

Traffic Crash Patterns: Assessing the Influence of Human Factors

Asheer Ali (22993810)

16.01.2025

1 Introduction

Traffic crashes are a big concern that needs attention and required a comprehensive analysis in order to identify the critical issues. This report investigates to main questions.

- How do driver characteristics correlate with injury severity in crashes?
- What types of vehicles are most involved in crashes?

Knowing these patterns may make roads safer and reduce serious accidents.

2 Data Sources

2.1 Descriptions of Data Sources

- **Traffic Crashes - People:** This dataset contains details of individuals involved in traffic incidents, including demographics, safety equipment use, and injury severity [1].

[4]:

| | PERSON_ID | PERSON_TYPE | | CRASH_RECORD_ID | VEHICLE_ID | CRASH_DATE | SEAT_NO | CITY | STATE | ZIPCODE | SEX | ... | EMS_RUN_ |
|---|-----------|-------------|---|-----------------|---------------------------|------------|---------|------|-------|---------|-----|-----|----------|
| 0 | O749947 | DRIVER | 81dc0de2ed92aa62baccab641fa377be7feb1cc47e6554... | 834816.0 | 09/28/2019 03:30:00 AM | NaN | CHICAGO | IL | 60651 | M | ... | | N |
| 1 | O871921 | DRIVER | af84fb5c8d996fcd3aefd36593c3a02e6e7509eeb27568... | 827212.0 | 04/13/2020 10:50:00 PM | NaN | CHICAGO | IL | 60620 | M | ... | | N |
| 2 | O10018 | DRIVER | 71162af7bf22799b776547132ebf134b5b438dcf3dac6b... | 9579.0 | 11/01/2015 05:00:00 AM | NaN | NaN | NaN | NaN | X | ... | | N |
| 3 | O10038 | DRIVER | c21c476e2ccc41af550b5d858d22aaac4ffc88745a1700... | 9598.0 | 11/01/2015 08:00:00 AM | NaN | NaN | NaN | NaN | X | ... | | N |
| 4 | O10039 | DRIVER | eb390a4c8e114c69488f5fb8a097fe629f5a92fd528cf4... | 9600.0 | 11/01/2015 10:15:00 AM | NaN | NaN | NaN | NaN | X | ... | | N |

5 rows × 29 columns

Figure 1: First 5 rows of traffic crashes people dataset

- **Traffic Crashes - Vehicles:** Provides records of vehicles involved in crashes, including type, direction, and damage details [2].

[6]:

| | CRASH_UNIT_ID | | CRASH_RECORD_ID | CRASH_DATE | UNIT_NO | UNIT_TYPE | NUM_PASSENGERS | VEHICLE_ID | CMRC_VEH_I | MAKE |
|---|---------------|---|---------------------------|------------|------------|-----------|----------------|------------|------------|----------|
| 0 | 1727162 | f5943b05f46b8d4148a63b7506a59113eae0cf1075aabc... | 12/21/2023 08:57:00 AM | 2 | PEDESTRIAN | NaN | NaN | NaN | NaN | NaN |
| 1 | 1717556 | 7b1763088507f77e0e552c009a6bf89a4d6330c7527706... | 12/06/2023 03:24:00 PM | 1 | DRIVER | NaN | 1634931.0 | NaN | NaN | NISSAN |
| 2 | 1717574 | 2603ff5a88f0b9b54576934c5ed4e4a64e8278e005687b... | 12/06/2023 04:00:00 PM | 2 | DRIVER | NaN | 1634978.0 | NaN | NaN | CHRYSLER |
| 3 | 1717579 | a52ef70e33d468b855b5be44e8638a564434dcf99c0edf... | 12/06/2023 04:30:00 PM | 1 | DRIVER | NaN | 1634948.0 | NaN | NaN | SUBARU |
| 4 | 1720118 | 609055f4b1a72a44d6ec40ba9036cefd7c1287a755eb6c... | 12/10/2023 12:12:00 PM | 1 | DRIVER | NaN | 1637401.0 | NaN | NaN | TOYOTA |

5 rows × 71 columns

Figure 2: First 5 rows of traffic crashes vehicle dataset

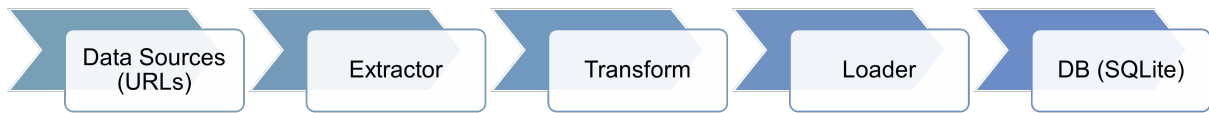


Figure 3: ETL Pipeline

2.2 Structure and Quality of Data Sources

- **People Dataset:** Contains individual-level data with fields for demographics, safety equipment use, and injury severity. Missing values exist but can be handled through removing the rows with Nan values, as the nan values are not present that much in the selected columns. [3].
- **Vehicle Dataset:** Vehicle-level data with fields for type, damage, and direction. Data quality is high, with minimal missing values.

2.3 Licenses and Permissions

Both datasets are publicly available under open-data licenses, allowing use with proper attribution [1, 2].

3 Data Pipeline

The data pipeline is implemented using Python and consists of the following steps:

- **Extractor:** Downloads CSV files from the given URLs.
- **Transformer:** Processes the data with:
 - Removing unnecessary columns.
 - Handling missing values through imputation.
 - Standardizing date formats for consistency.
- **Loader:** Stores the cleaned datasets in an SQLite database for efficient access.

4 Results and Limitations

4.1 Results

- Cleaned datasets stored in SQLite database.
- Data ready for analysis to address project questions about injury severity and crash patterns.

4.2 Limitations

- Missing data for certain fields may impact analysis accuracy.

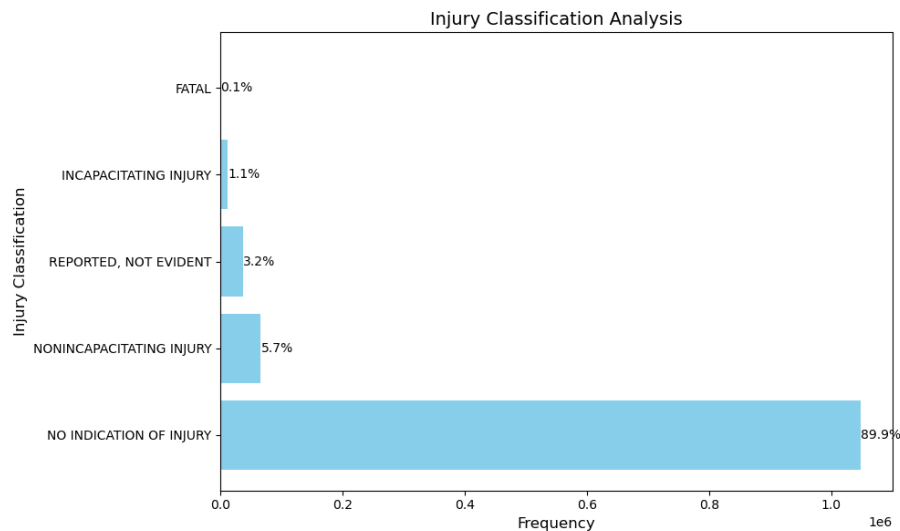


Figure 4: Injury Classification Analysis

5 Questions

5.1 How do driver characteristics correlate with injury severity in crashes?

To find out how driver characteristics affect injury severity in crashes, we use the age, sex, and injury classification columns for our analysis and plot 3 graphs in as shown in figures 4, 5a, and 5b.

5.1.1 Key Findings

Injury Types: As Shown in 4 Most crashes did not show any signs of injury (89. 9%). A small number of crashes had minor injuries (5. 7%) or nonevident injuries (3. 2%). Fatal crashes were very rare (0.1%).

This means most crashes are not very severe, but even rare severe crashes need attention.

Age and Injury Severity: When I checked the link between age and injury severity, I found:

- Severe injuries, like fatal or incapacitating ones, happened to drivers of all ages, but the average age was about 40 years.
- Younger drivers had more crashes with "No Indication of Injury," but age alone does not fully explain injury severity.

Gender and Severe Crashes: For severe crashes (fatal and incapacitating):

- Male drivers were in 61.7% of these crashes.
- Female drivers were in 38.1%.
- A small percentage (0.2%) of crashes had other gender recorded.

Male drivers seem to have more severe crashes 5b, which could be due to behavior or exposure.

Safety programs can focus on younger and male drivers to lower severe crashes. It is also important to study how other factors, like road conditions or vehicle types, might affect crashes.

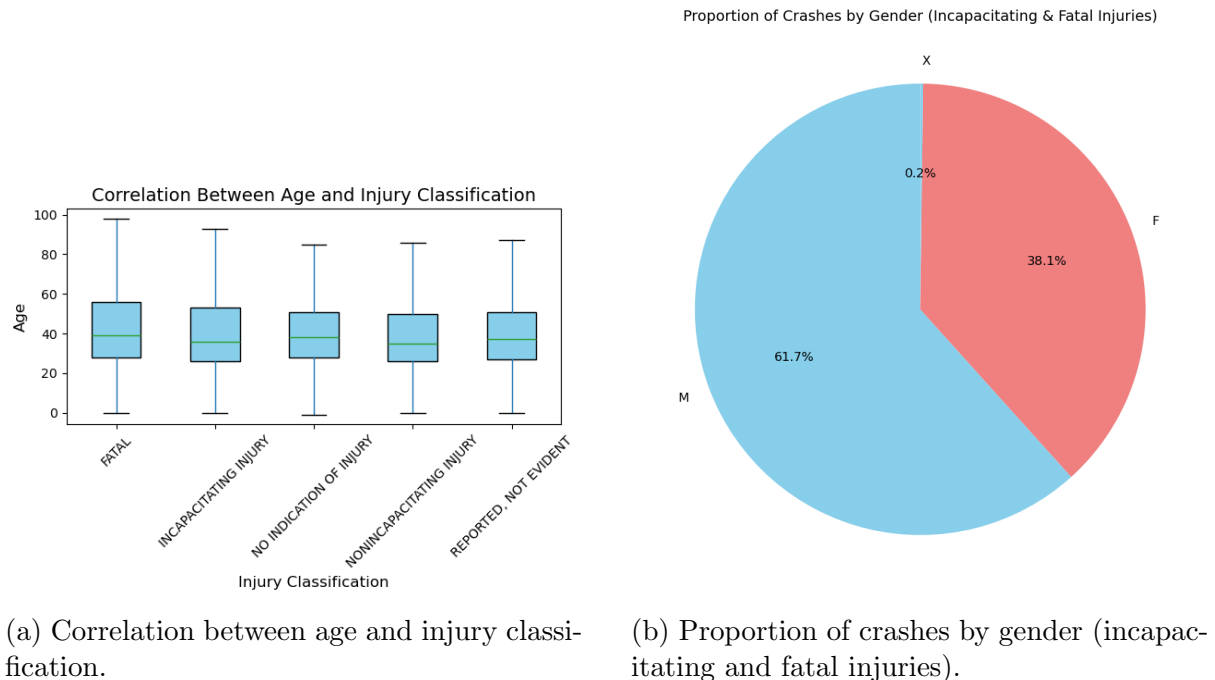


Figure 5: Comparison of Top Causes of Death and Their Trends

5.2 What types of vehicles are most involved in crashes?

5.2.1 Findings:

Top 3 Vehicles with Most Crashes in 2015:

- **NISSAN ROGUE:** 2,329 crashes.

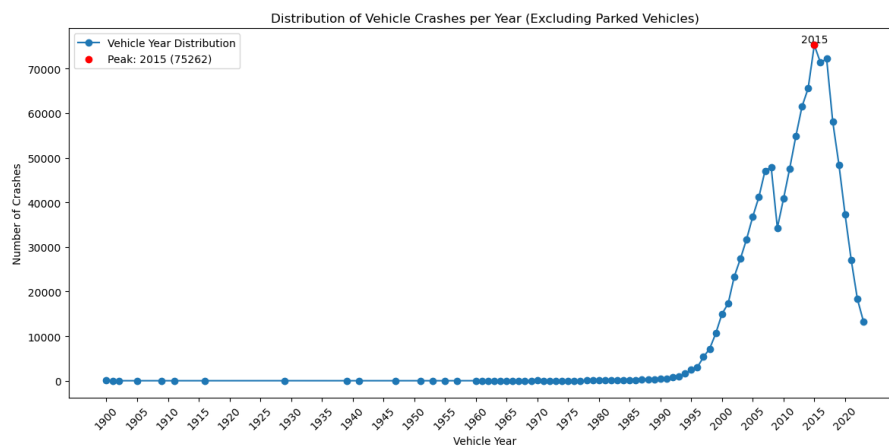


Figure 6: Distribution of Vehicle Crashes per Year (Excluding Parked Vehicles)

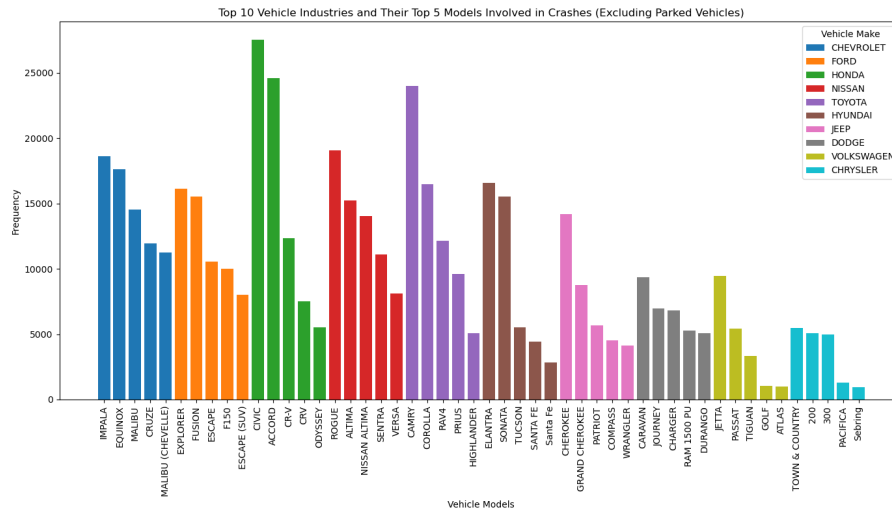


Figure 7: Top 10 Vehicle Industries and Their Top 5 Models Involved in Crashes (Excluding Parked Vehicles)

- **CHRYSLER 200:** 2,000 crashes.
- **JEEP CHEROKEE:** 1,884 crashes.

These vehicles had the most accidents in one year.

5.2.2 Vehicle with Most Crashes Overall:

- **HONDA CIVIC:** 27,536 crashes in total.

This is the car with the highest number of crashes ever.

5.2.3 Observations:

Cars from big brands like Nissan, Chrysler, and Honda are in more crashes. This may be because they are very popular. In 2015, some models had their highest crash numbers.

6 Conclusion

There are links between certain death rates and indicators of chronic diseases although the strength of these associations varies per condition. While cancer and mental health have weaker association with the death rates, conditions like diabetes and cardiovascular diseases have a large positive correlation with the death rates. Furthermore, the examination of the leading causes of deaths from 2020 to 2023 indicated that, although chronic conditions continued to rank among the primary causes of death, year-to-year variations may have been impacted by outside variables or other causes.

Although the deeper dive into the datasets highlighted the associations among the death rates and disease prevalence, this research does not support the idea of causality. It can be difficult to conclude the precise role of the chronic diseases alone because of the potential influence of outside factors on the trends illustrated. Moreover, demographics, lifestyle, and healthcare quality may all have an independent role on the death rates.

Following are the limitations:

- The analysis assumes linear relationships, which may not fully capture more complex patterns or interactions between variables.
- The dataset only considers shared years between datasets, which may exclude relevant trends over longer periods.
- Outliers and noise in the data were not deeply analyzed, potentially affecting the robustness of the results.
- The focus was on the top causes of death, which might overlook important secondary contributors.

7 Conclusion

The data reveals key insights into traffic crash patterns and their relationships with driver characteristics, vehicle types, and crash severity. Most crashes do not involve injuries, with 89.9% showing no indication of injury. However, attention must be given to severe crashes, even though they are rare. Drivers of all ages are involved in severe injuries, but the average age is around 40 years. Gender analysis highlights that male drivers are involved in a majority of severe crashes, indicating possible behavioral or exposure-related factors.

Popular vehicle models such as the Honda Civic, Nissan Rogue, and Chrysler 200 frequently appear in crash records. While the high crash numbers for these models may be linked to their popularity, further investigation into vehicle safety features and usage conditions is essential.

Limitations

- The analysis assumes linear relationships, which might not capture complex interactions.
- Missing data and limited variables like road conditions and vehicle maintenance were not deeply explored.
- Results are based on available datasets and may not represent all contributing factors to crashes.

Overall, the findings suggest targeting safety programs for male drivers and younger drivers to mitigate severe crashes and exploring additional factors such as road and vehicle conditions for a comprehensive safety strategy.

8 References

References

- [1] Traffic crashes - people dataset, 2024. Available online: catalog.data.gov.
- [2] Traffic crashes - vehicles dataset, 2024. Available online: catalog.data.gov.

-
- [3] Asheer Ali. Advanced data engineering project repository, 2024. Available at GitHub: <https://github.com/asheerali/advanced-data-engineering/>.