Analyzing the Impact of Weather on Traffic Collision Dynamics in Los Angeles

1 Question

Traffic accidents are a serious concern worldwide. This study explores how weather conditions, such as temperature, humidity, visibility, and wind speed, might influence the number of accidents in Los Angeles City. By analyzing data from January to June 2020, I aim to find insight of the key question: What is the impact of the weather on the frequency of traffic collisions in Los Angeles? The insights gained will help traffic safety experts and city planners develop strategies to improve road safety and minimize weather-related risks for all drivers.

2 Data Sources

2.1 Traffic Collision Data

- **Source**: Los Angeles Police Department (LAPD)
- **Period**: January to June 2020
- **Structure**: I have used two data columns: the time when the accident occurred and the date of the incident and based on them included the number of accidents that happened per day (collision count) Based on that, I tried to find the correlation with the weather data.
  - **date**: The date is used in this format in MM/DD/YYYY
  - **time**: Indicates the time of the accident
- **License**: CC0 1.0 Universal

2.2 Weather Data

- **Source**: Visual Crossing
- **Period**: January To June 2020
- **Structure**: The data includes the following columns: datetime, temp, humidity, windspeed, and visibility
  - **date**: The date is used in this format in MM/DD/YYYY
  - **time**: Represents the time of the accident

  - **temp**: Temperature in degree Celcius
  - **humidity**: Humidity percentage
  - **windspeed**: Wind speed in Kmph
  - **visibility**: Visibility in Kilometers.

- **License**: Visual Crossing Corporation

3 Analysis

3.1 Data Integration

This study combines data from two different data sources, as mentioned above; traffic collision data and weather data, by aligning both datasets based on the date. This integration process involved several key steps. First, I standardized the date formats across both datasets to ensure consistency. After that, the collision count and the weather data were merged, creating a unified dataset. The resulting dataset included daily traffic flow information along with the corresponding weather parameters, such as temperature, humidity, wind speed, and visibility. This integration enabled a comprehensive analysis of how weather conditions might influence traffic collisions.

3.1.1 Descriptive Statistics

I analyzed the basic descriptive statistics to get a clear understanding of the distributions and key metrics in the traffic flow and weather data. This helped me spot any anomalies or patterns within the dataset.

3.1.2 Correlation Analysis

I calculated the correlation matrix to assess the strength and relationships between weather parameters and the number of traffic collisions, as illustrated in Figure 1. The correlation coefficient ranges from -1 to 1, where values close to 1 indicate a strong positive correlation, values near -1 represent a strong negative correlation, and values around 0 suggest little to no correlation.

4 Results

So from my analysis, I found some insights that could be helpful; I did not consider all the parameters that could give more insight on the relationship, but at the same time the analysis would look confusing.

4.1 Impact of Temperature:

There is a slight negative correlation, indicating that higher temperatures are modestly associated with fewer collisions. This could suggest that people travel less or adjust their driving behavior during warmer conditions, potentially lowering the risk of accidents.

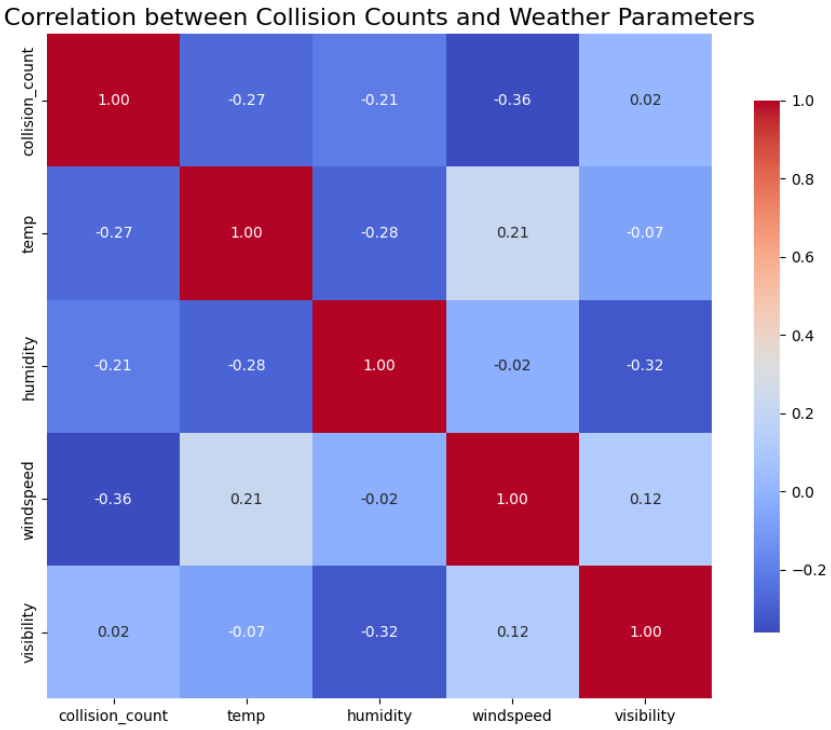

Figure 1: Correlation Heatmap

4.2 Impact of Humidity:

A slight negative correlation is observed, showing that as humidity increases, the number of collisions decreases slightly. This might imply that higher humidity—often linked to rain or fog—encourages more cautious driving, which could reduce accidents. I thought of including the precipitation attribute from weather data but after seeing this, and from the analysis, I decided not to include it.

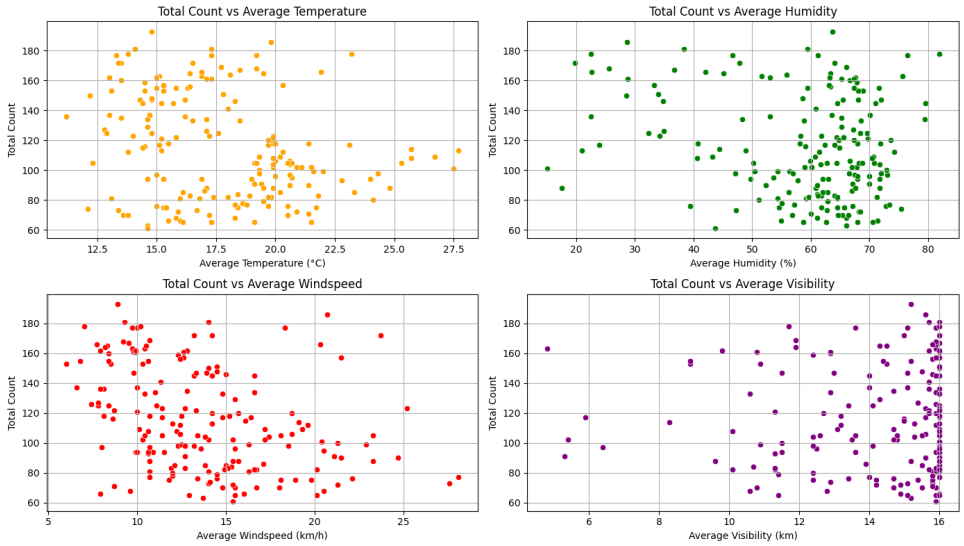

Figure 2: Correlation of total collision count and weather parameters

4.3 Impact of Wind-speed:

This moderately negative correlation suggests that higher wind speeds are more strongly associated with fewer collisions. One possible explanation is that severe wind conditions discourage travel, or when people are driving, they get more careful and slow down their speed, consequently fewer accidents.

4.4 Impact of Visibility:

There is no significant correlation between visibility and collision counts. This suggests that other factors, such as road conditions or driver behavior, likely play a more substantial role in influencing traffic accidents than visibility alone.

4.5 Interpretation of the Analysis

The analysis offers useful insights for improving road safety in Los Angeles. A moderate negative correlation between wind speed and collision counts suggests that strong winds might discourage travel, leading to fewer accidents. Similarly, slight negative correlations with temperature and humidity indicate that warmer or more humid weather may encourage safer driving behavior.

Interestingly, visibility showed no significant impact on collision rates, highlighting that other factors like road conditions or driver behavior may play a bigger role. By understanding these patterns, the city can work toward reducing accidents and making roads safer for everyone.

5 Limitations

5.1 Correlation vs Causation

While the analysis identifies relationships between weather conditions and traffic collisions, it doesn't prove from the parameters I considered that weather directly causes collision-count changes. Other factors could be at play.

5.2 Time of Day Effects

The study looks at daily data. Weather impacts might differ during peak and off-peak hours.

5.3 Exclusion of Other Data

The analysis doesn't include pedestrian activity, which could offer a fuller picture of how weather affects overall urban mobility.

5.4 Human Behavior

Factors like cautious driving in bad weather or decisions to stay home during extreme conditions are not accounted for, which may influence the results.