Analysis of Queensland Road Crashes (2018)

By TRISTAN HAYES

Agenda

- Introduction
- Data Integrity & Cleaning
- Visualisations & Analysis
- Major Insights
- Conclusion

Introduction

- Importance of Traffic Crash Analysis
  - Understanding crash data is crucial for improving road safety and identifying high-risk factors.
- Objective of the Analysis
  - Investigate relationships between various related variables including:
    - Speed Limit
    - Crash Severity
    - Crash Type
    - Time Trends
    - Geographical Distribution

Introduction

- **Key Questions Addressed**
  - How do speed limits impact crash severity?
  - Which crash types lead to the most severe outcomes?
  - What are the time-based trends in crash frequency and severity?
  - Which suburbs have the highest number of crashes, and what are their severity distributions?
  - Are there any notable geospatial patterns?
- **Methodology**
  - Used Python for data processing and visualization.
    - Pandas, matplotlib, seaborn, Plotly Express
   - Performed data quality checks to ensure accuracy

Data Integrity & Cleaning

- Importance of Data Integrity
  - Ensuring accuracy is crucial to avoid misleading visualisations and conclusions
- Data cleaning process
  - Removal of duplicates from all datasets (Crash Facts, Date Dimensions, Location Dimensions)
  - Categorical data checking to identify inconsistencies in labelling
    - Identified and corrected a minor error where an entry in the Crash Severity column was labelled "Medical Treatment." instead of "Medical Treatment"
  - Speed Limits were originally stored in unordered strings, converted to categorical data type with predefined ordering to assist visualisation interpretations

Crash Frequency Analysis

- Raw numbers - does not account for traffic volume differences
- 60km/h zones have the highest crash frequency
- 70km/h zones have the lowest crash frequency

Distribution of Crashes by Speed Limit
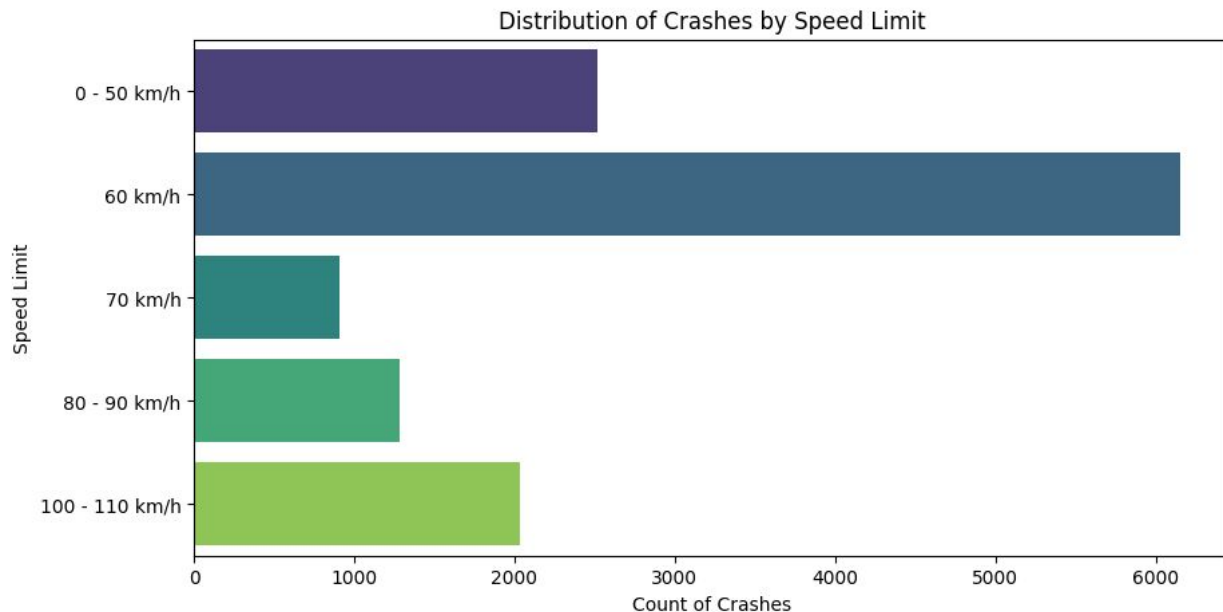
Crash Severity vs Speed Limit

- Higher speed zones (100-110km/h & 80km/h) have a higher proportion of fatal crashes (5.06%).
- Lower speed zones show fewer fatalities, as expected due to reduced impact forces.
- 60km/h zones have more crashes but fewer severe injuries or fatalities.
- Fatality proportions in 0-50km/h zones are similar to 60km/h, possibly due to pedestrian-vehicle collisions.

Crash Severity by Speed Limit

Heatmap of Crash Severity by Speed Limit
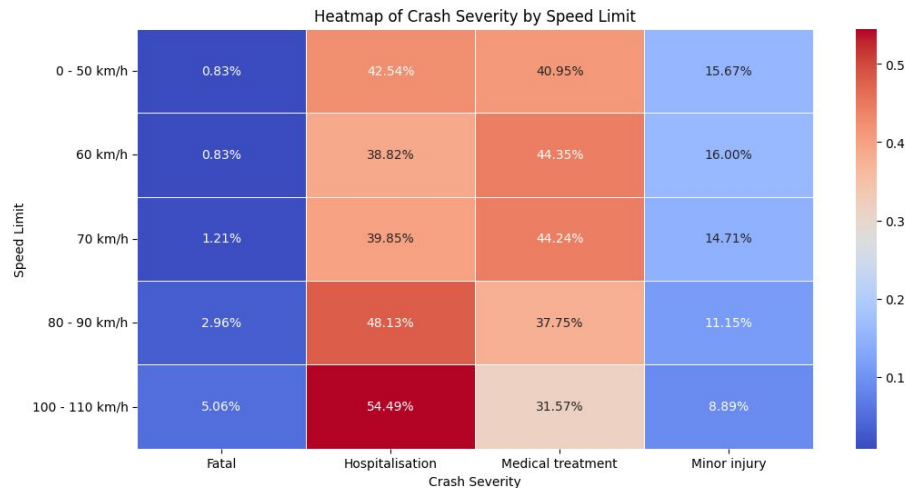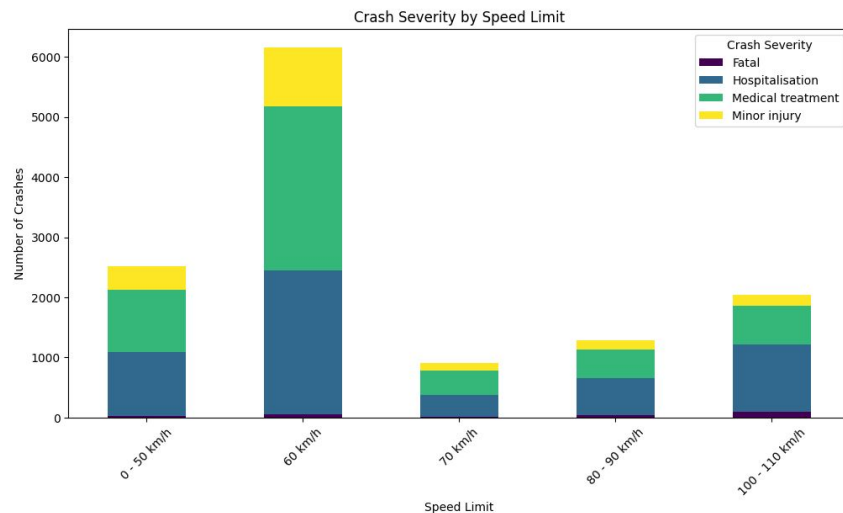
Crash Severity vs Crash Type

- Pedestrian crashes are highly fatal → No protective safety features.
- Single-vehicle crashes have a high fatality rate → Collisions with barriers, trees, etc
- Multi-vehicle crashes have the lowest fatality rates → Modern safety features & nearby assistance

Crash Severity by Crash Type

Heatmap of Crash Severity by Crash Type
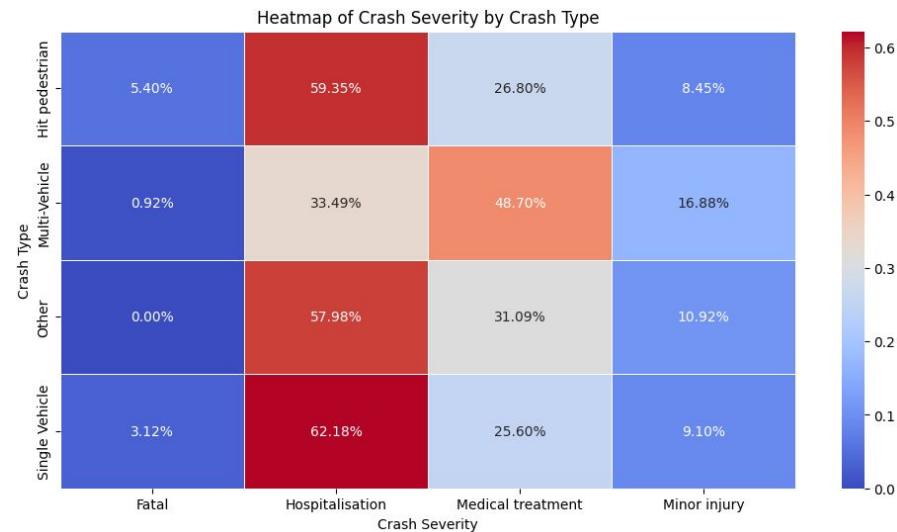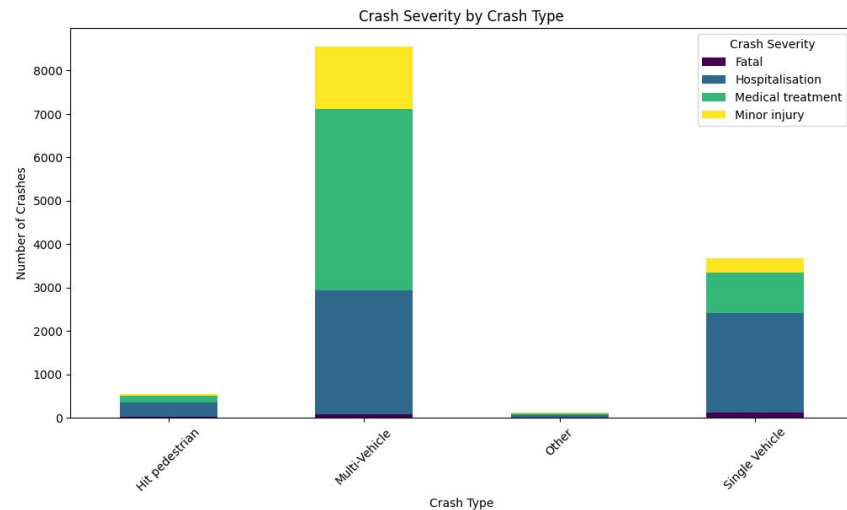
Monthly Crash Trends

- April shows significant drop in crashes → difficult to justify, possible data loss
- Ignoring the outlier month there is no consistent trend beyond expected random variations


Monthly Crash Severity in 2018
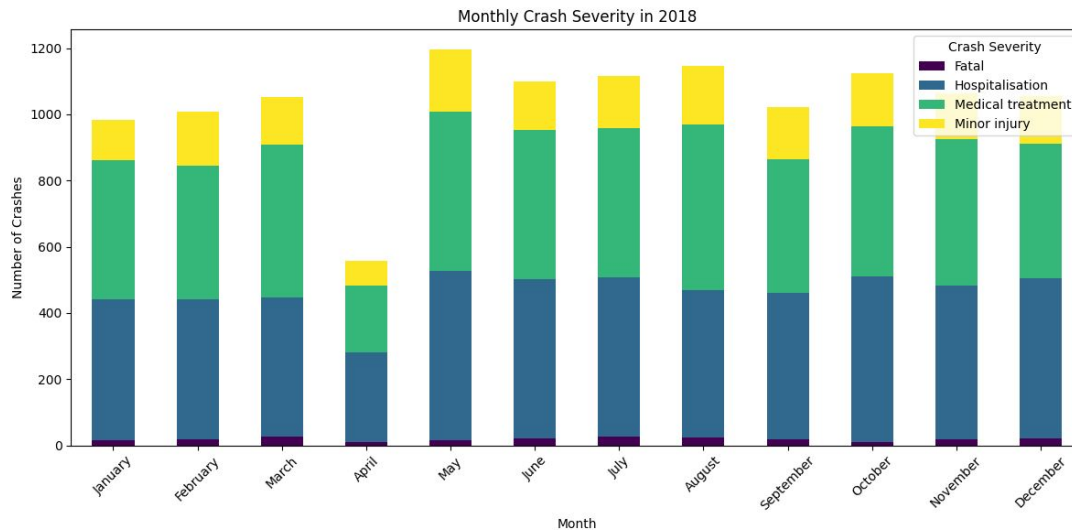
Top 10 Suburbs with Highest Crash Frequency

- Contributing factors → High traffic flow, urbanisation, road design
- High fatality proportion in Woolloongabba

Top 10 Suburbs with the Most Crashes by Severity
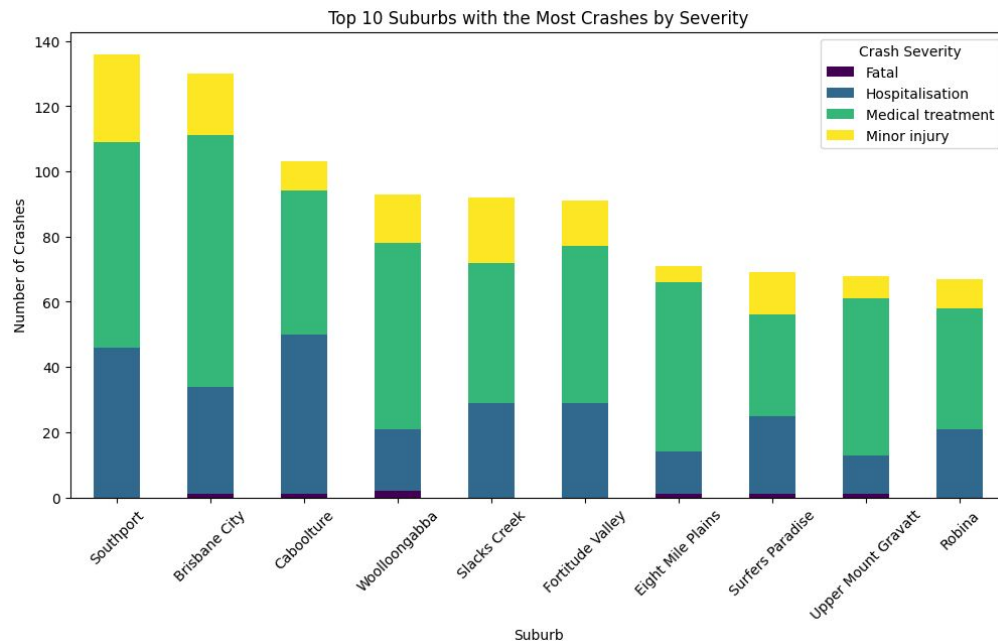
Geographical Heat Map

- Crash density increases with urbanisation & traffic flow
- Brisbane City (2nd highest crash frequency suburb) shows concentrated crash zones → road design, lane width, traffic congestion

Geographic Heatmap of Crashes
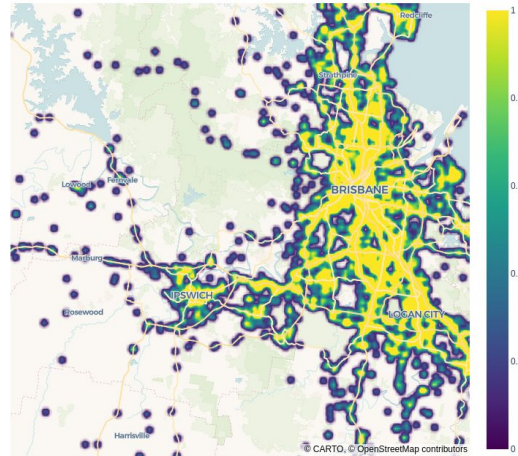
Geographic Heatmap of Crashes

Major Insights

- Crash Frequency & Speed Zones
  - 60km/h zones have highest crash frequency → likely due to high traffic flow
  - 70km/h zones have lowest crash frequency
- Crash Severity & Speed Limits
  - Higher speed zones (100 - 110km/h) have highest proportion of fatal crashes
  - 0-50km/h has similar fatal crash proportions → potentially pedestrian incidents
- Crash Type & Severity
  - Single-vehicle and pedestrian crashes are the most fatal
  - Multi-vehicle crashes have lower fatality rates → modern safety features, assisting drivers
- Monthly Crash Trends
  - No consistent trend apart from expected random deviations
  - April is an outlier → potential data loss

Conclusion

- Recommendations
  - Improve road infrastructure & traffic management in high-crash areas
  - Improve pedestrian safety measures in the required zones
  - Review speed limits & enforcements in high-crash areas
- Future Considerations
  - Driver behaviour, weather conditions, road design
  - Multi-year trends for deeper insights and analysis

Q&A

Thank You!