Name: Samarth Gayakhe

Reg. No: 21BCE8174

Branch: CSE with AI and ML

Campus: VIT-AP

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
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']
```

```
data = sns.load_dataset('car_crashes')
df = pd.DataFrame(data)
df.head()
```

```
   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

2   18.6     6.510    5.208          15.624       17.856       899.47

3   22.4     4.032    5.824          21.056       21.280       827.34

4   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
```

```
df.info()
```

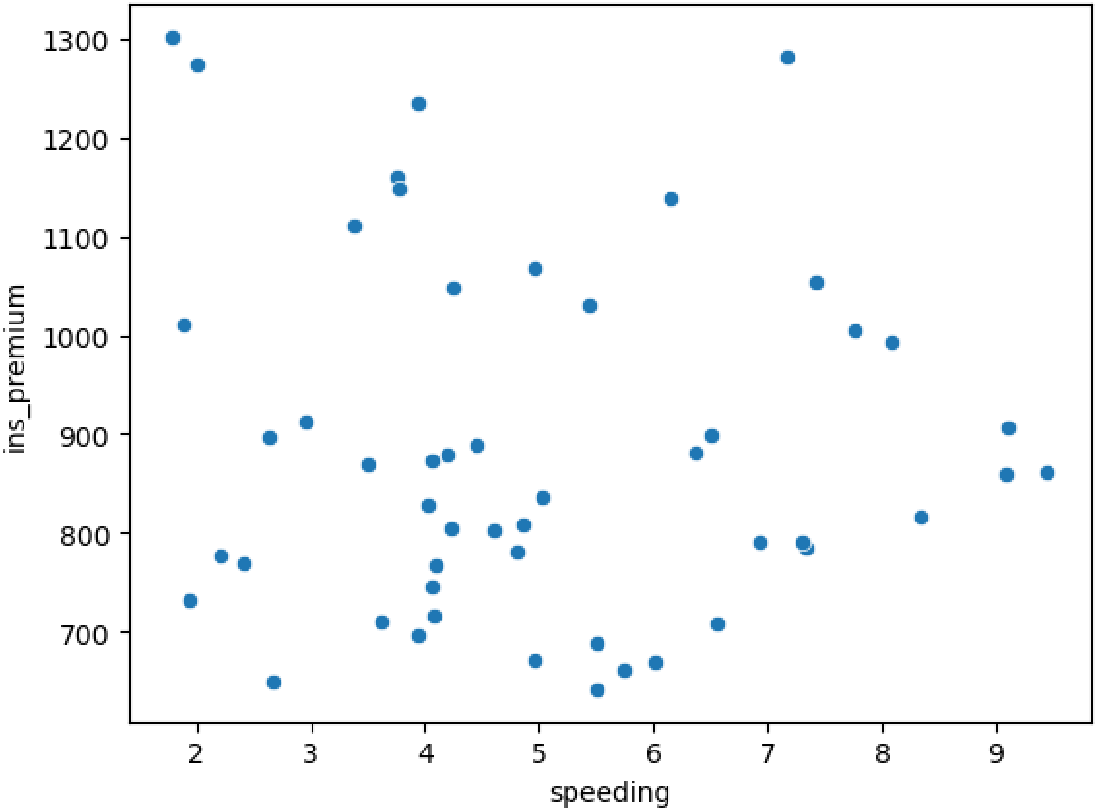
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 8 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   total           51 non-null     float64
 1   speeding        51 non-null     float64
 2   alcohol         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.shape
```

```
(51, 8)
```

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

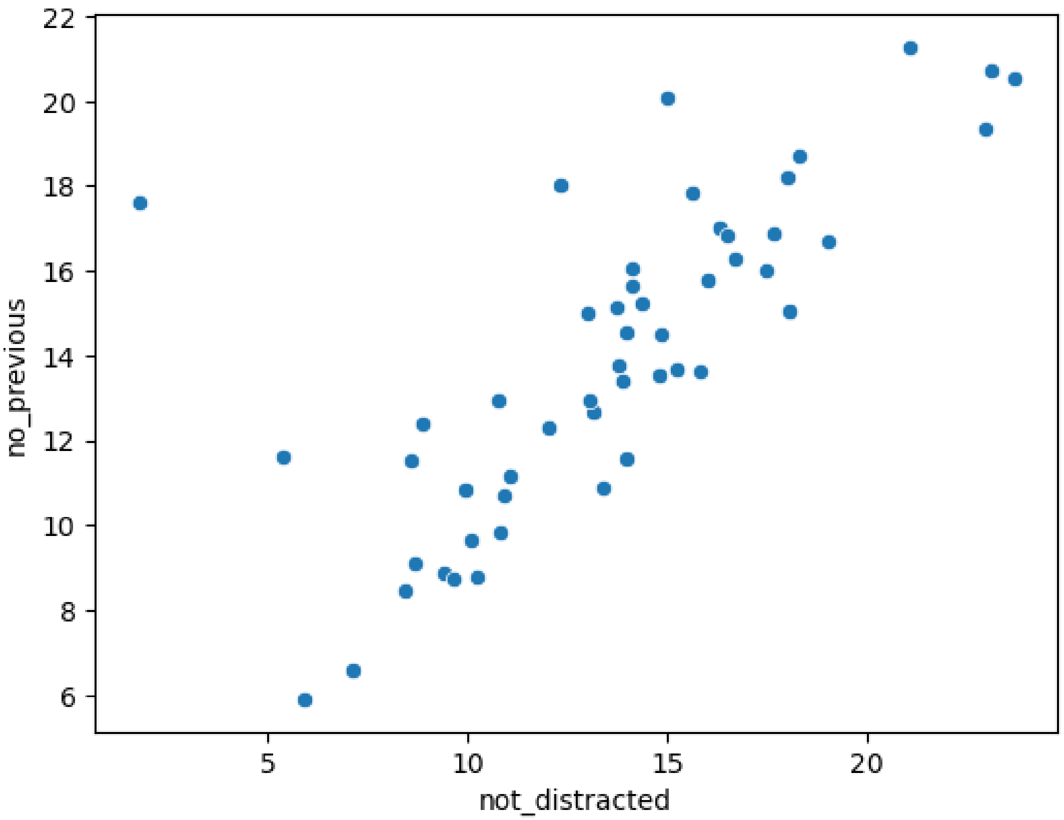
```
<Axes: xlabel='speeding', ylabel='ins_premium'>
```



inference: from the above graph we can see that there is near to zero correlation between Percentage Of Drivers Involved In Fatal Collisions Who Were Speeding and the Car Insurance Premiums.

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

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

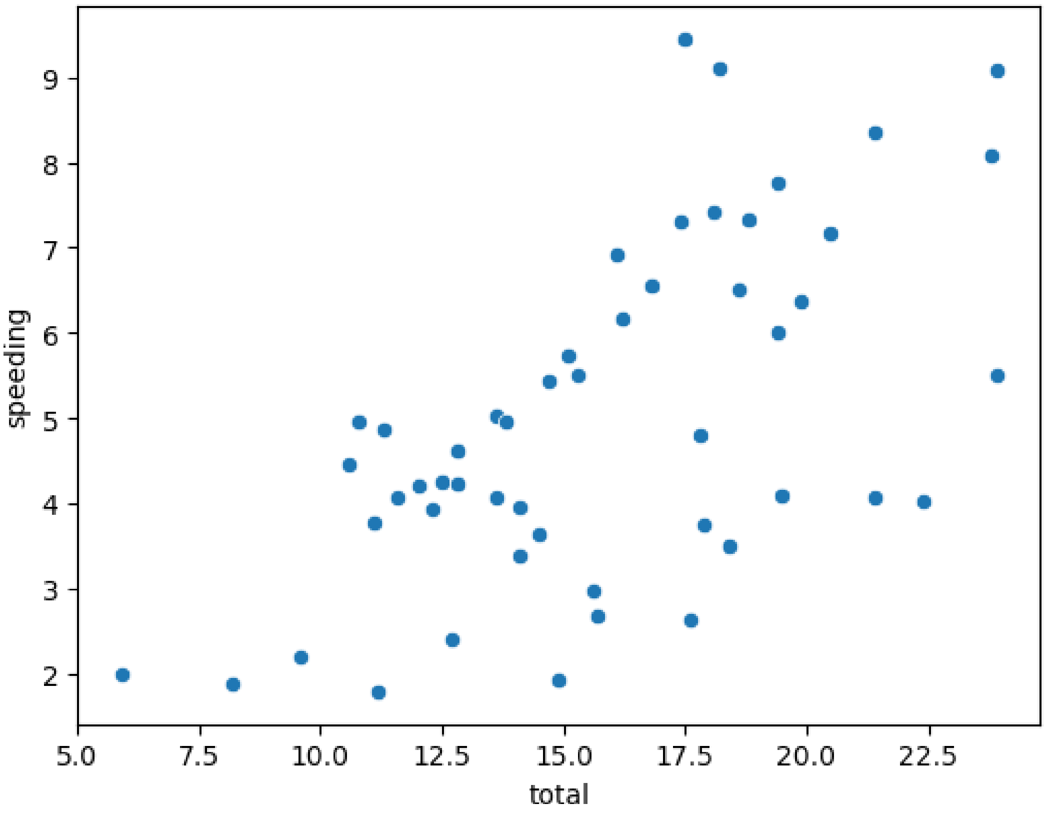


inference: from the above graph we can infere that the Percentage Of Drivers Involved In Fatal Collisions Who Were Not Distracted and Drivers Who Had Not Been Involved In Any Previous Accidents have a positive corelation.

i.e the drivers who were not distraced are more likely not to be involved in any previous accidents

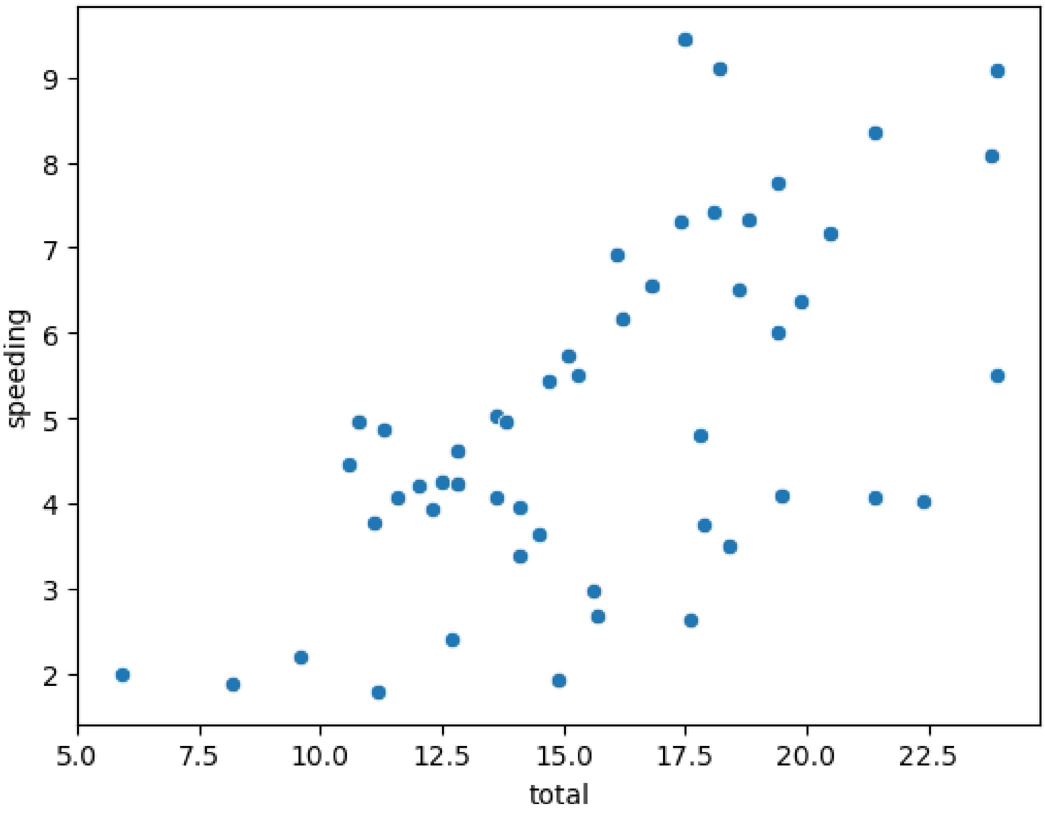
```
sns.scatterplot(x = 'total', y = 'speeding', data = df )
```

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



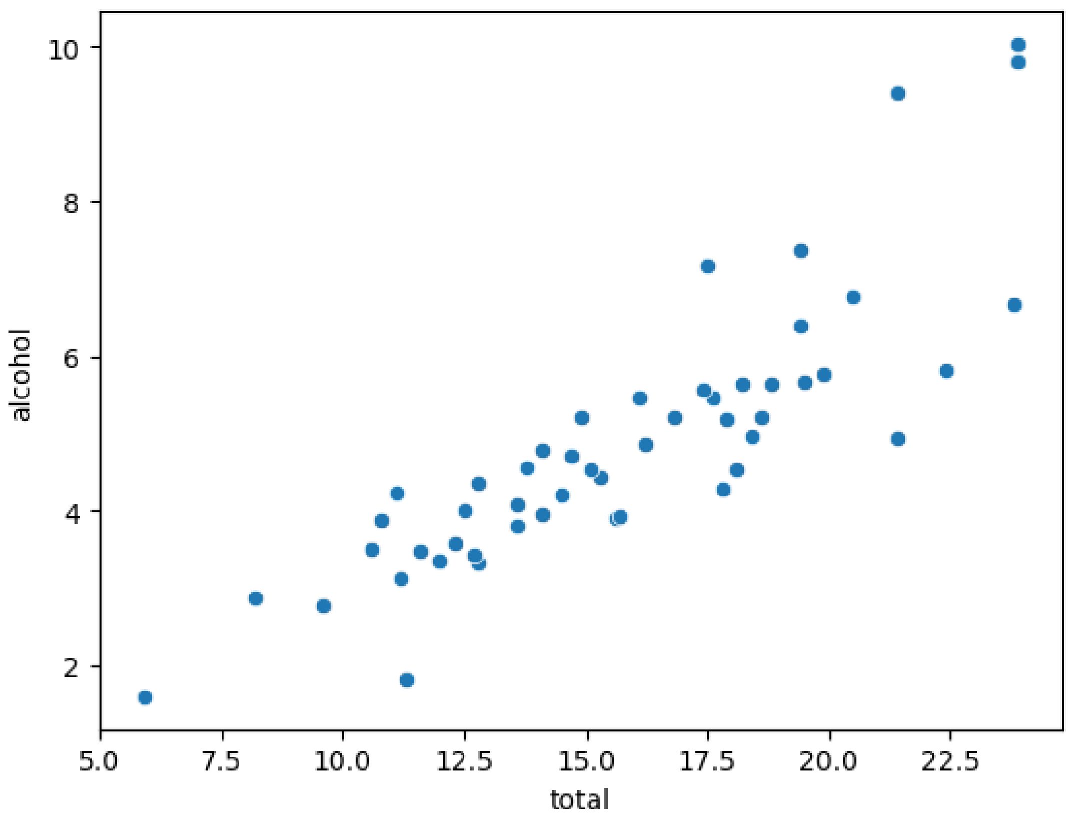
```
sns.scatterplot(x = 'total', y = 'speeding', data = df )
```

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



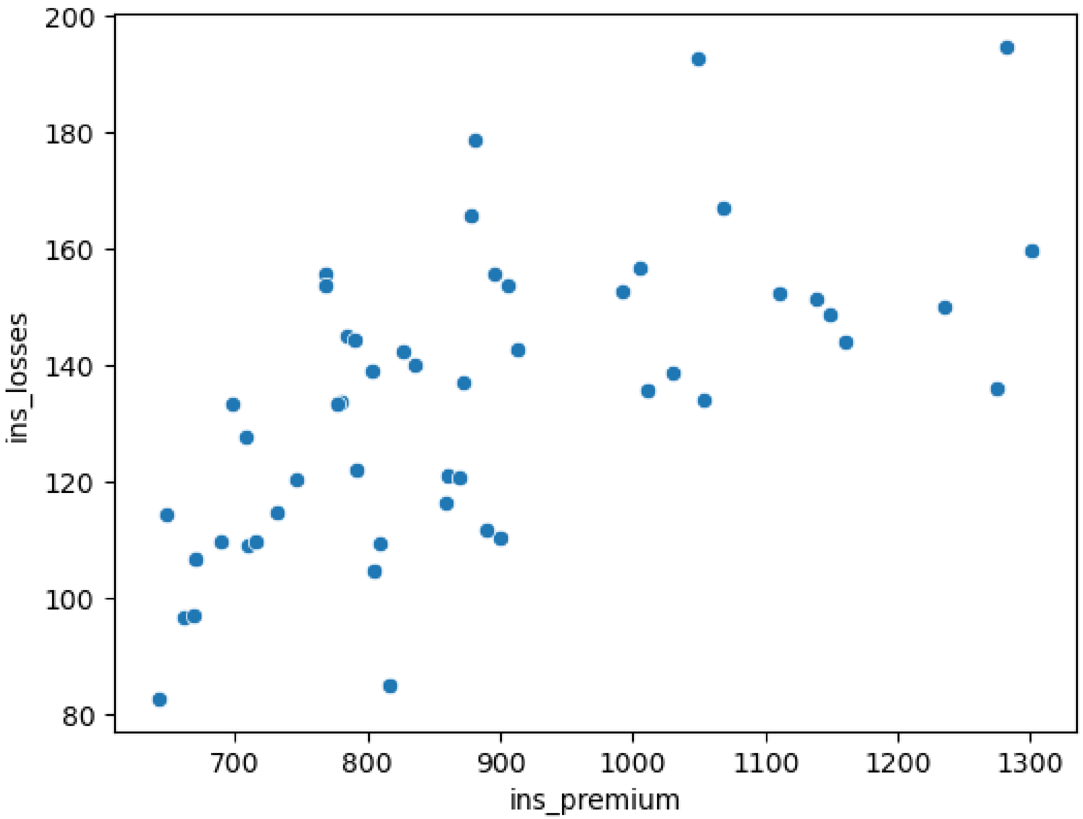
```
sns.scatterplot(x = 'total', y = 'alcohol', data = df )
```

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



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

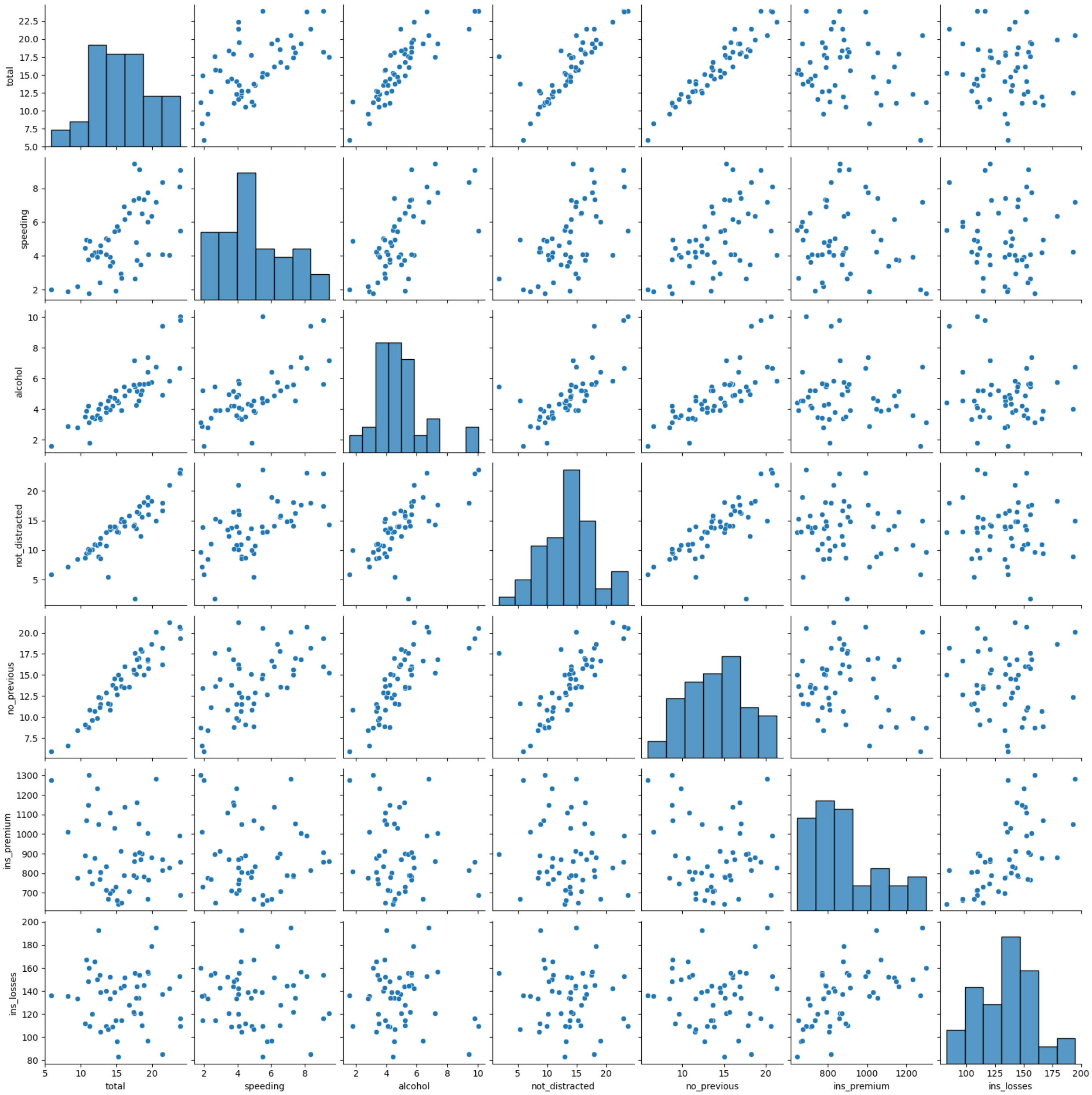
```
<Axes: xlabel='ins_premium', ylabel='ins_losses'>
```



inference: From the above graph we can see that the highere the premium the more losses are incurred by the Insurance companies

```
sns.pairplot(df)
```

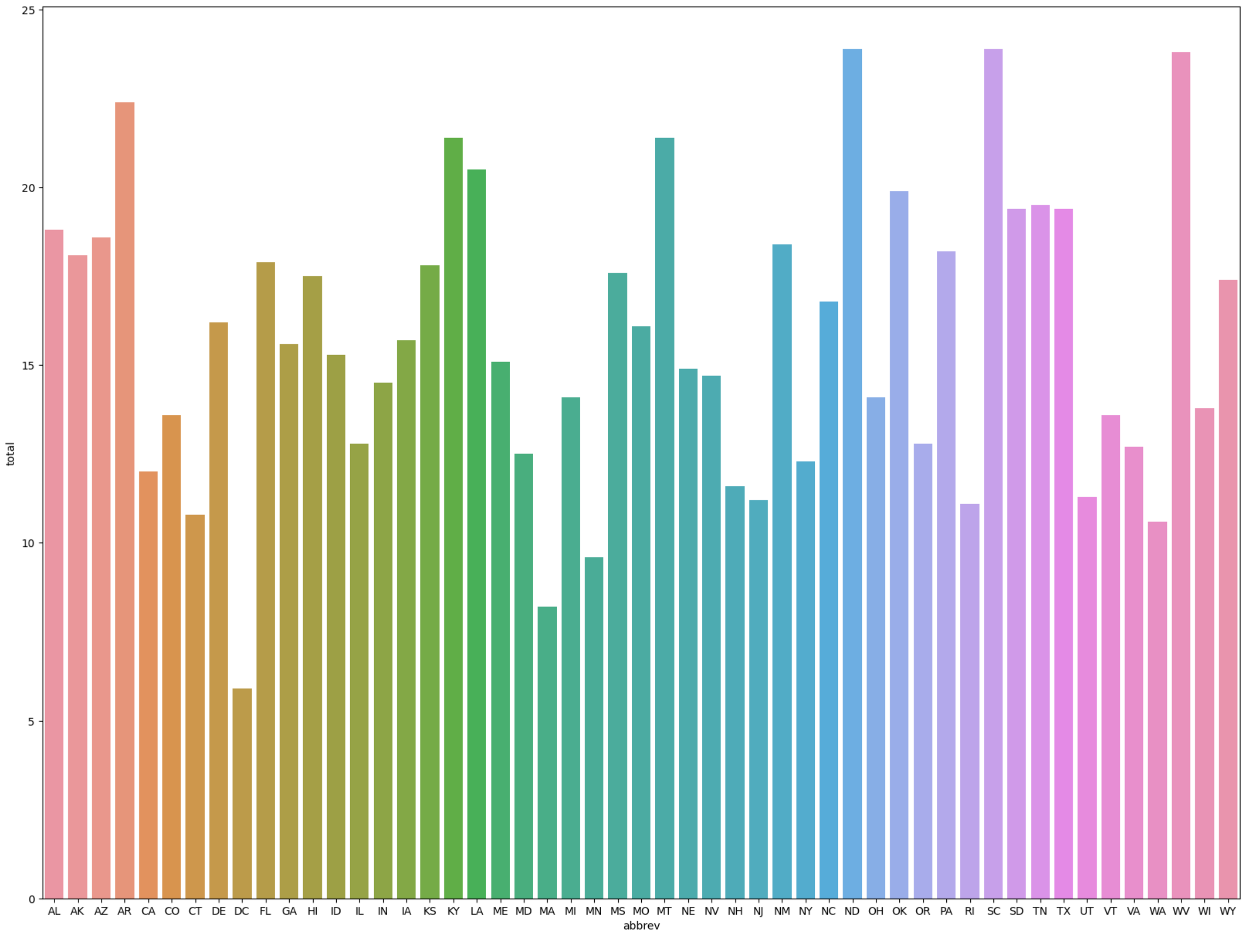
```
<seaborn.axisgrid.PairGrid at 0x14267b690>
```



inference: In the above figure we can see that different features ploted with every other feature

```
plt.subplots(figsize = (20,15))
sns.barplot(x = 'abbrev', y = 'total', data = df)
```

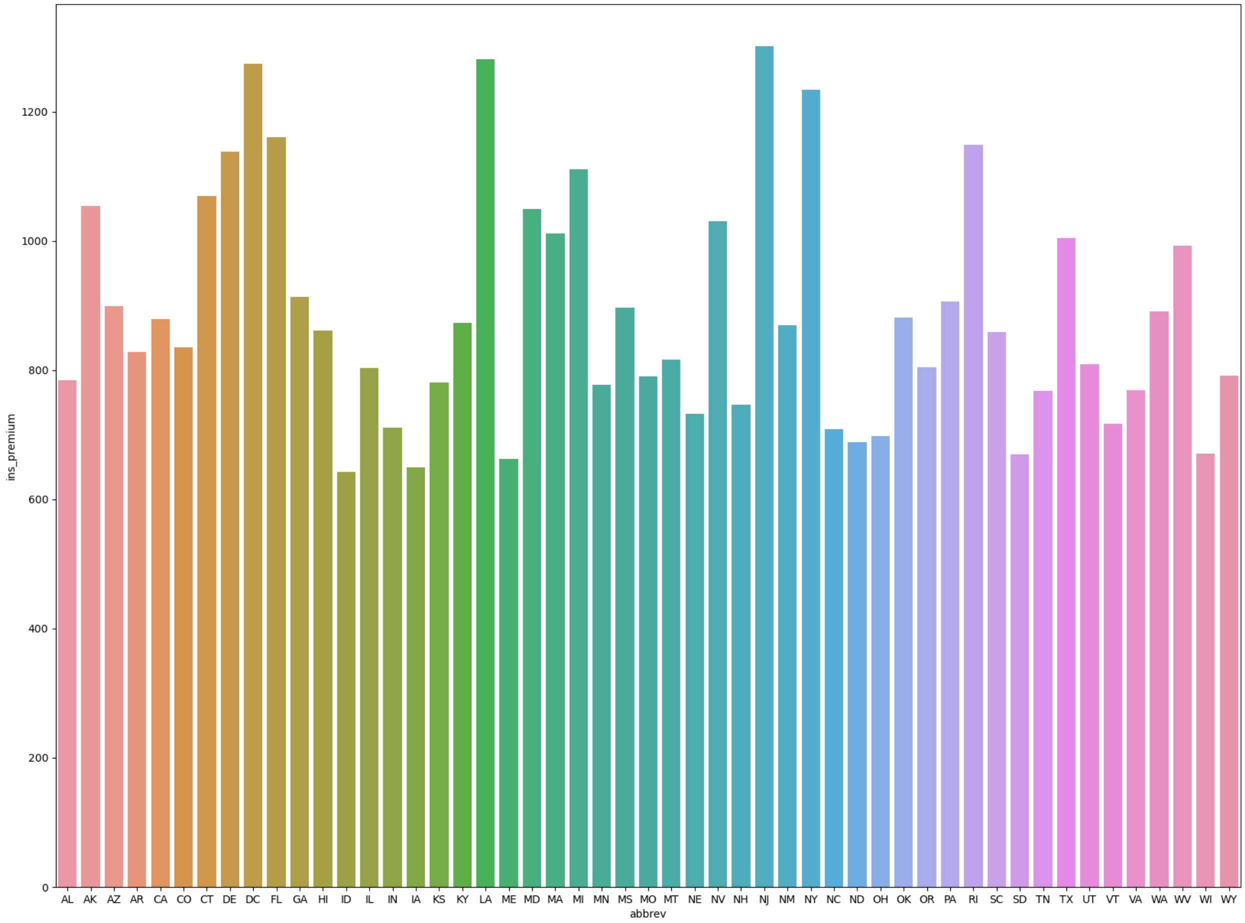
```
<Axes: xlabel='abbrev', ylabel='total'>
```



inference: from the above graph we can see total number of drivers involved in fatal collisions per billion miles in then different states of USA

```
plt.subplots(figsize = (20,15))
sns.barplot(x = 'abbrev', y = 'ins_premium', data = df)
```

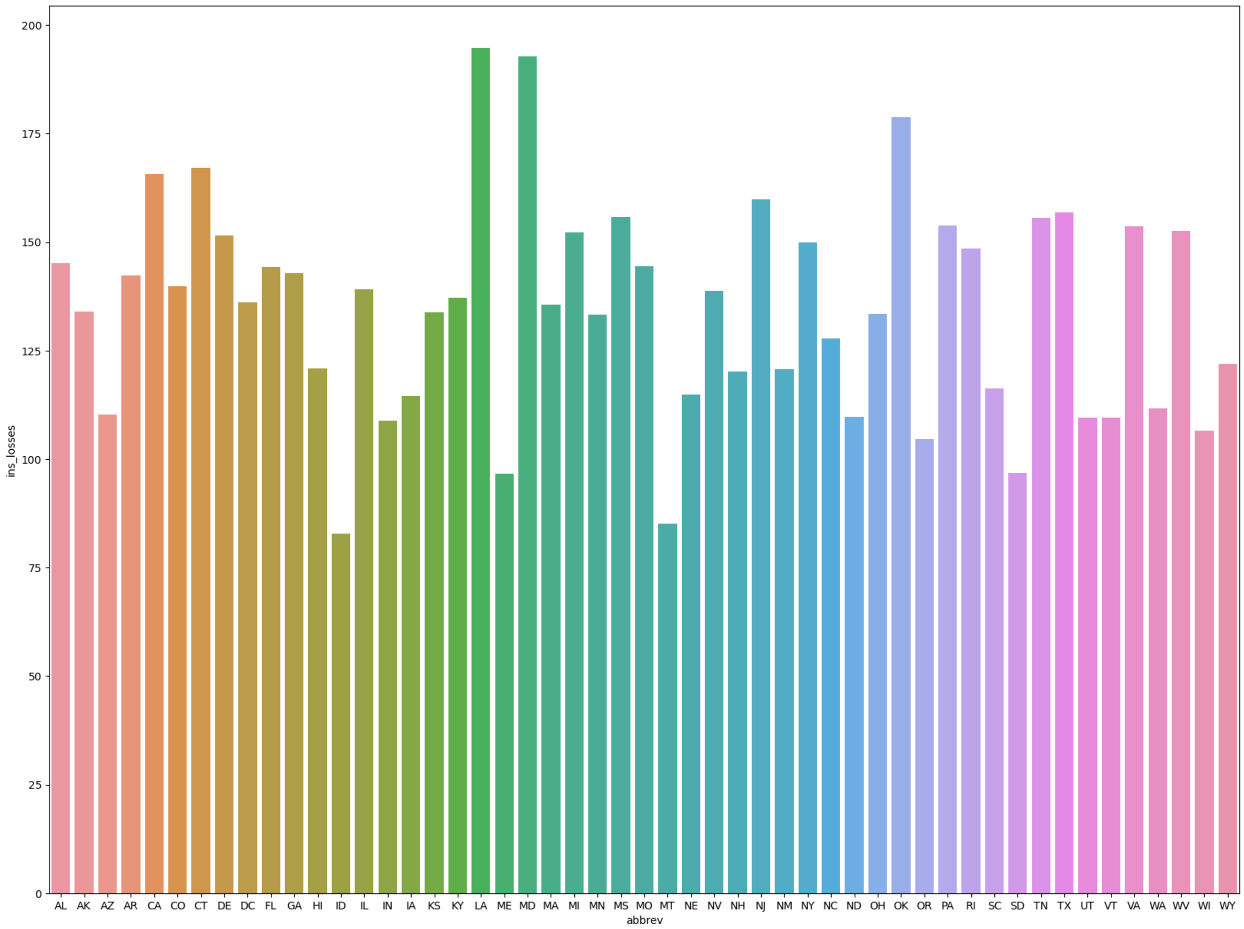
```
<Axes: xlabel='abbrev', ylabel='ins_premium'>
```



inference: from the above graph we can see the insurance premium paid in the differnt states of USA

```
plt.subplots(figsize = (20,15))
sns.barplot(x = 'abbrev', y = 'ins_losses', data = df)
```

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



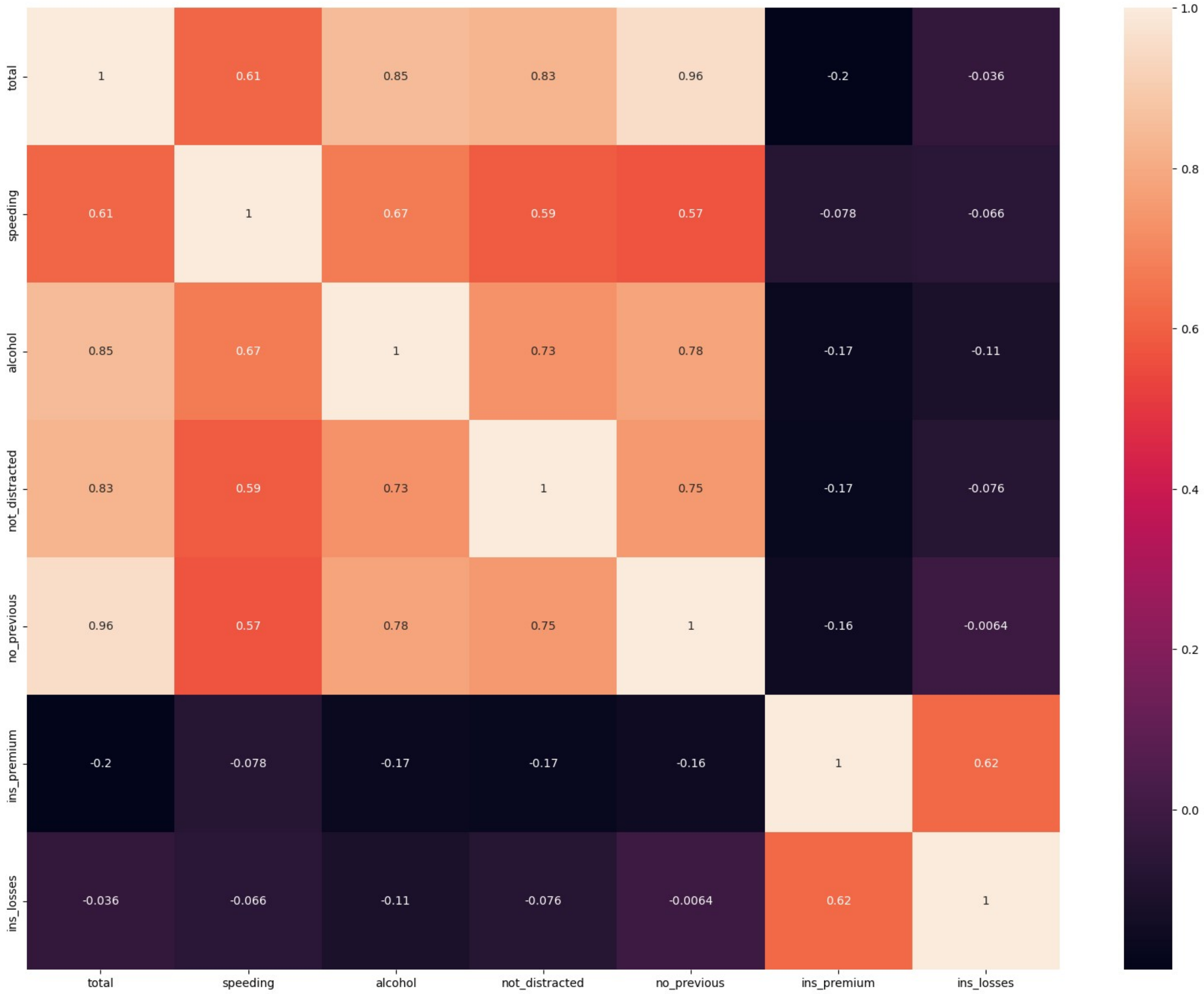
inference: from the above graph we can see the losses incurred by the insurance companies in the differnt states of USA

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

```
/var/folders/m8/dg41v9m11bdcfq4q15h80_l40000gn/T/
ipykernel_83242/658818363.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()
```

```
plt.subplots(figsize = (20,15))
sns.heatmap(corr, annot = True)
```

```
<Axes: >
```



inference: from the above heatmap we can see the Correlation with the diffrent features