

```
[54]: # Importing the libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [55]: # Taking car_crashes dataset from seaborn library and loading it
crashes = sns.load_dataset("car_crashes")
crashes
```

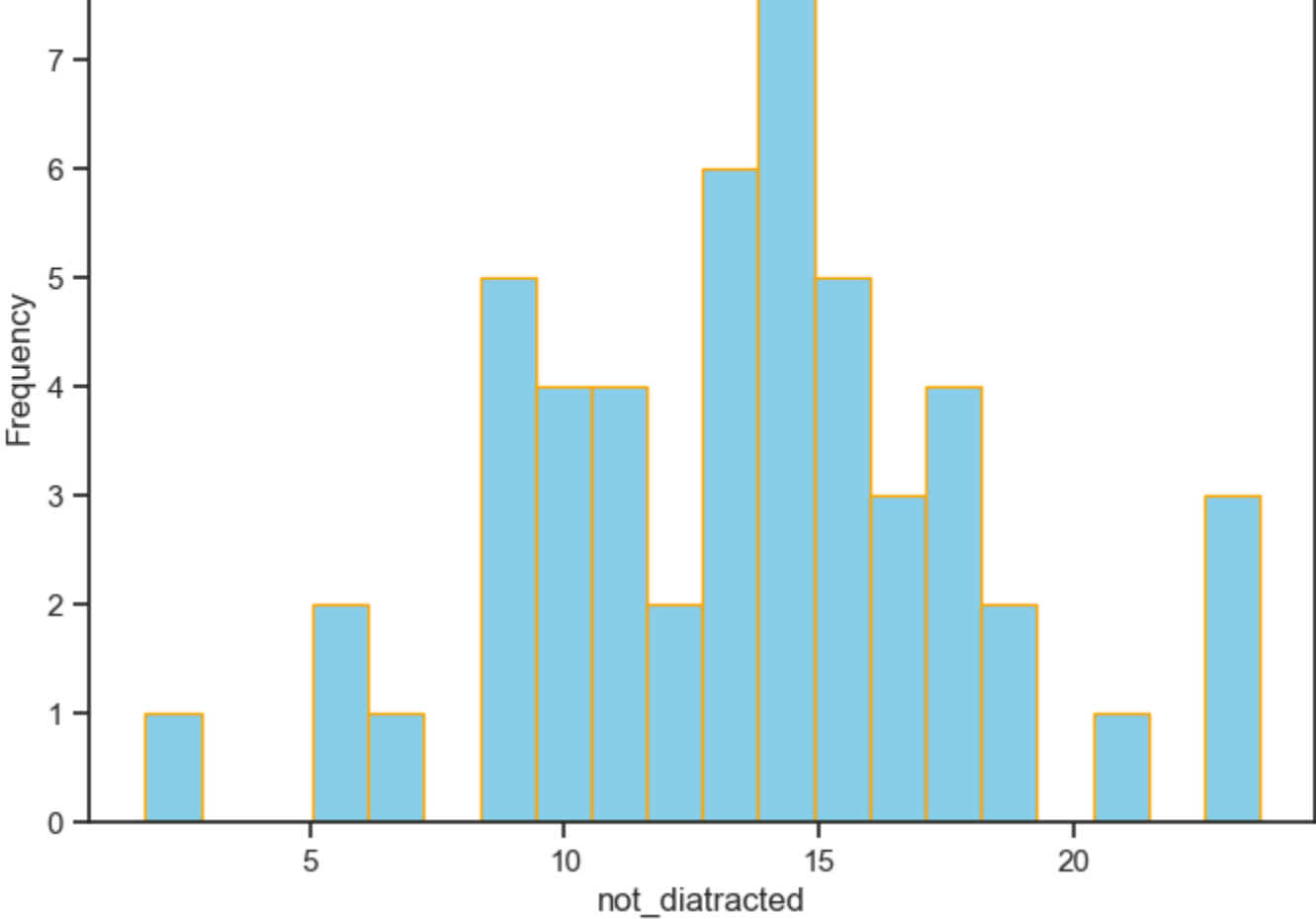
Out [55]:

	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	2.689	3.925	15.229	13.059	648.06	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	20.954	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	15.960	15.795	767.61	155.57	TN
43	19.4	7.760	7.372	17.654	16.678	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	4.988	4.954	5.382	11.992	670.31	106.62	WI
50	17.4	7.308	5.958	14.094	15.660	791.14	122.04	WY

Data Visualization

```
In [56]: # 1. Histogram
plt.figure(figsize=(8,6))
plt.hist(crashes['not_distracted'], bins=20, color='skyblue', edgecolor='orange')
plt.title('Histogram for Level of distraction')
plt.xlabel('not_distracted')
plt.ylabel('frequency')
plt.show()

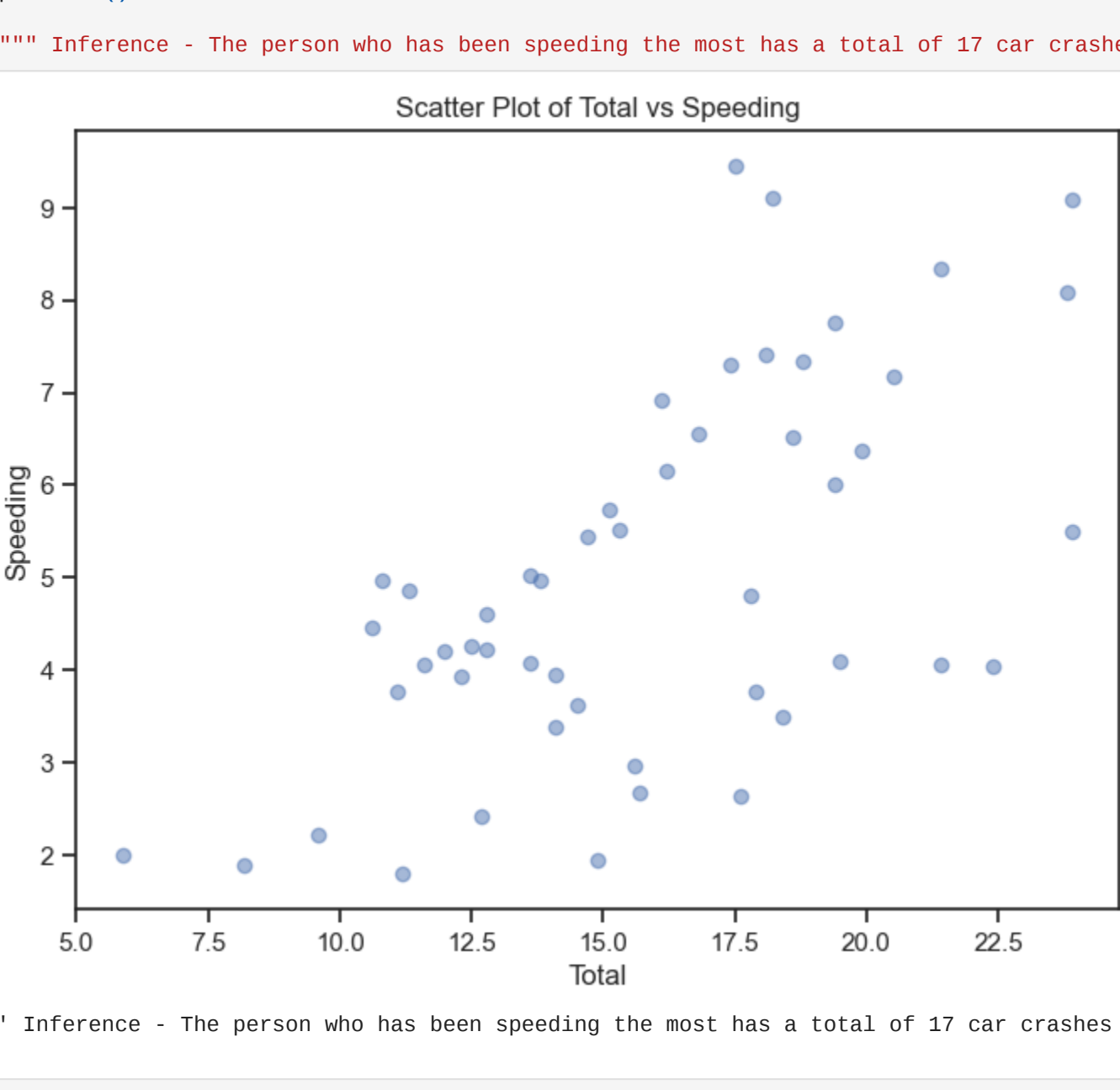
''' Inference based on graph given below - The maximum frequency of people who weren't distracted is 8 '''
```



```
Out [56]: ''' Inference based on graph given below - The maximum frequency of people who weren't distracted is 8 '''
```

```
In [57]: # 2. Scatter Plot
plt.figure(figsize=(8,6))
plt.scatter(crashes['total'], crashes['speeding'], marker='o', alpha=0.5)
plt.title('Scatter Plot of Total vs Speeding')
plt.xlabel('Total')
plt.ylabel('Speeding')
plt.show()

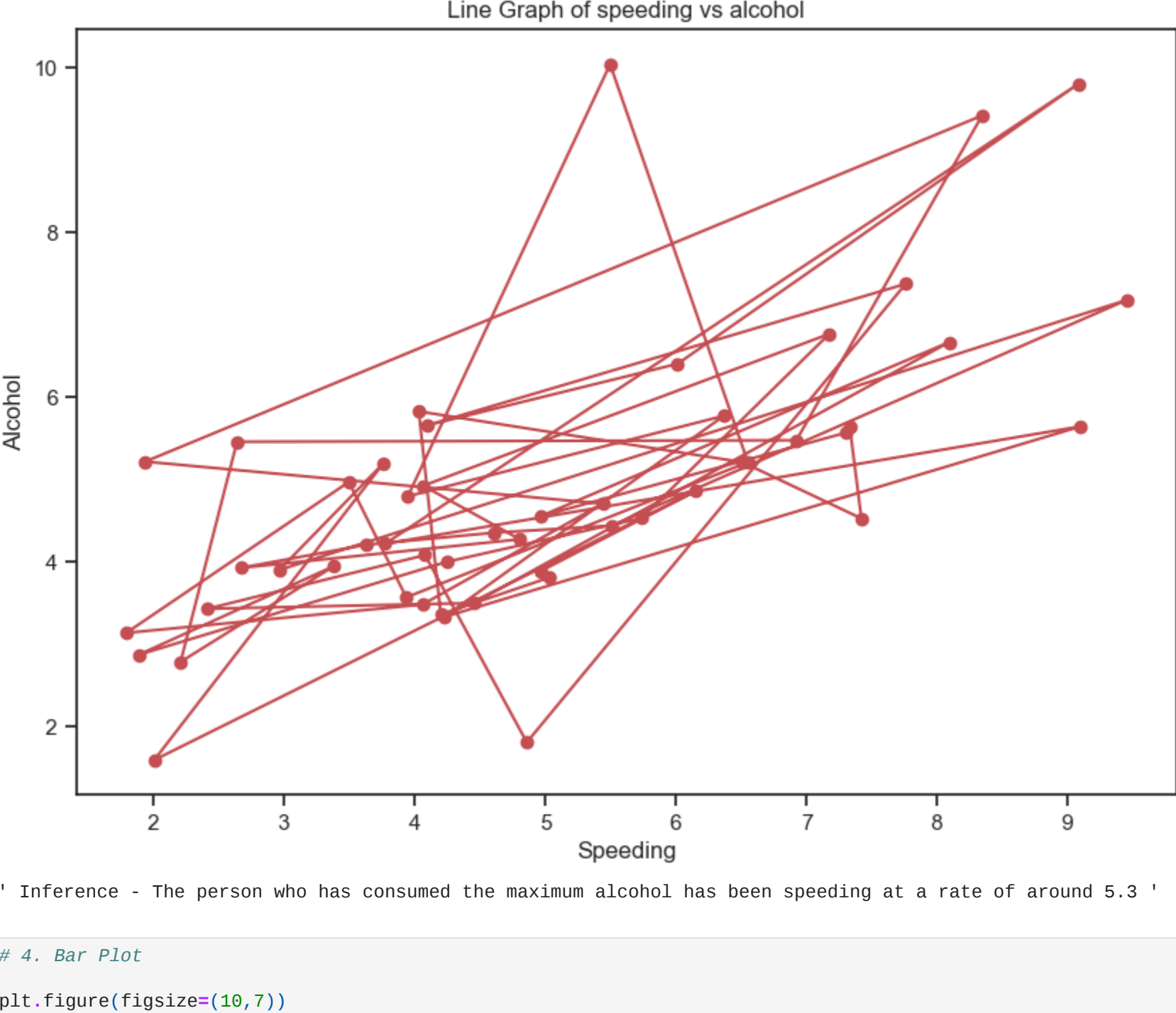
''' Inference - The person who has been speeding the most has a total of 17 car crashes '''
```



```
Out [57]: ' Inference - The person who has been speeding the most has a total of 17 car crashes '
```

```
In [58]: # 3. Line Graph
plt.figure(figsize=(10,7))
plt.plot(crashes['speeding'], crashes['alcohol'], marker='o', linestyle='-', color='r', label='Line Label')
plt.title('Line Graph of Speeding vs Alcohol')
plt.xlabel('Speeding')
plt.ylabel('Alcohol')
plt.show()

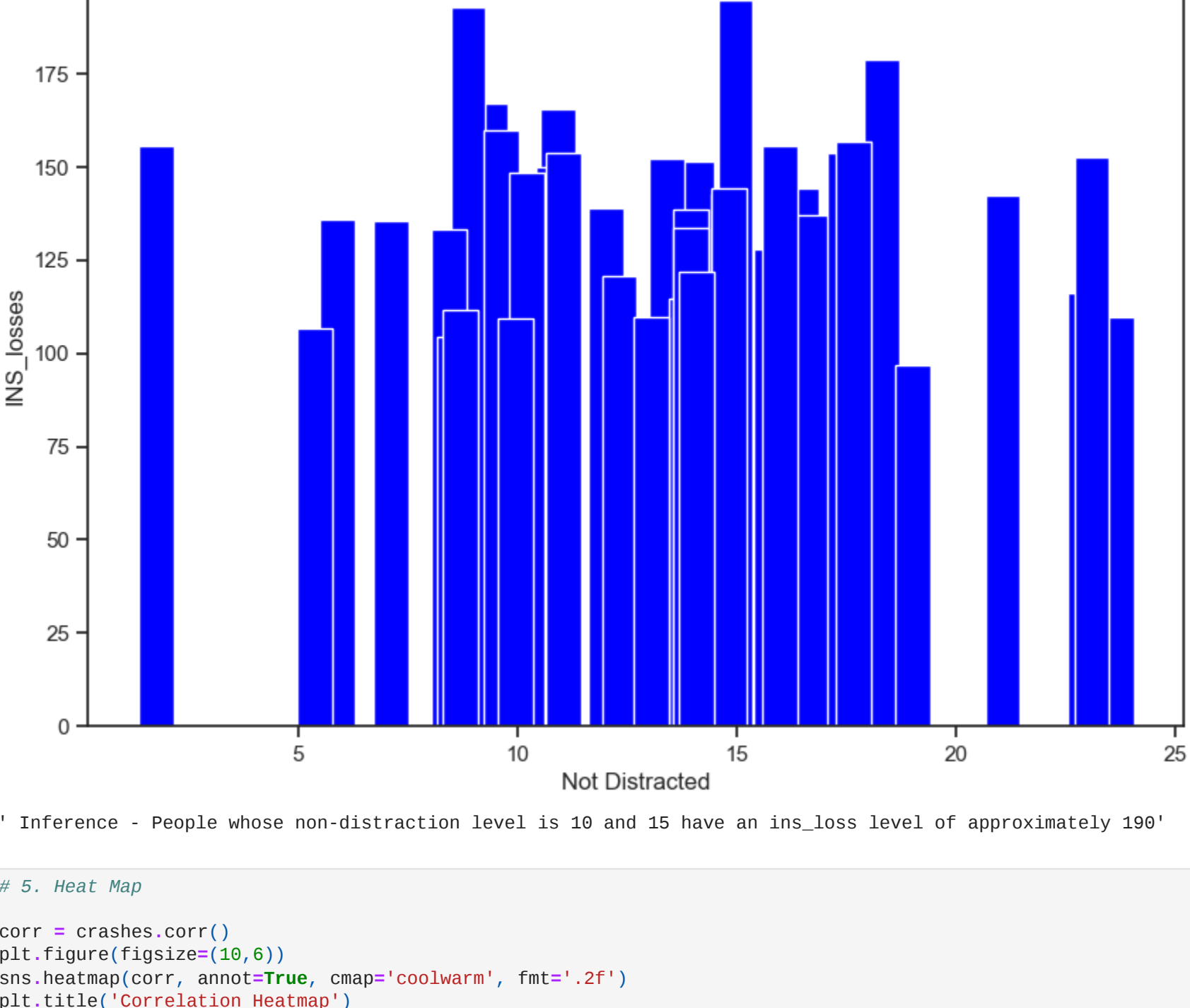
''' Inference - The person who has consumed the maximum alcohol has been speeding at a rate of around 5.3 '''
```



```
Out [58]: ' Inference - The person who has consumed the maximum alcohol has been speeding at a rate of around 5.3 '
```

```
In [59]: # 4. Bar Plot
plt.figure(figsize=(10,7))
plt.bar(crashes['not_distracted'], crashes['ins_losses'], color='blue')
plt.title('Bar Plot of not_distracted vs ins_losses')
plt.xlabel('not_distracted')
plt.ylabel('ins_losses')
plt.show()

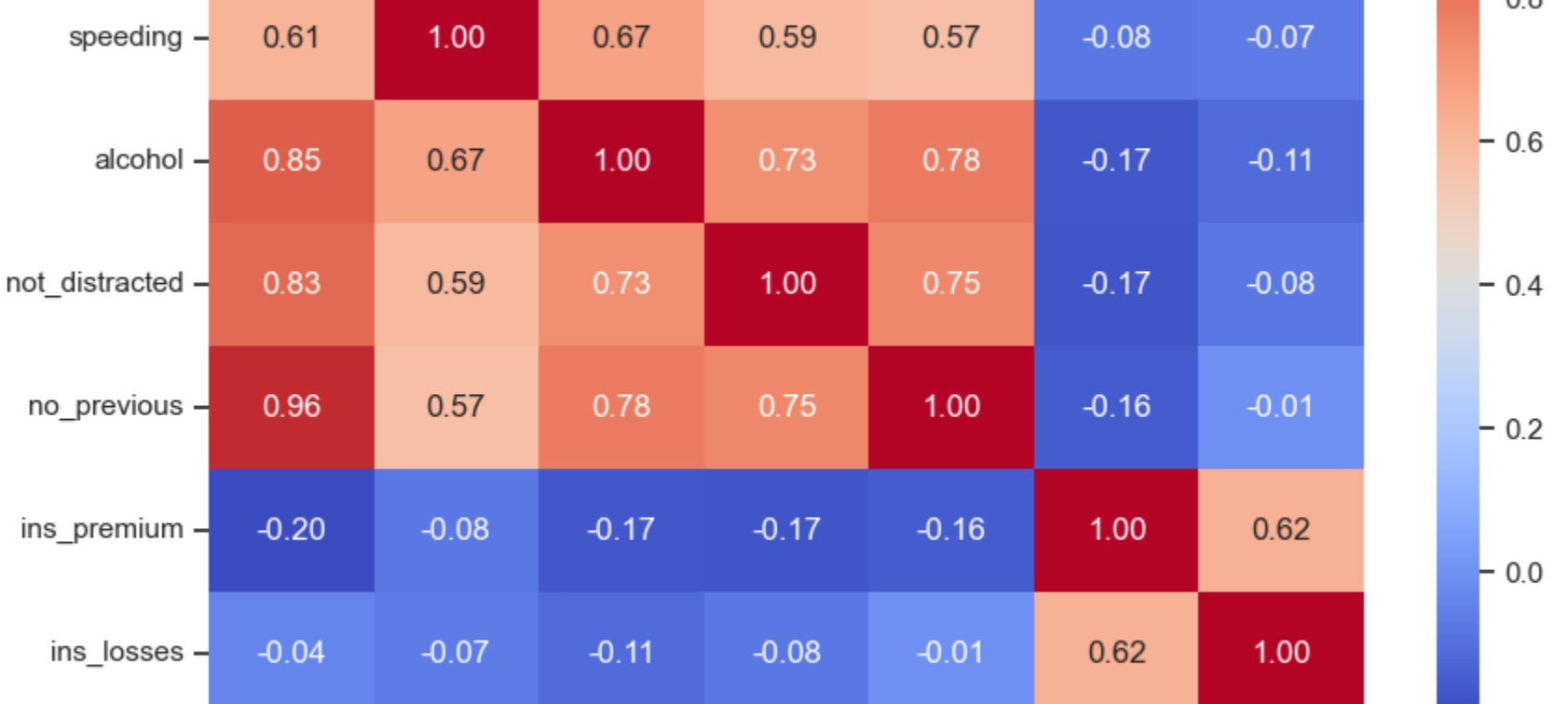
''' Inference - People whose non-distraction level is 10 and 15 have an ins_loss level of approximately 190'''
```



```
Out [59]: ' Inference - People whose non-distraction level is 10 and 15 have an ins_loss level of approximately 190'
```

```
In [60]: # 5. Heat Map
corr = crashes.corr()
plt.figure(figsize=(10,6))
sns.heatmap(corr, annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Heatmap')
plt.show()

''' Inference - The ins_premium values are initially negative and gradually increase and become positive '''
```



```
Out [60]: ' Inference - The ins_premium values are initially negative and gradually increase and become positive '
```

```
In [61]: # 6. Box Plot
plt.figure(figsize=(10, 7))
sns.boxplot(x=crashes['no_previous'], y=crashes['ins_premium'], palette='Set3')
plt.title('no_previous vs ins_premium')
plt.xlabel('no_previous')
plt.ylabel('ins_premium')
plt.show()

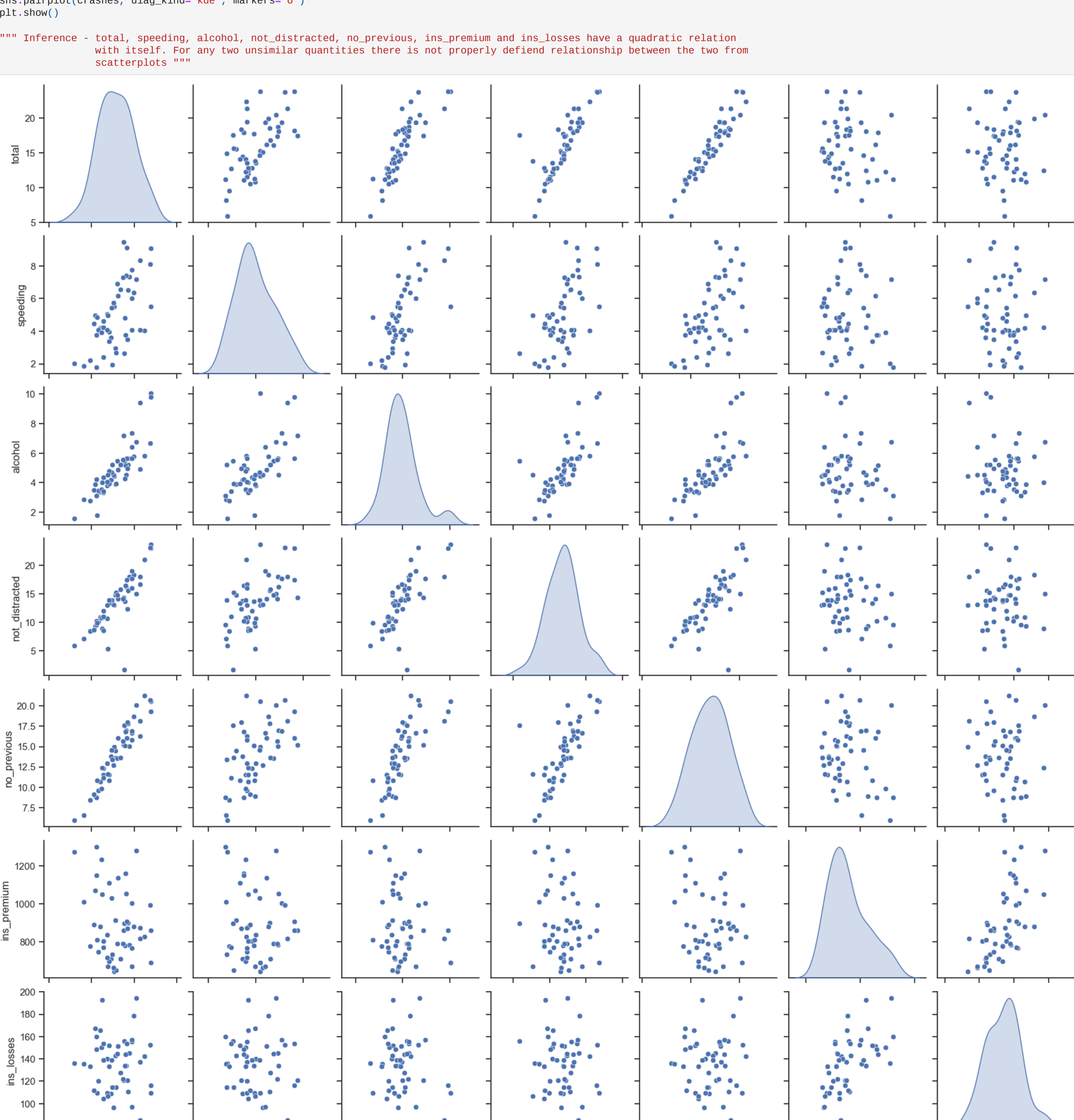
''' Inference - There is no well defined relationship between the ins_premium and no_previous'''
```



```
Out [61]: ' Inference - There is no well defined relationship between the ins_premium and no_previous'
```

```
In [62]: # 7. Pair Plot
sns.set(style='ticks')
sns.pairplot(crashes, diag_kind='kde', markers='o')
plt.show()

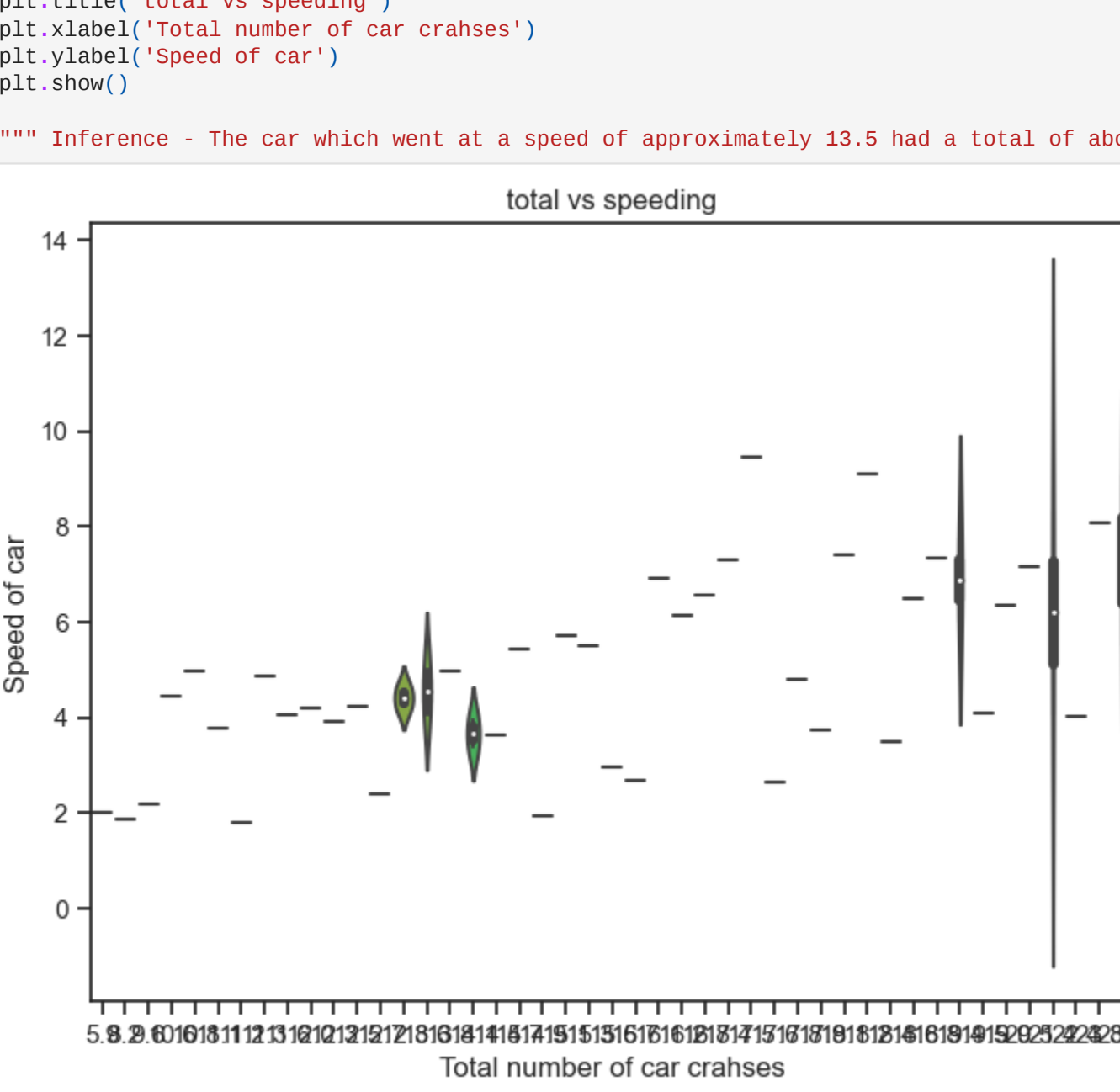
''' Inference - total, speeding, alcohol, not_distracted, no_previous, ins_premium and ins_losses have a quadratic relation with itself. For any two unsimilar quantities there is not properly defined relationship between the two from scatterplots '''
```



```
Out [62]: ' Inference - total, speeding, alcohol, not_distracted, no_previous, ins_premium and ins_losses have a quadratic relation with itself. For any two unsimilar quantities there is not properly defined relationship between the two from scatterplots '
```

```
In [64]: # 8. Violin Plot
plt.figure(figsize=(8, 6))
y='speeding', data=crashes
sns.violinplot(x='total', y='speeding', data=crashes)
plt.title('total vs speeding')
plt.xlabel('total number of car crashes')
plt.ylabel('speed of car')
plt.show()

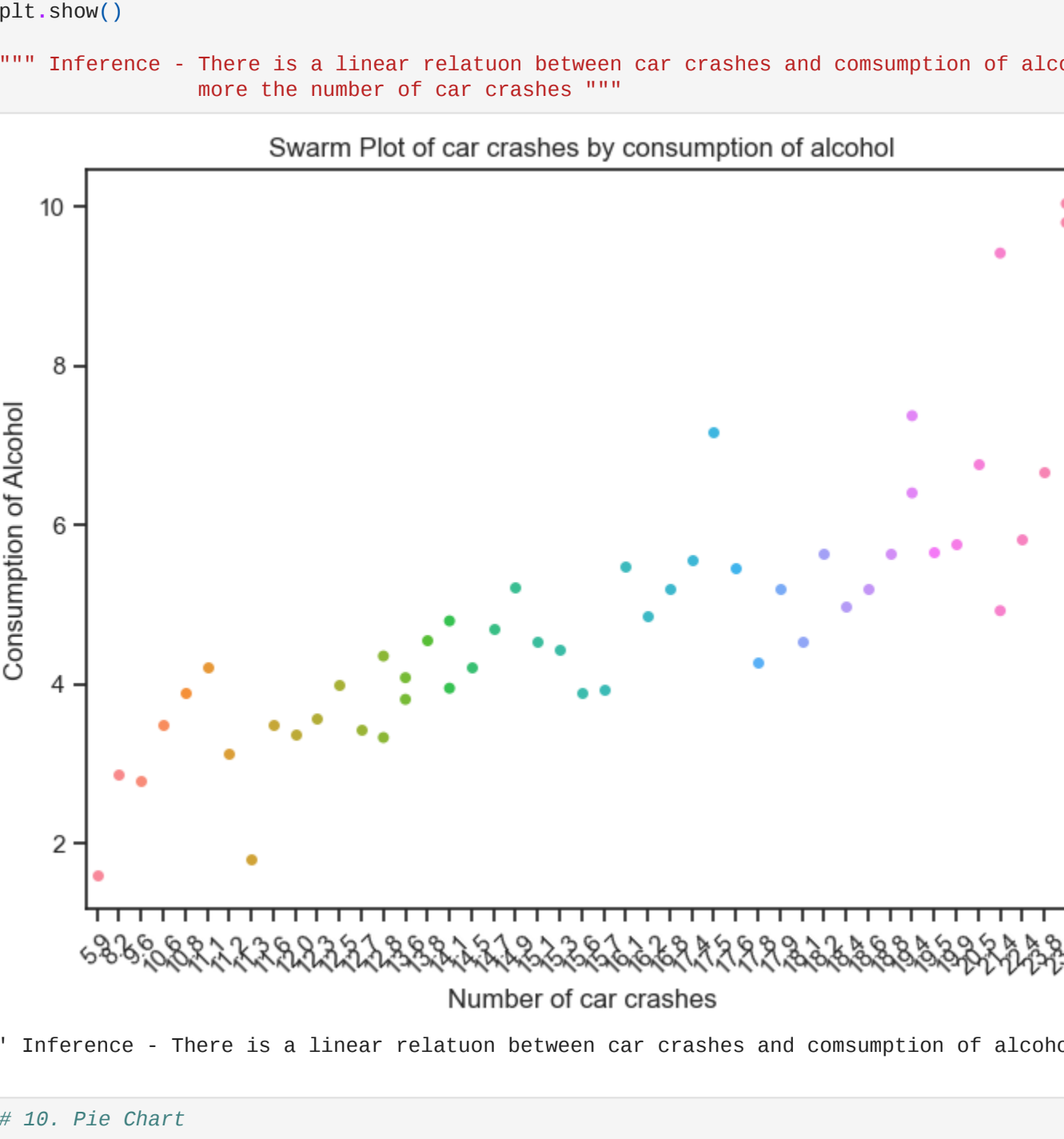
''' Inference - The car which went at a speed of approximately 13.5 had a total of about 9 car crashes '''
```



```
Out [64]: ' Inference - The car which went at a speed of approximately 13.5 had a total of about 9 car crashes '
```

```
In [65]: # 9. Swarm Plot
plt.figure(figsize=(8, 6))
sns.swarmplot(x='total', y='alcohol', data=crashes)
plt.title('Swarm Plot of car crashes by consumption of alcohol')
plt.xlabel('number of car crashes')
plt.ylabel('consumption of alcohol')
plt.xticks(rotation=45)
plt.show()

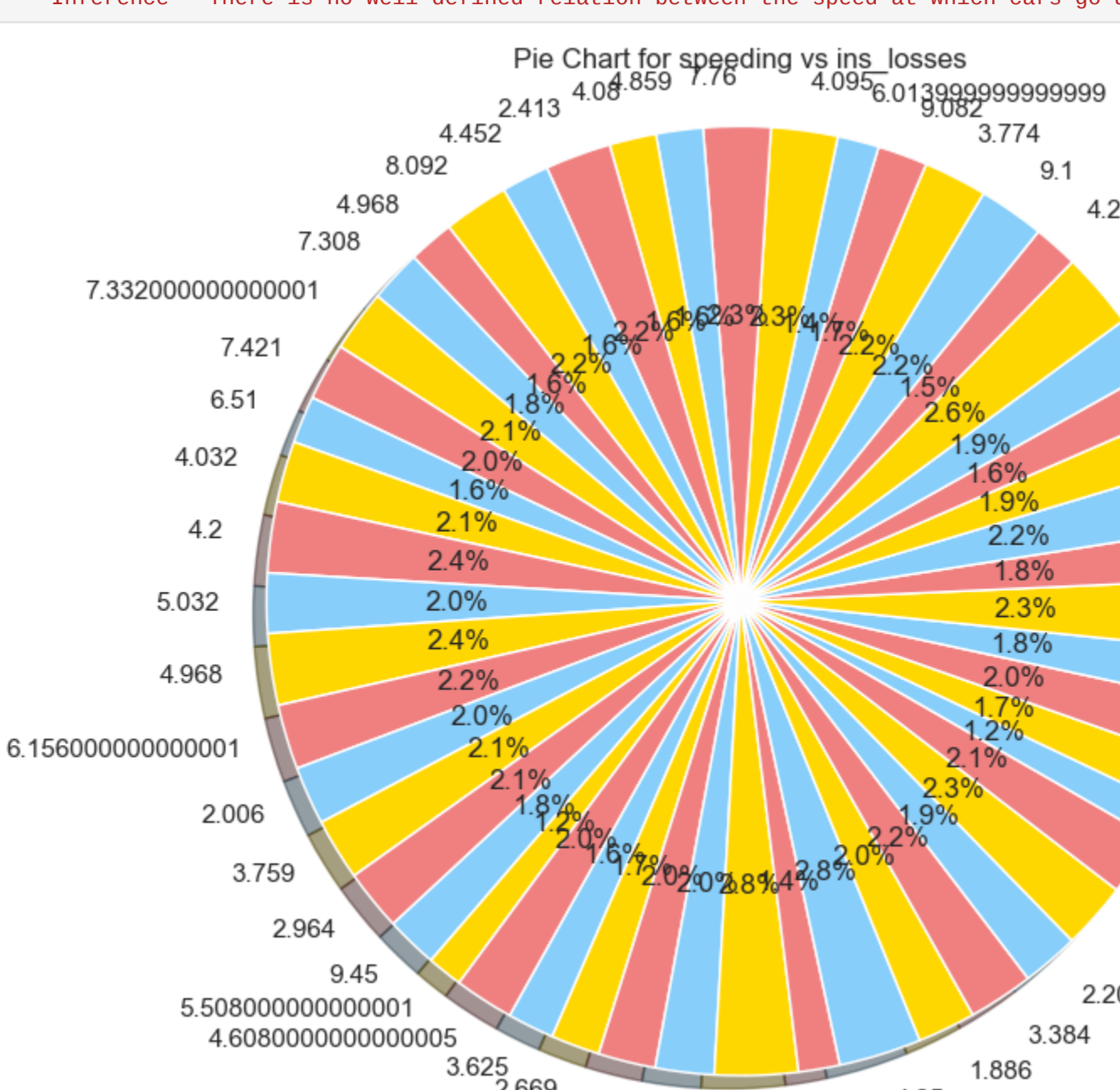
''' Inference - There is a linear relation between car crashes and consumption of alcohol. More the consumption of alcohol, more the number of car crashes '''
```



```
Out [65]: ' Inference - There is a linear relation between car crashes and consumption of alcohol. More the consumption of alcohol, more the number of car crashes '
```

```
In [66]: # 10. Pie Chart
plt.figure(figsize=(8, 8))
labels = crashes['speeding']
sizes = crashes['ins_losses']
colors = ['gold', 'lightcoral', 'lightskyblue']
plt.pie(sizes, labels=labels, colors=colors, autopct='%1.1f%%', shadow=True, startangle=140)
plt.title('Pie chart for speeding vs ins_losses')
plt.axis('equal')
plt.show()

''' Inference - There is no well defined relation between the speed at which cars go and the ins_losses '''
```



```
Out [66]: ' Inference - There is no well defined relation between the speed at which cars go and the ins_losses '
```

```
In [ ]:
```