# INTRODUCTION TO DATA MANAGEMENT

### PROJECT REPORT

(Project Semester January-April 2025)

## **CRASH REPORTING**

Submitted by

# Ritik Raushan

Registration No:- 12300915

Section:- K23GW

Course Code:- INT127

Under the Guidance of

Baljinder Kaur 27952

Discipline of CSE/IT

**Lovely School of Computer Science Engineering** 

Lovely Professional University, Phagwara

**CERTIFICATE** 

This is to certify that Ritik Raushan bearing Registration no. 12300915 has completed

INT127 project titled, "Crash Reporting" under my guidance and supervision. To the best of

my knowledge, the present work is the result of his/her original development, effort and study.

**Baljinder Kaur** 

**Professor** 

**School of Computer Science Engineering** 

Lovely Professional University

Phagwara, Punjab.

Date: 13 April, 2025

**DECLARATION** 

I, Ritik Raushan, student of B.tech under CSE/IT Discipline at, Lovely Professional

University, Punjab, hereby declare that all the information furnished in this project report is

based on my own intensive work and is genuine.

Date: 13 April, 2025

Signature

Registration No. 12300915

Ritik Raushan

### 1. Introduction

This report presents an analysis of car crash incidents based on the data visualized in the provided dashboard. The dashboard offers insights into various aspects of these incidents, including the count of crashes on different routes, the types of collisions, the average speed under different surface conditions, and the nature of injuries sustained. The analysis aims to identify key trends and patterns that can contribute to a better understanding of the factors influencing car crashes.

### 2. Source of Dataset

The dataset used for this analysis, "Crash Reporting - Drivers Data," was sourced from Montgomery County's crash reporting system. This publicly available dataset can be accessed through the following link:

Link: <a href="https://catalog.data.gov/dataset/crash-reporting-drivers-data">https://catalog.data.gov/dataset/crash-reporting-drivers-data</a>

The data encompasses a period from 2015 to 2023. It comprises over 40 features, offering a comprehensive view of crash incidents. Key attributes include:

- Collision Type
- Weather Conditions
- Crash Date
- Speed Limits
- Extent of Damage

The primary purpose of this dataset is to facilitate the study of crash patterns and to determine the key factors that influence road safety within Montgomery County. This analysis leverages the aggregated and visualized information from this dataset as presented in the dashboard.

# 3. Dataset Preprocessing

The dataset was preprocessed to ensure it was suitable for analysis and visualization. The cleaning process involved identifying and addressing missing values in key fields, such as route type and collision type. Incomplete or ambiguous records were either corrected, imputed, or removed based on their relevance to the analysis. Duplicate entries were also detected and eliminated to prevent inflated counts. Data transformation steps included standardizing and categorizing route types for uniform representation and reclassifying collision types into predefined categories for consistency. To generate meaningful insights, crash data was aggregated by route type and collision type, which formed the basis for the "Count of Crashes on Different Routes" and "Count of Collision Types" metrics on the dashboard. Finally, the cleaned and

transformed dataset was cross-verified against raw data samples to ensure accuracy and consistency, and statistical summaries were generated to confirm that preprocessing steps did not introduce any significant biases or errors. These efforts ensured the dataset was reliable and ready for effective visualization and analysis.

# 4. Analysis on Dataset

Based on the visualizations presented in the dashboard, the following analysis can be made: General Description:

The dashboard provides a multi-faceted view of car crash incidents, focusing on crash frequency across different routes, the types of collisions that occur, the severity of injuries resulting from these crashes, and the relationship between average speed and surface conditions. Filters for surface condition, collision type, and crash date allow for more granular exploration of the data.

Specific Requirements & Analysis Results:

#### Count of Crashes on Different Routes:

- o Visualization: A bar chart titled "Count of Crashes on different Route" is displayed.
- Analysis: The chart shows a significantly higher number of crashes occurring on "Other" routes compared to other categories like "Bicycle Route", "Busway", "Chartered Bus", "Expressway", "Interstate", "Local Road", "Major Arterial", "Minor Arterial", "Other Freeway or Expressway", "Other State Highway", "US Highway", and "Unknown". The "Other" category appears to have approximately 100,000 crashes, while most other categories have significantly fewer, often below 20,000. This suggests that the "Other" route category warrants further investigation to understand the specific characteristics contributing to the higher crash frequency.

### • Count of Collision Type:

- Visualization: A bar chart titled "Count of Collision Type" is presented.
- Analysis: The most frequent collision type appears to be "ANGLE" with a count of approximately 60,000. Other notable collision types include "REAR END" (around 40,000) and "SIDESWIPE, SAME DIRECTION" (around 20,000). Less frequent collision types include "SAME DIR REAR LEFT TURN", "SAME DIR REAR RIGHT TURN", "OPPOSITE DIRECTION SIDESWIPE", "HEAD ON", "ANGLE MEETS RIGHT TURN", and "UNKNOWN", all with counts below 10,000. This highlights "ANGLE" collisions as a significant area of concern.

### • Types of Injury:

Visualization: A pie chart titled "Types of Injury" is displayed.

Analysis: The pie chart shows the distribution of injury types. The largest proportion is "POSSIBLE INJURY" (approximately 43%), followed by "SUSPECTED MINOR INJURY" (around 34%) and "SUSPECTED SERIOUS INJURY" (around 19%). "FATAL INJURY" represents a smaller proportion (around 4%). This indicates that while a significant number of crashes result in possible or minor injuries, serious and fatal injuries are also a concerning aspect.

### Average Speed on Surface Condition:

- o Visualization: A line chart titled "Average Speed on Surface Condition" is shown.
- Analysis: The chart plots average speed against different surface conditions ("DRY", "ICE", "ICE/FROST", "MUD, DIRT, GRAVEL", "OTHER", "SAND", "SLUSH", "SNOW", "UNKNOWN", "WATER (STANDING, MOVING)", and "WATER/ICE"). The average speed appears to be relatively consistent across most surface conditions, generally ranging between 30 and 40. However, there might be slight variations depending on the specific conditions. It's important to note that this chart displays average speed and might not capture the impact of speed variations within each surface condition category.

#### Filters:

- Surface Condition: A filter allows users to select specific surface conditions (e.g., Dry, Ice,
   Snow) to analyze the data for those conditions only.
- Collision Type: A filter enables the selection of specific collision types (e.g., Angle, Rear End) to focus the analysis on those types.
- Crash Date: A date filter allows users to analyze crash data within a specific date range (from 01/13/2015 to 01/13/2023 as visible in the dropdown).

#### **Visualization:**

The dashboard effectively uses various chart types (bar charts, pie chart, and line chart) and filters to present the car crash data in an understandable format. The use of color and clear labels enhances the readability of the visualizations. The interactive filters allow for a more dynamic exploration of the relationships between different variables.

## 5. Conclusion

The analysis of the car crash dashboard reveals several key insights:

• A disproportionately high number of crashes occur on routes categorized as "Other," suggesting a need for further investigation into the characteristics of these routes.

- "Angle" collisions are the most frequent type of crash, indicating a potential focus area for safety interventions.
- While most injuries are classified as possible or minor, a significant percentage still involve serious or fatal outcomes.
- The average speed appears relatively consistent across different surface conditions, but further analysis of speed distributions might be necessary.

The interactive filters provide valuable tools for further exploration of the data and for identifying specific patterns related to surface conditions, collision types, and timeframes.

# 6. Future Scope

To enhance this analysis and gain deeper insights, the following could be considered:

- Detailed Data Source Information: Obtaining more information about the data source, collection methodology, and specific definitions of categories (e.g., "Other" routes) is crucial.
- Geospatial Analysis: If location data is available, incorporating maps to visualize crash hotspots could be highly beneficial.
- Temporal Analysis: Analyzing trends over time, including seasonal variations and the impact of any safety interventions, would provide valuable insights.
- Driver and Vehicle Characteristics: Incorporating data on driver demographics, vehicle types, and contributing factors (e.g., driver impairment, weather conditions) could lead to more targeted safety measures.
- Statistical Analysis: Applying statistical methods to identify significant correlations and causal relationships between different variables.
- Predictive Modeling: Developing models to predict the likelihood of crashes based on various factors.

## 7. References

### Dataset link:

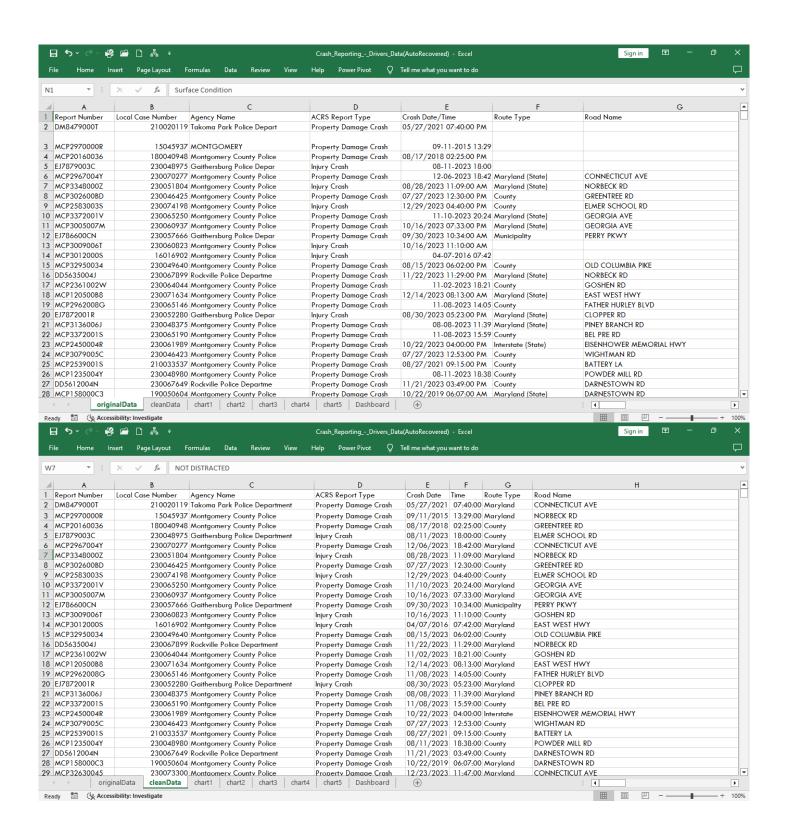
https://catalog.data.gov/dataset/crash-reporting-drivers-data

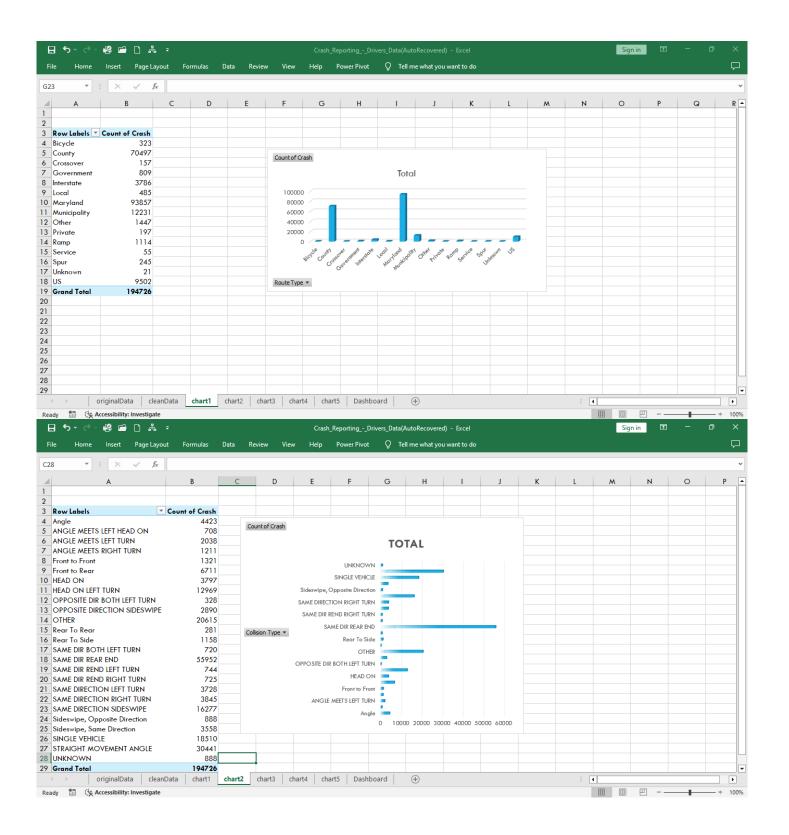
#### LinkedIn Link:-

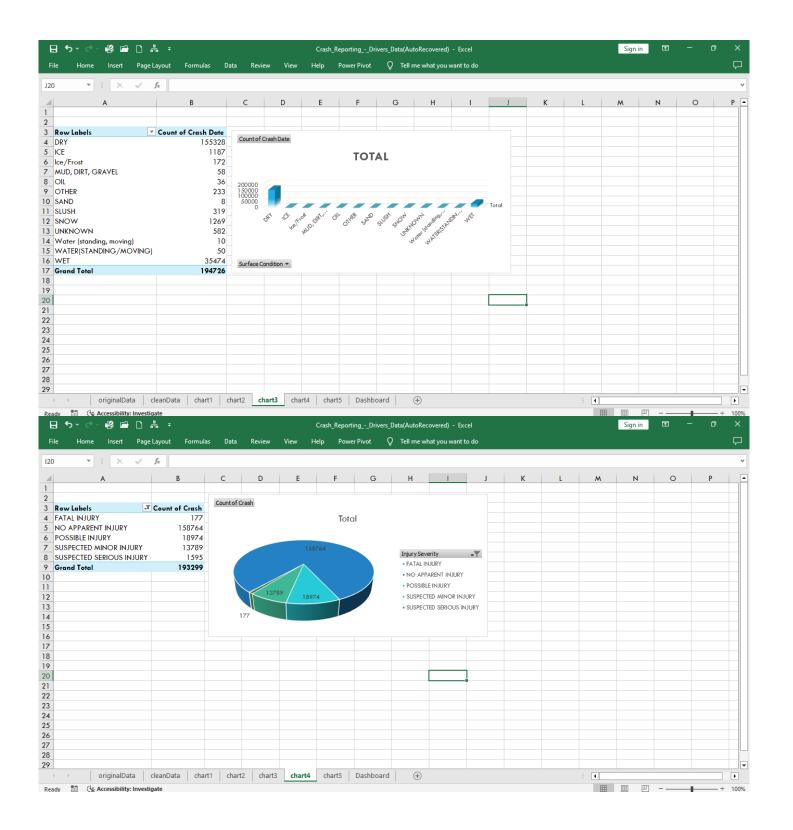
https://www.linkedin.com/posts/activity-7317615091070242817-

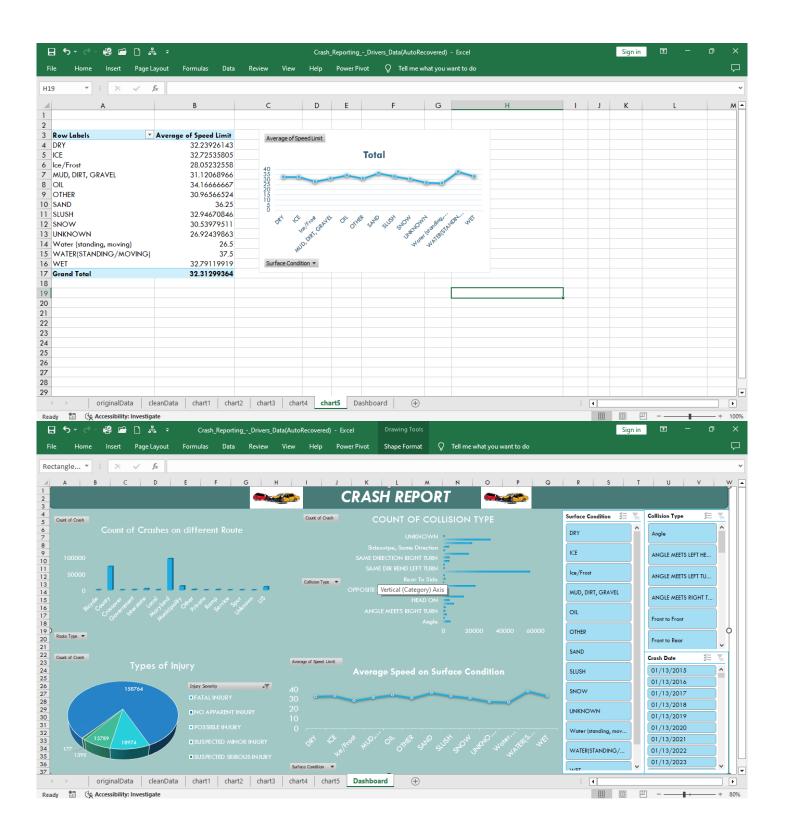
bz16?utm\_source=share&utm\_medium=member\_desktop&rcm=ACoAAE3OhDwB-

V7AgxJNsGy0EL80ysUPWrDIxoU









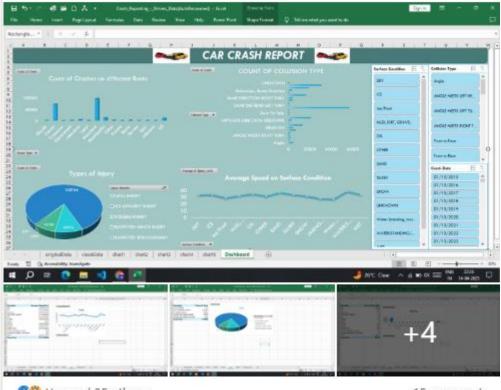




Data Management Project: Crash Reporting Analysis

Analyzed crash data (Montgomery County) focusing on route types, collision types
(Angle most frequent), injury severity, and speed vs. surface conditions. Key
takeaway: "Other" routes had a high crash count. Thanks to @BALJINDER KAUR for
guidance.

#DataManagement #CrashAnalysis #LPU #CSE #Project



CO You and 35 others

15 comments