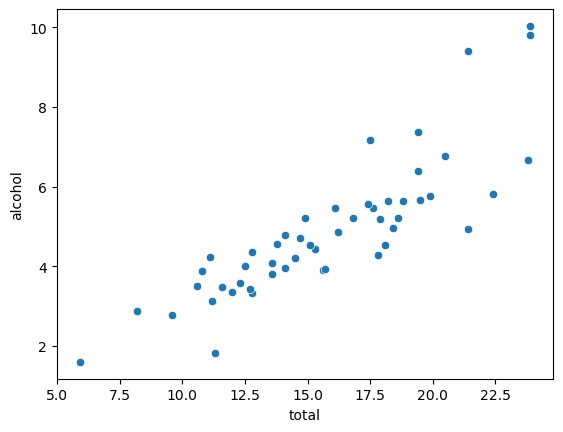


[10]: sns.scatterplot(x="total",y="alcohol",data=df)

*#inference*

*# alocohol content increases crashes also increases*

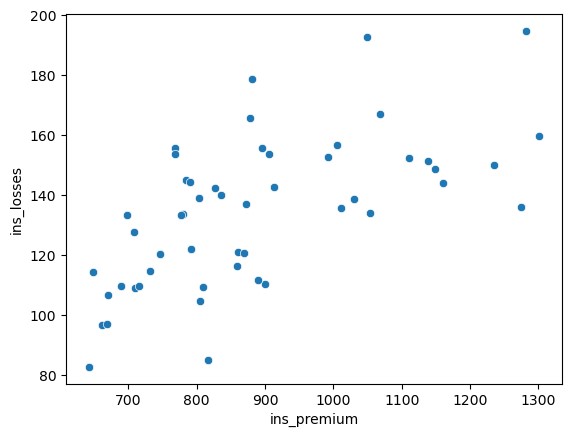
[10]: <Axes: xlabel='total', ylabel='alcohol'>



[14]: sns.scatterplot(x='ins\_premium',y='ins\_losses',data=df)

*# people who paid less insurance got less loss*

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



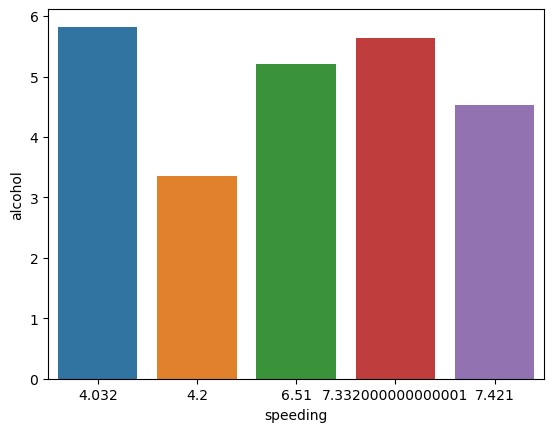
[19]: sns.barplot(data=x,x="speeding",y="alcohol",ci=**None**)

*# Inference* most of the drivers who drank more alcohol have droven **with** more speed C:\Users\hp\AppData\Local\Temp\ipykernel\_4456\1777853704.py:1: FutureWarning:

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

sns.barplot(data=x,x="speeding",y="alcohol",ci=None)

[19]: <Axes: xlabel='speeding', ylabel='alcohol'>



# [16]: sns.barplot(data=x,x="speeding",y="not\_distracted",ci=**None**)

*# inference*

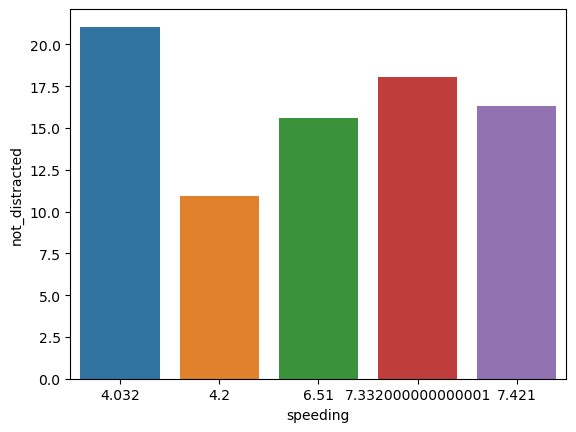
*# the persons who are driving with less speed are not distracted*

C:\Users\hp\AppData\Local\Temp\ipykernel\_4456\1143020830.py:1: FutureWarning:

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

sns.barplot(data=x,x="speeding",y="not\_distracted",ci=None)

[16]: <Axes: xlabel='speeding', ylabel='not\_distracted'>



[17]: sns.barplot(data=x,x="total",y="alcohol",ci=**None**)

*#Inference*

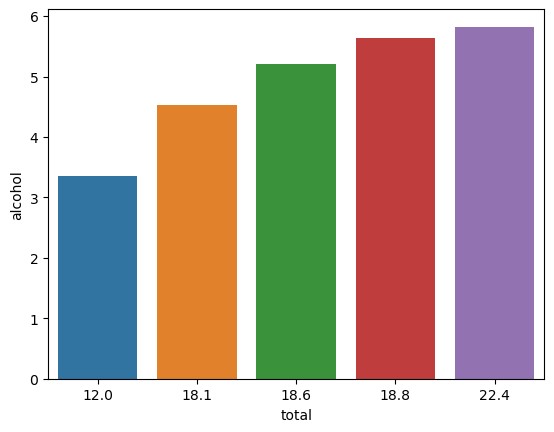
*# no of accidents increases as drinking more alcohol*

C:\Users\hp\AppData\Local\Temp\ipykernel\_4456\2713533087.py:1: FutureWarning:

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

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

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



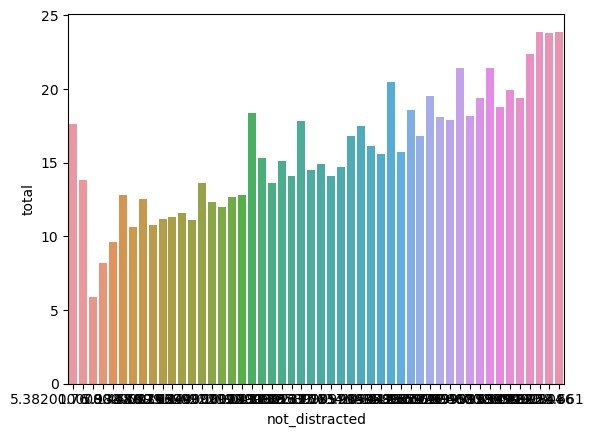
[22]: sns.barplot(data=df,x="not\_distracted",y="total",ci=**None**)

C:\Users\hp\AppData\Local\Temp\ipykernel\_4456\3023431553.py:1: FutureWarning:

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

sns.barplot(data=df,x="not\_distracted",y="total",ci=None)

[22]: <Axes: xlabel='not\_distracted', ylabel='total'>



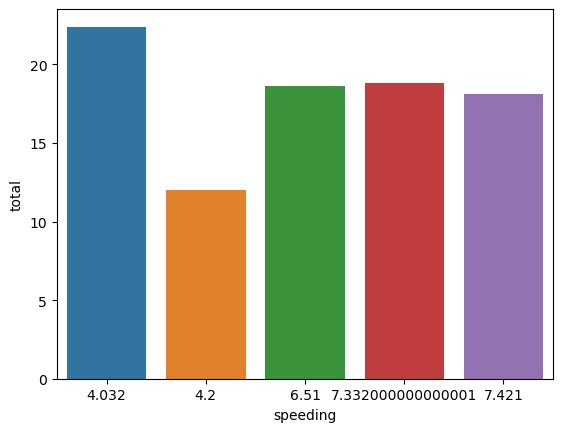
[44]: sns.barplot(data=x,x="speeding",y="total",ci=**None**)

C:\Users\hp\AppData\Local\Temp\ipykernel\_4456\3678240287.py:1: FutureWarning:

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

sns.barplot(data=x,x="speeding",y="total",ci=None)

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



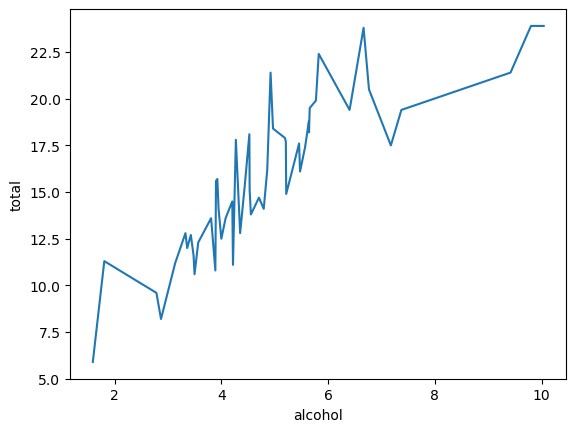
[14]: sns.lineplot(x='alcohol',y='total',data=df,ci=**None**)

*# as the alcohol consumption increases total crashes also increases* C:\Users\hp\AppData\Local\Temp\ipykernel\_2372\2372459405.py:1: FutureWarning:

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

sns.lineplot(x='alcohol',y='total',data=df,ci=None)

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



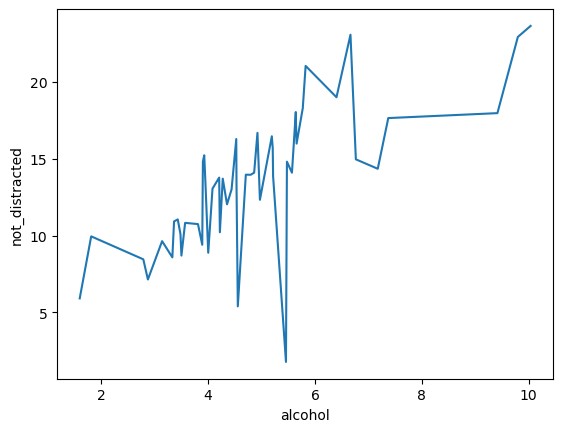
[47]: sns.lineplot(x='alcohol',y='not\_distracted',data=df,ci=**None**)

C:\Users\hp\AppData\Local\Temp\ipykernel\_4456\3085692794.py:1: FutureWarning:

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

sns.lineplot(x='alcohol',y='not\_distracted',data=df,ci=None)

[47]: <Axes: xlabel='alcohol', ylabel='not\_distracted'>



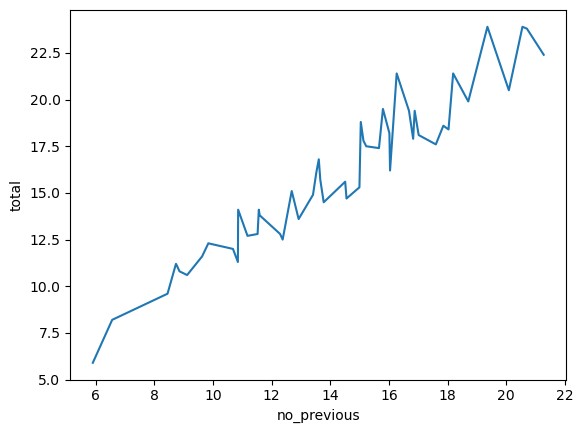
[20]: sns.lineplot(x='no\_previous',y='total',data=df,ci=**None**)

C:\Users\hp\AppData\Local\Temp\ipykernel\_4456\1903628411.py:1: FutureWarning:

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

sns.lineplot(x='no\_previous',y='total',data=df,ci=None)

[20]: <Axes: xlabel='no\_previous', ylabel='total'>



[15]: sns.lineplot(x='speeding',y='total',data=df,ci=**None**)

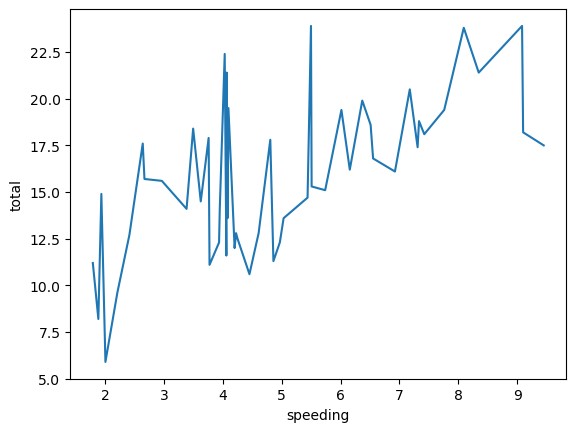
*# In most of the cases as the speed increases total crashes also increases*

C:\Users\hp\AppData\Local\Temp\ipykernel\_2372\2174020548.py:1: FutureWarning:

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

sns.lineplot(x='speeding',y='total',data=df,ci=None)

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



[41]: sns.lineplot(x='no\_previous',y='alcohol',data=df,ci=**None**)

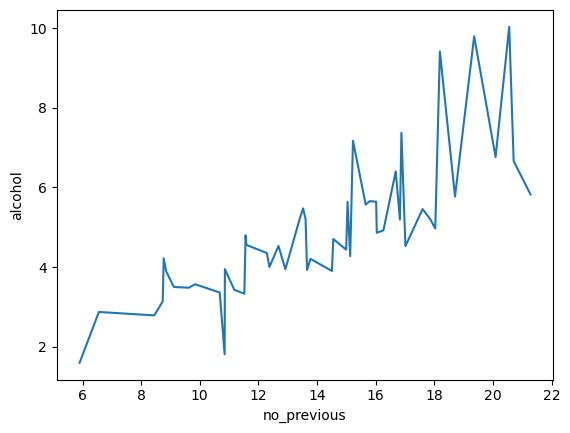
*# the person with no previous drank more alcohol*

C:\Users\hp\AppData\Local\Temp\ipykernel\_4456\3484007986.py:1: FutureWarning:

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

sns.lineplot(x='no\_previous',y='alcohol',data=df,ci=None)

[41]: <Axes: xlabel='no\_previous', ylabel='alcohol'>



[16]:

sns

.

distplot(df[

"

total

"

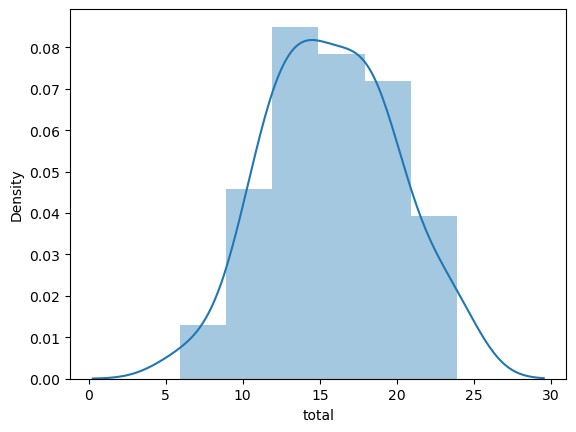
])

C:\Users\hp\AppData\Local\Temp\ipykernel\_2372\1102674835.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 sns.distplot(df["total"])

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



[40]:

sns

.

distplot(df[

"

no\_previous

"

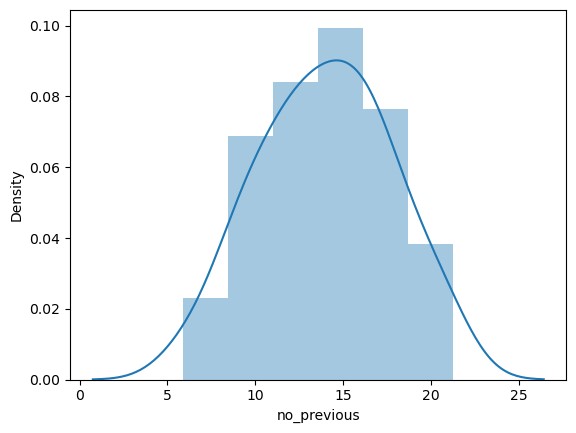
])

C:\Users\hp\AppData\Local\Temp\ipykernel\_4456\1806622040.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 sns.distplot(df["no\_previous"])

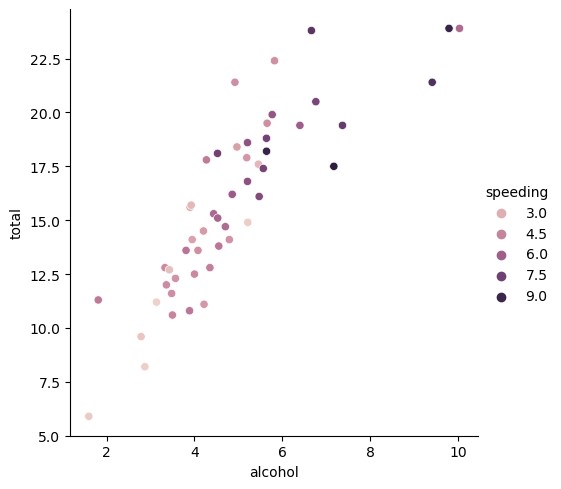
[40]: <Axes: xlabel='no\_previous', ylabel='Density'>



[17]: sns.relplot(x='alcohol',y='total',data=df,hue="speeding")

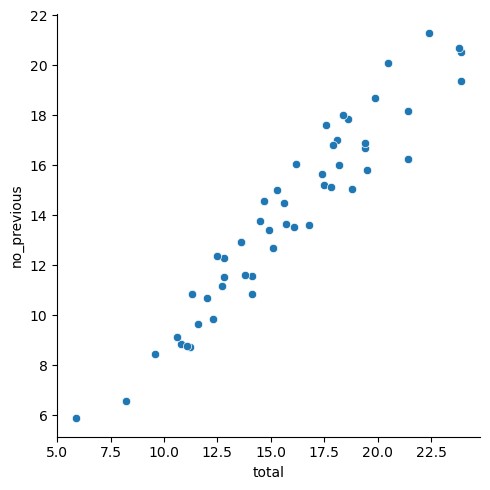
*# As alcohol consumption increases total crashes also increases*

[17]: <seaborn.axisgrid.FacetGrid at 0x1bb520b0490>



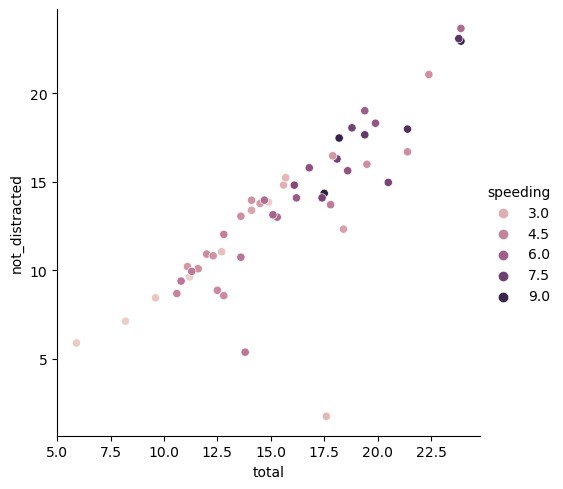
[45]: sns.relplot(x='total',y='no\_previous',data=df)

[45]: <seaborn.axisgrid.FacetGrid at 0x2d41ca32c90>



# [25]: sns.relplot(x='total',y='not\_distracted',data=df,hue="speeding")

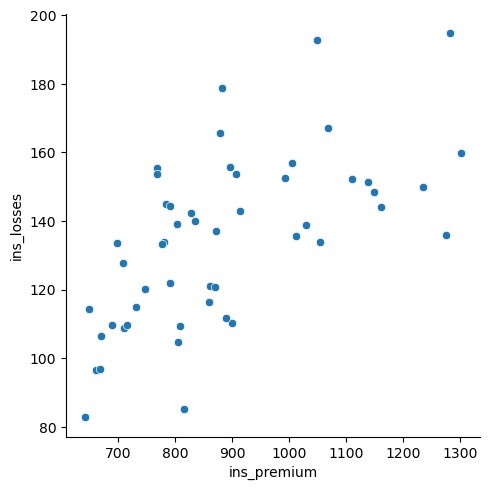
[25]: <seaborn.axisgrid.FacetGrid at 0x2d40f89a690>



# [26]: sns.relplot(x='ins\_premium',y='ins\_losses',data=df)

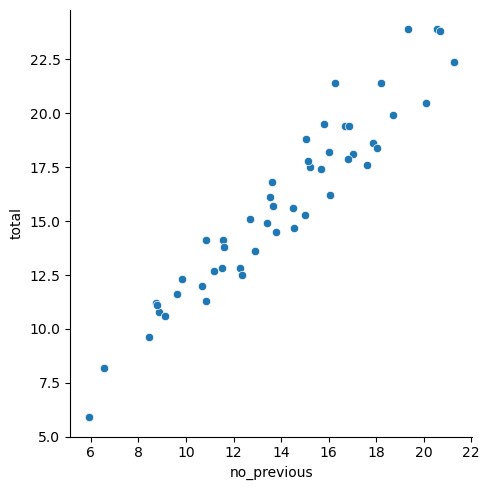
*#people who paid less insurance they faced more loss*

[26]: <seaborn.axisgrid.FacetGrid at 0x2d40fb56990>



[35]: sns.relplot(x='no\_previous',y='total',data=df)

[35]: <seaborn.axisgrid.FacetGrid at 0x2d412865050>



[42]:

sns

.

relplot(x

=

'

speeding

'

,y

=

'

total

'

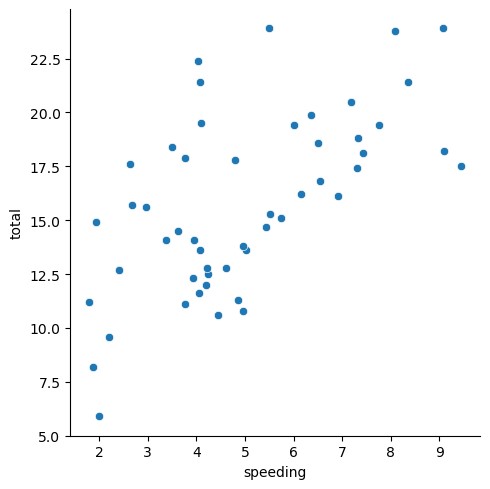
,data

=

df)

*# people who speed*

[42]: <seaborn.axisgrid.FacetGrid at 0x2d411a82750>



[18]:

df[

"

speeding

"

]

.

value\_counts()

[18]: 4.968 2 7.332 1 9.100 1 5.439 1 4.060 1 1.792 1 3.496 1 3.936 1 6.552 1 5.497 1 3.948 1 6.368 1 4.224 1 3.774 1 8.346 1

9.082 1

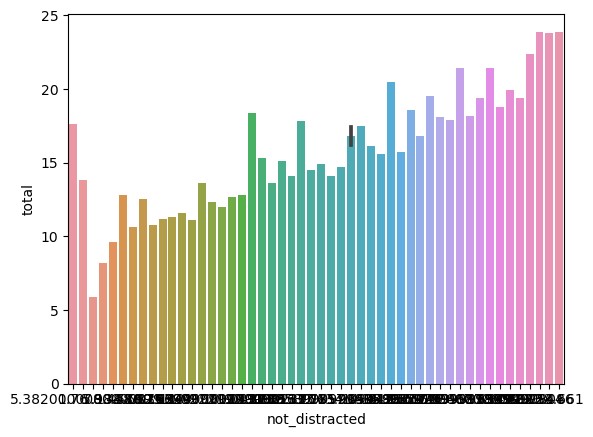
6.014 1 4.095 1 7.760 1 4.859 1 4.080 1 2.413 1 4.452 1 8.092 1 1.937 1 6.923 1 7.421 1 2.640 1 6.510 1 4.032 1 4.200 1 5.032 1 6.156 1 2.006 1 3.759 1 2.964 1 9.450 1 5.508 1 4.608 1 3.625 1 2.669 1 4.806 1 4.066 1 7.175 1 5.738 1 4.250 1 1.886 1 3.384 1 2.208 1

7.308 1

Name: speeding, dtype: int64

# [39]: sns.barplot(data=df,x='not\_distracted',y='total')

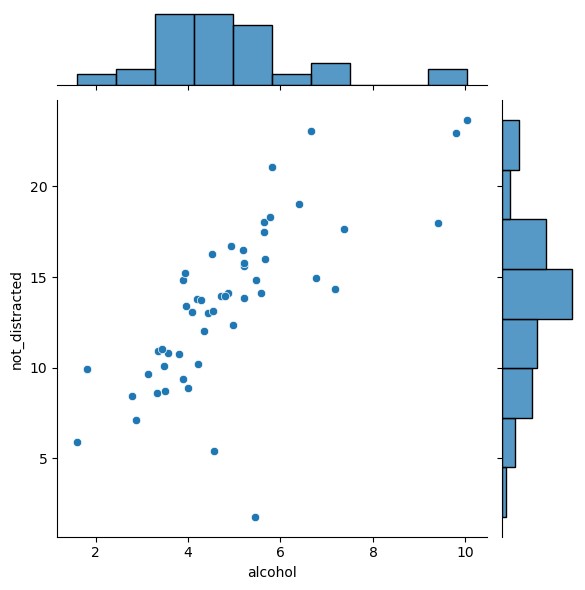
[39]: <Axes: xlabel='not\_distracted', ylabel='total'>



# [28]: sns.jointplot(x="alcohol",y="not\_distracted",data=df)

*# people who consumed more alcohol at 10 they are not\_distracted*

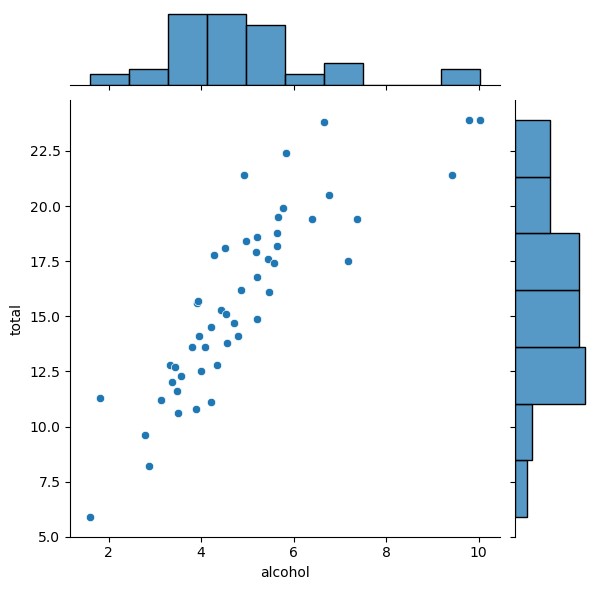
[28]: <seaborn.axisgrid.JointGrid at 0x2d40fbaf450>



[36]: sns.jointplot(x="alcohol",y="total",data=df)

*# with increase in alcohol consumption crashes also increased*

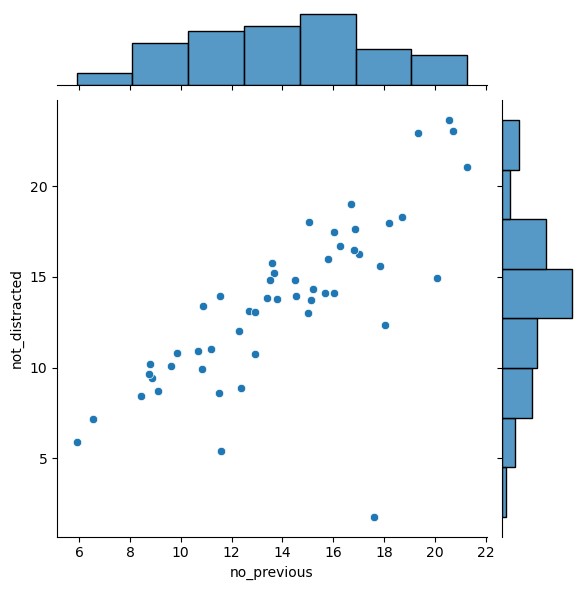
[36]: <seaborn.axisgrid.JointGrid at 0x1bb0a0123d0>



# [29]: sns.jointplot(x="no\_previous",y="not\_distracted",data=df)

*# as n0\_previous increases not distracted also increases in most of the cases*

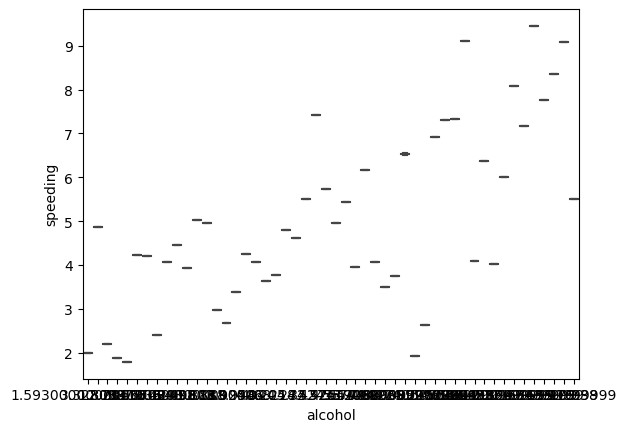
[29]: <seaborn.axisgrid.JointGrid at 0x2d4101371d0>



[32]: sns.boxplot(x="alcohol",y="speeding",data=df)

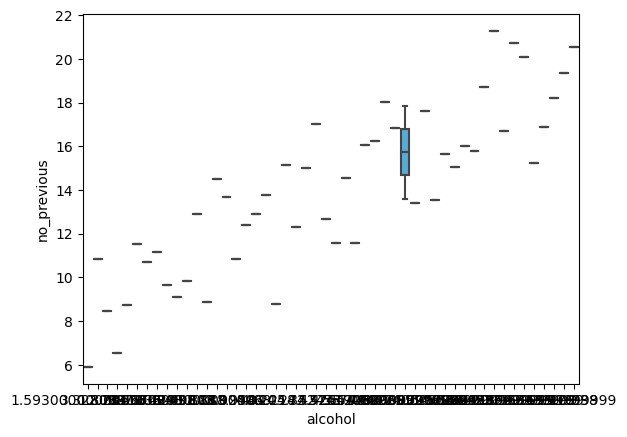
*# people who drank more alcohol have droven with more speed*

[32]: <Axes: xlabel='alcohol', ylabel='speeding'>



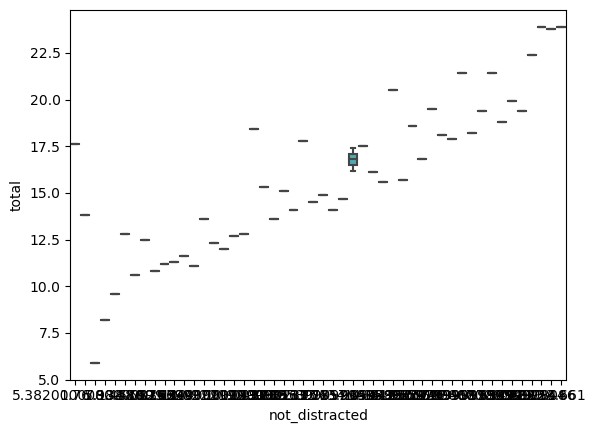
[33]: sns.boxplot(x="alcohol",y="no\_previous",data=df)

[33]: <Axes: xlabel='alcohol', ylabel='no\_previous'>



# [34]: sns.boxplot(x="not\_distracted",y="total",data=df)

[34]: <Axes: xlabel='not\_distracted', ylabel='total'>



[29]:

corr

=

df

.

corr()

corr

C:\Users\hp\AppData\Local\Temp\ipykernel\_2372\3311646455.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.

corr = df.corr()

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| [29]: total speeding alcohol not\_distracted no\_previous \ | | | | | | |
| total 1.000000 0.611548 0.852613 | | | | | 0.827560 | 0.956179 |
| speeding 0.611548 1.000000 0.669719 | | | | | 0.588010 | 0.571976 |
| alcohol 0.852613 0.669719 1.000000 | | | | | 0.732816 | 0.783520 |
| not\_distracted 0.827560 0.588010 0.732816 | | | | | 1.000000 | 0.747307 |
| no\_previous 0.956179 0.571976 0.783520 | | | | | 0.747307 | 1.000000 |
| ins\_premium -0.199702 -0.077675 -0.170612 | | | | | -0.174856 | -0.156895 |
| ins\_losses -0.036011 -0.065928 -0.112547  ins\_premium ins\_losses  total -0.199702 -0.036011 speeding -0.077675 -0.065928 | | | | | -0.075970 | -0.006359 |
| alcohol | -0.170612 | -0.112547 |
| not\_distracted | -0.174856 | -0.075970 |
| no\_previous | -0.156895 | -0.006359 |
| ins\_premium | 1.000000 | 0.623116 |
| ins\_losses | 0.623116 | 1.000000 |

[39]:

sns

.

heatmap(corr,annot

=

**True**

)

[39]: <Axes: >

