- Assignment-2
- Name:-Pothabattula Swamy
- Reg no:-21BCE9404

```
import seaborn as sns
import matplotlib.pyplot as plt
```

```
df=sns.load_dataset("car_crashes")
```

```
df
```

	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses	ab
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	

```
df.shape
```

```
(51, 8)
```

16	17.8	4.806	4.272	13.706	15.130	780.45	133.80	

```
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
```

```
df.head()
```

	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

```
df.tail()
```

	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses	abbrev
46	12.7	2.413	3.429	11.049	11.176	768.95	153.72	VA
47	10.6	4.452	3.498	8.692	9.116	890.03	111.62	WA
48	23.8	8.092	6.664	23.086	20.706	992.61	152.56	WV
49	13.8	4.968	4.554	5.382	11.592	670.31	106.62	WI
50	17.4	7.308	5.568	14.094	15.660	791.14	122.04	WY

```
df.describe()
```

	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses
count	51.000000	51.000000	51.000000	51.000000	51.000000	51.000000	51.000000
mean	15.790196	4.998196	4.886784	13.573176	14.004882	886.957647	134.493137
std	4.122002	2.017747	1.729133	4.508977	3.764672	178.296285	24.835922
min	5.900000	1.792000	1.593000	1.760000	5.900000	641.960000	82.750000
25%	12.750000	3.766500	3.894000	10.478000	11.348000	768.430000	114.645000
50%	15.600000	4.608000	4.554000	13.857000	13.775000	858.970000	136.050000
75%	18.500000	6.439000	5.604000	16.140000	16.755000	1007.945000	151.870000
max	23.900000	9.450000	10.038000	23.661000	21.280000	1301.520000	194.780000

```
df.isnull().any()
```

```
total             False
speeding          False
alcohol           False
not_distracted    False
no_previous       False
ins_premium       False
ins_losses        False
abbrev            False
dtype: bool
```

```
df.isna().sum()
```

```
total             0
speeding          0
alcohol           0
not_distracted    0
no_previous       0
ins_premium       0
ins_losses        0
abbrev            0
dtype: int64
```

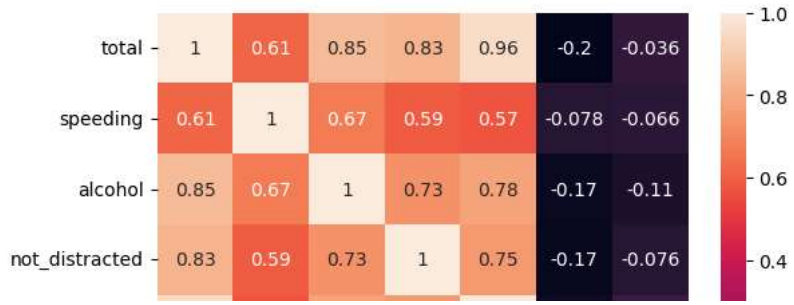
```
corr=df.corr()
corr
```

```
<ipython-input-24-7d5195e2bf4d>:1: FutureWarning: The default value of numeric_only in Dat
  corr=df.corr()
```

	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_
total	1.000000	0.611548	0.852613	0.827560	0.956179	-0.199702	-0
speeding	0.611548	1.000000	0.669719	0.588010	0.571976	-0.077675	-0
alcohol	0.852613	0.669719	1.000000	0.732816	0.783520	-0.170612	-0
not_distracted	0.827560	0.588010	0.732816	1.000000	0.747307	-0.174856	-0
no_previous	0.956179	0.571976	0.783520	0.747307	1.000000	-0.156895	-0
ins_premium	-0.199702	-0.077675	-0.170612	-0.174856	-0.156895	1.000000	0
ins_losses	-0.036011	-0.065928	-0.112547	-0.075970	-0.006359	0.623116	1

```
sns.heatmap(corr,annot=True)
```

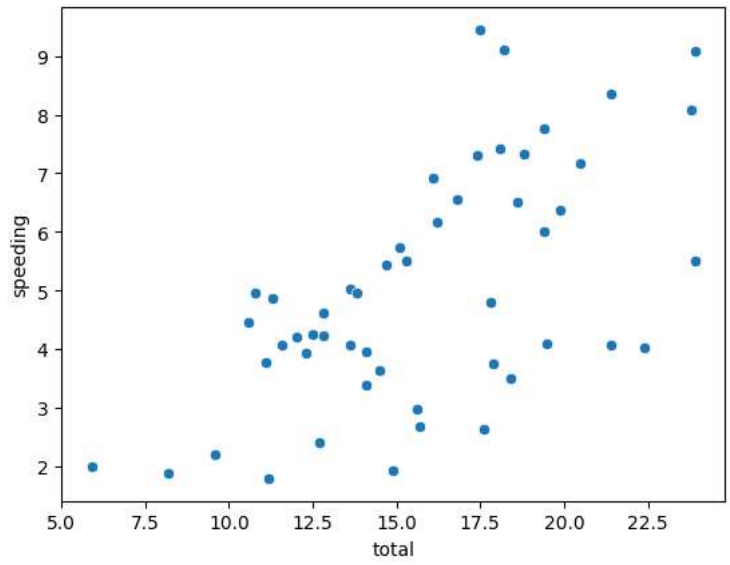
```
<Axes: >
```



The total no of drivers in fatal collisions are depend on percentage of involved in fatal collisions who consuming alcohol,who were not distracted and who do not have previous

```
sns.scatterplot(x='total',y='speeding',data=df)
```

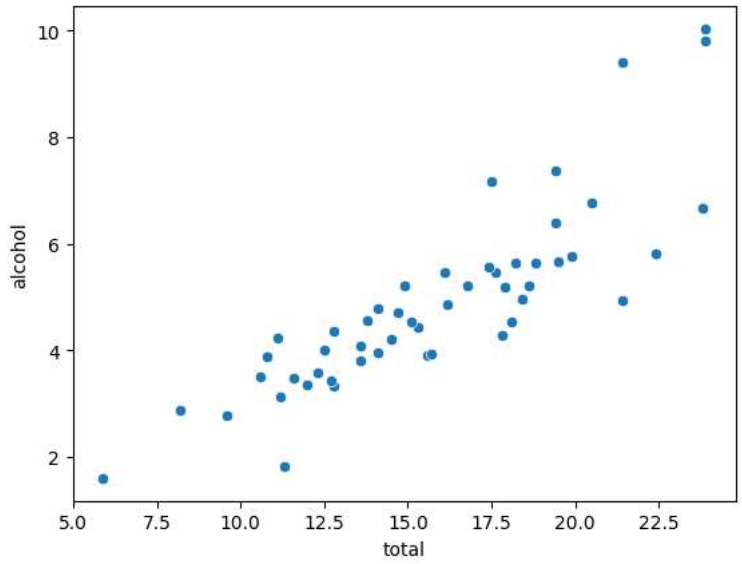
```
<Axes: xlabel='total', ylabel='speeding'>
```



From the above graph we can say that total no of drivers in fatal collisions increases with increasing in percentage of involved in fatal collisions who were speeding

```
sns.scatterplot(x='total',y='alcohol',data=df)
```

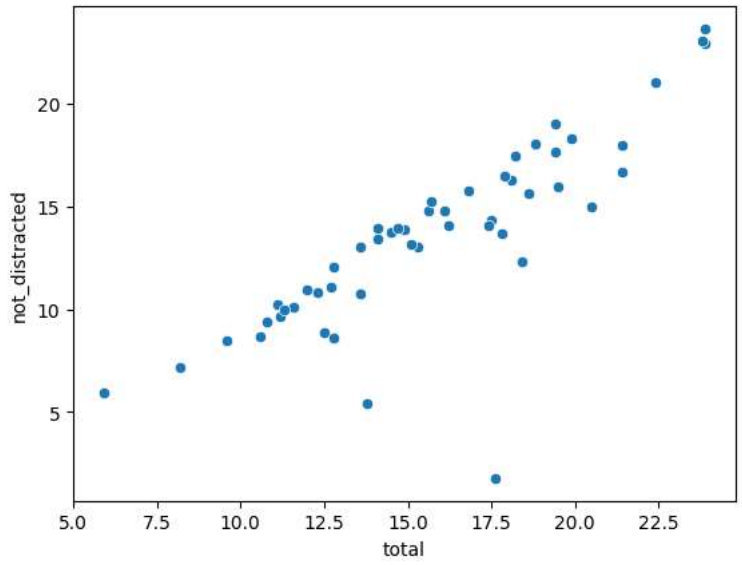
```
<Axes: xlabel='total', ylabel='alcohol'>
```



From the above graph we can say that total no of drivers in fatal collisions increases with increasing in percentage of involved in fatal collisions who are consuming alcohol

```
sns.scatterplot(x='total',y='not_distracted',data=df)
```

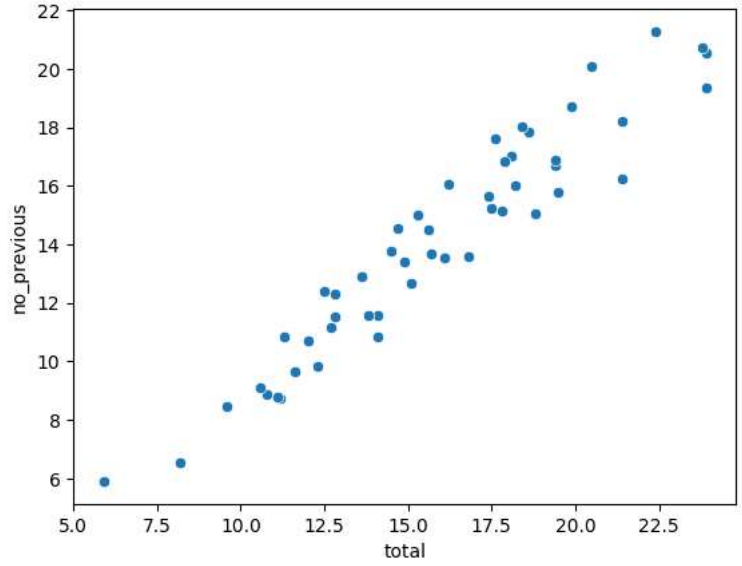
```
<Axes: xlabel='total', ylabel='not_distracted'>
```



From the above graph we can say that total no of drivers in fatal collisions increases with increasing in percentage of involved in fatal collisions who were not distracted

```
sns.scatterplot(x='total',y='no_previous',data=df)
```

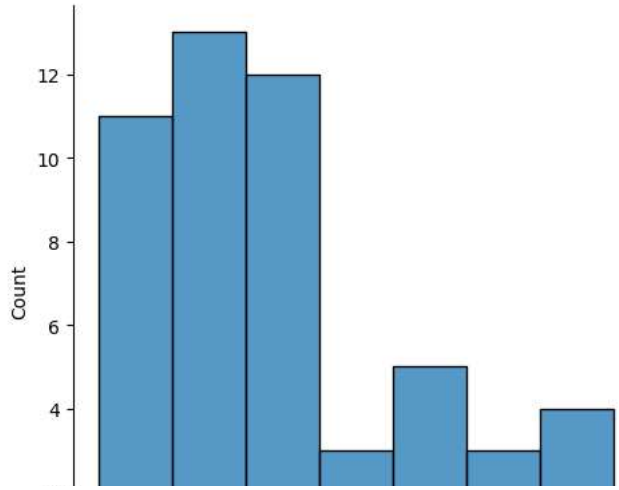
```
<Axes: xlabel='total', ylabel='no_previous'>
```



From the above graph we can say that total no of drivers in fatal collisions increases with increasing in percentage of involved in fatal collisions who do not have previous

```
sns.displot(df['ins_premium'])
```

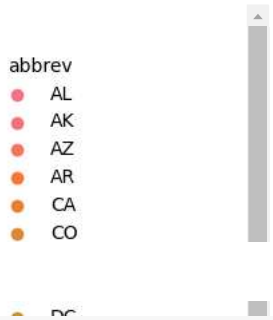
```
<seaborn.axisgrid.FacetGrid at 0x7d2d1474ea10>
```



Insurance premium lies between 700 to 1300

```
sns.relplot(x='alcohol',y='speeding',data=df,hue='abbrev')
```

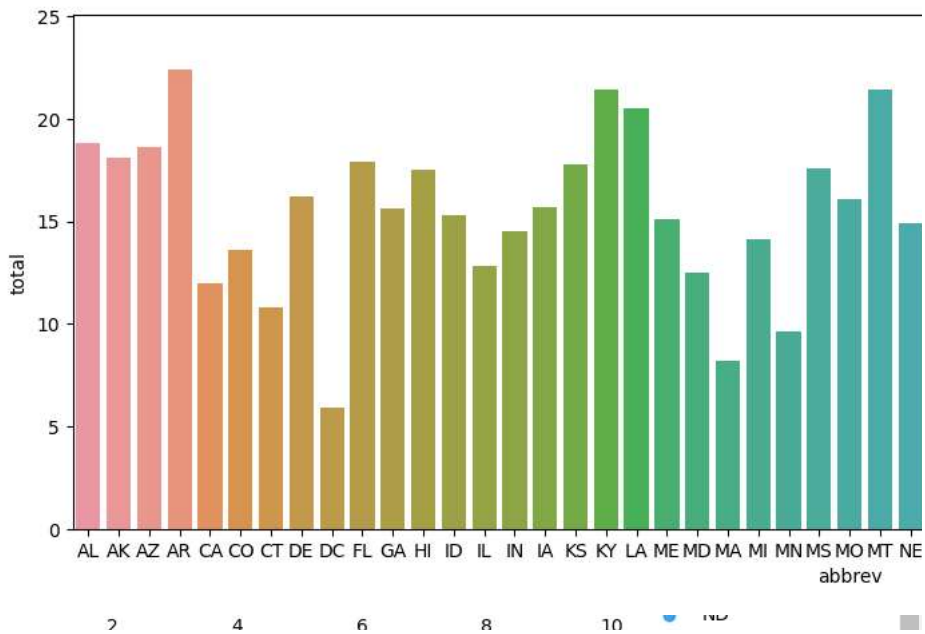
```
<seaborn.axisgrid.FacetGrid at 0x7d2d138ffe50>
```



Speeding increases with increase in consumption of alcohol

```
plt.figure(figsize=(15,5))
sns.barplot(x='abbrev',y='total',data=df)
```

```
<Axes: xlabel='abbrev', ylabel='total'>
```



The total of highest collisions occured in states ND,SC,WV

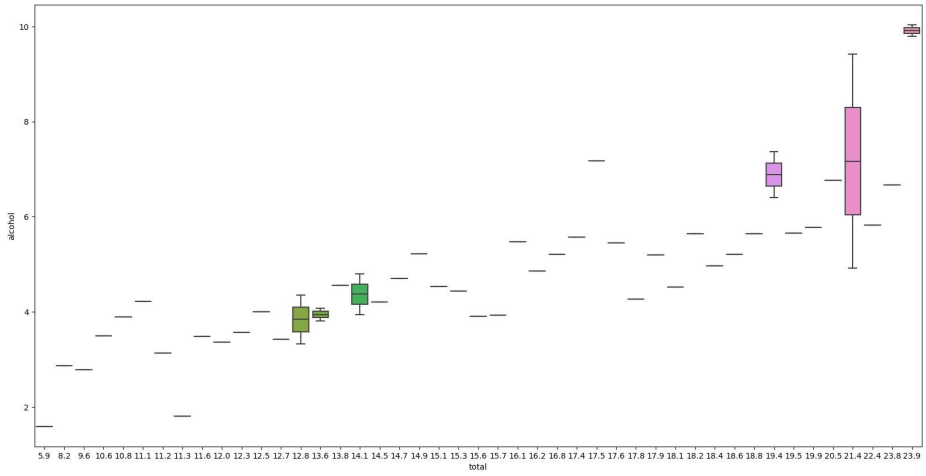
```
sns.jointplot(x='ins_premium',y='ins_losses',data=df)
```

```
<seaborn.axisgrid.JointGrid at 0x7d2d12428040>
```

From the above plot we can say that insurance losses and insurance premium are directly related

```
plt.figure(figsize=(20,10))
sns.boxplot(x='total',y='alcohol',data=df)
```

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



From the above box plot we can say that there are no outliers in our data