assignment-2

September 13, 2023

Importing Libraries

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Loading Dataset

[18]:
```
dataset=pd.read_csv("Car_crashes.csv")
```

[19]:
```
dataset
```

[19]:

	State
0	Alabama
1	Alaska
2	Arizona
3	Arkansas
4	California
5	Colorado
6	Connecticut
7	Delaware
8	District of Columbia
9	Florida
10	Georgia
11	Hawaii
12	Idaho
13	Illinois
14	Indiana
15	Iowa
16	Kansas
17	Kentucky
18	Louisiana
19	Maine
20	Maryland
21	Massachusetts
22	Michigan
23	Minnesota

24	Mississippi	
25	Missouri	
26	Montana	
27	Nebraska	
28	Nevada	
29	New Hampshire	
30	New Jersey	
31	New Mexico	
32	New York	
33	North Carolina	
34	North Dakota	
35	Ohio	
36	Oklahoma	
37	Oregon	
38	Pennsylvania	
39	Rhode Island	
40	South Carolina	
41	South Dakota	
42	Tennessee	
43	Texas	
44	Utah	
45	Vermont	
46	Virginia	
47	Washington	
48	West Virginia	
49	Wisconsin	
50	Wyoming	

	Number of drivers involved in fatal collisions per billion miles \
0	18.8
1	18.1
2	18.6
3	22.4
4	12.0
5	13.6
6	10.8
7	16.2
8	5.9
9	17.9
10	15.6
11	17.5
12	15.3
13	12.8
14	14.5
15	15.7
16	17.8
17	21.4

18	20.5
19	15.1
20	12.5
21	8.2
22	14.1
23	9.6
24	17.6
25	16.1
26	21.4
27	14.9
28	14.7
29	11.6
30	11.2
31	18.4
32	12.3
33	16.8
34	23.9
35	14.1
36	19.9
37	12.8
38	18.2
39	11.1
40	23.9
41	19.4
42	19.5
43	19.4
44	11.3
45	13.6
46	12.7
47	10.6
48	23.8
49	13.8
50	17.4

	Percentage Of Drivers Involved In Fatal Collisions Who Were Speeding \
0	39
1	41
2	35
3	18
4	35
5	37
6	46
7	38
8	34
9	21
10	19
11	54

12	36
13	36
14	25
15	17
16	27
17	19
18	35
19	38
20	34
21	23
22	24
23	23
24	15
25	43
26	39
27	13
28	37
29	35
30	16
31	19
32	32
33	39
34	23
35	28
36	32
37	33
38	50
39	34
40	38
41	31
42	21
43	40
44	43
45	30
46	19
47	42
48	34
49	36
50	42

	Percentage Of Drivers Involved In Fatal Collisions Who Were Alcohol-Impaired \
0	30
1	25
2	28
3	26
4	28

5	28
6	36
7	30
8	27
9	29
10	25
11	41
12	29
13	34
14	29
15	25
16	24
17	23
18	33
19	30
20	32
21	35
22	28
23	29
24	31
25	34
26	44
27	35
28	32
29	30
30	28
31	27
32	29
33	31
34	42
35	34
36	29
37	26
38	31
39	38
40	41
41	33
42	29
43	38
44	16
45	30
46	27
47	33
48	28
49	33
50	32

\	Percentage Of Drivers Involved In Fatal Collisions Who Were Not Distracted
0	96
1	90
2	84
3	94
4	91
5	79
6	87
7	87
8	100
9	92
10	95
11	82
12	85
13	94
14	95
15	97
16	77
17	78
18	73
19	87
20	71
21	87
22	95
23	88
24	10
25	92
26	84
27	93
28	95
29	87
30	86
31	67
32	88
33	94
34	99
35	99
36	92
37	67
38	96
39	92
40	96
41	98
42	82
43	91
44	88

45	96
46	87
47	82
48	97
49	39
50	81

	Percentage Of Drivers Involved In Fatal Collisions Who Had Not Been Involved In Any Previous Accidents \
0	80
1	94
2	96
3	95
4	89
5	95
6	82
7	99
8	100
9	94
10	93
11	87
12	98
13	96
14	95
15	87
16	85
17	76
18	98
19	84
20	99
21	80
22	77
23	88
24	100
25	84
26	85
27	90
28	99
29	83
30	78
31	98
32	80
33	81
34	86
35	82
36	94
37	90

38		88
39		79
40		81
41		86
42		81
43		87
44		96
45		95
46		88
47		86
48		87
49		84
50		90

	Car Insurance Premiums ($)	\
0	784.55	
1	1053.48	
2	899.47	
3	827.34	
4	878.41	
5	835.50	
6	1068.73	
7	1137.87	
8	1273.89	
9	1160.13	
10	913.15	
11	861.18	
12	641.96	
13	803.11	
14	710.46	
15	649.06	
16	780.45	
17	872.51	
18	1281.55	
19	661.88	
20	1048.78	
21	1011.14	
22	1110.61	
23	777.18	
24	896.07	
25	790.32	
26	816.21	
27	732.28	
28	1029.87	
29	746.54	
30	1301.52	
31	869.85	

32	1234.31
33	708.24
34	688.75
35	697.73
36	881.51
37	804.71
38	905.99
39	1148.99
40	858.97
41	669.31
42	767.91
43	1004.75
44	809.38
45	716.20
46	768.95
47	890.03
48	992.61
49	670.31
50	791.14

	Losses incurred by insurance companies for collisions per insured driver ($)
0	145.08
1	133.93
2	110.35
3	142.39
4	165.63
5	139.91
6	167.02
7	151.48
8	136.05
9	144.18
10	142.80
11	120.92
12	82.75
13	139.15
14	108.92
15	114.47
16	133.80
17	137.13
18	194.78
19	96.57
20	192.70
21	135.63
22	152.26
23	133.35
24	155.77
25	144.45

```
26                                                  85.15
27                                                 114.82
28                                                 138.71
29                                                 120.21
30                                                 159.85
31                                                 120.75
32                                                 150.01
33                                                 127.82
34                                                 109.72
35                                                 133.52
36                                                 178.86
37                                                 104.61
38                                                 153.86
39                                                 148.58
40                                                 116.29
41                                                  96.87
42                                                 155.57
43                                                 156.83
44                                                 109.48
45                                                 109.61
46                                                 153.72
47                                                 111.62
48                                                 152.56
49                                                 106.62
50                                                 122.04
```

[20]:
```
dataset.head()
```

[20]:
```
         State  \
0      Alabama
1       Alaska
2      Arizona
3     Arkansas
4   California

   Number of drivers involved in fatal collisions per billion miles  \
0                                                 18.8
1                                                 18.1
2                                                 18.6
3                                                 22.4
4                                                 12.0

   Percentage Of Drivers Involved In Fatal Collisions Who Were Speeding  \
0                                                 39
1                                                 41
2                                                 35
3                                                 18
```

```
4                                                  35

   Percentage Of Drivers Involved In Fatal Collisions Who Were Alcohol-Impaired
\
0                                                  30
1                                                  25
2                                                  28
3                                                  26
4                                                  28

   Percentage Of Drivers Involved In Fatal Collisions Who Were Not Distracted  \
0                                                  96
1                                                  90
2                                                  84
3                                                  94
4                                                  91

   Percentage Of Drivers Involved In Fatal Collisions Who Had Not Been Involved
In Any Previous Accidents  \
0                                                  80
1                                                  94
2                                                  96
3                                                  95
4                                                  89

   Car Insurance Premiums ($)  \
0                      784.55
1                     1053.48
2                      899.47
3                      827.34
4                      878.41

   Losses incurred by insurance companies for collisions per insured driver ($)
0                                             145.08
1                                             133.93
2                                             110.35
3                                             142.39
4                                             165.63
```

[21]:
```
dataset.tail()
```

[21]:
```
            State  \
46       Virginia
47     Washington
48  West Virginia
49      Wisconsin
50        Wyoming
```

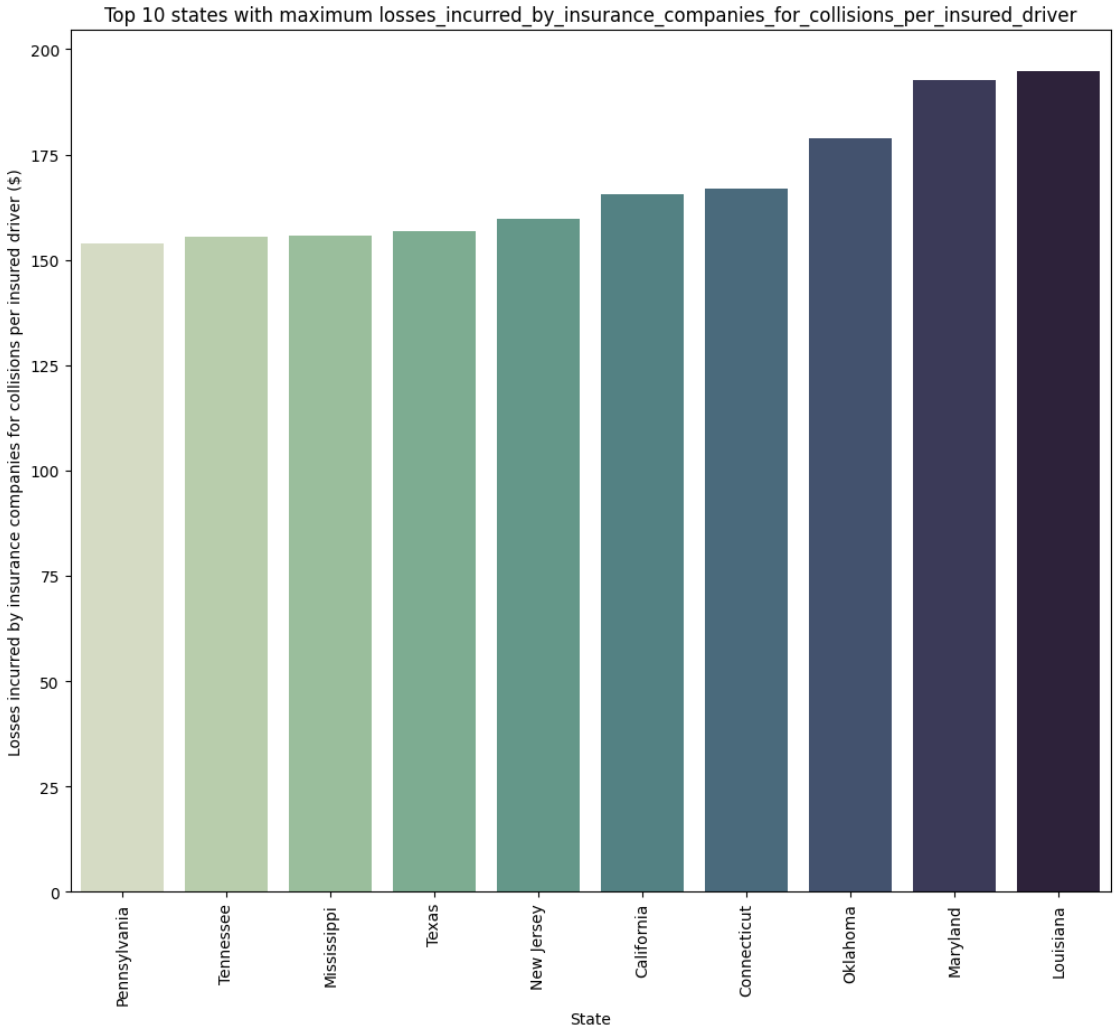
```
    Number of drivers involved in fatal collisions per billion miles  \
46                                                  12.7
47                                                  10.6
48                                                  23.8
49                                                  13.8
50                                                  17.4

    Percentage Of Drivers Involved In Fatal Collisions Who Were Speeding  \
46                                                   19
47                                                   42
48                                                   34
49                                                   36
50                                                   42

    Percentage Of Drivers Involved In Fatal Collisions Who Were Alcohol-Impaired
\
46                                                   27
47                                                   33
48                                                   28
49                                                   33
50                                                   32

    Percentage Of Drivers Involved In Fatal Collisions Who Were Not Distracted
\
46                                                   87
47                                                   82
48                                                   97
49                                                   39
50                                                   81

    Percentage Of Drivers Involved In Fatal Collisions Who Had Not Been Involved
In Any Previous Accidents  \
46                                                   88
47                                                   86
48                                                   87
49                                                   84
50                                                   90

    Car Insurance Premiums ($)  \
46                      768.95
47                      890.03
48                      992.61
49                      670.31
50                      791.14

    Losses incurred by insurance companies for collisions per insured driver ($)
```

46	153.72
47	111.62
48	152.56
49	106.62
50	122.04

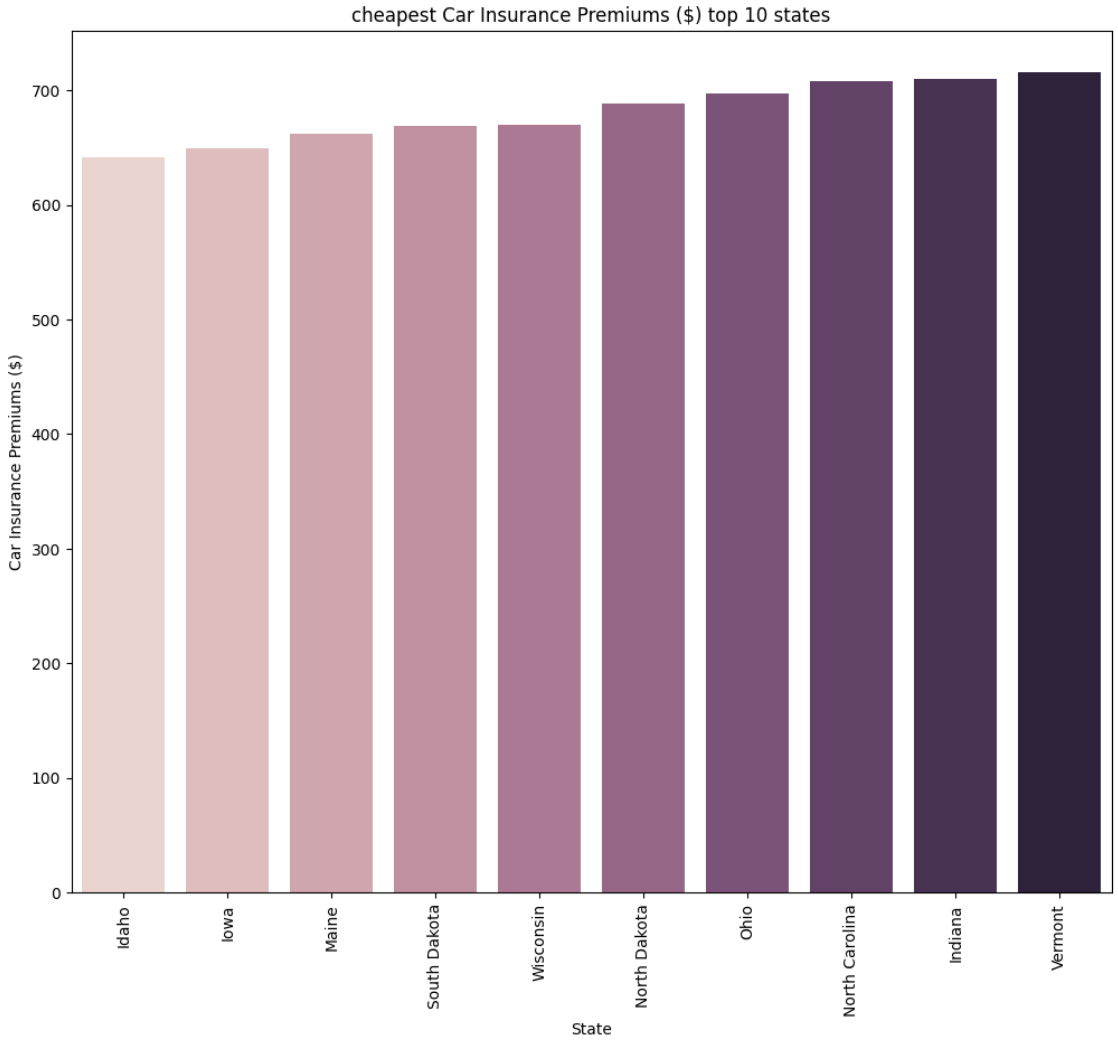
Data Visualisation

1) Univariate

[34]:
```
plt.figure(figsize=(12,10))
d2 = dataset[["State",'Number of drivers involved in fatal collisions per␣
 ↪billion miles','Percentage Of Drivers Involved In Fatal Collisions Who Were␣
 ↪Speeding','Percentage Of Drivers Involved In Fatal Collisions Who Were␣
 ↪Alcohol-Impaired','Percentage Of Drivers Involved In Fatal Collisions Who␣
 ↪Were Not Distracted','Percentage Of Drivers Involved In Fatal Collisions Who␣
 ↪Had Not Been Involved In Any Previous Accidents','Car Insurance Premiums␣
 ↪($)','Losses incurred by insurance companies for collisions per insured␣
 ↪driver ($)']].sort_values(by='Losses incurred by insurance companies for␣
 ↪collisions per insured driver ($)').tail(10)
sns.barplot(x=d2['State'], y=d2['Losses incurred by insurance companies for␣
 ↪collisions per insured driver ($)'],palette=sns.
 ↪cubehelix_palette(len(d2['State']),start=0.5,rot=-0.75))
plt.xticks(rotation= 90)
plt.xlabel('State')
plt.ylabel('Losses incurred by insurance companies for collisions per insured␣
 ↪driver ($)')
plt.title('Top 10 states with maximum␣
 ↪losses_incurred_by_insurance_companies_for_collisions_per_insured_driver')
plt.show()
```



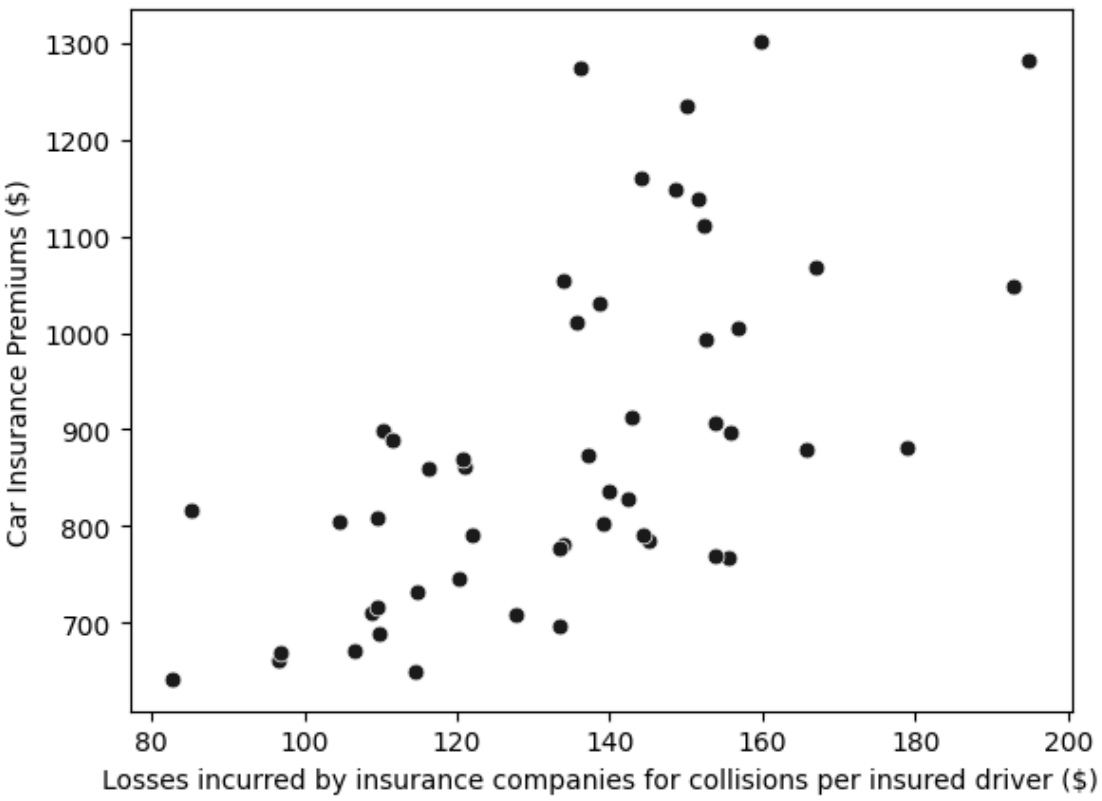
[36]:
```
plt.figure(figsize=(12,10))
d3 = dataset[["State","Car Insurance Premiums ($)"]].sort_values(by="Car␣
 ↪Insurance Premiums ($)").head(10)
sns.barplot(x=d3['State'], y=d3['Car Insurance Premiums ($)'],palette=sns.
 ↪cubehelix_palette(len(d3['State'])))
plt.xticks(rotation= 90)
plt.xlabel('State')
plt.ylabel('Car Insurance Premiums ($)')
plt.title('cheapest Car Insurance Premiums ($) top 10 states')
plt.show()
```



2) Bivariate

[39]:
```
sns.scatterplot(x="Losses incurred by insurance companies for collisions per␣
↪insured driver ($)", y="Car Insurance Premiums ($)", data=dataset, s=40,␣
↪color=".1")
```

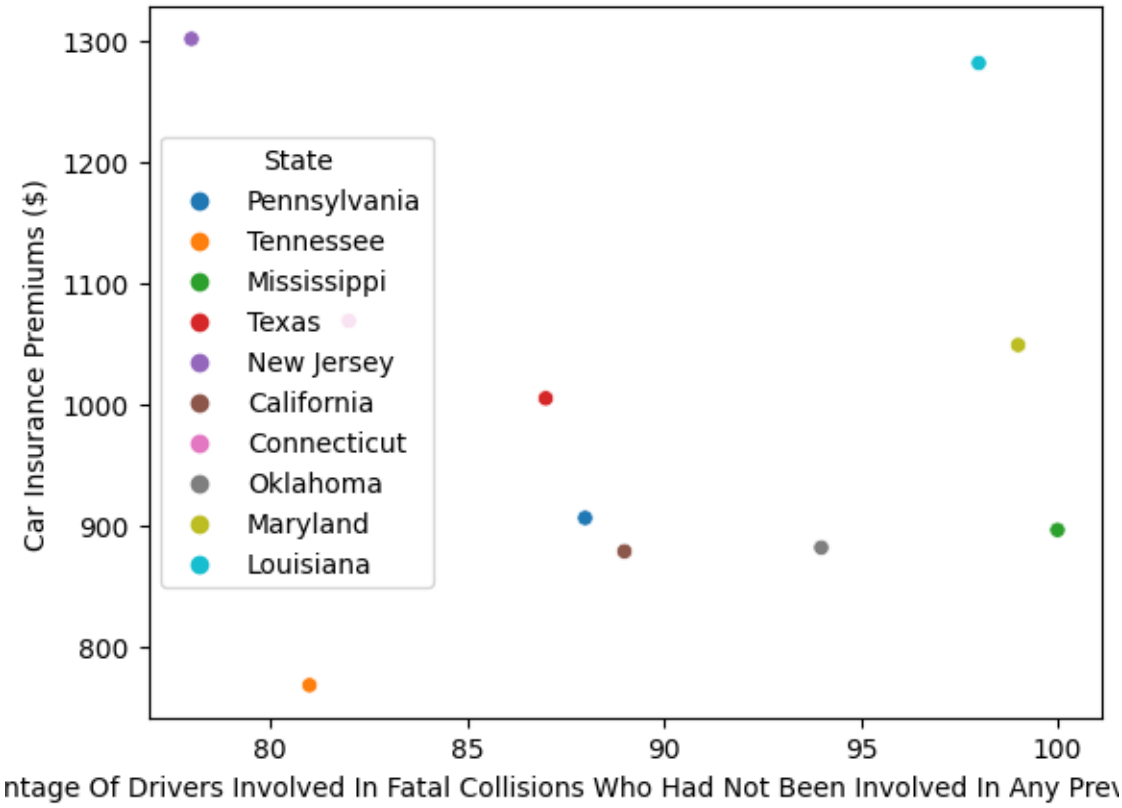
[39]:
```
<Axes: xlabel='Losses incurred by insurance companies for collisions per insured
driver ($)', ylabel='Car Insurance Premiums ($)'>
```



The scatterplot shows a linear relationship between car insurance premium and the losses incurred

3) Multivariate

[45]:
```
d5 = dataset[["State",'Number of drivers involved in fatal collisions per␣
 ↪billion miles','Percentage Of Drivers Involved In Fatal Collisions Who Were␣
 ↪Speeding','Percentage Of Drivers Involved In Fatal Collisions Who Were␣
 ↪Alcohol-Impaired','Percentage Of Drivers Involved In Fatal Collisions Who␣
 ↪Were Not Distracted','Percentage Of Drivers Involved In Fatal Collisions Who␣
 ↪Had Not Been Involved In Any Previous Accidents','Car Insurance Premiums␣
 ↪($)','Losses incurred by insurance companies for collisions per insured␣
 ↪driver ($)']].sort_values(by='Losses incurred by insurance companies for␣
 ↪collisions per insured driver ($)').tail(10)
ax = sns.scatterplot(x=d5["Percentage Of Drivers Involved In Fatal Collisions␣
 ↪Who Had Not Been Involved In Any Previous Accidents"], y=d5["Car Insurance␣
 ↪Premiums ($)"],
                     hue=d5["State"])
```



Car insurance premiums are really high when % of drivers not involved in previous accident is below 85%. This also means that there are higher chances that the insurance company will have to bear the cost for the same insured driver again and if the insurance premium is low in those states then more likely that the company will incur huge losses. Thus to reduce the loss margins the insurance premiums are high in those states which again signifies that the states have a lot of high risk drivers