MOLAKA RAGHU Vit-ap

In [1]:

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

In [ ]:

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

Out[ ]:

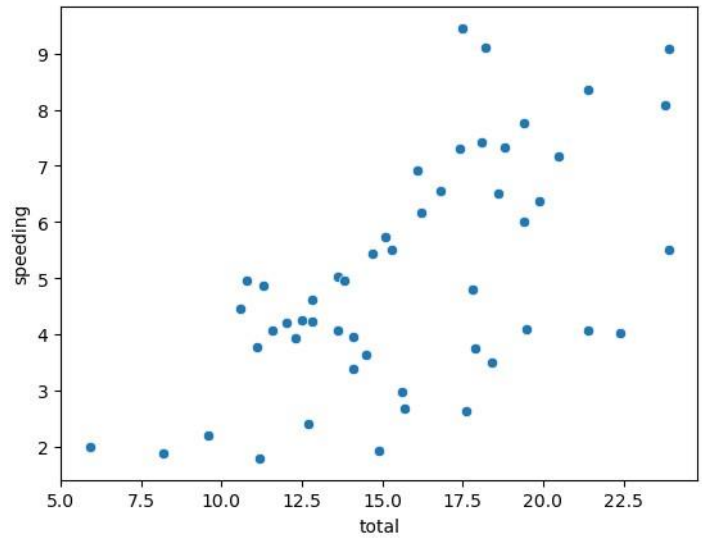
	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses	abbrev
0	18.8	7.332	5.640	18.048	15.040	784.55	145.08	AL
1	18.1	7.421	4.525	16.290	17.014	1053.48	133.93	AK
2	18.6	6.510	5.208	15.624	17.856	899.47	110.35	AZ
3	22.4	4.032	5.824	21.056	21.280	827.34	142.39	AR
4	12.0	4.200	3.360	10.920	10.680	878.41	165.63	CA
5	13.6	5.032	3.808	10.744	12.920	835.50	139.91	CO
6	10.8	4.968	3.888	9.396	8.856	1068.73	167.02	CT
7	16.2	6.156	4.860	14.094	16.038	1137.87	151.48	DE
8	5.9	2.006	1.593	5.900	5.900	1273.89	136.05	DC
9	17.9	3.759	5.191	16.468	16.826	1160.13	144.18	FL
10	15.6	2.964	3.900	14.820	14.508	913.15	142.80	GA
11	17.5	9.450	7.175	14.350	15.225	861.18	120.92	HI
12	15.3	5.508	4.437	13.005	14.994	641.96	82.75	ID
13	12.8	4.608	4.352	12.032	12.288	803.11	139.15	IL
14	14.5	3.625	4.205	13.775	13.775	710.46	108.92	IN
15	15.7	2.669	3.925	15.229	13.659	649.06	114.47	IA
16	17.8	4.806	4.272	13.706	15.130	780.45	133.80	KS
17	21.4	4.066	4.922	16.692	16.264	872.51	137.13	KY
18	20.5	7.175	6.765	14.965	20.090	1281.55	194.78	LA
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

In [3]:

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

Out[3]:

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



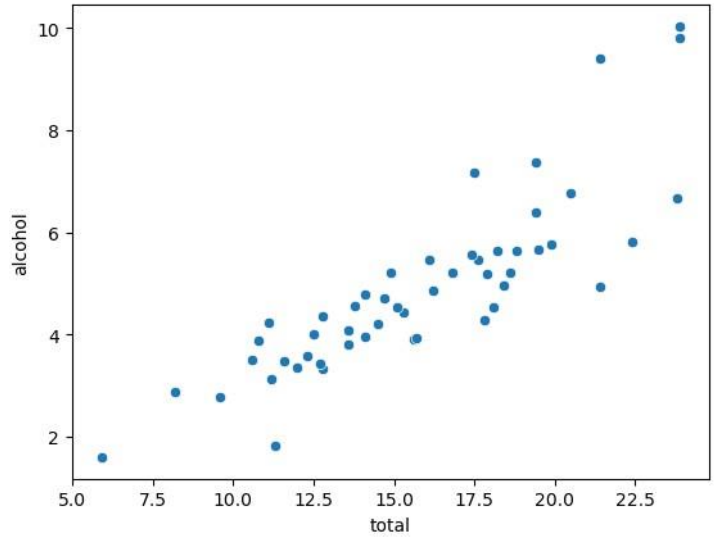
inference: from the plot we can say that as speeding increases total crashes is also increasing

In [4]:

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

Out[4]:

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



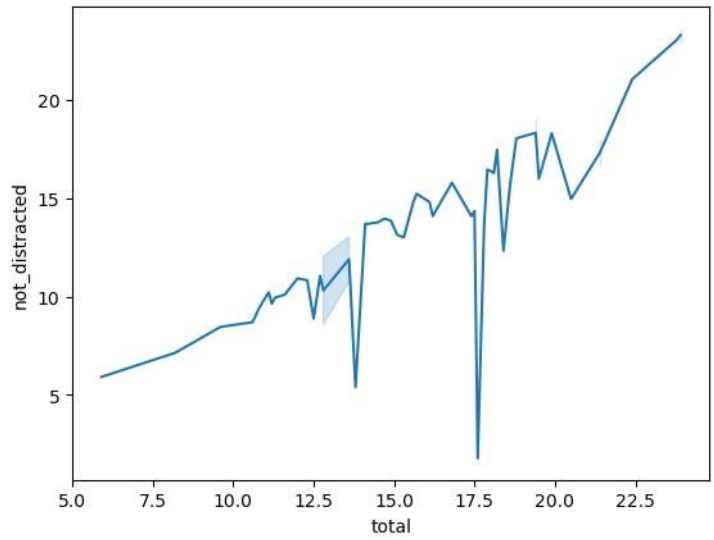
inference: from the plot we can say that as alcohol increases total crashes is also increasing

In [5]:

```
sns.lineplot(x="total", y="not_distracted",data=df)
```

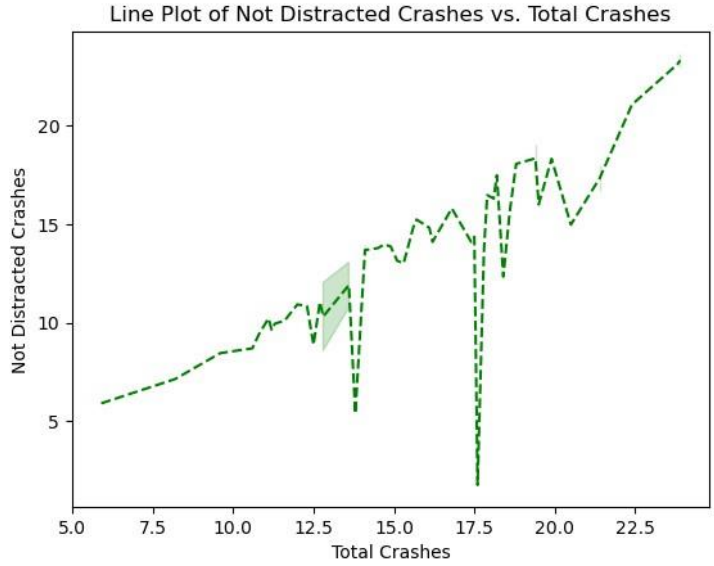
Out[5]:

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



inference: from the plot we can say that as not_distracted increases total crashes is also increasing

In [6]:
```
sns.lineplot(x="total", y="not_distracted", data=df, linestyle="--", color='green') plt.xlabel('Total Crashes')
plt.ylabel('Not Distracted Crashes')
plt.title('Line Plot of Not Distracted Crashes vs. Total Crashes') plt.show()
```



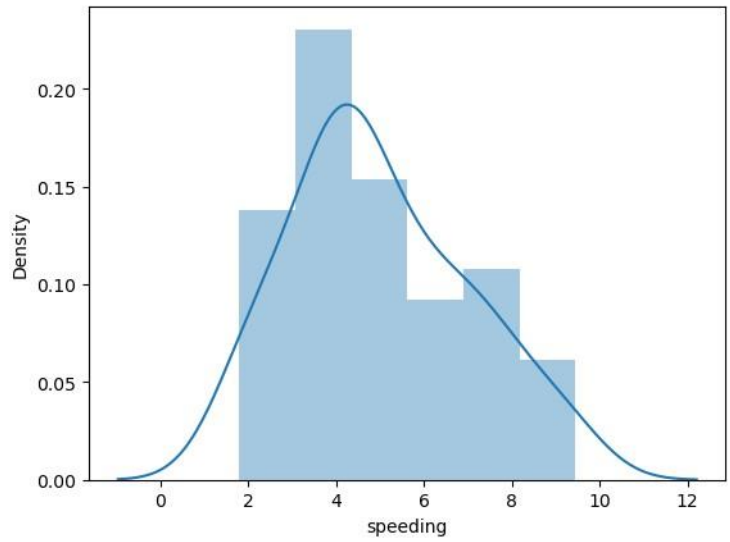
inference: from the plot we can say that as not_Distracted increases total crashes is also increasing

In [7]:
```
sns.distplot(df["speeding"])
```

```
C:\Users\mb419\AppData\Local\Temp\ipykernel_10092\2127910581.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["speeding"])
```

Out[7]:

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



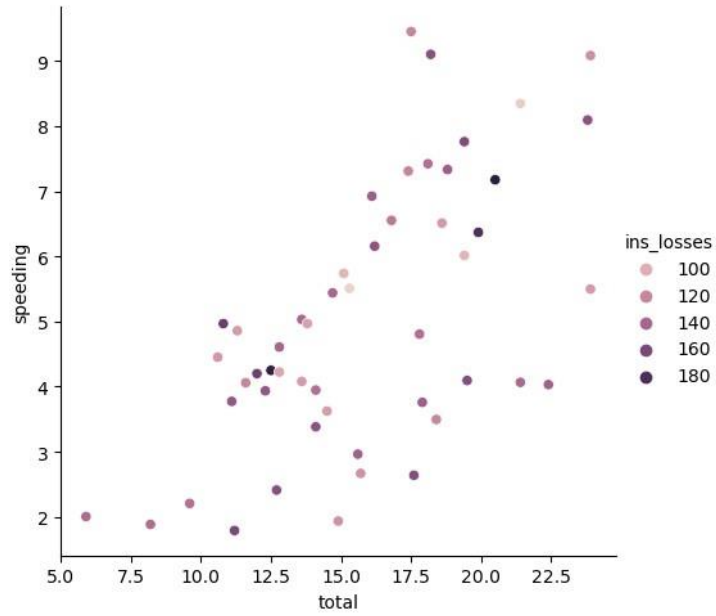
inference: from the plot we can say that speeding is high at 5

In [8]:

```
sns.relplot(x="total", y="speeding", data=df , hue="ins_losses")
```

Out[8]:

```
<seaborn.axisgrid.FacetGrid at 0x1b1d7edab50>
```



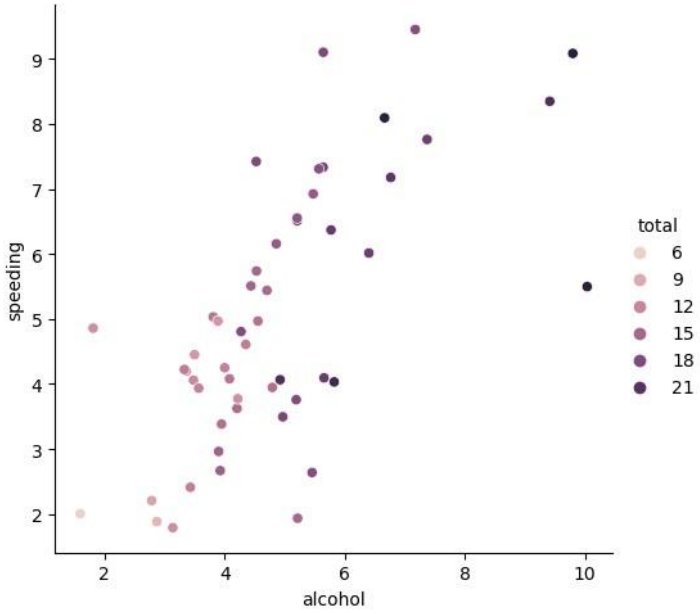
inference: from the plot we can say that as speeding increases total crashes is also increasing

In [9]:

```
sns.relplot(x="alcohol", y="speeding", data=df , hue="total")
```

Out[9]:

```
<seaborn.axisgrid.FacetGrid at 0x1b1d9411f50>
```



inference: from the plot we can say that as speeding increases alcohol is also increasing

In [10]:

```
df["total"].value_counts()
```

Out[10]:

```
14.1    2
12.8    2
13.6    2
21.4    2
19.4    2
23.9    2
14.9
14.7
11.6
11.2
18.4
12.3
16.8
19.9
17.6
18.2
11.1
19.5
11.3
12.7
10.6
23.8
13.8
16.1
18.8
9.6
18.1
18.6
22.4
12.0
10.8
16.2
5.9
17.9
15.6
17.5
15.3
14.5
15.7
17.8
20.5
15.1
12.5
8.2
17.4Name
:  total,
dtype:
int64
```

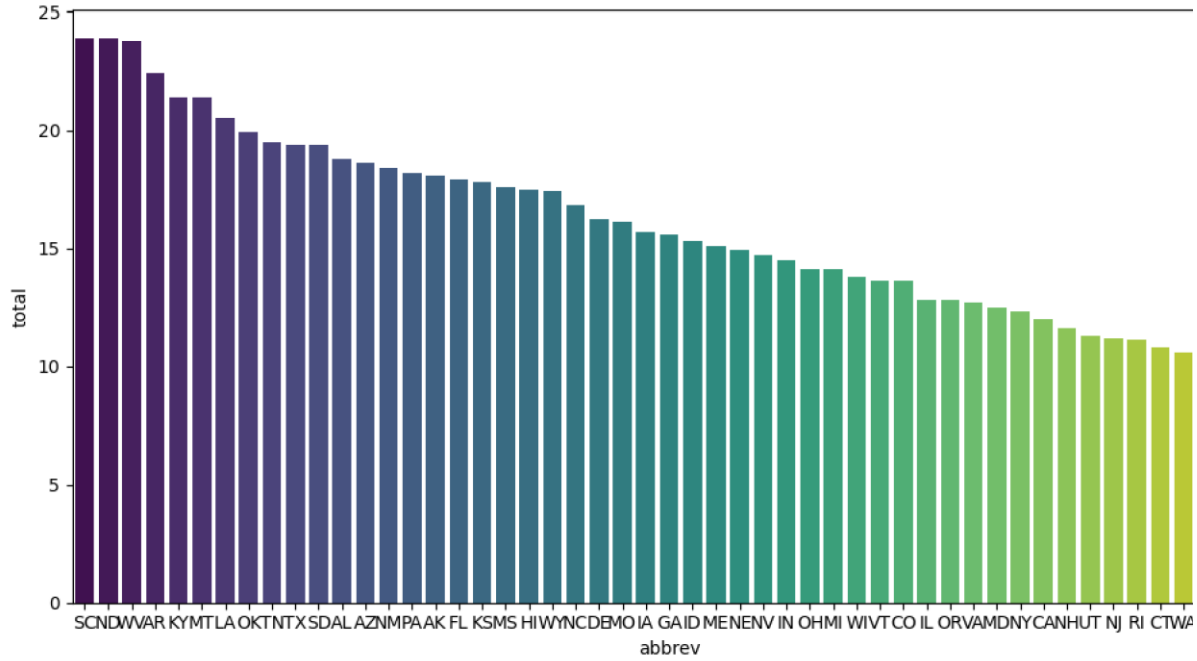
Error! Bookmark not defined.

In [11]:

```
df= df.sort_values(by='total', ascending=False) plt.figure(figsize=(12, 6))
sns.barplot(x='abbrev', y='total', data=df, palette='viridis')
```

Out[11]:

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



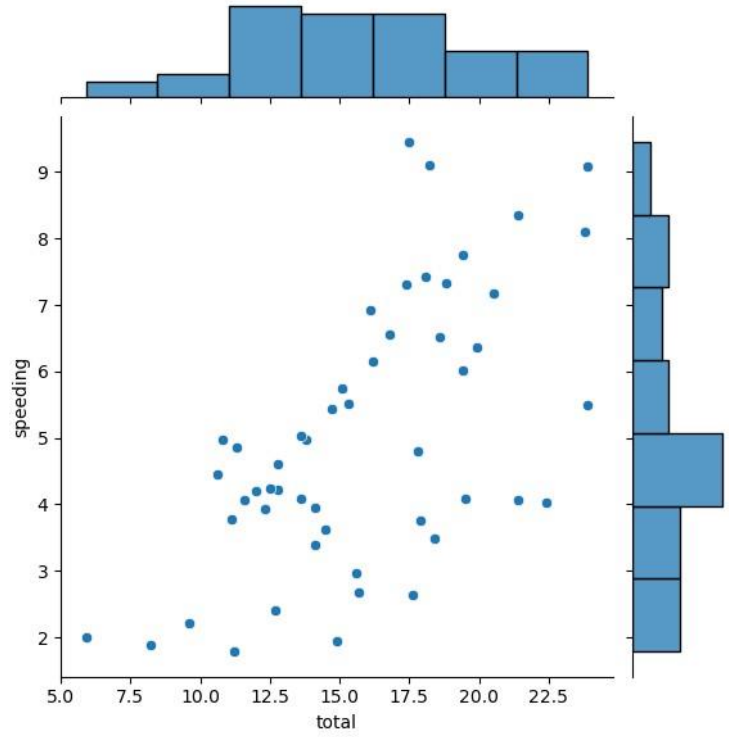
inference: from the plot we can say that as abbrev 'SC' have higher total crashes among the remaining abbrev and the least one is 'DC'

In [12]:

```
sns.jointplot(x="total",y="speeding",data=df)
```

Out[12]:

```
<seaborn.axisgrid.JointGrid at 0x1b1d959a750>
```



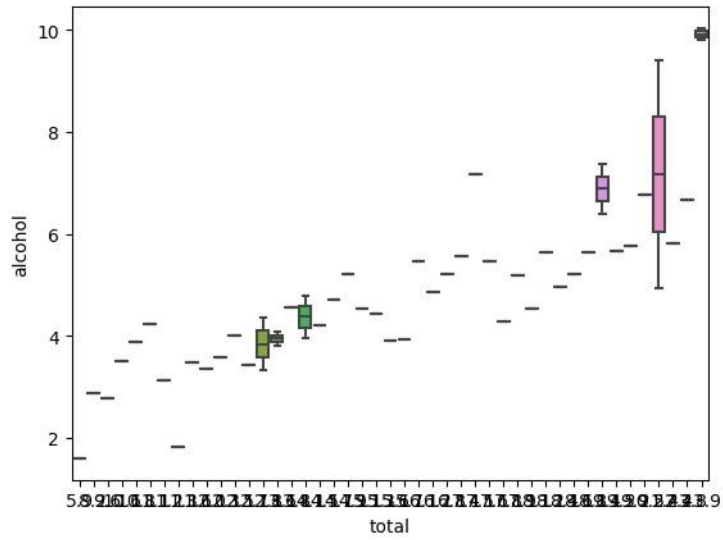
inference: from the plot we can say that as speeding and total crashes are increased at (12.5 , 4.1)

In [13]:

```
sns.boxplot(x="total",y="alcohol",data=df)
```

Out[13]:

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



inference: from the plot we can say that as alcohol increases total crashes is also increasing

In [14]:

```
corr=df.corr() corr
C:\Users\mb419\AppData\Local\Temp\ipykernel_10092\3182140910.py:1: FutureWarning: The default value of numeric_o
corr=df.corr()
```

Out[14]:

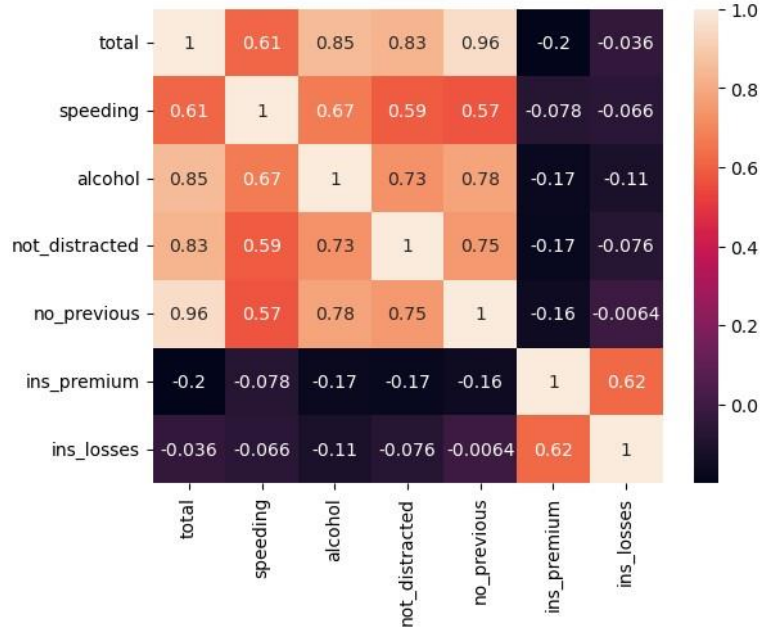
	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses
total	1.000000	0.611548	0.852613	0.827560	0.956179	-0.199702	-0.036011
speeding	0.611548	1.000000	0.669719	0.588010	0.571976	-0.077675	-0.065928
alcohol	0.852613	0.669719	1.000000	0.732816	0.783520	-0.170612	-0.112547
not_distracted	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

In [15]:

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

Out[15]:

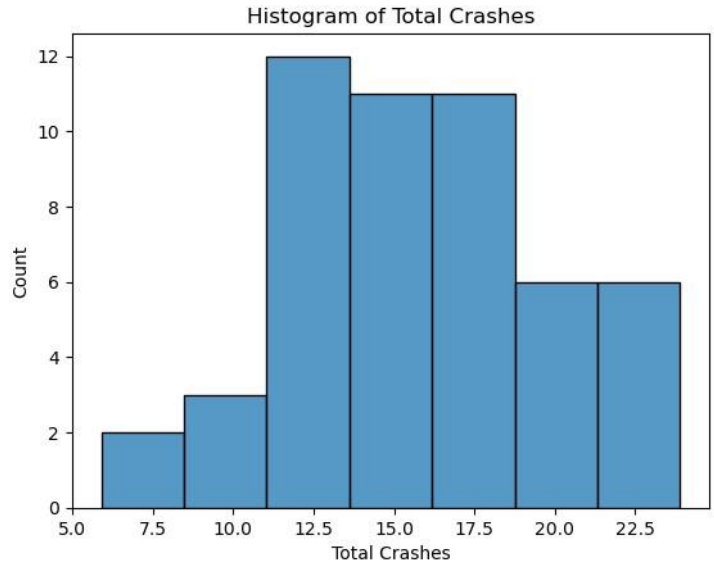
```
<Axes: >
```



inference: from the plot we can say that >0.5 is highly correlated <0.5 is less correlated

In [16]:

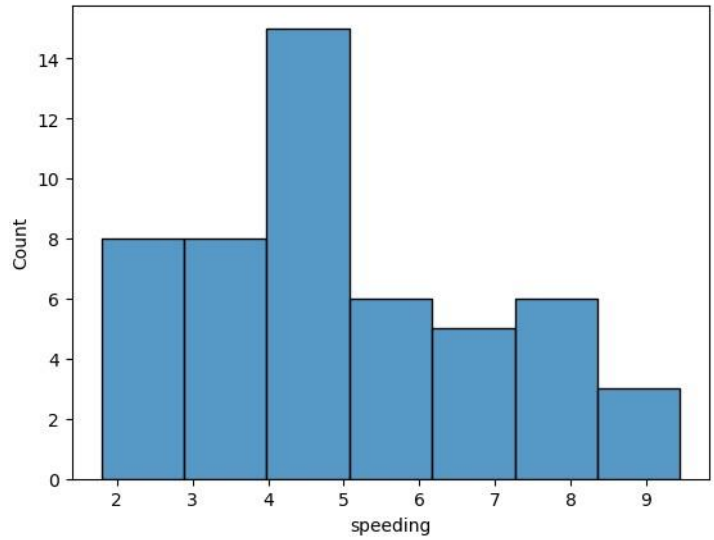
```
sns.histplot(df['total']) plt.xlabel('Total Crashes') plt.title('Histogram of Total Crashes')
plt.show()
```



inference: from the plot we can say that total crashes are high at 12.5 & equal at 15,17.5

In [17]:

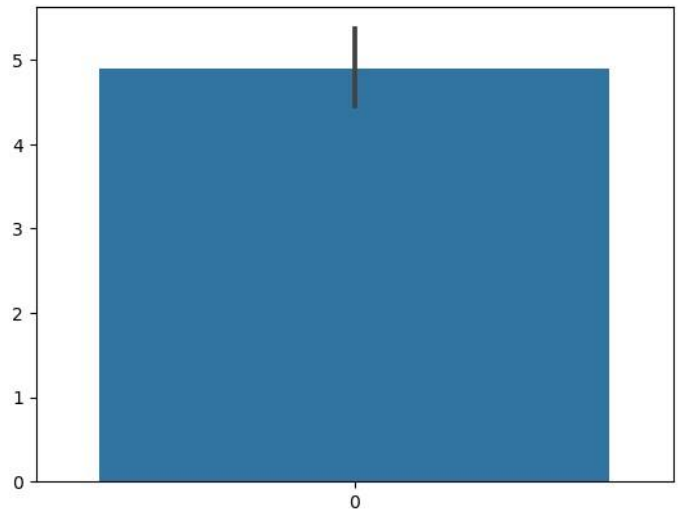
```
sns.histplot(df['speeding']) plt.show()
```



inference: from the plot we can say that speeding is very high in between 4,5

In [18]:

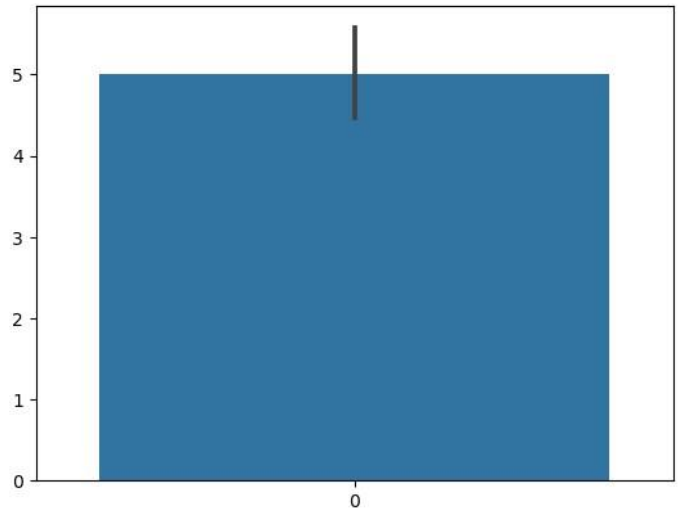
```
sns.barplot(df['alcohol']) plt.show()
```



inference: from the plot we can say that alcohol is just below the 5

In [19]:

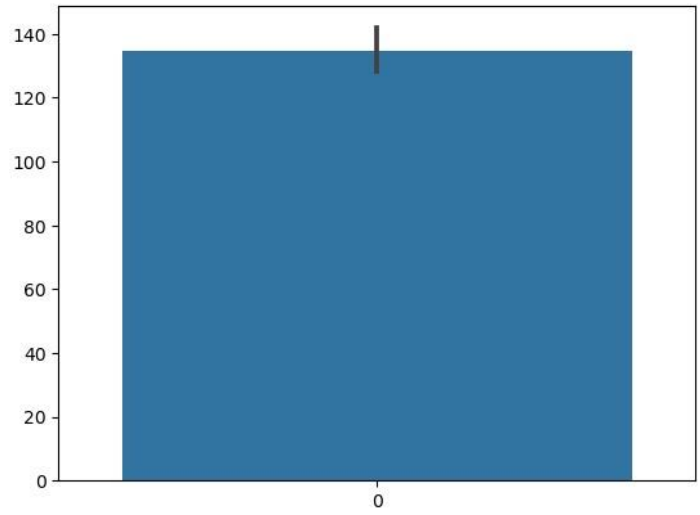
```
sns.barplot(df['speeding']) plt.show()
```



inference: from the plot we can say that speeding is equal to 5

In [20]:

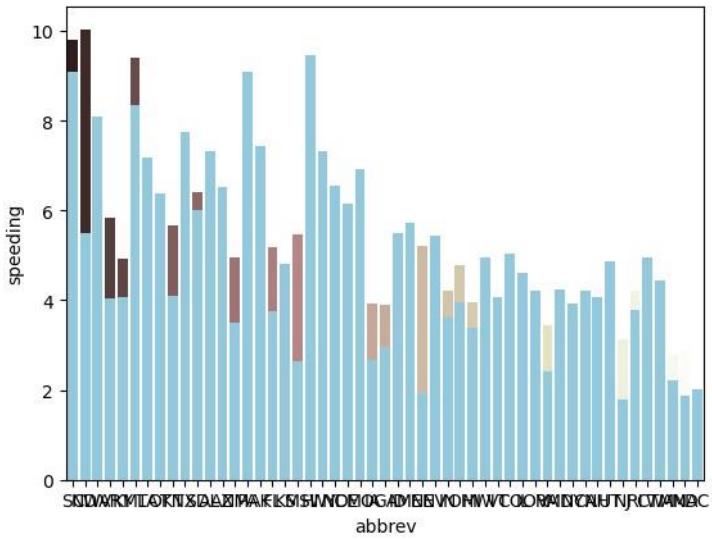
```
sns.barplot(df['ins_losses']) plt.show()
```



inference: from the plot we can say that ins_losses is at 137

In [21]:

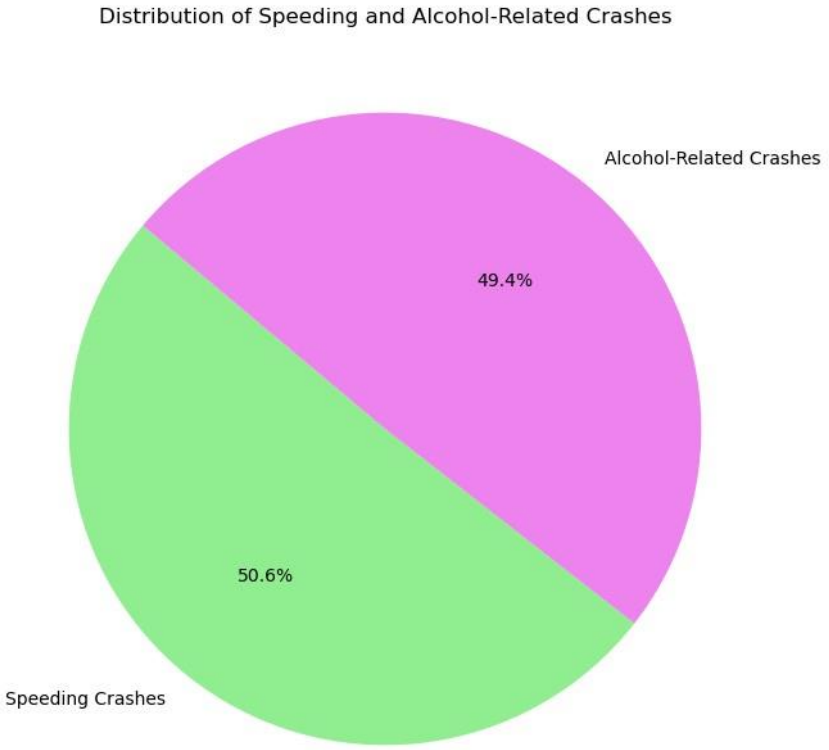
```
sns.barplot(x='abbrev', y='alcohol', data=df, palette='pink')
sns.barplot(x='abbrev', y='speeding', data=df, color='skyblue') plt.show()
```



inference: Here we plot the mixed bar graph of abbrev,alochol & abbrev,speeding

In [22]:

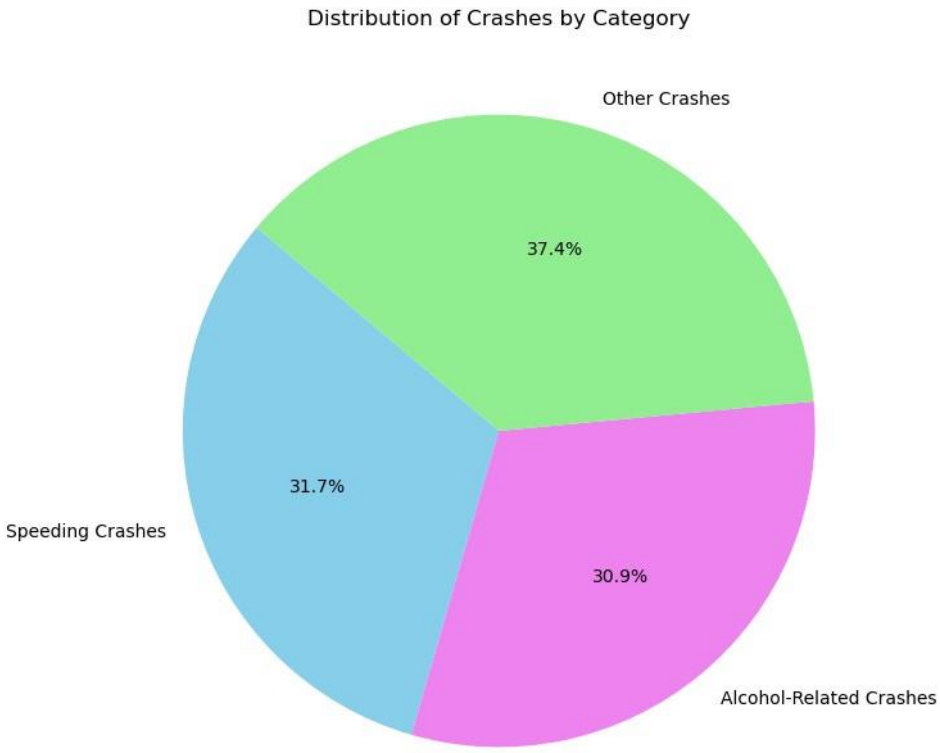
```
total_speeding_crashes = df['speeding'].sum() total_alcohol_crashes = df['alcohol'].sum() labels = ['Speeding Crashes', 'Alcohol-Related Crashes'] sizes = [total_speeding_crashes, total_alcohol_crashes] colors = ['lightgreen', 'violet'] plt.figure(figsize=(8, 8)) plt.pie(sizes, labels=labels, colors=colors, autopct='%1.1f%%', startangle=140) plt.title('Distribution of Speeding and Alcohol-Related Crashes') plt.show()
```



inference: from the pie chart we can say that as speeding crashes slightly greater than the Alcohol related Crashes

In [23]:

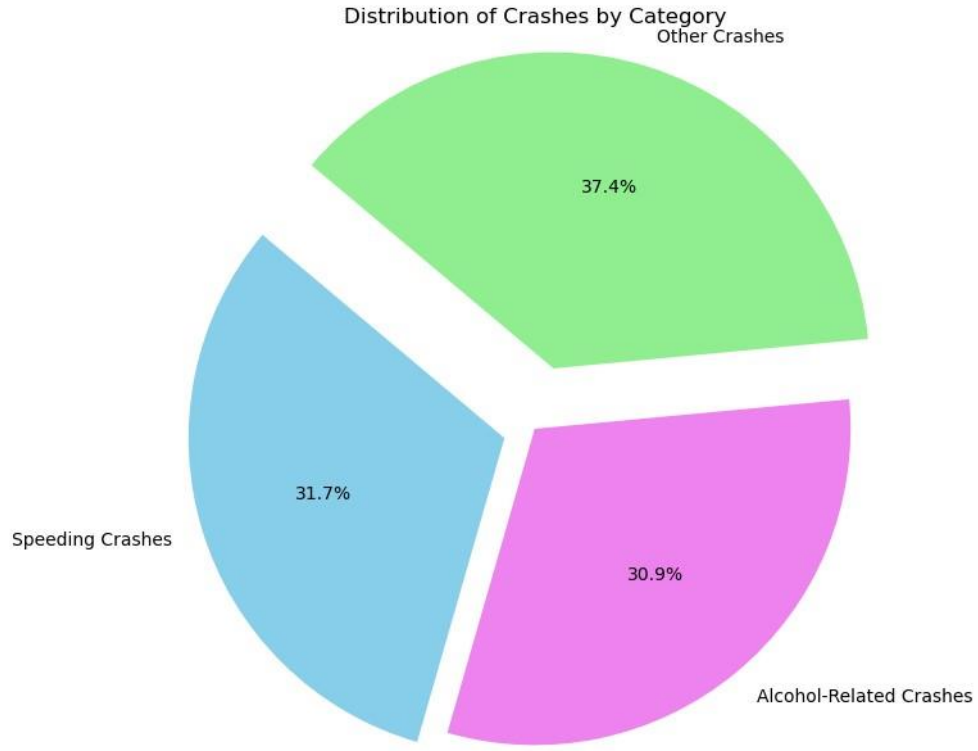
```
total_speeding_crashes = df['speeding'].sum() total_alcohol_crashes = df['alcohol'].sum() total_other_crashes = df['total'].sum() - (total_speeding_crashes + total_alcohol_crashes)
labels = ['Speeding Crashes', 'Alcohol-Related Crashes', 'Other Crashes']
sizes = [total_speeding_crashes, total_alcohol_crashes, total_other_crashes]
colors = ['skyblue', 'violet', 'lightgreen'] plt.figure(figsize=(8, 8))
plt.pie(sizes, labels=labels, colors=colors, autopct='%1.1f%%', startangle=140)
plt.title('Distribution of Crashes by Category') plt.show()
```



inference: from the plot we can say that as other reason crashes greater than the Alcohol related Crashes & speeding crashes. Speeding crashes slightly greater than the Alcohol related Crashes

In [24]:

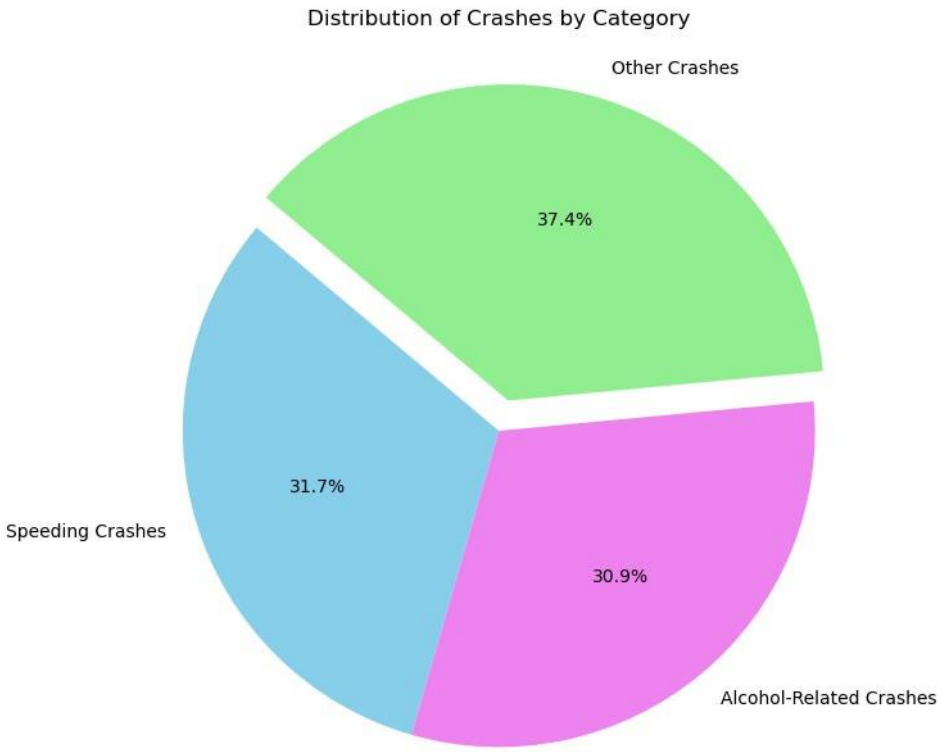
```
total_speeding_crashes = df['speeding'].sum() total_alcohol_crashes = df['alcohol'].sum()
total_other_crashes = df['total'].sum() - (total_speeding_crashes + total_alcohol_crashes) labels
= ['Speeding Crashes', 'Alcohol-Related Crashes', 'Other Crashes'] sizes =
[total_speeding_crashes, total_alcohol_crashes, total_other_crashes] colors = ['skyblue',
'violet', 'lightgreen'] Explode=[0.1,0,0.2] plt.figure(figsize=(8, 8)) plt.pie(sizes,
labels=labels, colors=colors, autopct='%1.1f%%', startangle=140, explode=Explode)
plt.title('Distribution of Crashes by Category') plt.show()
```



inference: from the plot we can say that as other reason crashes greater than the Alcohol related Crashes & speeding crashes. Speeding crashes slightly greater than the Alcohol related Crashes. Here i explode the higher crashes at 0.2 and other one at 0.1 and the least one at same position.

In [25]:

```
total_speeding_crashes = df['speeding'].sum() total_alcohol_crashes = df['alcohol'].sum() total_other_crashes = df['total'].sum() - (total_speeding_crashes + total_alcohol_crashes) labels = ['Speeding Crashes', 'Alcohol-Related Crashes', 'Other Crashes'] sizes = [total_speeding_crashes, total_alcohol_crashes, total_other_crashes] colors = ['skyblue', 'violet', 'lightgreen'] Explode=[0,0,0.1] plt.figure(figsize=(8, 8)) plt.pie(sizes, labels=labels, colors=colors, autopct='%1.1f%%', startangle=140, explode=Explode) plt.title('Distribution of Crashes by Category') plt.show()
```



inference: from the plot we can say that as other reason crashes greater than the Alcohol related Crashes & speeding crashes. Speeding crashes slightly greater than the Alcohol related Crashes. Here the higher crashes only explode at 0.1.