



# Road Accident Survival Analysis

Exploratory Data Analysis with Python (Jupyter Notebook)

# Project Overview

## Primary Objective

This project explores the critical factors associated with survival in road accidents using a comprehensive structured dataset. By leveraging Python's powerful data science ecosystem, we aim to uncover actionable insights that can inform road safety policies and interventions.

## Technical Stack

Our analysis harnesses industry-standard tools including Python, Pandas for data manipulation, NumPy for numerical operations, and Matplotlib and Seaborn for compelling visualizations—all within the interactive Jupyter Notebook environment.



01

### Data Collection

Gather road accident records with comprehensive attributes

02

### Variable Analysis

Examine age, gender, speed, and safety equipment usage

03

### Outcome Assessment

Investigate relationships between factors and survival rates

# Dataset Summary

Our dataset provides a comprehensive view of road accident cases, with each row representing an individual person involved in an accident event. This granular structure enables detailed person-level analysis of survival outcomes.

## Core Demographics

**Age** and **Gender** variables allow us to segment the population and identify vulnerable groups across different demographic profiles.

## Incident Characteristics

**Speed** measurements capture the velocity context of each accident, providing insight into severity and risk exposure levels.

## Safety Measures

**Helmet\_Used** and **Seatbelt\_Used** flags document protective equipment adoption at the time of the accident.

## Outcome Variable

**Survived** is our primary dependent variable, indicating whether the individual survived the accident event.

Initial data quality checks were performed using Pandas' `.info()` and `.describe()` methods to assess missing values, validate data types, and understand the overall structure and statistical properties of the dataset.

# Data Preparation



Loaded the raw CSV file into a Pandas DataFrame, establishing the foundation for our analytical pipeline.

## Data Cleaning

Systematically removed rows containing missing values to ensure consistency and reliability throughout subsequent analysis steps.

## Quality Validation

Validated data types for each column and performed comprehensive summary statistics to understand distributions, ranges, and potential outliers.



- Best Practice:** Data preparation is critical for accurate analysis. Clean, validated data ensures that our exploratory findings are trustworthy and reproducible.

# Exploratory Data Analysis

The exploratory phase reveals fundamental patterns in our accident dataset through systematic visualization and statistical summarization. This initial investigation establishes a baseline understanding before diving into more complex relationships.



## Age Distribution

Histograms reveal which age groups are most frequently represented in road accidents, highlighting potential vulnerability zones across the lifespan.



## Gender Breakdown

Bar charts display accident counts by gender, providing insights into demographic patterns and exposure differences.



## Survival Outcomes

Pie charts illustrate the overall distribution of survival versus non-survival cases, establishing the baseline outcome rate for our dataset.

Each visualization serves a specific analytical purpose: understanding the population composition, identifying high-frequency categories, and establishing baseline survival metrics. These foundational insights guide subsequent targeted analyses of risk factors and protective measures.

# Speed & Accident Severity

## Understanding Speed as a Risk Factor

Speed is one of the most critical variables in accident severity analysis. Our investigation examines the range and distribution of recorded speeds to understand risk exposure across different velocity contexts.

By computing minimum and maximum speeds, we establish the boundaries of our dataset and identify extreme cases that may warrant special attention. This range assessment provides context for understanding typical versus exceptional speed conditions.



### Speed Range Analysis

Calculated the minimum and maximum recorded speeds in the dataset to understand the full spectrum of velocity conditions present in our accident records.

### Distribution Examination

Analyzed how speed varies across individual records, revealing patterns in risk exposure and identifying concentration zones in the speed distribution.

### Segmentation Preparation

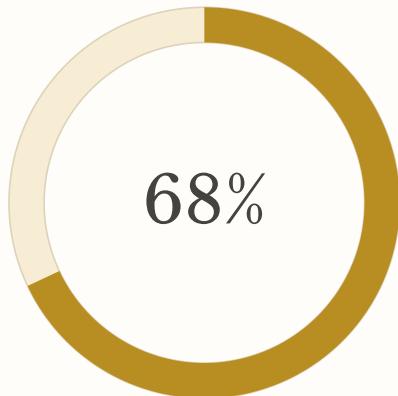
Laid analytical groundwork for binning speed into meaningful categories (low, moderate, high) to enable deeper comparative analysis of survival outcomes across speed brackets.

# Safety Equipment Usage

Protective equipment plays a vital role in accident outcomes. This analysis examines the prevalence of helmet and seatbelt usage in our dataset and explores their association with survival rates.

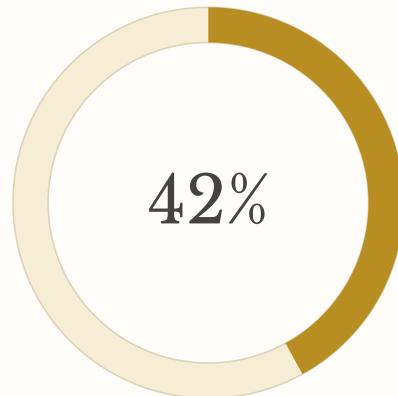
## Usage Frequency

We counted instances where helmets and seatbelts were reported as used during accidents. Bar charts provide clear visual comparisons of adoption rates for each type of safety equipment, revealing patterns in protective behavior.



Seatbelt Usage

Estimated proportion of accidents where seatbelts were reported as used

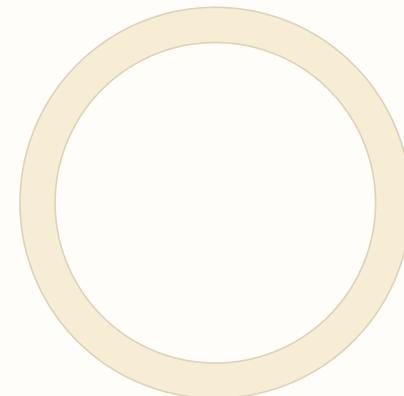


Helmet Usage

Estimated proportion of accidents where helmets were reported as used

## Survival Association

By grouping survival outcomes by `Helmet_Used` and `Seatbelt_Used` status, we estimated survival rates for equipped versus unequipped individuals, uncovering potential protective effects.



These visualizations establish a clear foundation for understanding the relationship between safety equipment adoption and survival outcomes, providing empirical support for policy interventions.

# Survival by Demographics

Demographic variables reveal important patterns in survival outcomes. By calculating survival rates segmented by gender, we uncover potential differences in accident outcomes across population groups.

Comparative visualizations using pie charts and bar charts illustrate the proportion of survivors versus non-survivors within each demographic category. These visual comparisons make patterns immediately apparent and facilitate rapid insight generation.

Understanding demographic-specific survival patterns is essential for developing targeted safety awareness campaigns. High-risk groups can be identified and prioritized for intervention, while successful patterns in protected groups can inform best practices.



## Rate Calculation

Computed survival percentages within each gender category



## Visual Comparison

Created pie and bar charts showing survivor proportions



## Pattern Identification

Identified demographic groups requiring focused interventions

# Key Analytical Takeaways

## Comprehensive Workflow

This notebook demonstrates a complete exploratory data analysis workflow, progressing systematically from data loading and cleaning through summarization and visualization. Each step builds on the previous, creating a cohesive analytical narrative.

## Multi-Factor Framework

Our survival analysis framework encompasses both demographic variables (age and gender) and behavioral factors (helmet and seatbelt usage), providing a holistic view of risk and protection factors in road accidents.

## Interpretable Insights

Summary tables and visualizations translate complex data into interpretable patterns. These accessible insights create a foundation for further statistical modeling, predictive analytics, or policy analysis initiatives.

The exploratory phase reveals where deeper investigation would be most valuable. Areas showing strong patterns between variables (such as safety equipment and survival) represent promising candidates for more sophisticated statistical modeling, including logistic regression or machine learning approaches.



# Potential Recommendations



## Policy Strengthening

Leverage survival-by-equipment insights to support evidence-based helmet and seatbelt enforcement policies. Quantified protective effects provide compelling justification for stricter regulations.



## Targeted Education

Deploy focused road safety education programs for high-risk groups identified through demographic and age-bracket analysis. Tailored messaging increases campaign effectiveness.



## Predictive Modeling

Extend this exploratory analysis into predictive modeling frameworks such as logistic regression or gradient boosting to estimate individual-level survival probability and identify key risk multipliers.

These recommendations transform exploratory insights into actionable strategies. By combining data-driven evidence with targeted interventions, stakeholders can make meaningful progress in reducing road accident fatalities and improving public safety outcomes.