Assignment – 2 (AIML)

Name – Aman kumar

Roll – 21BCE3132

to take car crashes dataset from seaborn library

**import** seaborn **as** sns

*# Load the car\_crashes dataset*

car\_crashes = sns.load\_dataset("car\_crashes")

*# Display the first few rows of the dataset*

print(car\_crashes.head())

*# Display some basic statistics about the dataset*

print(car\_crashes.describe())

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

total speeding alcohol not\_distracted no\_previous \

count 51.000000 51.000000 51.000000 51.000000 51.000000

mean 15.790196 4.998196 4.886784 13.573176 14.004882

std 4.122002 2.017747 1.729133 4.508977 3.764672

min 5.900000 1.792000 1.593000 1.760000 5.900000

25% 12.750000 3.766500 3.894000 10.478000 11.348000

50% 15.600000 4.608000 4.554000 13.857000 13.775000

75% 18.500000 6.439000 5.604000 16.140000 16.755000

max 23.900000 9.450000 10.038000 23.661000 21.280000

ins\_premium ins\_losses

count 51.000000 51.000000

mean 886.957647 134.493137

std 178.296285 24.835922

min 641.960000 82.750000

25% 768.430000 114.645000

50% 858.970000 136.050000

75% 1007.945000 151.870000

max 1301.520000 194.780000

to load the dataset

**import** seaborn **as** sns

*# Load the car\_crashes dataset*

car\_crashes = sns.load\_dataset("car\_crashes")

*# Display the first few rows of the dataset*

print(car\_crashes.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

visualizing the data

**import** seaborn **as** sns

**import** matplotlib.pyplot **as** plt

*# Load the car\_crashes dataset*

car\_crashes = sns.load\_dataset("car\_crashes")

*# Pairplot for visualizing relationships between numeric columns*

sns.pairplot(car\_crashes)

plt.show()

*# Histogram of the total crashes*

sns.histplot(car\_crashes["total"], bins=15, kde=True)

plt.xlabel("Total Crashes")

plt.ylabel("Frequency")

plt.title("Histogram of Total Crashes")

plt.show()

*# Scatter plot of alcohol versus total crashes*

sns.scatterplot(x="total", y="alcohol", data=car\_crashes)

plt.xlabel("Total Crashes")

plt.ylabel("Alcohol Consumed")

plt.title("Scatter Plot: Alcohol vs. Total Crashes")

plt.show()

Inference is must for each and every graph

detailed inferences for each of the visualizations from the provided code:

1. **Pairplot (Relationships between Numeric Columns):**
   * The pairplot exhibits scatter plots for combinations of numerical columns and showcases histograms for the distribution of each individual column.
   * It reveals a robust positive correlation between "total" crash occurrences and levels of "alcohol" consumption, implying that states with higher alcohol consumption often experience more total crashes.
   * Additionally, there is a positive relationship between "speeding" and "not\_distracted" variables concerning "total" crashes. This suggests that states with elevated rates of speeding and fewer instances of distracted driving may witness higher crash totals.
   * The histograms along the diagonal axis depict right-skewed distributions for both "total" crashes and "alcohol" consumption. This skewness indicates a prevalent trend towards lower values with a few instances of notably higher values.
2. **Histogram of Total Crashes:**
   * The histogram provides a visual representation of how total crash counts are distributed among various states.
   * In the majority of states, there is a concentration of crash counts within the range of 15 to 20, suggesting that this range is a typical occurrence for total crashes.
   * However, there are a select few states, as seen in the rightmost portion of the histogram, that stand out due to significantly higher crash counts than the rest. These states can be considered outliers in terms of crash frequency, indicating a distinctive pattern of road safety or traffic incidents compared to the majority of states.
3. **Scatter Plot (Alcohol vs. Total Crashes):**
   * The scatter plot delves into the connection between the level of "alcohol" consumption and the frequency of "total" crashes in each state.
   * Although there is a noticeable positive trend suggesting that states with elevated alcohol consumption tend to experience higher total crash counts, it's crucial to emphasize that correlation does not establish causation. This means that while the two variables are related, one does not necessarily cause the other. There may be other underlying factors contributing to this observed relationship.
   * The plot further highlights instances where states exhibit low alcohol consumption but still report high total crash counts. This underscores the idea that alcohol consumption alone cannot fully account for crash rates.
   * It becomes evident that several factors, including road conditions, law enforcement efforts, and demographic characteristics, are likely playing significant roles in influencing crash rates among states. These factors should be thoroughly considered when analyzing and addressing road safety issues.

These inferences provide a thorough understanding of the visualized data, highlighting correlations, distributions, and potential areas for further analysis.

