```
from google.colab import files
uploaded = files.upload()
```

Choose Files No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving car_crashes.csv to car_crashes.csv

```
#import required libraries

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


df = pd.read_csv('/content/car_crashes.csv')
df.head()
```

	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses	abbre
0	18.8	7.332	5.640	18.048	15.040	784.55	145.08	A
1	18.1	7.421	4.525	16.290	17.014	1053.48	133.93	Al
2	18.6	6.510	5.208	15.624	17.856	899.47	110.35	A
3	22.4	4.032	5.824	21.056	21.280	827.34	142.39	Al
4	12.0	4.200	3.360	10.920	10.680	878.41	165.63	C

```
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.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()
```

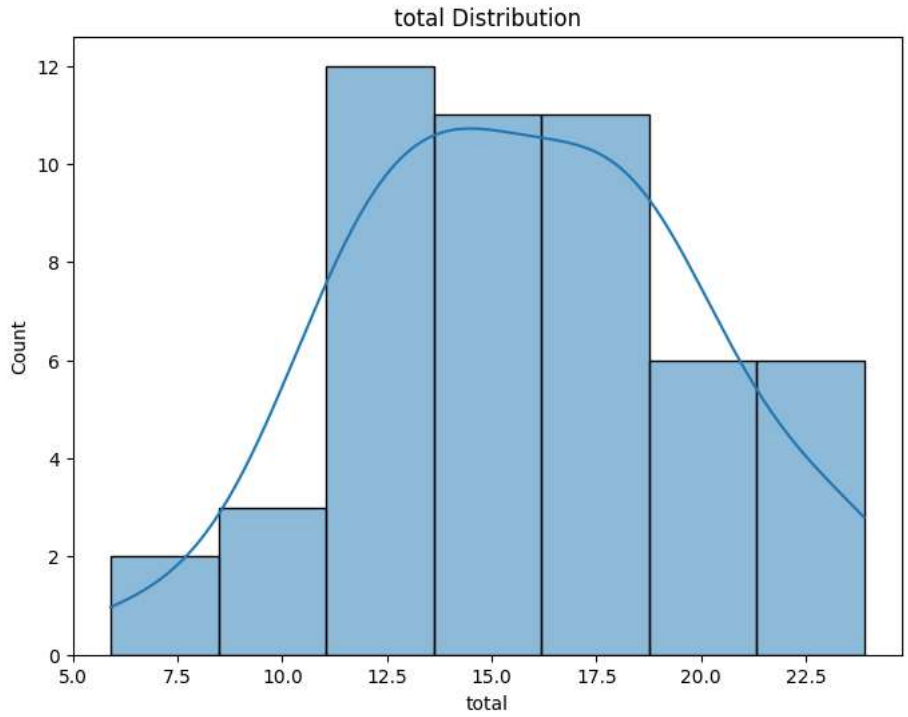
```
total             0
speeding          0
alcohol           0
not_distracted    0
no_previous       0
ins_premium       0
ins_losses        0
abbrev            0
dtype: int64
```

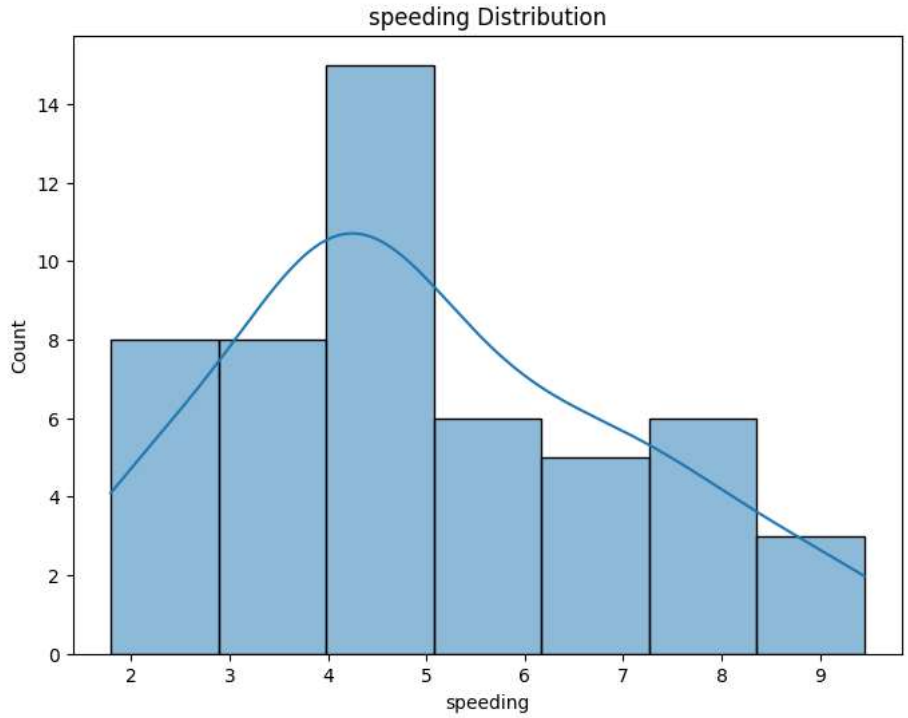
Univariate

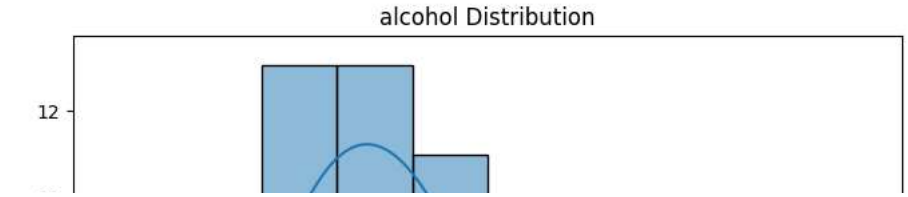
```
numerical_cols = ['total','speeding','alcohol','not_distracted','no_previous','ins_premium','ins_losses']


#histplot

for i in numerical_cols:
    plt.figure(figsize=(8, 6))
    sns.histplot(df[i], kde=True)
    plt.title(f'{i} Distribution')
    plt.show()
```

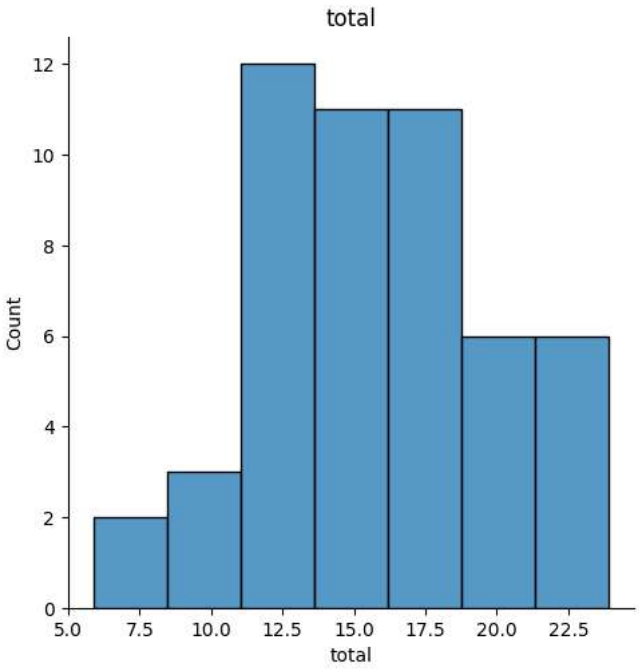

total Distribution
Count
total
0
2
4
6
8
10
12
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5

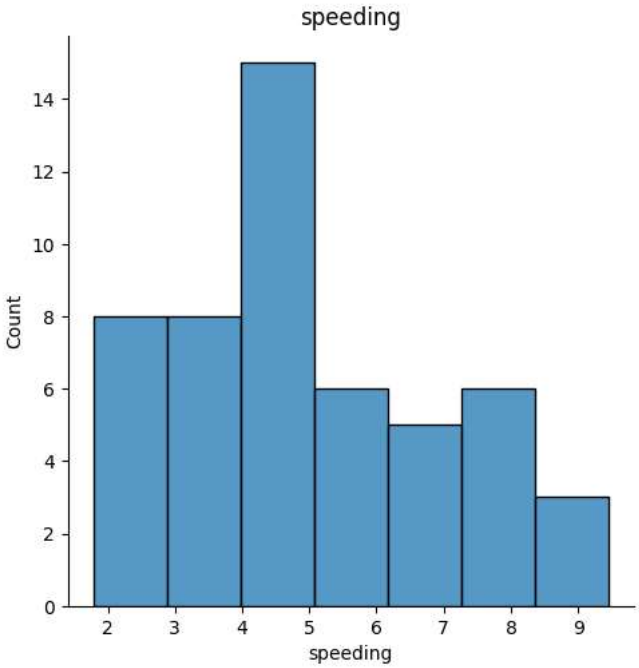

speeding Distribution
Count
speeding
0
2
4
6
8
10
12
14
2
3
4
5
6
7
8
9

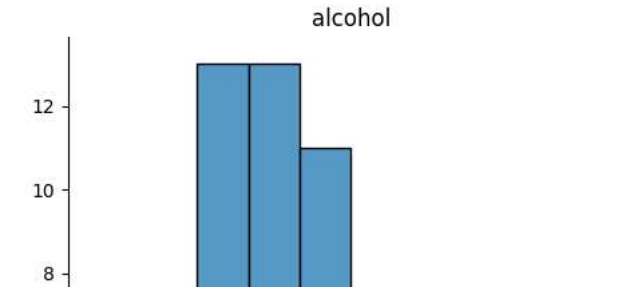

alcohol Distribution
12

```
#displot

for i in df.columns:
  sns.displot(df[i])
  plt.title(i)
  plt.show()
```

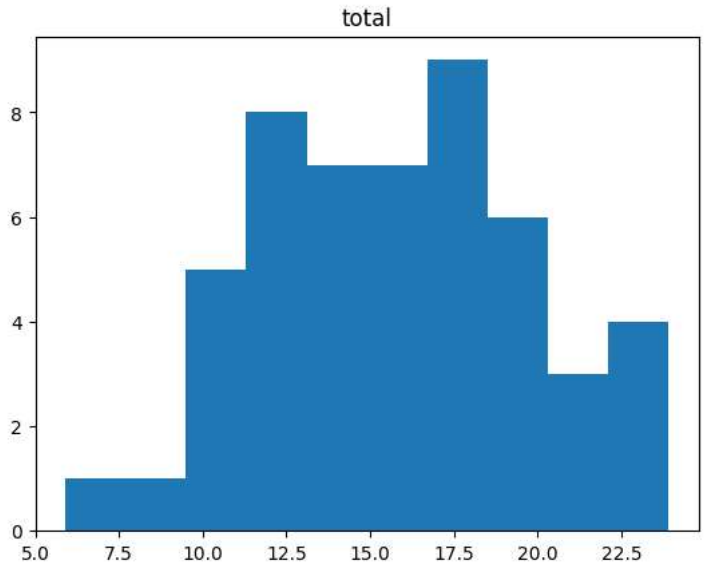

total
Count
total
0
2
4
6
8
10
12
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5

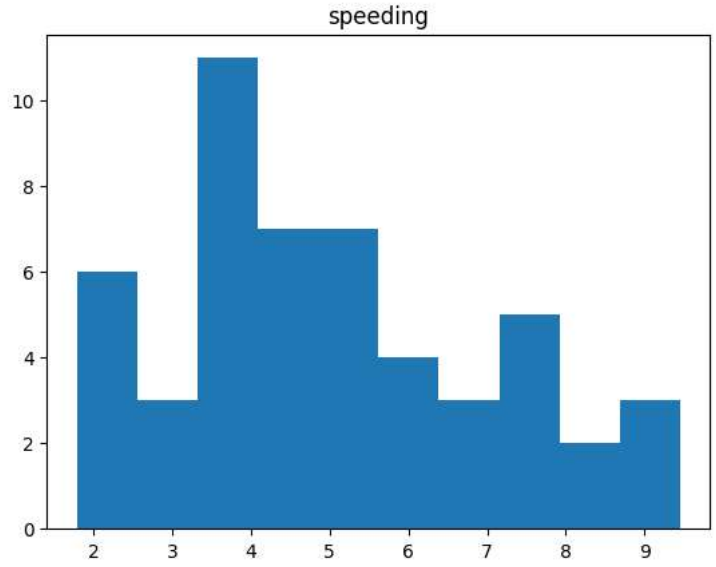

speeding
Count
speeding
0
2
4
6
8
10
12
14
2
3
4
5
6
7
8
9

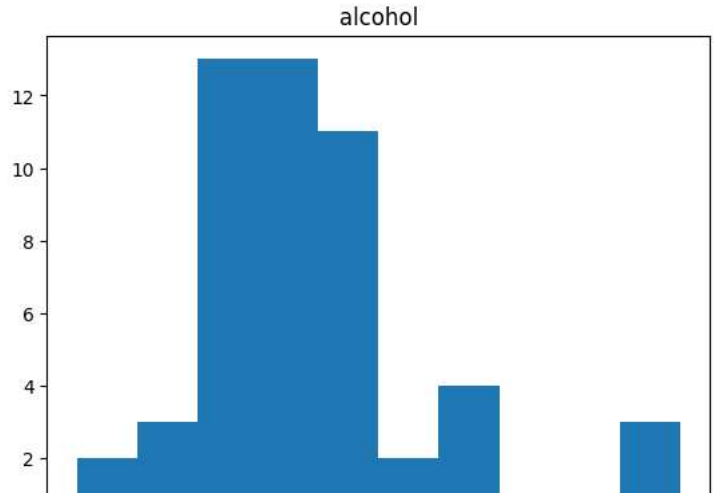

alcohol
8
10
12

```
#Histogram

for i in df.columns:
  plt.hist(df[i])
  plt.title(i)
  plt.show()
```

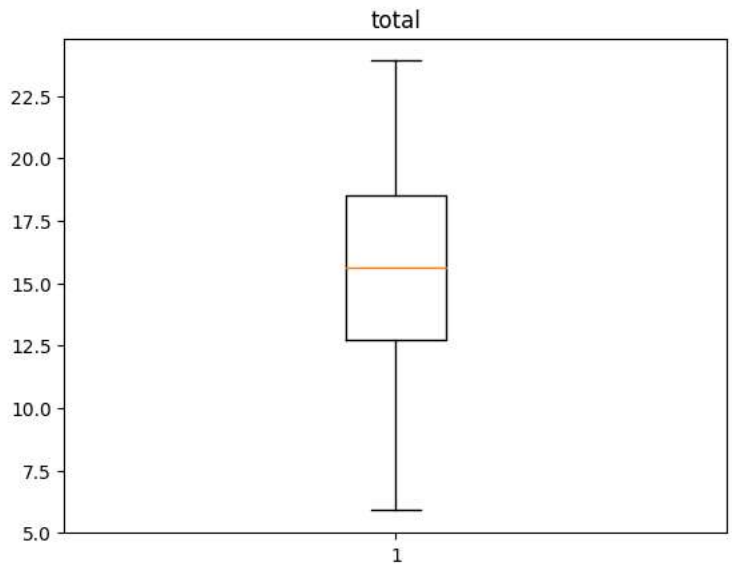

total
0
2
4
6
8
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5

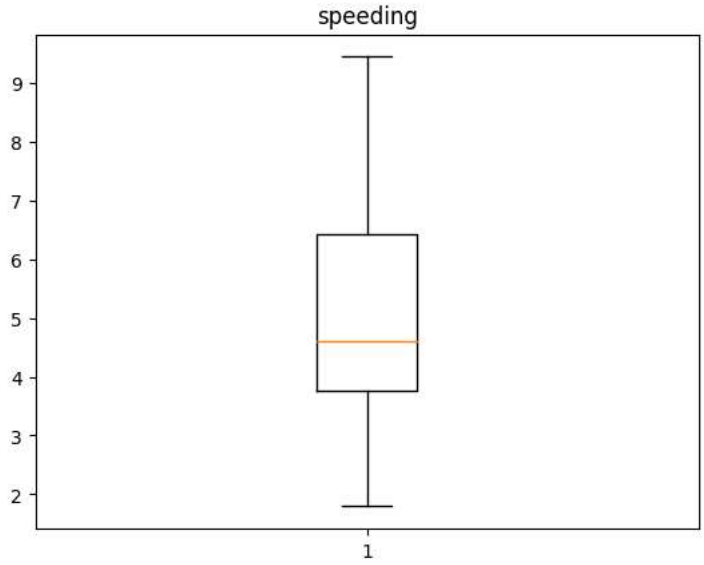

speeding
0
2
4
6
8
10
2
3
4
5
6
7
8
9

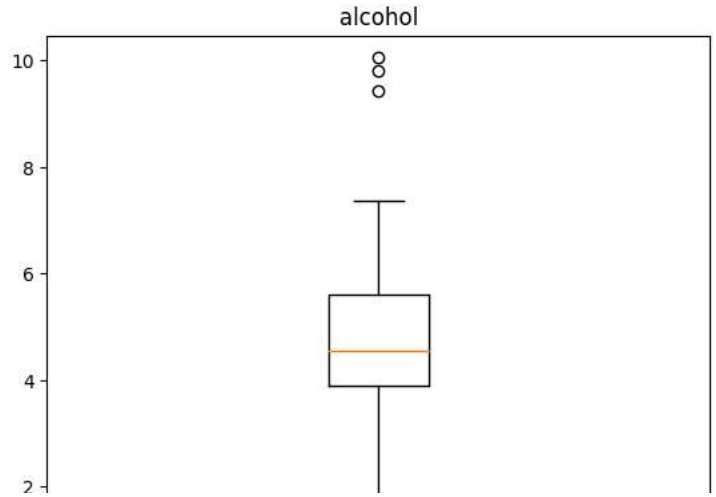

alcohol
2
4
6
8
10
12

```
# Boxplot

for i in numerical_cols:
  plt.boxplot(df[i])
  plt.title(i)
  plt.show()
```

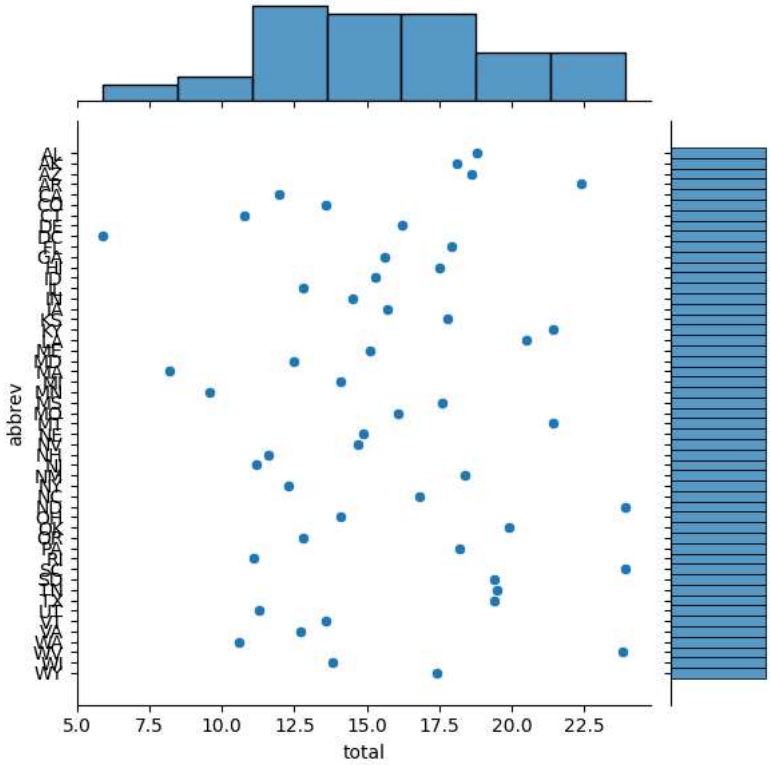

total
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
1

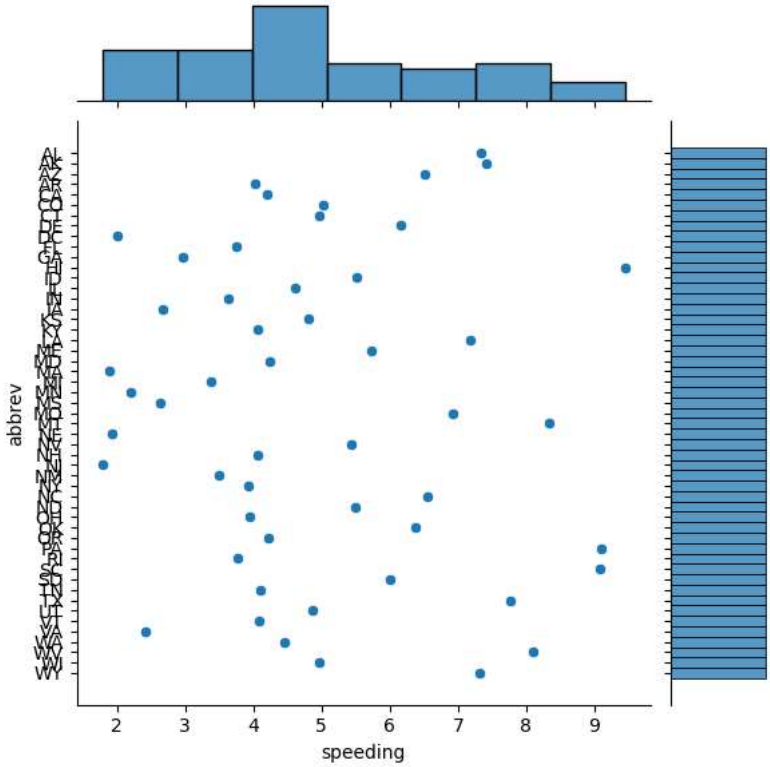

speeding
9
8
7
6
5
4
3
2
1


alcohol
10
8
6
4
2

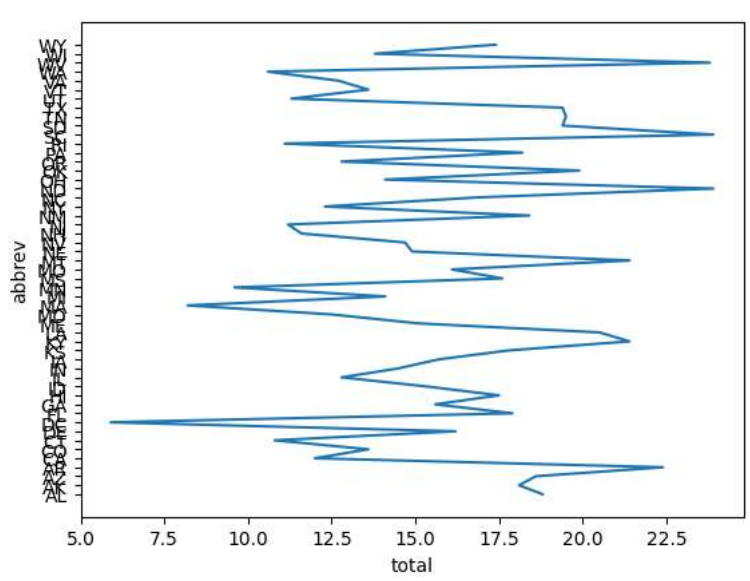
Bi variate

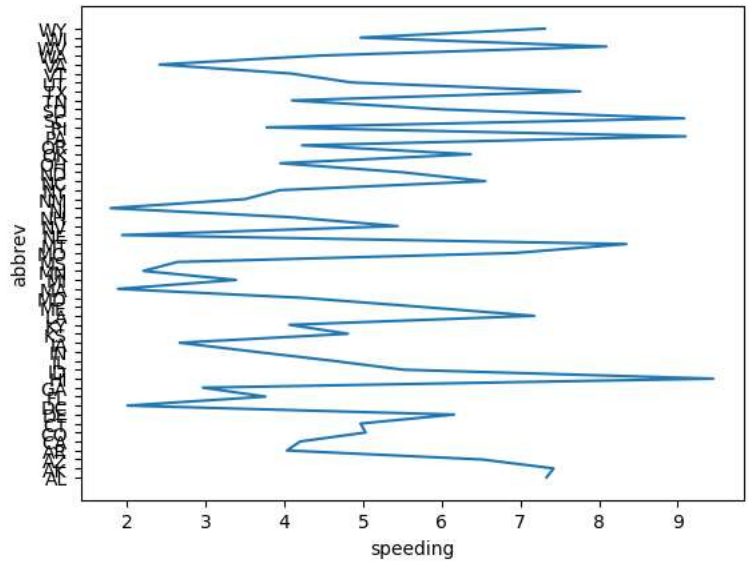
```
#jointplot
for i in df.columns:
  sns.jointplot(x= df[i],y =df['abbrev'],data=df)
```

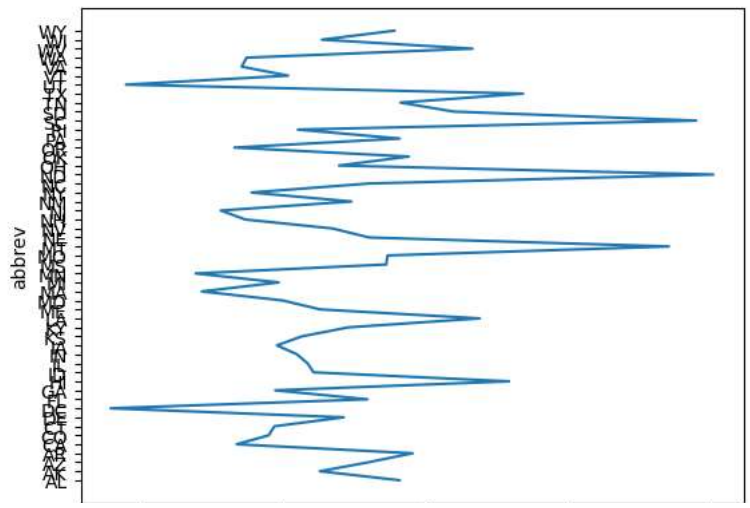

abbrev
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
total


abbrev
2
3
4
5
6
7
8
9
speeding

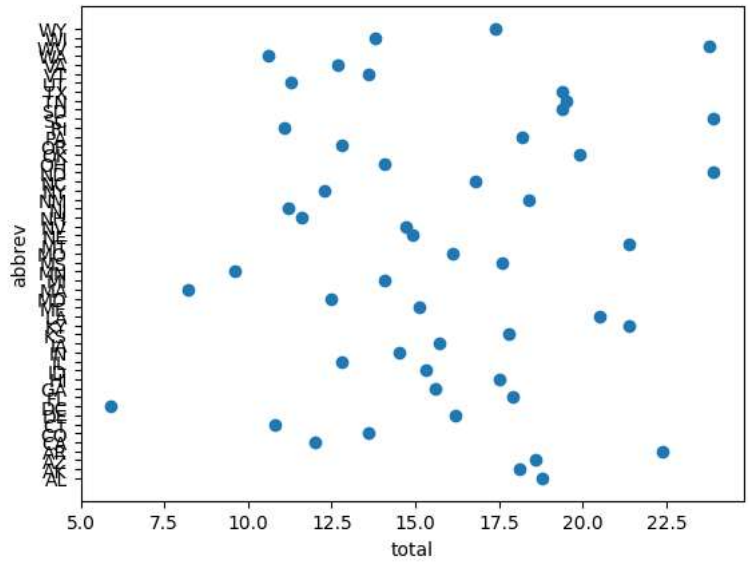
```
#Line Chart
for i in df.columns:
  plt.plot(df[i],df['abbrev'])
  plt.xlabel(i)
  plt.ylabel("abbrev")
  plt.show()
```

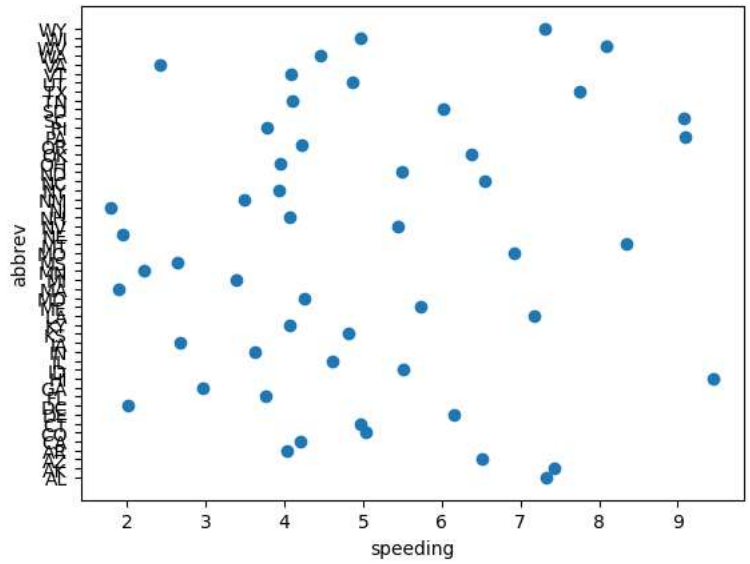

abbrev
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
total

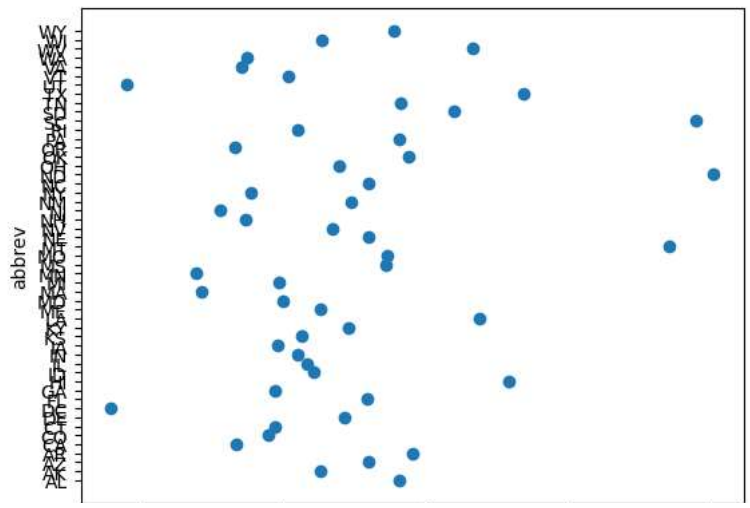

abbrev
2
3
4
5
6
7
8
9
speeding


abbrev

```
#Scatterplot
for i in df.columns:
  plt.scatter(df[i],df['abbrev'])
  plt.xlabel(i)
  plt.ylabel("abbrev")
  plt.show()
```


abbrev
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
total


abbrev
2
3
4
5
6
7
8
9
speeding

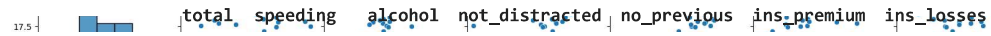

abbrev

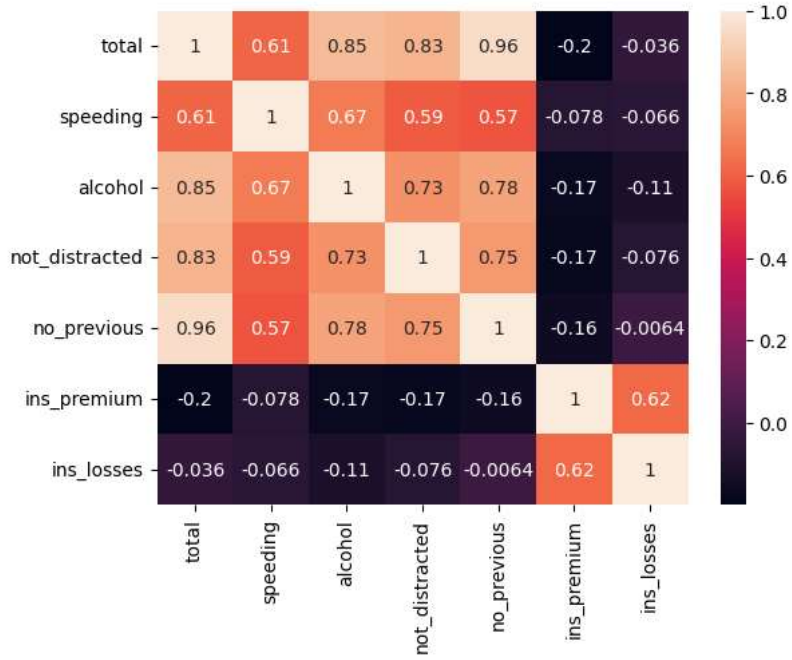
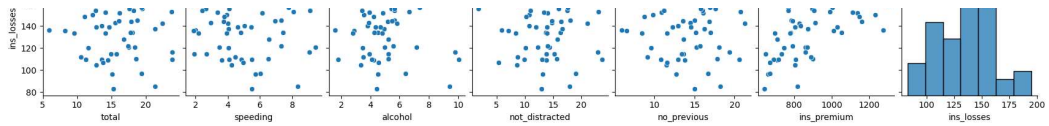
Multi variate

```
# Pairplot
sns.pairplot(df)
```

```
df.corr()
```

```
<ipython-input-25-2f6f6606aa2c>:1: FutureWarning: The default value of numeric_only in DataFrame.corr i
  df.corr()
```




```
#heatmap-correlation graph
sns.heatmap(df.corr(),annot=True)
```

```
<ipython-input-26-c9eba097a1ac>:2: FutureWarning: The default value of numeric_only in DataFrame.corr i
  sns.heatmap(df.corr(),annot=True)
<Axes: >
```