Final Project Report: Weather and Crash Correlation in U.S. Cities

By James Aretakis and Nnamdi Nwaenia

Research Question:

How does the weather affect the average number of car crashes in certain locations around the U.S.?

Overview

For our project, we wanted to look into the relationship between weather conditions, including wind speed and rain, and car crash frequency. We took two major cities in two different geographical locations to compare whether these places have varying results. We integrated the crash reports along with the weather reports to make an accurate analysis and conclusion on how weather affects crash rates. The two cities we chose were **Austin, TX,** and **Chicago, IL, because they are both very populated and known for their downtown areas.**

Data Collection

We collected data from **two APIs**:

1. **Open-Meteo Archive API**: Provided historical daily weather data (temperature, precipitation, windspeed).

2. **Crash Dataset** (converted into SQLite): Supplied detailed car crash records including date, injuries, fatalities, and location.

Each dataset was pulled and inserted into SQLite databases with proper normalization. We gathered **well over 100 rows of data**.

These data sets were extremely hard to find. While we originally had a used car API to try and see how the used car market is affected by weather, the data was too minimal and was very hard to get full access to.

Data Analysis & Calculations

We calculated:

- Average crashes per day under different weather conditions
- Crash rates on rainy vs dry days
- Crash frequency based on windy conditions
- Daily crash trends over time and their relation to weather
- City-by-city comparisons to examine whether the weather impact varies by location

Visualizations

All graphs were created using **Matplotlib**, as required. Here are the key plots:

1. **Average Crashes Per Day – Rainy vs Dry Conditions**
   Shows significantly more crashes on rainy days compared to dry days.

2. **Crash Frequency by Weather Condition**
   Highlights that moderate and heavy rain days lead to higher crash rates.

3. **Average Crashes by City**
   Provides a baseline comparison showing that Chicago consistently has higher daily crash averages than Austin.

4. **Crashes on Rainy & Windy Days vs Dry & Calm Days**
   Combining wind and rain data to show that the highest crash rates occur when both are present.

Key Findings

- **Rain is a significant factor** in increasing the average number of crashes. This is originally what we predicted, just because we thought that rain would make it harder to see while driving and cause hydroplaning along with other issues. Although this isn't surprising, it is pretty interesting that there is a clear and direct correlation between these

two datapoints.

- **Wind contributes** to crash frequency, especially when combined with rain. This was a little surprising considering wind does not seem to have a large effect while driving, but it is no surprise that when it is combined with rain, it makes it much harder to see the road.

- **Chicago has a higher average crash rate** than Austin, even under similar weather conditions. The magnitude of the difference is very surprising considering they are both big cities. Chicago, being a more densely populated city, could be the reason for this statistic. Also, the type of roads used in Austin and Chicago very quite greatly. Chicago has narrower roads and follows a structured grid system, while Austin has more of a sprawl layout and wider roads, which may be why they have fewer accidents.

- Dry and calm days consistently show the **lowest crash averages**. Again, this shows to correlation with the opposite fact that rainy and windy days cause more crashes.

Challanges

Our biggest challenges included finding the correct APIs and combining the data so that they match. Especially when using data dependent on crashes there were a lot of struggles since often times they weren't public. We had to pivot many times from focusing on states to specific cities. We did use ChatGPT to help combine the data and make every variable match, so when it came to graphing, it was easy to go through. We changed our idea for this project many times, but in the end, we thought this would be very interesting to see how much of a correlation there actually was.

Technical Approach

- All data was stored and joined using **SQLi/te**
- Python was used for all data cleaning, merging, and visualization
- Columns were normalized to use **integers and categorical bins** instead of raw text
- The code was modular, repeatable, and structured per rubric's best practice
- Used AI to fix and refine code when we ran into major errors and problems

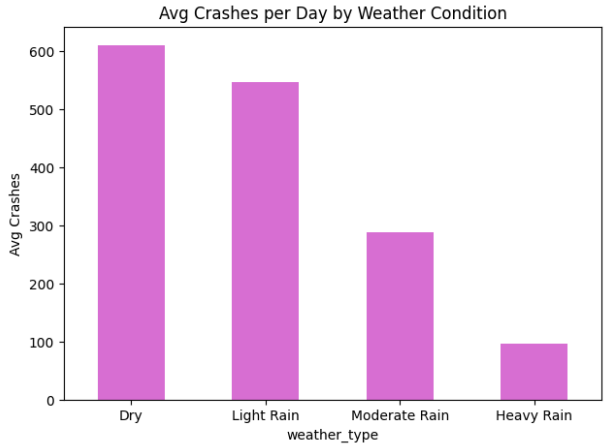
Contributions

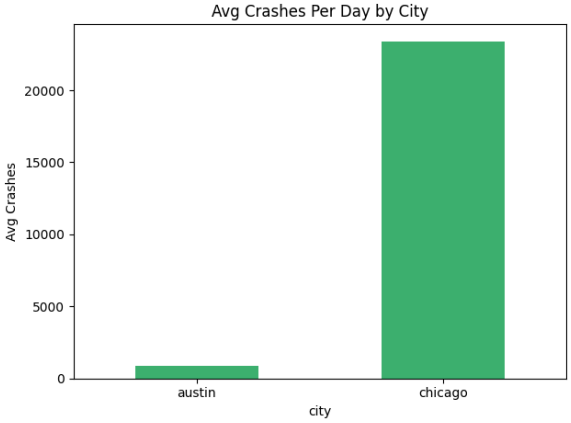
- **James**: API integration, crash database setup, weather-categorized analysis

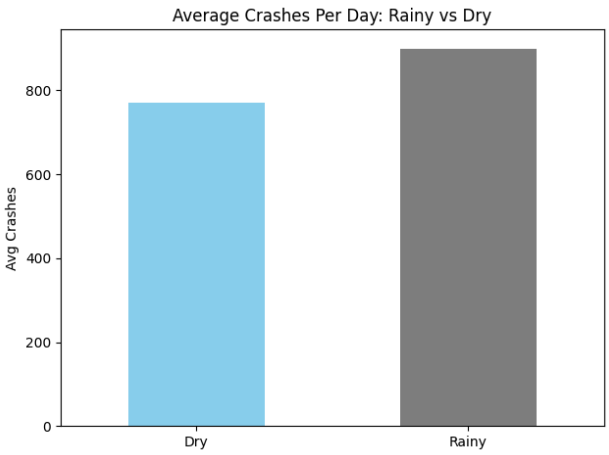
- **Nnamdi**: Data cleaning, normalization, graph development, and interpretation
- We both contributed equally to the final code, testing, and report writing.

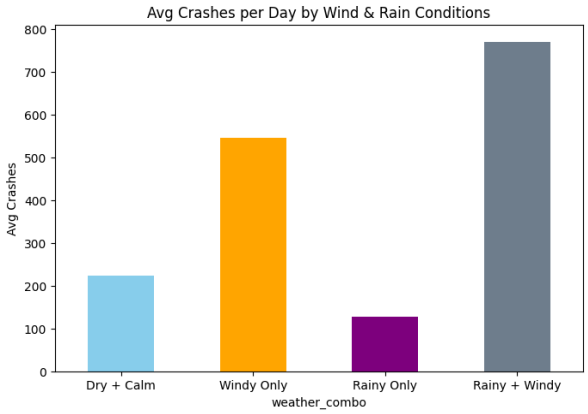
Conclusion

Weather—especially precipitation and wind—**strongly affects** the likelihood of traffic accidents. Through this project, we observed that the **most dangerous driving conditions occur when both rain and wind are present**, and these effects are even more pronounced in larger urban areas like Chicago. This analysis could help inform traffic safety planning and public policy in urban environments.






Average Crashes Per Day: Rainy vs Dry
Avg Crashes
0
200
400
600
800
Dry
Rainy


Avg Crashes per Day by Wind & Rain Conditions
Avg Crashes
0
100
200
300
400
500
600
700
800
Dry + Calm
Windy Only
Rainy Only
Rainy + Windy
weather_combo