Name: Darryll Fonseca Registration Number: 21BRS1541

```python
import seaborn as sns
import matplotlib.pyplot as plt

crash_data = sns.load_dataset("car_crashes")

crash_data.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
```

```python
plt.figure(figsize=(8, 6))
sns.histplot(data=crash_data, x="total", kde=True)
plt.title("Distribution of Total Crashes")
plt.xlabel("Total Crashes")
plt.ylabel("Frequency")
plt.show()
```

Distribution of Total Crashes
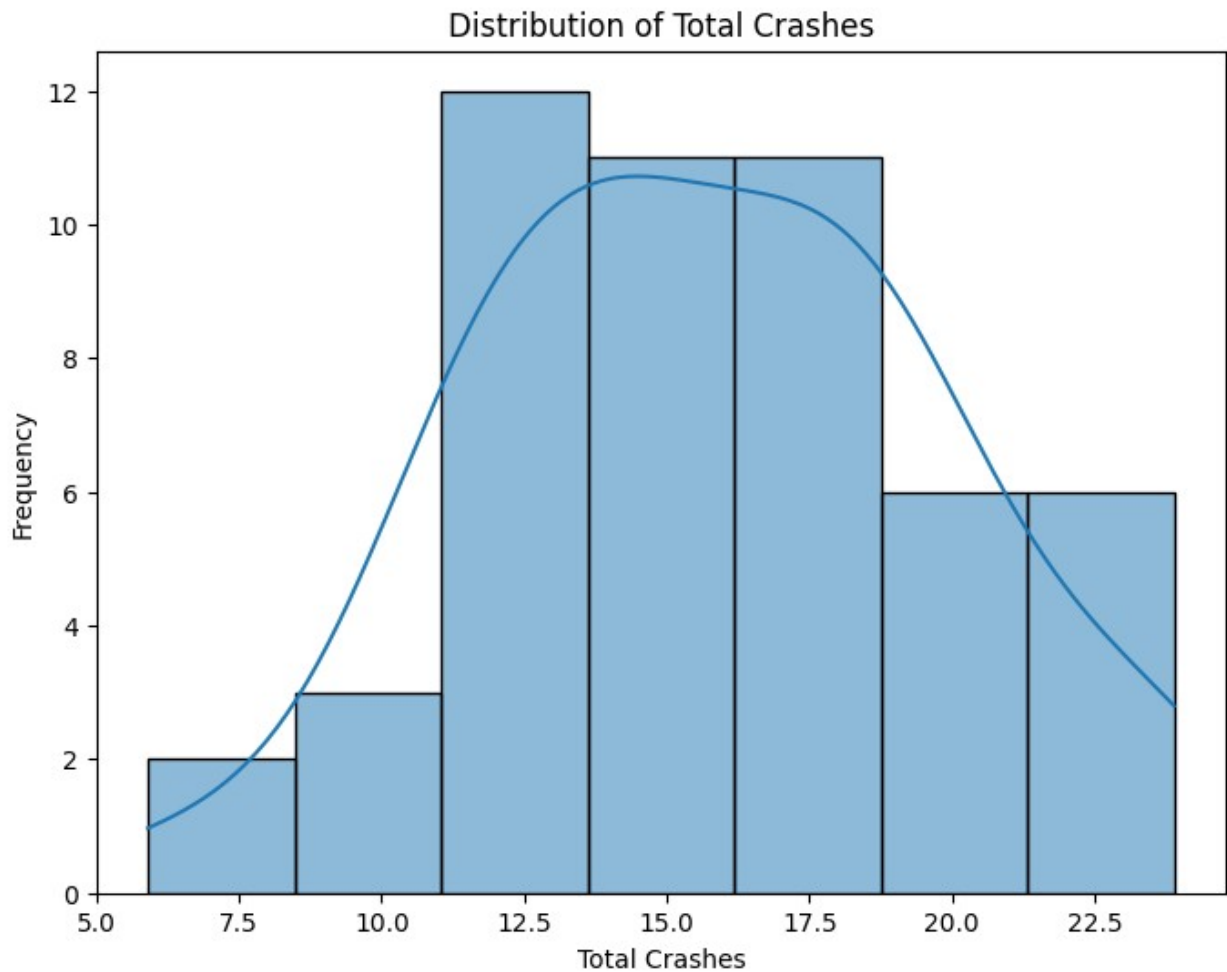
This histogram shows us how many states have different numbers of car crashes. Most states have a relatively low number of car crashes, but there are a few states where the number of car crashes is much higher than in the rest. These states with exceptionally high crash numbers stand out as outliers in the data.

```
sns.pairplot(crash_data)
plt.show()
```
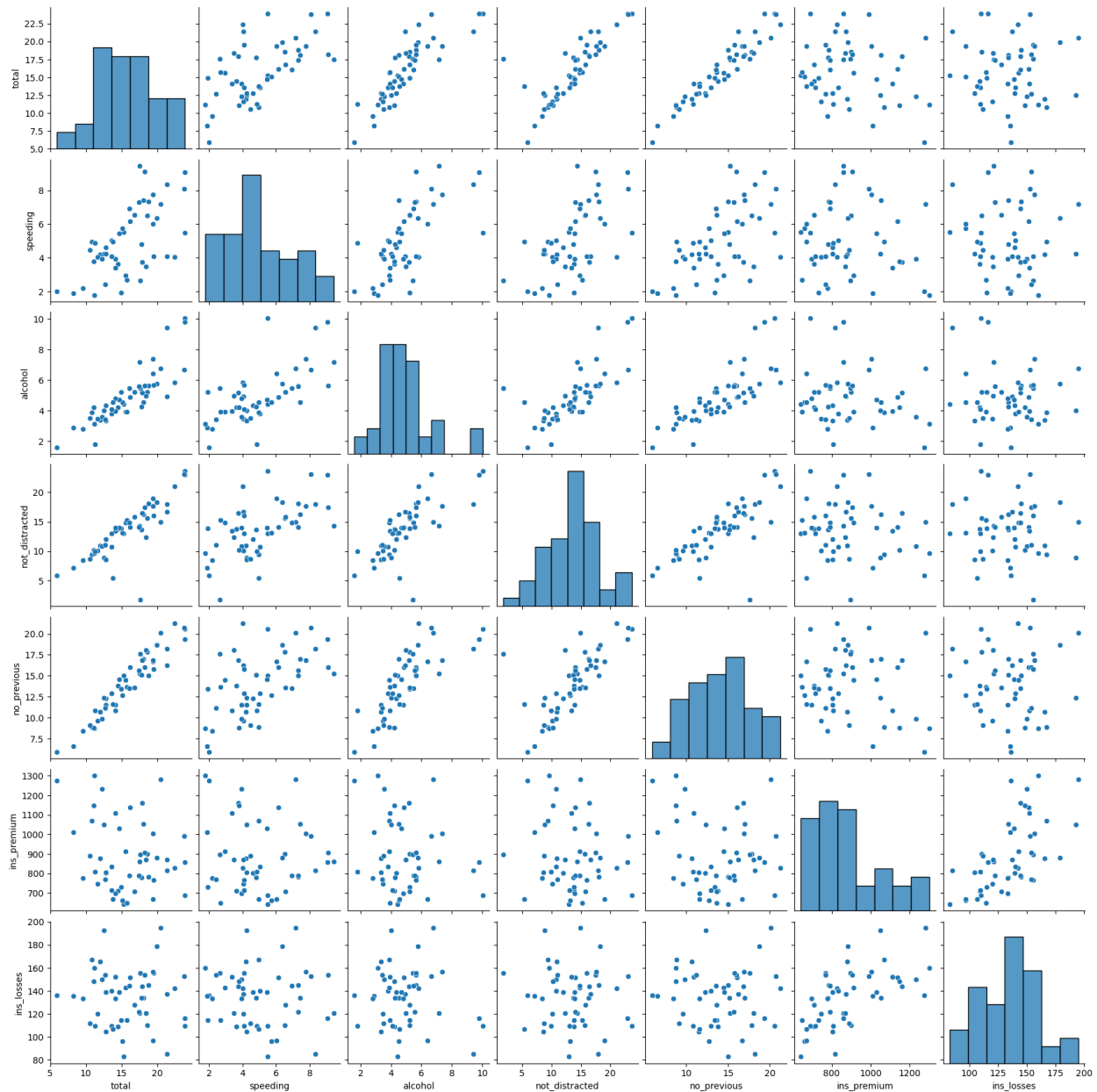
the pair plot helps us quickly identify potential correlations and patterns in the dataset, which can be valuable for understanding the relationships between different factors related to car crashes in various states. More dispersed the graph the less relation two of the points have

```
correlation_matrix = crash_data.corr()
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm")
plt.title("Correlation Heatmap")
plt.show()

C:\Users\DND\AppData\Local\Temp\ipykernel_17100\2382666951.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.
  correlation_matrix = crash_data.corr()
```


Correlation Heatmap
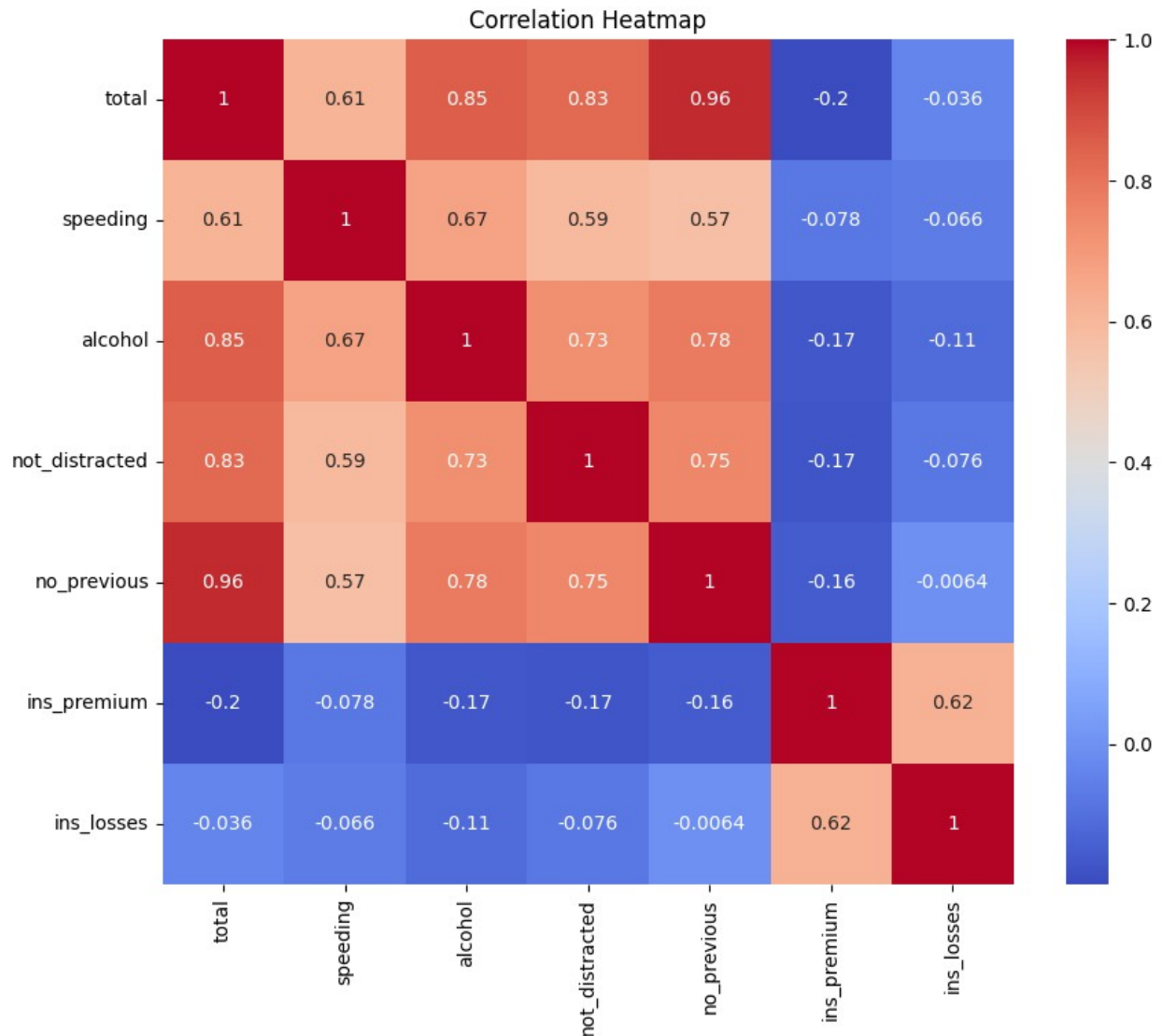
It helps identify variables that may have a significant impact on each other and can guide more in-depth analyses and modeling efforts. the warmer colours mean more direct relation while colder means an inverse relation.

```
sns.lmplot(data=crash_data, x="speeding", y="total")
plt.title("Regression Plot: Speeding vs. Total Crashes")
plt.show()
```

Regression Plot: Speeding vs. Total Crashes
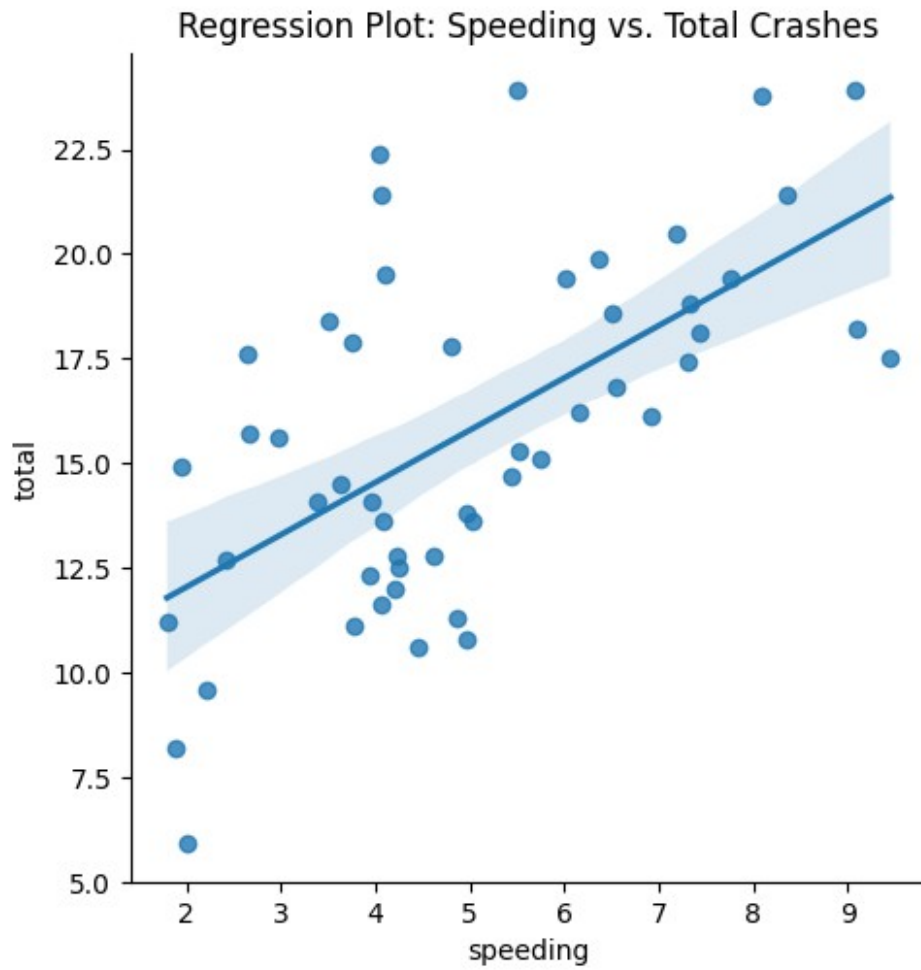
the regression plot provides evidence of a positive linear relationship between the percentage of speeding drivers and total crashes