Crash Data Analysis Report – New Zealand

Introduction

This report presents an exploratory data analysis (EDA) of New Zealand crash data with the goal of identifying patterns, risk factors, and trends in traffic accidents. The analysis aims to support data-driven road safety interventions and develop predictive insights for crash severity and high-risk zones.



Objectives

Predictive Modeling

* **Crash Severity Prediction** – Can we predict the severity of a crash based on environmental and road conditions?
* **Accident Hotspot Identification** – Which locations are more likely to experience accidents based on historical data?

Risk Factor Analysis

* **Weather Impact on Crashes** – How do weather conditions affect crash severity?
* **Speed Limit and Accidents** – Is there a correlation between speed limits and accident severity?
* **Time-Based Analysis** – Are there peak periods when crashes are more likely?

Anomaly Detection

* **Unusual Crash Patterns** – Are there abnormal crash cases that deviate from typical trends?
* **Vehicle Type Risk Analysis** – Are certain vehicle types more prone to severe accidents?

Geospatial Analysis

* **Urban vs. Rural Crashes** – Do crash patterns differ between urban and rural areas?
* **Intersection Accidents** – Are intersections more prone to accidents?

Traffic Control & Infrastructure

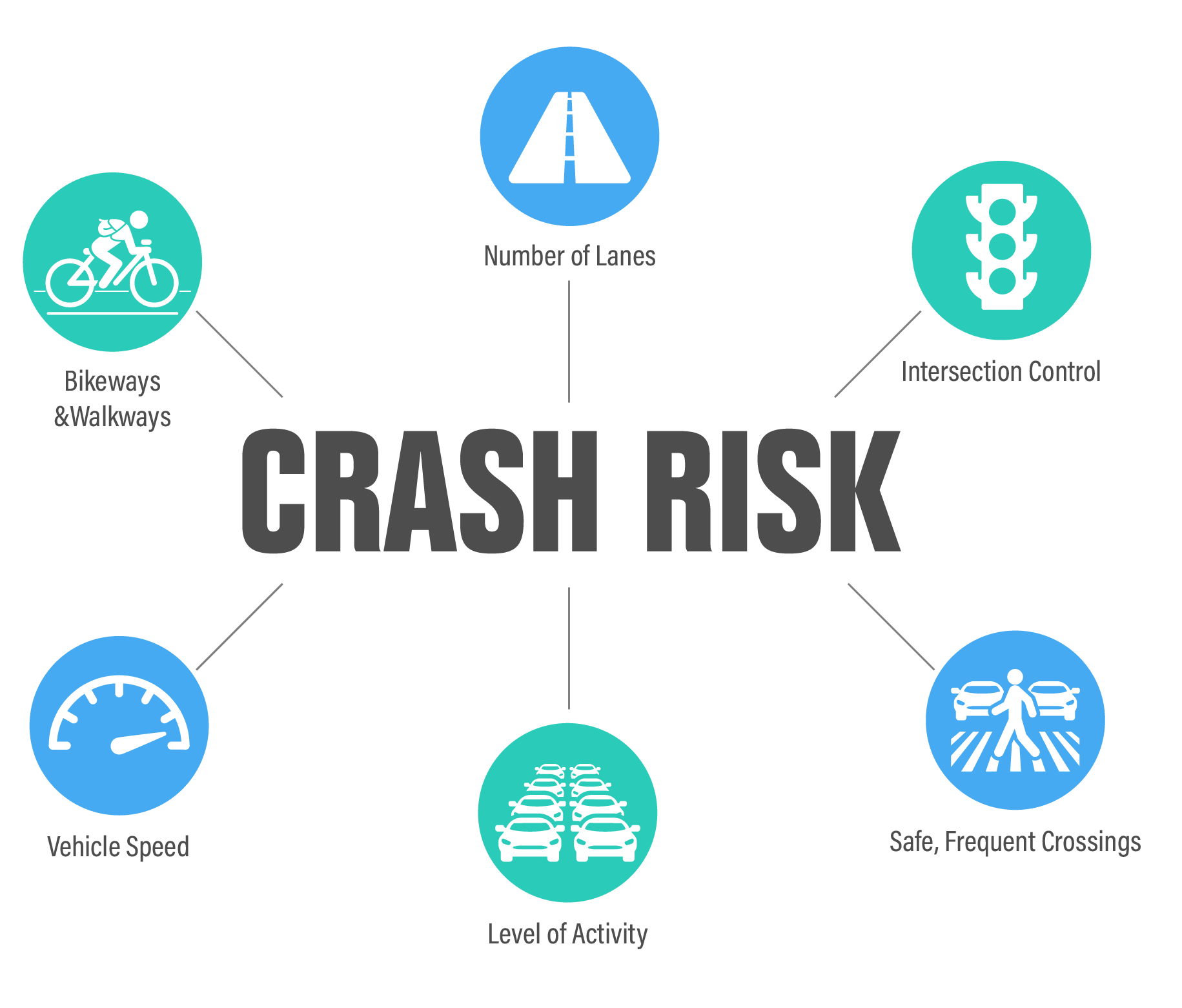
* **Traffic Control Effectiveness** – Do signs/lights reduce crash severity?
* **Road Surface & Conditions** – How do road types and elevation affect severity?

Dataset Overview

* **Source**: New Zealand Crash Analysis System (CAS)
* **Size**: Approx. **400K**+ rows (accidents)

Key features from the dataset

|  |  |  |
| --- | --- | --- |
| crashSeverity | speedLimit | weatherA |
| roadCharacter | vehicleType, | trafficControl |
| crashHour | crashMonth | urbanOrRural |
| intersectionType |  |  |



Approach and Methodology

Data Cleaning Summary

* + Filled missing values where appropriate.
  + Removed or flagged rows with incomplete geospatial data (for hotspot analysis).
  + Standardized column names and data types.
  + Verified class balance in target variable (crashSeverity).

Exploratory Data Analysis (EDA)

Crash Severity Distribution

* **Count plot** shows that the majority of crashes fall into the **"Minor"** and **"Non-Injury"** categories.
* **Insight**: Data is **imbalanced**, which will need to be addressed in any modeling tasks.

Accident Hotspot Identification

* **Top crash locations** visualized using frequency counts.
* **Latitude/Longitude columns** ready for geospatial mapping.
* **Insight**: Certain intersections and urban zones exhibit higher crash frequency.

Weather Impact on Crashes

* Bar plots show **"Rain"** and **"Bright Sun"** associated with more crashes, but **fog** and **ice** correlate with higher severity.
* **Insight**: Weather conditions are a significant factor in crash outcomes.

Speed Limit and Accidents

* **Box plots** show higher crash severity at increased speed limits.
* **Insight**: There’s a clear relationship between **speed** and **fatal/severe** crashes.

Time-Based Trends

* Peak crash times observed during **rush hours** and early evening.
* **Winter months** show a slight increase in severe crashes.
* **Insight**: Time of day and seasonality are important predictors.

Unusual Crash Patterns

* Some cells flagged **rare values** or **outlier cases** based on combinations of speed, weather, and crash severity.
* **Insight**: Potential to use anomaly detection (e.g., Isolation Forest) in the future.

Vehicle Type Risk Analysis

* **Motorcycles** and **trucks** are disproportionately involved in severe crashes.
* **Insight**: Certain vehicle types have significantly higher fatal crash rates.

Urban vs. Rural Analysis

* Rural roads show fewer crashes but **higher severity** outcomes.
* **Insight**: Emergency response time and road conditions may play a role.

Intersection Accidents

* Intersections have a higher count of **minor and non-injury crashes**, but fatal crashes occur more on **open roads**.
* **Insight**: Type of location significantly influences severity.

Traffic Control Effectiveness

* Crashes at **stop signs and traffic lights** show lower fatality rates than those with **no control**.
* **Insight**: Presence of control measures reduces crash severity.

Road Surface & Conditions

* **Unsealed** and **hilly roads** have higher rates of severe crashes.
* **Insight**: Road quality directly impacts crash outcomes.

Key Insights & Recommendations

* Weather, speed limit, and time are **strong predictors** of crash severity.
* Road design and surface conditions significantly affect outcomes.
* Urban areas have more frequent but less severe crashes.
* **Traffic control measures** and **intersection designs** are protective.
* High-risk vehicle types (e.g., motorcycles) may need focused awareness or policy interventions.

Limitations

* Data imbalance in severity classes may bias predictive models.
* Missing geospatial or environmental features in some rows.
* No driver behavior or demographic data included (e.g., age, intoxication).

Conclusion

This analysis reveals several clear patterns and risk factors in New Zealand’s crash data. It provides a strong foundation for developing predictive models and targeting traffic safety interventions. Further work can focus on anomaly detection, machine learning-based crash severity prediction, and GIS-based hotspot mapping.

Contact

For further inquiries or collaboration, please reach out via:

Email: [sadesanya@gmail.com](mailto:sadesanya@gmail.com)

LinkedIn: [www.linkedin.com/in/olu-adesanya](http://www.linkedin.com/in/olu-adesanya)