```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns


df = sns.load_dataset("car_crashes")


df
```

19	15.1	5.738	4.530	13.137	12.684	661.88	96.57	ME
20	12.5	4.250	4.000	8.875	12.375	1048.78	192.70	MD
21	8.2	1.886	2.870	7.134	6.560	1011.14	135.63	MA
22	14.1	3.384	3.948	13.395	10.857	1110.61	152.26	MI
23	9.6	2.208	2.784	8.448	8.448	777.18	133.35	MN
24	17.6	2.640	5.456	1.760	17.600	896.07	155.77	MS
25	16.1	6.923	5.474	14.812	13.524	790.32	144.45	MO
26	21.4	8.346	9.416	17.976	18.190	816.21	85.15	MT
27	14.9	1.937	5.215	13.857	13.410	732.28	114.82	NE
28	14.7	5.439	4.704	13.965	14.553	1029.87	138.71	NV
29	11.6	4.060	3.480	10.092	9.628	746.54	120.21	NH
30	11.2	1.792	3.136	9.632	8.736	1301.52	159.85	NJ
31	18.4	3.496	4.968	12.328	18.032	869.85	120.75	NM
32	12.3	3.936	3.567	10.824	9.840	1234.31	150.01	NY
33	16.8	6.552	5.208	15.792	13.608	708.24	127.82	NC
34	23.9	5.497	10.038	23.661	20.554	688.75	109.72	ND
35	14.1	3.948	4.794	13.959	11.562	697.73	133.52	OH
36	19.9	6.368	5.771	18.308	18.706	881.51	178.86	OK
37	12.8	4.224	3.328	8.576	11.520	804.71	104.61	OR
38	18.2	9.100	5.642	17.472	16.016	905.99	153.86	PA
39	11.1	3.774	4.218	10.212	8.769	1148.99	148.58	RI
40	23.9	9.082	9.799	22.944	19.359	858.97	116.29	SC
41	19.4	6.014	6.402	19.012	16.684	669.31	96.87	SD

42	19.5	4.095	5.655	15.990	15.795	767.91	155.57	TN
43	19.4	7.760	7.372	17.654	16.878	1004.75	156.83	TX
44	11.3	4.859	1.808	9.944	10.848	809.38	109.48	UT
45	13.6	4.080	4.080	13.056	12.920	716.20	109.61	VT
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.head()
```

	total	speeding	alcohol	not distracted	no previous	ins premium	ins losses	abbrev

```
df.shape
```

```
(51, 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
```

```
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

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

```
<ipython-input-8-7d5195e2bf4d>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is
  corr=df.corr()
```

	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	0.623116
ins_losses	-0.036011	-0.065928	-0.112547	-0.075970	-0.006359	0.623116	1.000000

```
plt.subplots(figsize=(10,5))
```

```
(<Figure size 1000x500 with 1 Axes>, <Axes: >)
```


```
df.speeding.value_counts()
```

```
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
```

```
df.alcohol.value_counts()
```

```
5.208     2
5.640     1
4.218     1
4.704     1
3.480     1
3.136     1
4.968     1
3.567     1
10.038    1
4.794     1
5.771     1
3.328     1
5.642     1
```

```
9.799     1
9.416     1
6.402     1
5.655     1
7.372     1
1.808     1
4.080     1
3.429     1
3.498     1
6.664     1
4.554     1
5.215     1
5.474     1
4.525     1
5.456     1
5.824     1
3.360     1
3.808     1
3.888     1
4.860     1
1.593     1
5.191     1
3.900     1
7.175     1
4.437     1
4.352     1
4.205     1
3.925     1
4.272     1
4.922     1
6.765     1
4.530     1
4.000     1
2.870     1
3.948     1
2.784     1
5.568     1
Name: alcohol, dtype: int64
```

Handling Null Values

```
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.isnull().sum()
```

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

```
# There is no null values so no need to handle it
```

Seperate dependent and independent variables

```
x=df.iloc[:,1:13]
```

```
x.head()
```

	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses	abbrev
0	7.332	5.640	18.048	15.040	784.55	145.08	AL
1	7.421	4.525	16.290	17.014	1053.48	133.93	AK
2	6.510	5.208	15.624	17.856	899.47	110.35	AZ
3	4.032	5.824	21.056	21.280	827.34	142.39	AR
4	4.200	3.360	10.920	10.680	878.41	165.63	CA

```
y=df.iloc[:,7:8]
```

```
y.head()
```

	abbrev
0	AL
1	AK
2	AZ
3	AR
4	CA

```
x.shape
```

```
(51, 7)
```

```
y.shape
```

```
(51, 1)
```

Encoding

```
from sklearn.preprocessing import LabelEncoder
```

```
le=LabelEncoder()
```

```
x["not_distracted"]=le.fit_transform(x["not_distracted"])
```

```
x["not_distracted"]
```

```
0     43
1     37
2     34
3     46
4     15
5     13
6      8
7     28
8      2
9     38
10    31
11    29
12    19
13    17
14    24
15    33
16    23
17    39
18    32
19    21
20     7
21     3
22    22
```

```
23     4
24     0
25    30
26    42
27    25
28    27
29    11
30     9
31    18
32    14
33    35
34    49
35    26
36    44
37     5
38    40
39    12
40    47
41    45
42    36
43    41
44    10
45    20
46    16
47     6
48    48
49     1
50    28
Name: not_distracted, dtype: int64
```

```
x.head()
```

	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses	abbrev
0	42	5.640	43	15.040	784.55	145.08	1
1	43	4.525	37	17.014	1053.48	133.93	0

```
x["not_distracted"].value_counts()
```

```
28    2
43    1
40    1
27    1
11    1
9     1
18    1
14    1
35    1
49    1
26    1
44    1
5     1
12    1
42    1
47    1
45    1
36    1
41    1
10    1
20    1
16    1
6     1
48    1
25    1
30    1
37    1
0     1
34    1
46    1
15    1
13    1
8     1
```

```
2     1
38    1
31    1
29    1
19    1
17    1
24    1
33    1
23    1
39    1
32    1
21    1
7     1
3     1
22    1
4     1
1     1
Name: not_distracted, dtype: int64
```

```
x["not_distracted"].nunique()
```

```
50
```

```
x.head()
```

	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses	abbrev
0	42	5.640	43	15.040	784.55	145.08	1
1	43	4.525	37	17.014	1053.48	133.93	0
2	37	5.208	34	17.856	899.47	110.35	3
3	16	5.824	46	21.280	827.34	142.39	2
4	21	3.360	15	10.680	878.41	165.63	4

```
x.alcohol.value_counts()
```

```
5.208     2
5.640     1
4.218     1
4.704     1
3.480     1
3.136     1
4.968     1
3.567     1
10.038    1
4.794     1
5.771     1
3.328     1
5.642     1
9.799     1
9.416     1
6.402     1
5.655     1
7.372     1
1.808     1
4.080     1
3.429     1
3.498     1
6.664     1
4.554     1
5.215     1
5.474     1
4.525     1
5.456     1
5.824     1
3.360     1
3.808     1
3.888     1
4.860     1
1.593     1
5.191     1
3.900     1
7.175     1
4.437     1
4.352     1
```

```
4.205     1
3.925     1
4.272     1
4.922     1
6.765     1
4.530     1
4.000     1
2.870     1
3.948     1
2.784     1
5.568     1
Name: alcohol, dtype: int64
```

```
x.shape
```

```
(51, 7)
```

```
Alcohol=pd.get_dummies(x["alcohol"],drop_first=True)
```

```
Alcohol
```

20	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
21	0	0	1	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
23	0	1	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	1	0	0	...	0	0	0	0	0	0	0
30	0	0	0	1	0	0	0	0	0	0	...	0	0	0	0	0	0	0
31	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	1	...	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
36	0	0	0	0	0	0	0	0	0	0	...	1	0	0	0	0	0	0
37	0	0	0	0	1	0	0	0	0	0	...	0	0	0	0	0	0	0
38	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
39	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
40	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
41	0	0	0	0	0	0	0	0	0	0	...	0	0	1	0	0	0	0
42	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0

											...							
43	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	1
44	1	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
45	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
46	0	0	0	0	0	0	1	0	0	0	...	0	0	0	0	0	0	0
47	0	0	0	0	0	0	0	0	1	0	...	0	0	0	0	0	0	0
48	0	0	0	0	0	0	0	0	0	0	...	0	0	0	1	0	0	0
49	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
50	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0

51 rows x 49 columns

```
x=pd.concat([x,alcohol],axis=1)
```

```
x
```

	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses	abbrev	1.808	2.784	2.87	...	5.770999999999995
0	42	5.640	43	15.040	784.55	145.08	1	0	0	0	...	
1	43	4.525	37	17.014	1053.48	133.93	0	0	0	0	...	
2	37	5.208	34	17.856	899.47	110.35	3	0	0	0	...	
3	16	5.824	46	21.280	827.34	142.39	2	0	0	0	...	
4	21	3.360	15	10.680	878.41	165.63	4	0	0	0	...	
5	29	3.808	13	12.920	835.50	139.91	5	0	0	0	...	
6	28	3.888	8	8.856	1068.73	167.02	6	0	0	0	...	
7	35	4.860	28	16.038	1137.87	151.48	8	0	0	0	...	
8	3	1.593	2	5.900	1273.89	136.05	7	0	0	0	...	
9	12	5.191	38	16.826	1160.13	144.18	9	0	0	0	...	
10	8	3.900	31	14.508	913.15	142.80	10	0	0	0	...	
11	49	7.175	29	15.225	861.18	120.92	11	0	0	0	...	
12	32	4.437	19	14.994	641.96	82.75	13	0	0	0	...	
13	25	4.352	17	12.288	803.11	139.15	14	0	0	0	...	
14	11	4.205	24	13.775	710.46	108.92	15	0	0	0	...	
15	7	3.925	33	13.659	649.06	114.47	12	0	0	0	...	
16	26	4.272	23	15.130	780.45	133.80	16	0	0	0	...	
17	18	4.922	39	16.264	872.51	137.13	17	0	0	0	...	
18	40	6.765	32	20.090	1281.55	194.78	18	0	0	0	...	
19	33	4.530	21	12.684	661.88	96.57	21	0	0	0	...	
20	23	4.000	7	12.375	1048.78	192.70	20	0	0	0	...	
21	1	2.870	3	6.560	1011.14	135.63	19	0	0	1		