NAME-AMAR DEEP SINGH

In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [10]:
```
sns.get_dataset_names()
```

Out[10]:
```
['anagrams',
 'anscombe',
 'attention',
 'brain_networks',
 'car_crashes',
 'diamonds',
 'dots',
 'dowjones',
 'exercise',
 'flights',
 'fmri',
 'geyser',
 'glue',
 'healthexp',
 'iris',
 'mpg',
 'penguins',
 'planets',
 'seaice',
 'taxis',
 'tips',
 'titanic']
```

In [4]:
```
dataset= sns.load_dataset('car_crashes')
```

In [14]:
```
dataset.head()
```

Out[14]:

	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

In [15]:
```
#getting the information of the dataset
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 8 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
  0total           51 non-null     float64
  1speeding        51 non-null     float64
  2alcohol         51 non-null     float64
  3not_distracted  51 non-null     float64
  4no_previous     51 non-null     float64
  5ins_premium     51 non-null     float64
  6ins_losses      51 non-null     float64
  7abbrev          51 non-null     object dtypes: float64(7), object(1)
 memory usage: 3.3+ KB
```

[16]:
```
#getting the shape
dataset.shape
```

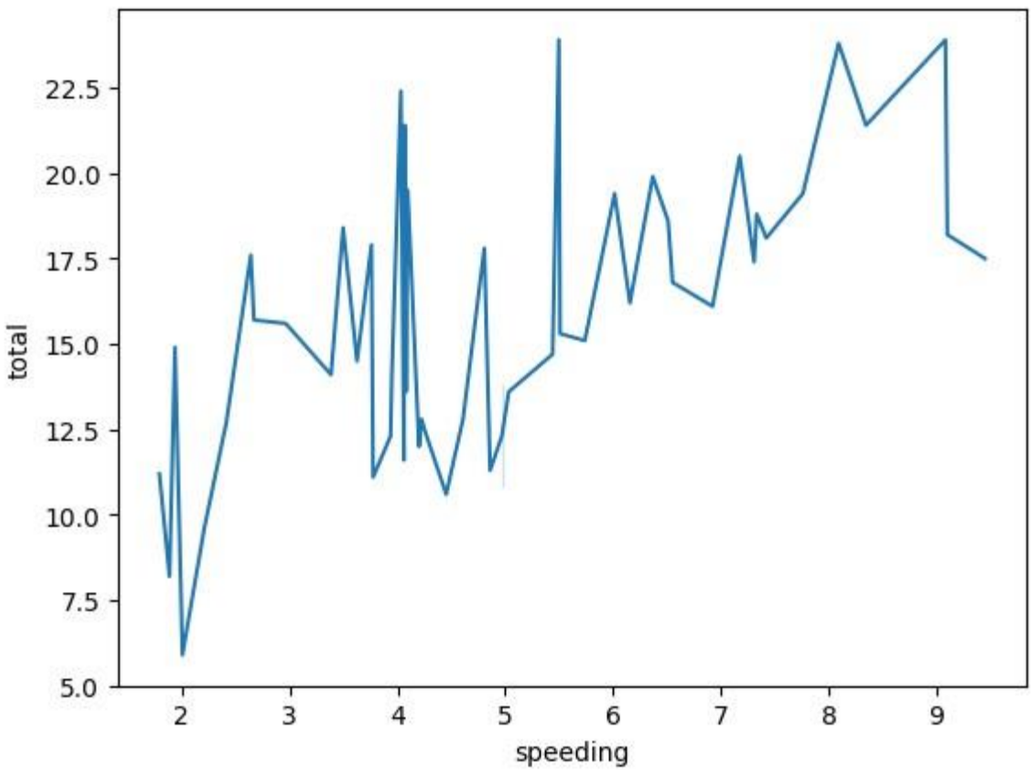
In

Out[16]: (51, 8)

In [20]:
```
# plotting lineplot
sns.lineplot(data=dataset, y='total', x='speeding')
```

```
C:\Users\hp\AppData\Local\Programs\Python\Python310\lib\site-packages\seaborn
\_oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will
be removed in a future version. Use isinstance(dtype, CategoricalDtype) inste
ad
  if pd.api.types.is_categorical_dtype(vector):
C:\Users\hp\AppData\Local\Programs\Python\Python310\lib\site-packages\seaborn
\_oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will
be removed in a future version. Use isinstance(dtype, CategoricalDtype) inste
ad
  if pd.api.types.is_categorical_dtype(vector):
C:\Users\hp\AppData\Local\Programs\Python\Python310\lib\site-packages\seaborn
\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will
be removed in a future version. Convert inf values to NaN before operating in
stead.  with pd.option_context('mode.use_inf_as_na', True):
C:\Users\hp\AppData\Local\Programs\Python\Python310\lib\site-packages\seaborn
\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will
be removed in a future version. Convert inf values to NaN before operating in
stead.  with pd.option_context('mode.use_inf_as_na', True):
```

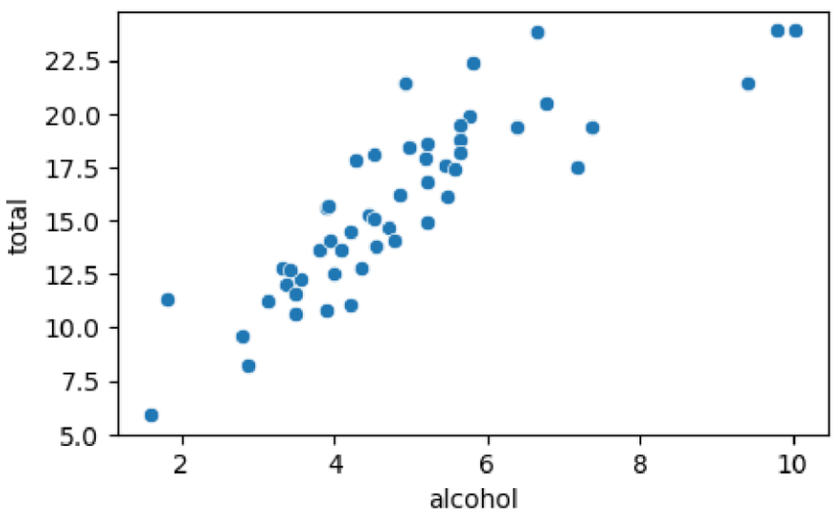
Out[20]: <AxesSubplot: xlabel='speeding', ylabel='total'>



Inference: · total -> Number of drivers involved in fatal collisions per billion miles (5.900–23.900) speeding -> Percentage Of Drivers Involved In Fatal Collisions Who Were Speeding (1.792–9.450)

this is a line plot between the speeding dircvers and total no of drivers involved in collosion we can clearly see that as speeding increases the no.of collesions increases drastically

In [10]:
```
#scatter plot
plt.figure(figsize=(5, 3))
sns.scatterplot(data=dataset, y='total', x='alcohol')
plt.show()
```

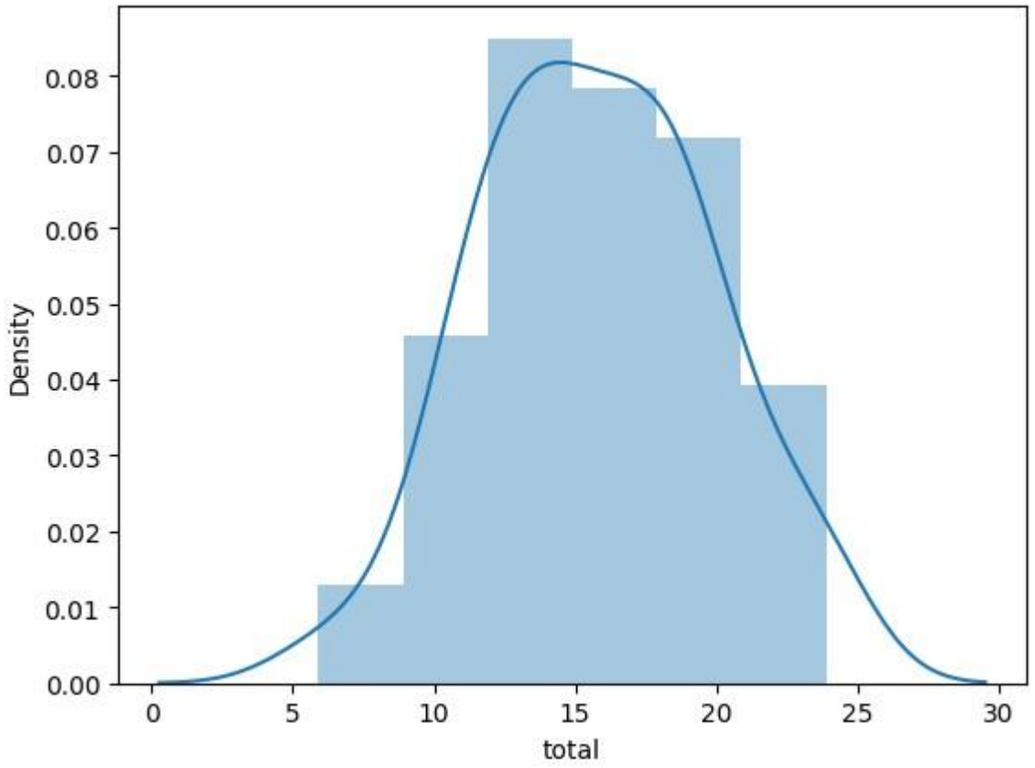


In [ ]:
```
Inference:
    The scatter plot of the 'total' and 'alcohol' columns in the car_crashes da
    reveals a positive correlation between these variables.
    This suggests that as the total number of car crashes increases, there tend
    be a corresponding rise in the number of incidents involving alcohol. Howev
    it's important to note that while this correlation exists, it doesn't simpl
    causation. Other factors may contribute to this relationship, and further
    analysis would be needed to establish any causal links. Additionally,
    the presence of a few outliers indicates that there are instances where the
    number of alcohol-related incidents deviates significantly from the general
    This scatter plot provides a valuable visual representation of the relation
    offering insights into potential areas of concern for traffic safety and al
```

In [30]:
```
#dist plot sns.distplot(dataset['total'])
```

```
C:\Users\hp\AppData\Local\Temp\ipykernel_17484\2994018954.py:2: 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 (https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751)

  sns.distplot(dataset['total'])
C:\Users\hp\AppData\Local\Programs\Python\Python310\lib\site-packages\seaborn\_oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will be removed in a future version. Use isinstance(dtype, CategoricalDtype) instead
  if pd.api.types.is_categorical_dtype(vector):
C:\Users\hp\AppData\Local\Programs\Python\Python310\lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use_inf_as_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating in stead.  with pd.option_context('mode.use_inf_as_na', True):
```

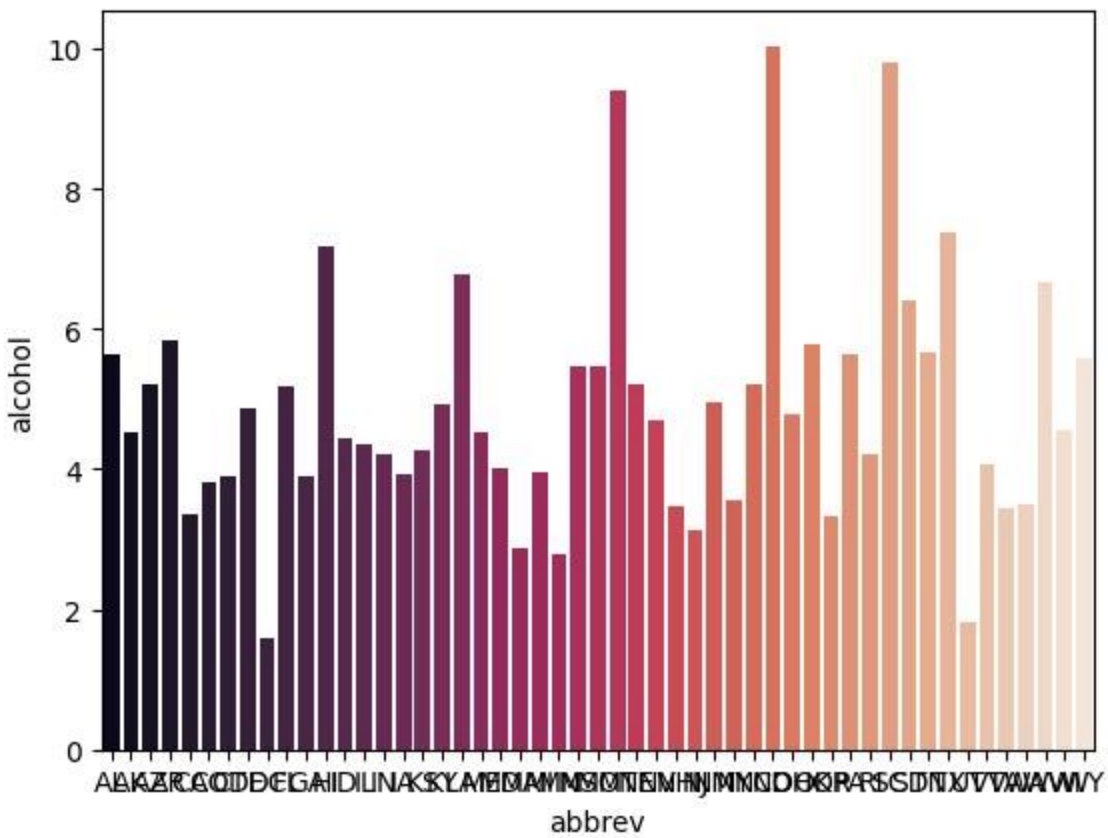
Out[30]: <AxesSubplot: xlabel='total', ylabel='Density'>



In [ ]:
```
Inference:

This plot illustrates the distribution of total car crashes. It appears to be s
indicating that there are more instances of lower crash counts. The majority of
suggesting that most areas experience a relatively low number of car crashes. H
counts, which are represented by the tail on the right side of the distribution
the frequency and severity of car accidents in the dataset.
```

In [9]:
```
#bar plot res = sns.barplot(x ='abbrev' , y='alcohol', data=dataset ,
palette="rocket" ) plt.show()
```

```
C:\Users\hp\AppData\Local\Programs\Python\Python310\lib\site-packages\seaborn
\_oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will
be removed in a future version. Use isinstance(dtype, CategoricalDtype) inste
ad
  if pd.api.types.is_categorical_dtype(vector):
C:\Users\hp\AppData\Local\Programs\Python\Python310\lib\site-packages\seaborn
\_oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will
be removed in a future version. Use isinstance(dtype, CategoricalDtype) inste
ad
  if pd.api.types.is_categorical_dtype(vector):
C:\Users\hp\AppData\Local\Programs\Python\Python310\lib\site-packages\seaborn
\_oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and
will be removed in a future version. Use isinstance(dtype, CategoricalDtype)
inste ad  if pd.api.types.is_categorical_dtype(vector):
```



In [ ]:

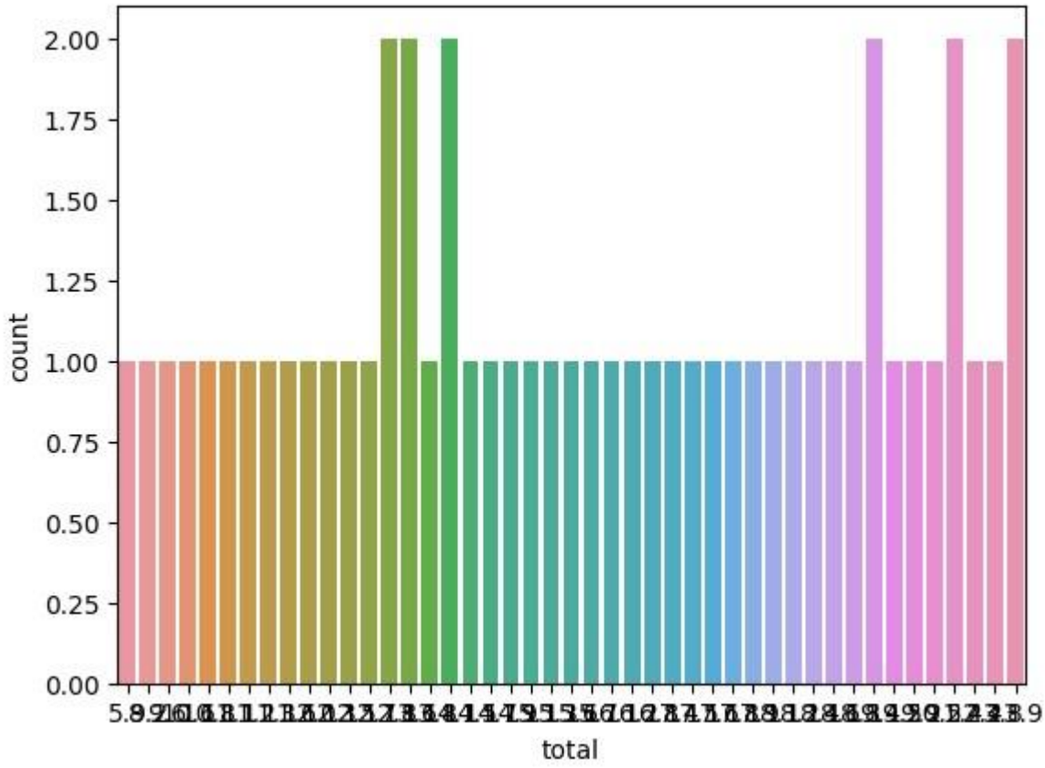
```
inference:
    This bar plot displays the average alcohol involvement in car crashes for d
    The vertical bars provide a visual comparison of the alcohol-related incide
    From the plot, it can be observed that some regions have notably higher ave
    This information can be crucial for identifying areas where stricter measur
    the risks associated with alcohol-related accidents.
```

In [18]:

```
#countplot

sns.countplot(x='total', data = dataset)
```

```
C:\Users\hp\AppData\Local\Programs\Python\Python310\lib\site-packages\seaborn
\_oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will
be removed in a future version. Use isinstance(dtype, CategoricalDtype) inste
ad
  if pd.api.types.is_categorical_dtype(vector):
C:\Users\hp\AppData\Local\Programs\Python\Python310\lib\site-packages\seaborn
\_oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will
be removed in a future version. Use isinstance(dtype, CategoricalDtype) inste
ad
  if pd.api.types.is_categorical_dtype(vector):
C:\Users\hp\AppData\Local\Programs\Python\Python310\lib\site-packages\seaborn
\_oldcore.py:1498: FutureWarning: is_categorical_dtype is deprecated and will
be removed in a future version. Use isinstance(dtype, CategoricalDtype) inste
ad
  if pd.api.types.is_categorical_dtype(vector):
```

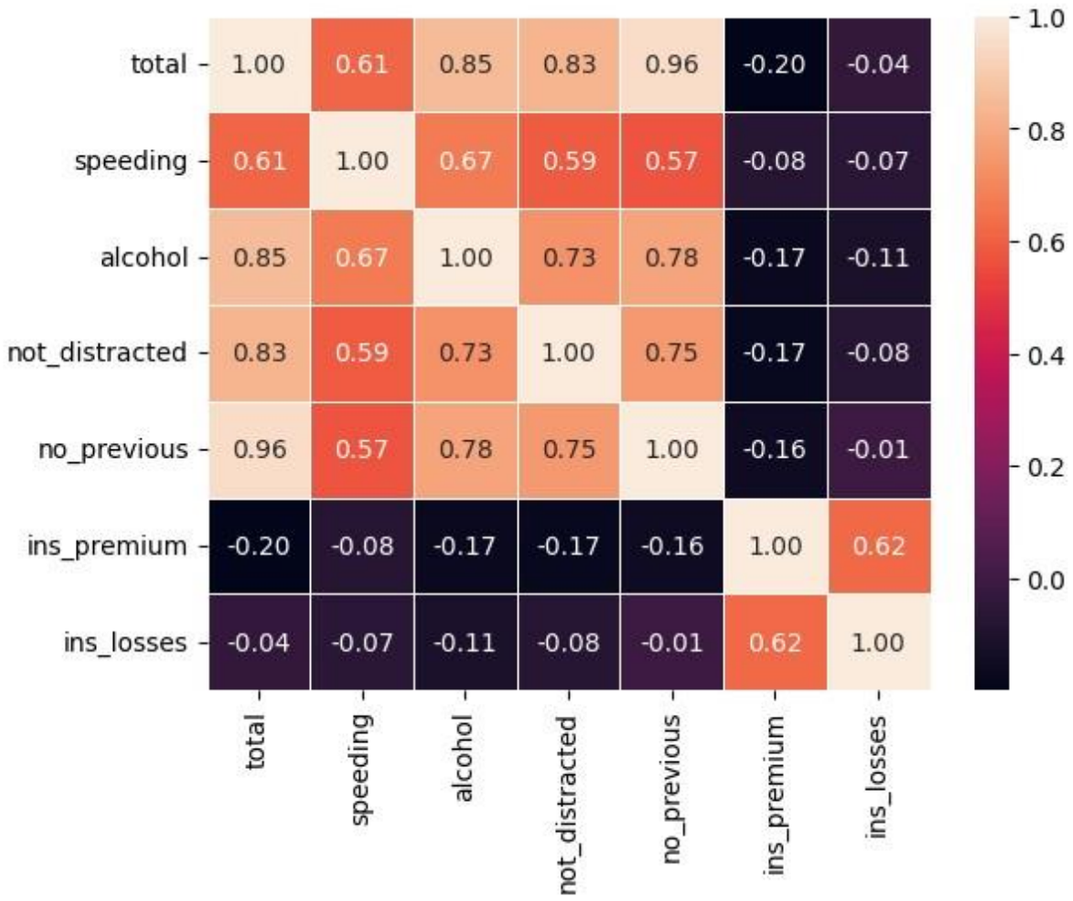
Out[18]: <AxesSubplot: xlabel='total', ylabel='count'>



inference: This count plot offers insight into the distribution of total car crash occurrences. It displays the frequency of different total crash counts in the dataset. From the plot, it can be observed that the majority of instances involve a relatively low number of car crashes. However, there are also instances of higher crash counts, although they are less frequent. This information is crucial for understanding the distribution and frequency of car accidents in the dataset, which may have implications for safety measures and policy decisions.

In [24]:
```
#co-relation  Heatmap numeric_data =
dataset.select_dtypes(include=['float64', 'int64'])
sns.heatmap(numeric_data.corr(), cbar=True, annot=True, fmt=".2f", linewidths=0
```

Out[24]: <AxesSubplot: >



In [ ]:
```
inference:     here the lighter colors indicate that they are positively
corealated i.e th     are directly proportional , whereas the darker colors
indicates that they a     corelated i.e they are inversly propotional.
```