National Highway Traffic Safety Administration (NHTSA) Collision Dataset Analysis

by

Dheeraj Kallakuri

Submitted to:

Dr. Jeffrey Wishart

ARIZONA STATE UNIVERSITY

November 2023

ABSTRACT

The goal of this project is to look at data from submitted crash reports of Automated Driving Systems (ADS) and Advanced Driver Assistance Systems Level 2 (ADAS L2) vehicles from across the country and summarize road conditions, injury severity, collision type, crash modeling features, and injury outcomes. It also offers fascinating data about the crash reports supplied. The study will also conduct a comparison analysis to determine whether the accident features are the same or dissimilar based on crash reports produced by human-driven vehicles in the Arizona region (Arizona crash reports) and ADS vehicles.

The information on ADS and ADAS Level 2 crash reports is obtained from the NHTSA website. For both types of cars, there are approximately 700-1000 crash data points. The NHTSA has established a certain structure for reporting the crash report. This project is entirely data-driven. The data set provided for the study has a large number of variables, so the relationship between these variables must first be understood. In terms of the use case, we must first consider the required fields of a dataset before performing statistical analysis to summarize the data. The NHTSA has also produced data visualizations of crashes based on month, automobile model, state, and so on. The initiative also intends to investigate whether the collision factors in these vehicles differ from those in human-driven cars. The goal of this effort is to remove dangerous vehicles from public highways.

The report does not provide in-depth crash reconstruction at the individual crash level, instead focusing on broad data trends.

TABLE OF CONTENTS

Page

[LIST OF FIGURES](#bookmark=id.uq1yqh1d81jv) iv

[**1 INTRODUCTION 1**](#_heading=h.cj0gqcln5yz2)

[1.1 General Order 1](#_heading=h.9dml5whs3px5)

[1.2 Difference between ADS and ADAS Level 2 2](#_heading=h.h2kzyln677sp)

[1.3 Data Summary 3](#_heading=h.hwyaa1vonb8p)

[**2 ANALYSIS 6**](#_heading=h.kx74ieluuxo)

[2.1 Data Cleaning and Preprocessing: 6](#_heading=h.jifsxg26jr9p)

[2.2 Exploratory Data Analysis (EDA): 6](#_heading=h.mq94t2qx4d35)

[2.3 Predictive Modeling: 6](#_heading=h.m46pascqaylt)

[2.4 Summarization: 7](#_heading=h.6c6obdfwdnvq)

[2.5 Natural Language Analysis: 7](#_heading=h.bk7pvq1dpqvt)

[2.6 Prediction of Crashing Partner: 7](#_heading=h.2e7szls0uilu)

[**3 REPORTING ENTITY ANALYSIS 8**](#_heading=h.3uauf2np2n0)

[3.1 Unique Report Versions Count 8](#_heading=h.nyivxhijwgy2)

[3.2 Report Versions and Reporting Entity Count 8](#_heading=h.msvwbwvzx06g)

[3.3 Unique Reporting Entity Count 9](#_heading=h.vo1pzoc6d4wj)

[3.4 Report Type Count 9](#_heading=h.d9bmljapperg)

[3.5 Report Type and Reporting Entity Count 10](#_heading=h.jv02robwrbxo)

[3.6 Monthly Report Submission Count 10](#_heading=h.mip9zcm9jej5)

[3.7 Monthly Incident Count 11](#_heading=h.b6yzu0e6bwho)

[3.8 Temporal Discrepancy Analysis 12](#_heading=h.togsf2wk05a)

[3.9 Make-Model Analysis 13](#_heading=h.jddmbb7w7a5w)

[3.10 Year Discrepancy Analysis 13](#_heading=h.qdcia8k9r8kk)

[3.11 Repeat Vehicle Incidents 14](#_heading=h.ot5y7pn1bwz1)

[**4 INCIDENT TIME ANALYSIS 15**](#_heading=h.ilbb1sq3siab)

[**5 GEOSPATIAL ANALYSIS 15**](#_heading=h.mdgcsbj3r8mq)

[5.1 State-Wise Analysis 15](#_heading=h.2mbgy22w4lxw)

[5.1 City-Wise Analysis 16](#_heading=h.nvecdv6yx53q)

[5.3 Permit Analysis 16](#_heading=h.fjk7glrotnvn)

[**6 CRASH FACTOR ANALYSIS 17**](#_heading=h.5mp3ws57w8fm)

[6.1 Mileage Analysis 17](#_heading=h.tyu4thxbwvhp)

[6.2 Driver/Operator Type 18](#_heading=h.dtuvhdjji9wa)

[6.3 Source Analysis 19](#_heading=h.palqisug1ac6)

[6.4 Road Infrastructure Analysis 20](#_heading=h.kmo4fhimnezd)

[6.5 Posted Speed Limit (MPH) 22](#_heading=h.vpe72dws26uw)

[6.6 Lighting Analysis 22](#_heading=h.ohjixt2tv1z4)

[6.7 Weather Analysis 23](#_heading=h.z09tao5qb18p)

[**7 CRASH ANALYSIS 24**](#_heading=h.9tpey0gap1o7)

[7.1 Crash With? Analysis 24](#_heading=h.orz51wdolwej)

[7.2 Vehicle Crash Trends: Passenger Vehicle (PV) 25](#_heading=h.yn6pzazgj0eg)

[7.3 Pre-Crash Movement Analysis 26](#_heading=h.84d9s4ww8ezt)

[7.4 Pre-Crash Speed of Subject Vehicle (SV) 27](#_heading=h.4f2lzkgq6yth)

[7.5 Overspeeding Analysis 28](#_heading=h.hdcczh8p8bkl)

[7.6 Correlation Analysis 28](#_heading=h.hm34z14b2en3)

[7.7 Contact Area Analysis 29](#_heading=h.j1bb741rxii5)

[7.8 Injury Analysis 30](#_heading=h.7x7c7tsc3pud)

[**8 NARRATIVE ANALYSIS 30**](#_heading=h.k1ylby76rqee)

[**9 SAFETY ANALYSIS 31**](#_heading=h.nh3h0iqd2fbt)

[9.1 Airbags Deployed in Crash Partner (CP) and Subject Vehicle (SV) 31](#_heading=h.1bk2xf29zuvb)

[9.2 Crash Partner (CP) or Subject Vehicle (SV) Towed 32](#_heading=h.h94766lp3w1h)

[9.3 Subject Vehicle (SV) Passenger Seat Belt 33](#_heading=h.vd4y527lvpo4)

[**10 PREDICTION 33**](#_heading=h.myvhs69bn2xz)

[10.1 Flowchart of Prediction 33](#_heading=h.nbu9bxr213h7)

[10.2 Model 34](#_heading=h.gyyaxs5xgupk)

[10.3 Confusion matrix 35](#_heading=h.xtm9k79afse6)

[10.4 Plot between Actual vs. Predicted 36](#_heading=h.uhnyvqkm6vs1)

[10.5 Graphical User Interface (GUI) 37](#_heading=h.jdzsubczyb3c)

[10.6 Future Work 37](#_heading=h.uh8pkh4e1cm5)

[**11 HUMAN VS. AUTOMATED: AZ VEHICLE ANALYSIS 38**](#_heading=h.jllv0a0jyxj)

[11.1 Total Number of Accidents: 38](#_heading=h.t0ngfxb7cd9o)

[11.2 Peak Location 38](#_heading=h.ivhflvmwnwd)

[11.3 Highest Road Type: 38](#_heading=h.f9k1o1izmg3s)

[11.4 Number of Speeding Accidents: 38](#_heading=h.6jn0zbq5w3mv)

[11.5 Crash With: 39](#_heading=h.3gf7jknld5nf)

[11.6 Peak Time of Accidents: 39](#_heading=h.okdnlm3pbw0o)

[11.7 Weather: 39](#_heading=h.8yufs8kmhmn3)

[11.8 Lightning: 39](#_heading=h.60pkmqdxcc24)

[**REFERENCES 40**](#_heading=h.58fwgeol6wxo)

[**ACKNOWLEDGMENTS 41**](#_heading=h.735bpkearkk3)

LIST OF FIGURES

| Figure | Page |
| --- | --- |
| [Figure 1: ADS Unique Report Version Count](#bookmark=id.s7vgsqruiw4c) | 8 |
| [Figure 2: ADAS L2 Unique Report Version Count](#bookmark=id.bzv9aevijbou) | 8 |
| [Figure 3: ADS Report Versions and Reporting Entity Count](#bookmark=id.lozhzndf7mrs) | 8 |
| [Figure 4: ADAS L2 Report Versions and Reporting Entity Count](#bookmark=id.58s5ey294i0g) | 8 |
| [Figure 5: ADS Reporting Entity Count](#bookmark=id.tlq0pjcig5ox) | 9 |
| [Figure 6: ADAS L2 Reporting Entity Count](#bookmark=id.ymex9cbbmscl) | 9 |
| [Figure 7: ADS Report Type Count](#bookmark=id.ilhu6k5z3ztn) | 9 |
| [Figure 8: ADAS L2 Report Type Count](#bookmark=id.jl3005oug1zu) | 9 |
| [Figure 9: ADS Report Type and Reporting Entity Count](#bookmark=id.1i2wdmesq6dl) | 10 |
| [Figure 10: ADAS L2 Report Type and Reporting Entity Count](#bookmark=id.a2katn8e87cj) | 10 |
| [Figure 11: ADS Monthly Report Submission Count](#bookmark=id.ywtn9if5z8w) | 10 |
| [Figure 12: ADAS L2 Monthly Report Submission Count](#bookmark=id.i5fuqan824ex) | 10 |
| [Figure 13: ADS Monthly Incident Count](#bookmark=id.vco3ddl9ducd) | 11 |
| [Figure 14: ADAS L2 Monthly Incident Count](#bookmark=id.zhn6zmguaq1e) | 11 |
| [Figure 15: ADS Temporal Discrepancy](#bookmark=id.248am1337601) | 12 |
| [Figure 16: ADAS L2 Temporal Discrepancy](#bookmark=id.681dluoybj4s) | 12 |
| [Figure 17: Delayed ADS Entities Report](#bookmark=id.7oyz1rwayejj) | 12 |
| [Figure 18: Delayed ADAS L2 Entities Report](#bookmark=id.uvx4z5p30qij) | 12 |
| [Figure 19: ADS Make-Model Analysis Count](#bookmark=id.nu2guc80bqnm) | 13 |
| [Figure 20: ADAS L2 Make-Model Analysis Count](#bookmark=id.3y7yv6uh8j1k) | 13 |
| [Figure 21: ADS Year Discrepancy](#bookmark=id.7hk2lh9xlcfb) | 13 |
| [Figure 22: ADAS L2 Year Discrepancy](#bookmark=id.3g0jt2rsa32o) | 13 |
| [Figure 23: ADS Repeat Vehicle Count](#bookmark=id.v45wp3ylxa19) | 14 |
| [Figure 24: ADAS L2 Repeat Vehicle Count](#bookmark=id.dyiedohgc7vv) | 14 |
| [Figure 25: ADS Incident Time Analysis](#bookmark=id.65nonr10bxta) | 15 |
| [Figure 26: ADAS L2 Incident Time Analysis](#bookmark=id.um26d4wb2ev8) | 15 |
| [Figure 27: ADS State Analysis](#bookmark=id.66zsqafdtunw) | 15 |
| [Figure 28: ADAS L2 State Analysis](#bookmark=id.31hi957ywu71) | 15 |
| [Figure 29: ADS City Analysis](#bookmark=id.z96v8l1snegk) | 16 |
| [Figure 30: ADAS L2 City Analysis](#bookmark=id.ymkjttdhhkoq) | 16 |
| [Figure 31: ADS Mileage Stats](#bookmark=id.5q8dba91tfc) | 17 |
| [Figure 32: ADAS L2 Mileage Stats](#bookmark=id.pwk94ome77su) | 17 |
| [Figure 33: ADS Mileage Distribution](#bookmark=id.4r9fbt5r8kip) | 17 |
| [Figure 34: ADAS L2 Mileage Distribution](#bookmark=id.f86t884wphgk) | 17 |
| [Figure 35: ADS Driver/Operator Type](#bookmark=id.5c5uf14aq19b) | 18 |
| [Figure 36: ADAS L2 Driver/Operator Type](#bookmark=id.nuqv5eiwl53a) | 18 |
| [Figure 37: ADS Source Analysis](#bookmark=id.3j1z55ujcnyp) | 19 |
| [Figure 38: ADAS L2 Source Analysis](#bookmark=id.jcf7x7q4947n) | 19 |
| [Figure 39: ADS Source Analysis Count](#bookmark=id.j00vh0oobl3h) | 19 |
| [Figure 40: ADAS L2 Source Analysis Count](#bookmark=id.d1yrldmg4r7e) | 19 |
| [Figure 41: ADS Road Infrastructure Analysis](#bookmark=id.ngiy8t9rky19) | 20 |
| [Figure 42: ADAS L2 Road Infrastructure Analysis](#bookmark=id.thoafjhl171m) | 20 |
| [Figure 43: ADS Road Infrastructure Count](#bookmark=id.kgt56ww40pby) | 21 |
| [Figure 44: ADAS L2 Road Infrastructure Count](#bookmark=id.vhusy5ydlzh2) | 21 |
| [Figure 45: ADS Posted Speed Limit Distribution](#bookmark=id.n191hksqzhve) | 22 |
| [Figure 46: ADAS L2 Posted Speed Limit Distribution](#bookmark=id.habs9s1gqy9w) | 22 |
| [Figure 47: ADS Lighting](#bookmark=id.7z3bbqwyhdw5) | 23 |
| [Figure 48: ADAS L2 Lighting](#bookmark=id.r5dxm7jv89q6) | 23 |
| [Figure 49: ADS Weather Analysis](#bookmark=id.d0oqusdl48si) | 23 |
| [Figure 50: ADAS L2 Weather Analysis](#bookmark=id.agic2slo1ije) | 23 |
| [Figure 51: ADS Weather Type Count](#bookmark=id.p1jkc0v9bzw5) | 24 |
| [Figure 52: ADAS L2 Weather Type Count](#bookmark=id.qdtjux3bvzrt) | 24 |
| [Figure 53: ADS Crash With? Count](#bookmark=id.b9uyrziah3) | 25 |
| [Figure 54: ADAS L2 Crash With? Count](#bookmark=id.pw77chkk9dgr) | 25 |
| [Figure 55: ADS-PV Crash Count](#bookmark=id.3frxizn4vqtl) | 25 |
| [Figure 56: ADAS L2-PV Crash Count](#bookmark=id.chwinzr2fdlc) | 25 |
| [Figure 57: ADS CP Pre-Crash Movement](#bookmark=id.lmf1bctj9n4) | 26 |
| [Figure 58: ADAS L2 CP Pre-Crash Movement](#bookmark=id.nr7593bgecs2) | 26 |
| [Figure 59: ADS SV Pre-Crash Movement](#bookmark=id.9n8nmpewa6tp) | 26 |
| [Figure 60: ADAS L2 CP Pre-Crash Movement](#bookmark=id.q2uk6xi31vwb) | 26 |
| [Figure 61: ADS Pre-Crash Speed (SV) Distribution](#bookmark=id.vvlw2rwc790d) | 27 |
| [Figure 62: ADAS L2 Pre-Crash Speed (SV) Distribution](#bookmark=id.z05jl55t0w0v) | 27 |
| [Figure 63: ADS Pre-Crash Speed (SV) Stats](#bookmark=id.yitkbtefis3f) | 27 |
| [Figure 64: ADAS L2 Pre-Crash Speed (SV) Stats](#bookmark=id.rr51nacqshp5) | 27 |
| [Figure 65: ADS Correlation Analysis](#bookmark=id.miz3ybfs4dk2) | 28 |
| [Figure 66: ADAS L2 Correlation Analysis](#bookmark=id.kcqcxe4jh038) | 28 |
| [Figure 67: ADS Contact Area Analysis](#bookmark=id.yf3ni1dvnpqg) | 29 |
| [Figure 68: ADAS L2 Contact Area Analysis](#bookmark=id.fv0rd1lxcsi4) | 29 |
| [Figure 69: ADS Injury Count](#bookmark=id.d4qbjr700etg) | 30 |
| [Figure 70: ADAS L2 Injury Count](#bookmark=id.llwqrtbswqb2) | 30 |
| [Figure 71: ADS CP Airbags Deployed](#bookmark=id.7e5wkujbwohd) | 31 |
| [Figure 72: ADAS L2 CP Airbags Deployed](#bookmark=id.caz0l7go0wpa) | 31 |
| [Figure 73: ADS SV Airbags Deployed](#bookmark=id.imnvfddkn5i5) | 31 |
| [Figure 74: ADAS L2 SV Airbags Deployed](#bookmark=id.1lowlr9zl84c) | 31 |
| [Figure 75: ADS CP Towed](#bookmark=id.satwqjsg4ki2) | 32 |
| [Figure 76: ADAS L2 CP Towed](#bookmark=id.pfpaa06f6rvi) | 32 |
| [Figure 77: ADS SV Towed](#bookmark=id.7kpy2yhfczvy) | 32 |
| [Figure 78: ADAS L2 SV Towed](#bookmark=id.45gvocgwzbfn) | 32 |
| [Figure 79: ADS SV Seat Belt](#bookmark=id.fp9224mgxgs1) | 33 |
| [Figure 80: ADAS L2 SV Seat Belt](#bookmark=id.x96v8tl3otrp) | 33 |
| [Figure 81: Prediction Flowchart](#bookmark=id.5yqti4fcd82) | 33 |
| [Figure 82: ADS Confusion Matrix](#bookmark=id.4ih8xlnptrw) | 35 |
| [Figure 83: ADAS L2 Confusion Matrix](#bookmark=id.se54hfx53srg) | 35 |
| [Figure 84: ADS Scatter Plot](#bookmark=id.c7m5chnwvi87) | 36 |
| [Figure 85: ADAS L2 Scatter Plot](#bookmark=id.ttnepypw0gv3) | 36 |
| [Figure 86: ADS Prediction GUI](#bookmark=id.77tcr4jcotd2) | 37 |
| [Figure 87: ADAS L2 Prediction GUI](#bookmark=id.l8c9xcfc3h41) | 37 |

# 

# INTRODUCTION

## 1.1 General Order

The National Highway Traffic Safety Administration (NHTSA) has recently issued a Standing General Order, referred to as the General Order [[3]](#bookmark=id.ll9ms0g0kmbq), which mandates specific manufacturers and operators to report certain types of crashes involving vehicles equipped with automated driving systems (ADS) or Society of Automotive Engineers (SAE) Level 2 advanced driver assistance systems (ADAS). The purpose of this General Order is to facilitate the timely and transparent reporting of real-world crashes associated with these advanced technologies.

Under the General Order, manufacturers and operators are required to provide the NHTSA with information on crashes involving vehicles utilizing automated driving systems or Level 2 ADAS. This reporting mechanism enables the NHTSA to promptly receive crucial data regarding incidents related to these advanced technologies. The aim is to enhance the agency's ability to respond effectively to crashes that may raise safety concerns about the use of automated driving systems and Level 2 ADAS.

The collected data will enable the NHTSA to conduct thorough investigations into incidents that may indicate potential safety issues with ADS and Level 2 ADAS technologies. If, during the investigation, the NHTSA identifies a safety defect, it will take appropriate action to address the concern. This action may involve removing vehicles with safety defects from public roads or implementing remedies to rectify the issues.

In essence, the General Order empowers the NHTSA to proactively monitor and address safety-related issues associated with the deployment of automated driving systems and Level 2 advanced driver assistance systems. By having access to real-world crash data, the NHTSA can take decisive actions to mitigate risks and uphold public safety in the rapidly evolving landscape of vehicle automation.

In June 2021, the National Highway Traffic Safety Administration (NHTSA) took a significant step by issuing the General Order. The primary objective of this order is to assess whether manufacturers of Automated Driving Systems (ADS) and Level 2 Advanced Driver Assistance Systems (ADAS), as well as the vehicles equipped with these technologies, are fulfilling their statutory obligations to guarantee that their products are free from defects that could pose unreasonable risks to motor vehicle safety. According to the NHTSA website, the revised updated general order will be issued in April 2023.

One key motivation behind the General Order is the recognition that, before its implementation, the NHTSA faced limitations in its sources of timely crash notifications. The information available to the agency was often inconsistent across different manufacturers, including those involved in the development of such advanced technologies. This lack of standardized and comprehensive data made it challenging for the NHTSA to gain a holistic understanding of real-world incidents involving vehicles with automated driving systems or Level 2 ADAS.

By mandating reporting from manufacturers and operators, the General Order aims to address these information gaps and enhance the NHTSA's ability to promptly receive and analyze crash data associated with ADS and Level 2 ADAS-equipped vehicles. This, in turn, facilitates a more thorough evaluation of whether these technologies are meeting safety standards and whether any defects exist that could jeopardize motor vehicle safety.

The most recent general order pdf available on the NHTSA website contains details regarding the manufacturing or testing companies that report. It also offers the guidelines and classifications that the crashed object falls under, as well as the process for reporting the crash report to the NHTSA.

In summary, the issuance of the General Order in June 2021 represents a proactive measure by the NHTSA to ensure that manufacturers of automated driving systems and Level 2 ADAS fulfill their safety obligations. It acknowledges the need for standardized and timely reporting of crash data to better assess the safety performance of these advanced technologies and respond effectively to potential safety concerns.

## 1.2 Difference between ADS and ADAS Level 2

ADS

An Automated Driving System implies a higher level of automation, typically encompassing a range of driving tasks. In a fully automated system, the vehicle can handle all aspects of driving without human intervention. This includes functions like steering, acceleration, and braking in various driving scenarios. In a vehicle equipped with a high-level ADS, the system is designed to handle complex driving scenarios and navigate through various environments without human input. This often includes the capability to handle lane changes, navigate intersections, and respond to dynamic traffic conditions. A well-developed ADS aims to minimize or eliminate the need for human intervention during most driving scenarios. The vehicle can handle challenging situations independently. The term "Automated Driving System" often implies a higher level of automation and may be associated with Society of Automotive Engineers (SAE) Levels 3, 4, or 5, where the vehicle can operate automated in certain conditions without human intervention.

ADAS Level 2

Advanced Driver Assistance Systems at Level 2 involve partial automation. While the system can assist with certain driving tasks, such as steering and acceleration, it still requires the active supervision of a human driver. The driver must remain engaged and be ready to take control when needed. ADAS Level 2 systems offer advanced driver assistance features, such as adaptive cruise control, lane-keeping assistance, and automated parking. However, these systems are not intended to provide fully automated driving. They support the driver but require constant supervision. Human intervention is a key aspect of Level 2 ADAS. The driver must actively monitor the driving environment and be prepared to take control when the system requests or encounters situations it cannot handle. Advanced Driver Assistance Systems at Level 2 fall within the category of partial automation, where the vehicle can assist with specific driving tasks but still requires the driver to be actively engaged.

## 1.3 Data Summary

The ADS-ADAS Level 2 Incident Report Sample Form is a comprehensive document designed to gather detailed information about incidents involving vehicles equipped with Automated Driving Systems (ADS) or Advanced Driver Assistance Systems (ADAS) at Level 2. This form is a crucial tool for reporting entities, such as manufacturers and operators, to provide the National Highway Traffic Safety Administration (NHTSA) with essential data regarding real-world incidents. The information collected helps the NHTSA assess the safety performance of these advanced technologies and respond effectively to potential safety concerns.

Here's an elaboration on the key fields typically found on such a report form:

**Incident Type**: Describe the nature of the incident, whether it's a collision, near-miss, system failure, or another type of incident.

**Location**: Specifies where the incident occurred, providing details such as the city, state, and specific location (e.g., intersection, highway).

**Weather Conditions**: Records the weather conditions at the time of the incident, which is crucial for understanding how environmental factors may have contributed to the event.

**Road Conditions**: Details the state of the road (e.g., dry, wet, icy) to assess the impact of road conditions on the incident.

**Factors Related to Speed**: Captures information related to the speed of the vehicle(s) at the time of the incident. Speed is a critical factor in understanding the dynamics of a collision.

**Incident Description**: Allows for a narrative description of the incident, providing a more detailed account of what transpired. This section may include information about the behavior of the automated system and the driver's actions.

**Vehicle Information**: Collects details about the vehicles involved, including make, model, year, and any specific details about the ADS or ADAS components.

These are some fields of the report. This content collectively contributes to a thorough understanding of the incident, enabling the NHTSA to conduct detailed investigations into the performance of ADS and Level 2 ADAS technologies. The static image of the incident report form available on the NHTSA website serves as a visual representation of the data fields and structure that reporting entities need to complete when submitting incident reports. The form aims to ensure that critical information is systematically captured for analysis and regulatory purposes.

The inclusion of crash data in the Arizona Department of Transportation (ADOT) database since 2010 represents a valuable and extensive repository of information on traffic incidents that have occurred within the state. This database is a comprehensive collection that covers various aspects of each crash event, offering insights into the dynamics of road incidents and supporting efforts to enhance road safety. The database encompasses crash data from 2010 onward, providing a decade-long perspective on-road incidents. This extensive time frame enables the analysis of trends, patterns, and changes in road safety over the years. The data covers all events that have taken place in Arizona. This wide geographic scope ensures that the database reflects the diversity of road conditions, traffic patterns, and environmental factors across the state. Information about the vehicles involved in each crash is documented. This includes details such as make, model, year, and potential information about the state of the vehicle after the crash. Analyzing this data can provide insights into vehicle safety and potential areas for improvement. The database contains information about individuals involved in the crashes. This may include drivers, passengers, pedestrians, or cyclists. Details about injuries sustained, if any, and other demographic information can be crucial for understanding the human impact of road incidents. Each entry in the database includes a description of the incident, offering a narrative that outlines the circumstances leading to the crash. This narrative can provide context on contributing factors, weather conditions, road layout, and other variables influencing the incident. The database may include information about contributing factors to the crashes, such as distracted driving, impaired driving, speeding, or weather-related conditions. Analyzing these factors helps in identifying areas for targeted interventions and safety improvements. This information plays a crucial role in promoting road safety, guiding policy decisions, and fostering a data-driven approach to transportation planning and management.

ADS and ADAS Level 2 Data

The dataset related to ADS contains about 703 data points, while the ADAS Level 2 dataset has 1154 data points. Each dataset comprises 137 columns, suggesting a comprehensive set of information for each recorded incident. The data covers information from the year 2021 up to **September 15, 2023,** providing a recent and ongoing perspective on incidents involving vehicles equipped with ADS and ADAS Level 2 technologies. The datasets likely include detailed information on accidents involving vehicles with these advanced technologies, such as incident type, location, weather conditions, road conditions, damages sustained, factors related to speed, and incident descriptions. These data points are crucial for analyzing the performance and safety of ADS and ADAS Level 2 systems. Certain sensitive information, such as precise location data and Criminal Background Investigation (CBI) related data, has been redacted to ensure privacy and compliance with regulations.

Human-Driven Vehicle Data in Arizona

The human-driven vehicle data for Arizona was provided by a sponsor of the project, and it covers a substantial time from 2010 to 2022 [[1]](#bookmark=id.uxli81kazo6n). This dataset is expansive, reflecting a comprehensive record of road incidents involving traditional, human-driven vehicles. Due to the large size of the dataset, the decision was made to focus on the 2020-2022 timeframe. This approach likely aimed to balance the need for relevant and recent data while mitigating the challenges associated with processing and managing extensive datasets. According to the information provided, approximately 100,000 road accidents are occurring each year in Arizona. This statistic underscores the importance of analyzing and addressing road safety issues in the state. The dataset contains more than 50 inputs, encompassing various aspects of road incidents. These inputs likely cover a range of details, including driver information, unit details (possibly referring to vehicles involved), and incident-specific data. The human-driven vehicle data is deemed crucial for accurate interpretation and analysis. The richness of the dataset allows for a comprehensive understanding of the factors contributing to road accidents and can serve as a valuable resource for safety assessments, policy formulation, and planning.

The following datasets are used for conducting detailed analyses, identifying trends, and predicting crashes for this project. Also interesting facts regarding the crash reports provided. The project will also conduct a comparison analysis to ascertain whether the collision elements are the same or distinct based on the crash reports produced due to human-driven vehicles(Arizona crash reports) and ADS vehicles in the Arizona region.

# ANALYSIS

## 2.1 Data Cleaning and Preprocessing:

Identify and handle missing values in the datasets for both ADS and ADAS Level 2 vehicles. This may involve imputation, removal of rows/columns, or other appropriate strategies. Ensure consistency in data types, convert variables to appropriate formats, and address any anomalies in the data structure. Respect privacy concerns by keeping redacted information (such as precise location data and CBI-related data) in a secure state, ensuring compliance with regulations.

## 2.2 Exploratory Data Analysis (EDA):

**Descriptive Statistics**: Compute basic descriptive statistics (mean, median, standard deviation, etc.) for key variables such as injury severity, road conditions, and crash modeling factors in ADS and ADAS Level 2 datasets.

**Data Visualization**: Utilize various visualization techniques (bar charts, pie charts, grouped bar graphs) to visually represent the distribution of injury severity, road conditions, and other factors in the datasets.

## 2.3 Predictive Modeling:

**Data Splitting**: Split the datasets into training and testing sets for model development and evaluation.

**Model Selection**: Choose an appropriate predictive modeling algorithm based on the nature of the problem (e.g., classification for predicting injury severity).

**Feature Selection**: Identify key features contributing to the prediction of injury severity, road conditions, or crash modeling factors.

**Model Training**: Train the predictive model using the training dataset, optimizing hyperparameters as needed.

**Model Evaluation**: Evaluate the model's performance using the testing dataset, considering metrics like accuracy and precision.

## 2.4 Summarization:

**Comparative Analysis**: Compare the results of injury severity, road conditions, and crash modeling factors between human-driven vehicles and ADS/ADAS Level 2 vehicles. Identify commonalities and differences.

**Numerical Summary:** Provide a numerical summary or index to quantify the severity of injuries, prevailing road conditions, and relevant crash modeling factors. This can aid in clear comparisons and conclusions.

## 2.5 Natural Language Analysis:

**Narrative Exploration**: Perform natural language analysis on the narrative descriptions of accidents. Extract key themes, commonalities, or patterns that may not be apparent in structured data.

## 2.6 Prediction of Crashing Partner:

**Target Variable Definition**: Defined the target variable for predicting the crashing partner (e.g., another vehicle, or pedestrian).

**Model Development:** Build a predictive model to identify the likely crashing partner based on relevant features from the dataset.

**Evaluation**: Evaluate the model's accuracy and performance in predicting the crashing partner.

With the help of these thorough techniques, the ADS-ADAS level data was cleaned, explored, and subjected to natural language analysis and predictive modeling to provide a more nuanced understanding of the variables influencing traffic incidents. The next section will cover the trends, discoveries, and comparisons between automated and human-driven cars.

# REPORTING ENTITY ANALYSIS

## 3.1 Unique Report Versions Count

| Figure 1: ADS Unique Report Version Count |  |
| --- | --- |
| Figure 1: ADS Unique Report Version Count | Figure 2: ADAS L2 Unique Report Version Count |

In the Incident Report scenario, when a Reporting Entity revises a previously submitted report, a fresh version is generated and assigned a sequential number through the portal. Figure 1 illustrates the distribution of ADS reporting entities creating various versions. The majority of reports belong to version 1, and only two incident reports are in version 4. Conversely, within ADAS L2, there are 883 reports in version 1, and approximately six reports have version 5.

## 3.2 Report Versions and Reporting Entity Count

|  |  |
| --- | --- |
| Figure 3: ADS Report Versions and Reporting Entity Count | Figure 4: ADAS L2 Report Versions and Reporting Entity Count |

The horizontal bar graphs above show how many report versions each reporting entity has created. Figure 3 shows that ADS Waymo has the most version 1s. Lucid and Transdev Alternative Services are the reporting entities that have version 4 of the incident report. Figure 4 shows that Tesla has the most version 1 and version 2 reports. Subaru and BMW are the reporting entities that created version 4s of incident reports, and four companies created version 5s of incident reports.

## 3.3 Unique Reporting Entity Count

|  |  |
| --- | --- |
| Figure 5: ADS Reporting Entity Count | Figure 6: ADAS L2 Reporting Entity Count |

The figures above represent the total number of incident reports submitted by reporting entities. In Figure 5, Waymo has the most incidents under ADS, while Tesla has the most incidents under ADAS L2.

## 3.4 Report Type Count

|  |  |
| --- | --- |
| Figure 7: ADS Report Type Count | Figure 8: ADAS L2 Report Type Count |

The Reporting Entity's preferred type of Incident Report. The following values are possible: 1-Day, 5-Day, Monthly, and Yearly. There are other options, such as no new or updated incident reports, a 10-day update, and an Update. 408 incidents are monthly report type under ADS and 840 incidents are 10-Day Update under ADAS L2. In ADS as per the reports submitted, there are 1 Day, Update, 5-day, 10-day Update, and monthly reports. Whereas in ADAS L2 there are monthly, 1-Day, Update, 5-Day, and 10-day Update report types.

## 3.5 Report Type and Reporting Entity Count

|  |  |
| --- | --- |
| Figure 9: ADS Report Type and Reporting Entity Count | Figure 10: ADAS L2 Report Type and Reporting Entity Count |

The following figures give the number of incident reports report types submitted by the reporting entity or manufacturer. In Figure 9 under ADS Waymo submits 5 types of reports 1-day, monthly, Update, 10-day, and 5-day. So among the submitted reports by Waymo, most of the incidents are monthly report types. Consequently in Figure 10 in ADAS L2, Tesla submits 10-day and 5-day reports frequently.

## 3.6 Monthly Report Submission Count

|  |  |
| --- | --- |
| Figure 11: ADS Monthly Report Submission Count | Figure 12: ADAS L2 Monthly Report Submission Count |

The above bar graphs do not provide an exact date of report submission, but rather a count of report submissions that occurred in each month from July 2021 to September 2023. Figure 11 shows that the majority of ADS reports were submitted in September 2023. In Figure 12 under ADAS L2, the largest number of report submissions occurred in February 2022.

## 3.7 Monthly Incident Count

|  |  |
| --- | --- |
| Figure 13: ADS Monthly Incident Count | Figure 14: ADAS L2 Monthly Incident Count |

The bar graphs above show the monthly count of incidents involving ADS and ADAS L2 vehicles. Figure 13 depicts the monthly count of ADS vehicle accidents from July 2021 to September 2023. The monthly accident count for ADAS L2 vehicles is displayed from August 2019 to September 2023. The most accidents in ADS occurred in August 2023, and the fewest occurred in September 2023, but this date is only considered until 15 September 2023. The number of crashes in ADS in January 2023 is 10, which is the second-lowest count. The majority of accidents in ADAS L2 occurred in December 2022. From August 2019 to June 2021, there were a total of 21 accidents, with the number of accidents being less than 5 per month. There was a significant increase in ADAS L2 vehicle crashes from June 2021 to July 2021.

## 3.8 Temporal Discrepancy Analysis

|  |  |
| --- | --- |
| Figure 15: ADS Temporal Discrepancy | Figure 16: ADAS L2 Temporal Discrepancy |

The following analysis looks at the time difference between the incident date and the report submission date. There could be a variety of reasons why a specific reporting entity took so long to complete and submit a report for a specific incident. The day count of 0 indicates that the following report was submitted in the same month as the incident. According to Figure 15 in ADS, the majority of reports were submitted within 31 days of the incident. The majority of the accident reports in the case of ADAS L2, as per Figure 16 were submitted within the same month of the incident. The incident reports were submitted within 31 days, which is second best.

|  |  |
| --- | --- |
| Figure 17: Delayed ADS Entities Report | Figure 18: Delayed ADAS L2 Entities Report |

The tables above show which incidents of reporting entities required more than 100 days for report submission. According to Figure 17, there were 13 incidents in ADS, with Zoox being the reporting entity with 7 incidents and taking 424 days to complete the report submission of a particular crash. In ADAS L2, there are 100 incidents where incident report submissions are completed after 100 days. However, only ten are shown in Figure 18. Honda and Tesla are the two companies with the highest number of incidents. The longest time taken by Tesla for an incident to be reported is 942 days.

## 3.9 Make-Model Analysis

|  |  |
| --- | --- |
| Figure 19: ADS Make-Model Analysis Count | Figure 20: ADAS L2 Make-Model Analysis Count |

This analysis determines which car of the reporting entity was involved in the most crashes. Many reporting entities had numbers models. The vehicles whose make-model combination has more than 15 incidents are shown in the figures above. The Jaguar I-Pace is involved in the most crashes (251) in Figure 19 of ADS. According to Figure 20 in ADAS L2, the Tesla Model 3 was the most involved in crashes (436).

## 3.10 Year Discrepancy Analysis

|  |  |
| --- | --- |
| Figure 21: ADS Year Discrepancy | Figure 22: ADAS L2 Year Discrepancy |

The differences between the incident year and the manufacturing year are defined in the following analysis. The number of years between incident and manufacturing years is shown in the bar graphs above. The value -1 represents the year difference, indicating that the following vehicles have not yet been released to customers but will be in the future. The reported incident occurred during the vehicle's testing phase. According to Figure 21, most ADS vehicles were involved in crashes during the same year of manufacture. According to Figure 22, a majority of crashes in ADAS L2 vehicles were reported after one year of manufacture

## 3.11 Repeat Vehicle Incidents

|  |  |
| --- | --- |
| Figure 23: ADS Repeat Vehicle Count | Figure 24: ADAS L2 Repeat Vehicle Count |

The following analysis counts the number of vehicles from the reporting entity that were involved in multiple crashes. Figure 23 shows that there were several ADS vehicles involved in multiple crashes, with General Motors vehicles being the most prevalent. In ADAS L2, Figure 24 on the other hand, five vehicles were involved in multiple crashes.

# INCIDENT TIME ANALYSIS

|  |  |
| --- | --- |
| Figure 25: ADS Incident Time Analysis | Figure 26: ADAS L2 Incident Time Analysis |

The following analysis depicts the number of accidents that occurred during each hour of the day. According to Figure 25, most of ADS incidents occurred around 1 pm. At 5 a.m., a few accidents were reported. On the other hand, Figure 26 shows that a majority of ADAS L2 crashes occurred at 10 p.m. during the night, with the fewest accidents occurring at 8 a.m.

# GEOSPATIAL ANALYSIS

## 5.1 State-Wise Analysis

|  |  |
| --- | --- |
| Figure 27: ADS State Analysis | Figure 28: ADAS L2 State Analysis |

The table below lists the top ten states where incidents have been reported. According to Figure 27, most of the incidents (469 in total) were reported in California. In the case of ADAS L2, the majority of the incidents (377 in total) were reported in California. Arizona is the second highest in the ADS case, and Texas is the highest in the ADAS L2 case.

## 5.1 City-Wise Analysis

|  |  |
| --- | --- |
| Figure 29: ADS City Analysis | Figure 30: ADAS L2 City Analysis |

The table above lists the top ten cities where incidents have been reported. According to Figure 29, most of the incidents (420 in total) were reported in San Francisco. In the case of ADAS L2, the majority of the incidents (42 in total) were reported in Los Angeles. Austin is the second highest in the ADS case, and Houston is the highest in the ADAS L2 case.

## 5.3 Permit Analysis

There is a field for the vehicle permit details, according to the incident report. If the vehicle has a State or local permit, the permit issuer's name must be included in the following report. In the case of ADS vehicles involved in crashes, 176 stated that they had a permit, 32 stated that they did not, and 5 stated that they did not know. The majority of the incidents involved vehicles with California permit plates. In ADAS L2, however, none of the reports mentioned having a state or local permit. Either it was No or unknown.

# CRASH FACTOR ANALYSIS

## 6.1 Mileage Analysis

|  |  |
| --- | --- |
| Figure 31: ADS Mileage Stats | Figure 32: ADAS L2 Mileage Stats |

The odometer reading of the subject vehicle in miles at the time of the incident, as reported by the Reporting Entity. The mileage factor of the reported incident vehicles was subjected to statistical analysis. According to figure 31 in ADS vehicle, the average mileage of the vehicle is 27k mpg, and there were only 690 incidents where the mileage of the vehicle was mentioned in incident reports, the minimum mileage is 0 mpg, and the maximum mileage is 44k mpg. In ADS vehicles, the median mileage is around 16k mpg.

Figure 32 shows that the average mileage in ADAS L2 vehicles is around 34k mpg. There were 1002 incidents where the mileage of the vehicle was mentioned in incident reports. The minimum mileage reported is 48 mpg, while the maximum mileage reported is around 840k mpg. In ADAS L2 median mileage is 27k mpg.

|  |  |
| --- | --- |
| Figure 33: ADS Mileage Distribution | Figure 34: ADAS L2 Mileage Distribution |

The following is the histogram distribution of the mileage of vehicles involved in crashes. As per Figure 33 most of the ADS vehicles involved in crashes has mileage in the range of 0-50k mpg and Figure 34 implies that ADAS L2 vehicle involved in crashes has milega in the range of 0-200k mpg.

## 6.2 Driver/Operator Type

The Reporting Entity's documentation of the individual responsible for the operation, fallback operation, or any component of the dynamic driving task (DDT) during a vehicle incident encompasses various categories. The possible values include "Consumer," denoting an individual operating a commercially available Level 2 ADAS/ADS without affiliations to a manufacturer at the incident time. "In-vehicle (Commercial/Test)" represents a non-consumer inside the subject vehicle, while "Remote (Commercial/Test)" signifies a non-consumer capable of remote driving or assistance. "In-Vehicle and Remote (Commercial/Test)" combines both categories. "None" indicates no individual responsibility for the DDT during the incident, and "Other, see Narrative" allows for custom descriptions. "Unknown" is applicable when the responsible individual's identity or role cannot be determined.

|  |  |
| --- | --- |
| Figure 35: ADS Driver/Operator Type | Figure 36: ADAS L2 Driver/Operator Type |

The following pie chart gives the operator-type distribution of incidents reported. In Figure 35 as per ADS most of the incidents were In-vehicle (Commercial/Test) type. and the second best was in Remote (Commercial/Test). As per Figure 36, in ADAS L2 the reported accidents were of consumer type.

## 6.3 Source Analysis

The source from which the Reporting Entity first learned of the incident, as reported by the Reporting Entity. There are about 8 options under source: Source - Complaint/Claim, Source – Telematics, Source - Law Enforcement, Source - Field Report, Source - Testing, Source - Media, Source - Other, and Source - Other Text.

|  |  |
| --- | --- |
| Figure 37: ADS Source Analysis | Figure 38: ADAS L2 Source Analysis |

The pie chart below depicts the distribution of the sources from which the reporting entity learned about the incident. The source of reporting the incident is Telematics, as shown in Figures 37 and 38. Telematics means that the Reporting Entity stated that it first learned of the incident through communication in the form of electronic data transmitted from the subject vehicle to the Reporting Entity.

|  |  |
| --- | --- |
| Figure 39: ADS Source Analysis Count | Figure 40: ADAS L2 Source Analysis Count |

The tables above show the number of different sources from which incidents were reported to the reporting entity. According to Figures 39 and 40, there were 502 incidents reported on ADS vehicles and 890 incidents were reported on ADAS L2 vehicles, respectively. The smallest number of non-zero numbers In ADS, there are seven reports whose source is a complaint/claim, whereas in ADAS L2, there are five reports from a testing source.

## 6.4 Road Infrastructure Analysis

The road type on which the subject vehicle was traveling at the time of the incident, as reported by the Reporting Entity. The possible Road types would be Highways/freeways, streets, intersections, Parking Lot, Traffic Circle, Unknown, etc. The environmental conditions on the roadway surface at the time of the incident, as reported by the Reporting Entity. Values that could be used - Dry, Snow / Slush / Ice, Wet, Other, and Unknown.

|  |  |
| --- | --- |
| Figure 41: ADS Road Infrastructure Analysis | Figure 42: ADAS L2 Road Infrastructure Analysis |

The graphs below depict the road infrastructure. The x-axis represents the type of roadway, and the y-axis represents the count. Each road type has a subplot of different roadway surfaces. Figure 41 shows that the majority of ADS vehicle crashes occurred on Dry-Intersections, while Figure 42 shows that the majority of ADAS L2 vehicle crashes occurred on Dry-Highway/Freeway.

|  |  |
| --- | --- |
| Figure 43: ADS Road Infrastructure Count | Figure 44: ADAS L2 Road Infrastructure Count |

The tables below show the number of possible combinations based on the road type and road surface mentioned in incident reports by reporting entities. In the case of ADS from Figure 43, the huge majority of incidents were reported on Intersection-Dry with a count of 300. Figure 44 in ADAS L2 Highway/Freeway-Dry with a count of 362. According to the above table, ADS vehicles are not commonly permitted on highways or freeways, and the vast majority of ADS and ADAS L2 incidents occurred on dry road surfaces. In the case of ADS vehicles, the most common road type where crashes occur is Interactions, while in the case of ADAS L2, it is Highway/Freeway.

## 6.5 Posted Speed Limit (MPH)

The speed limit in miles per hour posted on the roadway where the incident occurred, as reported by the Reporting Entity.

|  |  |
| --- | --- |
| Figure 45: ADS Posted Speed Limit  Distribution | Figure 46: ADAS L2 Posted Speed Limit  Distribution |

The histogram above depicts the distribution of posted speed limits at the incident locations. According to Figure 45, the majority of ADS vehicle incidents occurred where the posted speed limit was 25 mph, while the majority of ADAS L2 vehicles as per Figure 46 incidents occurred where the posted speed limit was 65 mph. The range of the posted speed limit in the case of ADS is 0-75 mph and in the case of ADAS L2 is 10-80 mph.

## 6.6 Lighting Analysis

The Reporting Entity's report of the lighting conditions at the time and location of the incident. Values that could be used -

Daylight: The incident occurred during the day.

Dusk / Dawn: The incident occurred at dusk or dawn.

Dark-Lightend: The incident occurred in the dark, but street lights were turned on.

Dark-Not Lightened: The incident occurred after dark, but there were no street lights.

Dark-Unknown Lighting: Although the incident occurred after dark, the presence of street lights is unknown.

|  |  |
| --- | --- |
| Figure 47: ADS Lighting | Figure 48: ADAS L2 Lighting |

According to Figures 47 and 48, the majority of ADS and ADAS L2 accidents occurred during the day. The second highest incidents in ADAS L2 vehicles occurred in unknown lighting conditions, either because the reporting entity was unaware of the lighting condition or because they specifically avoided it. In the case of ADS, the second highest incidents occurred in Dark-Lightened. The number of incidents that occur in Drak-Unknown Lighting in the case of ADS and ADAS L2 is less than 10.

## 6.7 Weather Analysis

The incident's weather or environmental conditions at the time and location. There are 8 types of weather included in the incident report: clear, snow, cloudy, fog/smoke, rain, severe wind, unknown, and other

|  |  |
| --- | --- |
| Figure 49: ADS Weather Analysis | Figure 50: ADAS L2 Weather Analysis |

The following pie chart shows th distribution of weather among incidents. From Figure 49 and Figure 50, the majority of accidents occurred when the weather was clear.

|  |  |
| --- | --- |
| Figure 51: ADS Weather Type Count | Figure 52: ADAS L2 Weather Type Count |

The table above shows the number of incidents that occurred during specific weather conditions. Figures 51 and 52 show that when the weather was clear, 592 crashes occurred with ADS vehicles and 502 with ADAS L2 vehicles. There were no incidents reported when the wind was strong. In the case of ADS, no incidents were reported in snow and fog/smoke weather. In contrast, few incidents occurred in snow and fog/smoke-type weather in ADAS L2. In ADAS L2, the second highest number of incidents occurred when the weather was unknown.

# CRASH ANALYSIS

## 7.1 Crash With? Analysis

The categorical description provided by the Reporting Entity of any vehicle, non-motorist, animal, or object with which the subject vehicle came into contact during the incident. Possible values include: Vehicles such as passenger cars, SUVs, vans, pickup trucks, motorcycles, buses, heavy trucks, and first responder vehicles are examples of these types of vehicles. Pedestrian, Cyclist, Non-Motorist: Other, Pole / Tree, Other Fixed Object, Animal, "Other, see Narrative," Unknown.

|  |  |
| --- | --- |
| Figure 53: ADS Crash With? Count | Figure 54: ADAS L2 Crash With? Count |

The above bar graphs depict the number of incidents in which subject vehicles collided with various subjects. According to Figure 53, ADS vehicles most commonly collide with passenger vehicles, and according to Figure 54, ADAS L2 vehicles collide most frequently with unknown subjects. In the ADAS L2 case, the third highest count is a collision with a passenger vehicle.

## 7.2 Vehicle Crash Trends: Passenger Vehicle (PV)

|  |  |
| --- | --- |
| Figure 55: ADS-PV Crash Count | Figure 56: ADAS L2-PV Crash Count |

The following bar graph gives the year-wise count of crashes between ADS/ADAS L2 - Passenger Vehicles. On the x-axis, we have the incident year in which crashes occurred with passenger vehicles, and on the y-axis, we have the count of the incidents. As per Figure 55 in ADS-PV most of the crashes that happened in the years 2022 and 2023(till September 15) are also close. From Figure 56 ADAS L2-PV the majority of crashes occurred in the year 2022 and before 2021 there were only two incidents submitted under the following filter.

## 7.3 Pre-Crash Movement Analysis

The Reporting Entity’s report should mention the movement of any Crash Partner (CP) or Subject Vehicle (SV) before the incident. The report should contain the actions taken by the vehicles that is CP and SV before getting into a collision. Subject Vehicle in our case would be the reporting entity’s vehicle that is ADS or ADAS L2 and the Crash partner with which the subject vehicle crashed.

|  |  |
| --- | --- |
| Figure 57: ADS CP Pre-Crash Movement | Figure 58: ADAS L2 CP Pre-Crash Movement |

|  |  |
| --- | --- |
| Figure 59: ADS SV Pre-Crash Movement | Figure 60: ADAS L2 SV Pre-Crash Movement |

The above figure depicts the pre-crash movements on the x-axis and the count on the y-axis. Figures 57, 58, 59, and 60 shown above Going straight is one of the top two most common pre-crash movements, regardless of whether the vehicle is the Crash Partner or the Subject Vehicle. According to Figure 58, the ADAS L2's CP with the greatest pre-crash movement is an unknown subject. In the ADS case, CP backing is the third-highest Pre-Crash Movement as per Figure 57. In Figure 59, the second highest SV Pre-Crash Movement is ADS stopped, and in Figure 60, the second highest SV Pre-Crash Movement is ADAS L2 Lane/Raod departure.

## 7.4 Pre-Crash Speed of Subject Vehicle (SV)

|  |  |
| --- | --- |
| Figure 61: ADS Pre-Crash Speed (SV)  Distribution | Figure 62: ADAS L2 Pre-Crash Speed (SV)  Distribution |

The histogram above depicts the distribution of Pre-Crash Speed of Subject Vehicle (SV). According to Figure 61, the majority of ADS vehicles had a pre-crash speed between 0-10 mph, while the majority of ADAS L2 vehicles as per Figure 62 had a pre-crash speed between 40-60 mph.

|  |  |
| --- | --- |
| Figure 63: ADS Pre-Crash Speed (SV)  Stats | Figure 64: ADAS L2 Pre-Crash Speed (SV)  Stats |

According to the statistical analysis on the pre-crash speed of SV, the mean speed in the case of ADS is 8mph, the highest speed is 70mph, and the lowest speed is 0. Figure 64 shows that the average speed for ADAS L2 is 45 mph, with the highest being 97 mph and the lowest being 0.

## 7.5 Overspeeding Analysis

This analysis gave the count of vehicles where the subject vehicle's precrash speed exceeded the posted speed limit, resulting in 237 accidents in ADAS L2 and 4 accidents in the context of ADS. This data suggests a link between incidents and instances in which the subject vehicle's speed before a crash exceeded the posted speed limit.

## 7.6 Correlation Analysis

This is a statistical technique that is used to assess the strength and direction of a linear relationship between two quantitative variables. A correlation analysis produces a correlation coefficient, which can range from -1 to 1. The correlation coefficient is a numerical measure that indicates the strength and direction of a linear relationship between two variables. It ranges from -1 to 1.

* A correlation coefficient of 1 indicates a perfect positive linear relationship.
* A correlation coefficient of -1 indicates a perfect negative linear relationship.
* A correlation coefficient of 0 indicates no linear relationship.

Correlations between numerical variables (Mileage, SV Pre-crash Speed, and Posted Speed Limit).

|  |  |
| --- | --- |
| Figure 65: ADS Correlation Analysis | Figure 66: ADAS L2 Correlation Analysis |

SV Precrash Speed (MPH) and Posted Speed Limit (MPH) have a stronger correlation in ADS but there is no certain co between any of these variables in the case of ADAS L2.

## 7.7 Contact Area Analysis

The location of any crash contact or damage resulting from the crash on the crash partner or the subject vehicle as reported by the Reporting Entity. There are in total 11 contact area points. the reporting entity must mention the spot of damage on the Crash Partner or the Subject Vehicle in the incident report. The contact areas are Bottom, Front, Front Left, Front Right, Left, Rear, Rear Left, Rear Right, Right, Top and Unknown.

|  |  |
| --- | --- |
| Figure 67: ADS Contact Area Analysis | Figure 68: ADAS L2 Contact Area Analysis |

The x-axis in the plots above represents the type of contact area, and the y-axis represents the count. Under each contact area, there are two subplots: blue for the crash vehicle and orange for the subject vehicle. As per Figure 67 under CP the most damaged part is the front of the vehicle and in SV is the rear part of the vehicle. In ADAS L2, Figure 68, the most damaged part in SV is the front left contact area and in CP is unknown. The rear of the ADS SV and the front of the CP are more damaged. As a result, the majority of accidents would be either SV backing/stationary or CP moving straight/stationary. assuming that SV and CP are not stationary. In ADAS L2 the most damaged part in SV is the front.

## 7.8 Injury Analysis

The highest confirmed or alleged crash injury severity level reported by the Reporting Entity as a result of the incident.

|  |  |
| --- | --- |
| Figure 69: ADS Injury Count | Figure 70: ADAS L2 Injury Count |

According to the bar graph analysis in Figures 69 and 70, No Injuries Reported is among the top two injury severity levels. Figure 70 ADAS L2 shows that the number of reported injuries by the reporting entity in the incident report is unknown, with a count of 954. The minor injuries reported in ADS vehicles is 62 whereas in ADAS L2 is 44.

# NARRATIVE ANALYSIS

The following Narrative analysis is done on the narration provided by the reporting entity in the incident report. Natural language processing is applied in it using a natural language kit.

In ADS

* The **lane** word is used 514 times so there are accidents involved while changing lanes either center, left, right, etc.
* The **sensor** word was used 57 times, and most of the time it was about the Waymo sensor which came into contact with the crash partner.
* The word **Waymo** is used more than 2000 times. there is a detailed narration about the incident with Waymo in ADS crashes.

In ADAS L2

* Most of the narration was **redacted** since it may contain confidential business information.
* The **lane** word is used 220 times so there are accidents involved while changing lanes.
* The word "**speed**" was used 35 times in the narration, and in each case, either the vehicle was traveling faster than the posted speed limit or it was not slowing down.

# SAFETY ANALYSIS

## 9.1 Airbags Deployed in Crash Partner (CP) and Subject Vehicle (SV)

|  |  |
| --- | --- |
| Figure 71: ADS CP Airbags Deployed | Figure 72: ADAS L2 CP Airbags Deployed |

## 

|  |  |
| --- | --- |
| Figure 73: ADS SV Airbags Deployed | Figure 74: ADAS L2 SV Airbags Deployed |

The Reporting Entity's report on whether the Crash Partner or Subject vehicle's airbags deployed during the incident. Yes, No, Not Applicable, and Unknown are all possible values. In Figures 71 and 73 bar graphs, there was no deployment of airbags either in the crash partner or subject vehicle in the majority of the reported incidents. So in the case of ADS crashes weren’t that major. Figure 72, on the other hand, depicts that in crashes involving ADAS L2 with its crash partner, airbag deployment was unknown in the majority of incidents, and the Subject vehicles ADAS L2 involved had incidents where the majority of incidents where airbags were deployed as per Figure 74.

## 9.2 Crash Partner (CP) or Subject Vehicle (SV) Towed

|  |  |
| --- | --- |
| Figure 75: ADS CP Towed | Figure 76: ADAS L2 CP Towed |

## 

|  |  |
| --- | --- |
| Figure 77: ADS SV Towed | Figure 78: ADAS L2 SV Towed |

The Reporting Entity's report on whether the Crash Partner / Other vehicle was towed from the incident site. Possible outcomes: Yes, No, Unknown, Not Applicable. In most incidents involving ADS vehicles depicted in Figures 75 and 77, none of the crash partners or subject vehicles were towed. The status of towing crash partners or subject vehicles is unknown in most incidents involving ADAS L2 vehicles depicted in Figures 78 and 79 bar graphs.

## 9.3 Subject Vehicle (SV) Passenger Seat Belt

|  |  |
| --- | --- |
| Figure 79: ADS SV Seat Belt | Figure 80: ADAS L2 SV Seat Belt |

The report of the Reporting Entity on whether all passengers in the subject vehicle were wearing seat belts at the time of the incident. Possible values include: Yes, that's right. "No, see Narrative," Unknown, no passengers are in the vehicle. According to the ADS bar graphs, more than half of the incidents had all passengers wearing seat belts. On the other hand, the majority of ADAS L2 passengers wearing seat belts is unknown.

# PREDICTION

## 10.1 Flowchart of Prediction

|  |
| --- |
| Figure 81: Prediction Flowchart |

The flowchart description of figure 81 as follows:

**Data Filtering**: Identify the columns that are relevant to your analysis or prediction task. Remove unnecessary columns or rows from the dataset.

**Data Cleaning and Preprocessing**: Check for missing values in the dataset and handle them appropriately (e.g., impute missing values or remove rows/columns). Convert categorical data into numerical format, as machine learning models typically require numerical input. Handle outliers or anomalies in the data. Standardize or normalize numerical features if needed.

**Creation of Dictionary for Unique Values**: Create a dictionary that maps unique categorical values to specific numerical codes. This step is crucial for converting categorical data into a format suitable for machine learning models.

**Mapping Categorical Values**: Apply the mapping from the dictionary to convert categorical values to numerical representations.

**Training Data Preparation**: Split the dataset into training and testing sets. Separate the features (input variables) from the target variable (output variable).

**Model Training**: Use a Random Forest classifier [[2]](#bookmark=id.1kkny570851y) for training. This ensemble learning method builds multiple decision trees and merges them to get a more accurate and stable prediction. Fit the model to the training data.

**Model Saving**: Save the trained model to a file. This allows you to reuse the model for future predictions without retraining.

**Testing and Prediction**: Load the saved model when needed. Use the model to make predictions on new, unseen data. Evaluate the model's performance on the testing set to assess its accuracy, precision, recall, etc.[[3] Standing General Order on Crash Reporting. (n.…](#bookmark=id.ll9ms0g0kmbq)

## 10.2 Model

The Inputs for the prediction are Vehicle Make, Mileage, State, Road Infrastructure, Weather, Lighting, Subject vehicle, and Crash Partner pre-crash movement, pre-crash speed of the subject vehicle and Incident time. The Output of the prediction will tell you with what subject the following vehicle is highly likely to crash. There are 2 predictions separate ADS and ADAS L2.

Results: ADS Prediction Accuracy is 74% and ADAS Pred Accuracy: is 60%.

## 10.3 Confusion matrix

|  |  |
| --- | --- |
| Figure 82: ADS Confusion Matrix | Figure 83: ADAS L2 Confusion Matrix |

Figures 82 and 83 are confusion matrices of ADS and ADAS L2 respectively.

A confusion matrix is a table that is often used to evaluate the performance of a classification model. It provides a summary of the predicted and actual classifications made by a classification algorithm. The matrix has four entries: true positive (TP), false positive (FP), true negative (TN), and false negative (FN).

Here's a breakdown of these terms:

True Positive (TP): The number of instances where the model correctly predicted the positive class.

True Negative (TN): The number of instances where the model correctly predicted the negative class.

False Positive (FP): The number of instances where the model incorrectly predicted the positive class (it predicted positive, but the actual class is negative).

False Negative (FN): The number of instances where the model incorrectly predicted the negative class (it predicted negative, but the actual class is positive).

## 10.4 Plot between Actual vs. Predicted

|  |  |
| --- | --- |
| Figure 84: ADS Scatter Plot | Figure 85: ADAS L2 Scatter Plot |

Figures 84 and 85 are Scatter Plot of predicted vs. actual values of ADS and ADAS L2 respectively.

In a scatter plot of predicted vs. actual values:

Above the Line: Points above the diagonal line indicate instances where the model overestimated the target variable. This implies that the predicted values are higher than the actual values.

Below the Line: Points below the diagonal line indicate instances where the model underestimated the target variable. This implies that the predicted values are lower than the actual values.

The diagonal line represents a perfect prediction where predicted values equal actual values. Deviations from this line suggest errors in the predictions.

## 10.5 Graphical User Interface (GUI)

|  |  |
| --- | --- |
| Figure 86: ADS Prediction GUI | Figure 87: ADAS L2 Prediction GUI |

The Python PyQt5 [[4]](#bookmark=id.g6w92fa1oai) library was used to create the following GUI. There are approximately 20 inputs, a few of which have dropdown and numerical input. If the given input refers to an ADS or ADAS L2 model, the output will be displayed.

## 10.6 Future Work

To create a user-friendly GUI for your model, focus on an intuitive design that accommodates users with varying levels of technical expertise. Implement clear navigation, and visually appealing interfaces, and provide interactive elements for ease of use. Additionally, ensure that the GUI facilitates seamless integration with different use cases by allowing users to input specific parameters or customize settings based on their requirements. This adaptability will enhance the model's accessibility and utility across diverse scenarios.

To improve the model's accuracy, prioritize the collection of more data related to the incidents under consideration. A larger and more diverse dataset can provide the model with a richer understanding of patterns and variations within the data, ultimately leading to enhanced predictive capabilities. Experiment with different machine learning algorithms to identify the most suitable one for your specific use case, as different algorithms may excel in various contexts. Moreover, explore complex queries and analyses to extract deeper insights from the data, and consider adding more fields to your dataset to capture additional details that could contribute to more precise incident predictions. By combining these strategies, you can develop a robust and versatile model that caters to a wide range of user needs while delivering accurate and insightful results.

# HUMAN VS. AUTOMATED: AZ VEHICLE ANALYSIS

## 11.1 Total Number of Accidents:

* ADS: 96 accidents
* ADAS L2: 25 accidents
* Human Driven: around 200,000 accidents

## 11.2 Peak Location

* ADS: The peak location for ADS accidents is Phoenix and Tempe.
* ADAS L2: The peak location for ADAS accidents is Phoenix.
* Human Driven: The peak location for accidents caused by human drivers is Flagstaff

## 11.3 Highest Road Type:

* ADS: The highest number of accidents for ADS occurs on dry surface streets.
* ADAS L2: The road surface type for ADAS accidents is unknown, but it happens on highways or freeways.
* Human Driven: Accidents caused by human drivers predominantly occur on dry surface-level roads.

## 11.4 Number of Speeding Accidents:

* ADS: 0 accidents related to speeding.
* ADAS L2: 2 accidents involving speeding.
* Human Driven: 1,538 accidents attributed to speeding.

## 

## 11.5 Crash With:

* ADS: Accidents involving ADS often result in crashes with passenger cars.
* ADAS L2: The crash partner is unknown for ADAS accidents.
* Human Driven: Accidents caused by human drivers frequently involve crashes with passenger cars.

## 11.6 Peak Time of Accidents:

* ADS: The peak time for ADS accidents is at 1 PM.
* ADAS L2: ADAS accidents tend to peak at midnight.
* Human Driven: The peak time for accidents caused by human drivers is between 3-4 PM.

## 11.7 Weather:

* ADS: clear
* ADAS L2: clear
* Human Driven: clear

## 11.8 Lightning:

* ADS: Daylight
* ADAS L2: Daylight and Unknown
* Human Driven: Daylight

# REFERENCES

[1] *Arizona Motor Vehicle Crash Facts | Department of Transportation*. (n.d.). ADOT. Retrieved December 1, 2023, from https://azdot.gov/mvd/services/statistics/arizona-motor-vehicle-crash-facts

[2] Breiman, L. (n.d.). *Random Forests*. SpringerLink. Retrieved December 1, 2023, from https://link.springer.com/article/10.1023/A:1010933404324

[3] *Standing General Order on Crash Reporting*. (n.d.). NHTSA. Retrieved December 1, 2023, from <https://www.nhtsa.gov/laws-regulations/standing-general-order-crash-reporting#overview>

[4] *PyQt5 · PyPI*. (n.d.). PyPI. Retrieved December 1, 2023, from <https://pypi.org/project/PyQt5/>

# ACKNOWLEDGMENTS

I would like to express my sincere gratitude to Prof. Brendan Russo of Northern Arizona University (NAU) and Prof. Jeffrey Wishart of Arizona State University (ASU) for their invaluable guidance and support throughout the course of this project. Their expertise and insights have played a pivotal role in shaping the direction of the research and enhancing the quality of the work. I am truly thankful for their time, encouragement, and commitment to academic excellence, which have significantly contributed to the successful completion of this report.