Analysis of Car Safety measures vs Fatality

A PROJECT REPORT
Submitted by
GROUP 12
RXB210045 - RAJESH BHATTACHARJEE
MXM210088 - MANEESH CHANDRA MANSHETTY
SXP210146 - SUPRIYA PADMAPRIYA SURESH
AXR210044 - AKANKSHA RAWAT
GXS210012 - GAUTHAM SHETTY

Executive Summary

ABC Car Manufacturing company wants to roll out a generic awareness campaign for all drivers to obey speed limits and follow simple safety measures like airbags and seat belts.

Analysis to find out whether Safety measures like Airbags and Seat belts play a critical role in saving lives in case of any major accident in conjunction with higher speed limit.

The initial investigation shows it is mostly due to the combined effect of exceeding speed limit and not using safety measures - Air bags and Seat Belts, causing most of the fatalities.

As per the project requirement, we will be analyzing the data to find a correlation between Driver conditions after collision and whether Safety measures as well as whether car has exceeded Speed limit.

The analysis will be also centered around fatality as that is the most unfortunate turn of event which can happen.

PROJECT INTRODUCTION

CONTEXT

ABC Car Manufacturing company wants to roll out a generic awareness campaign for all drivers to obey speed limits and follow simple safety measures like airbags and seat belts

OBJECTIVE

Predict fatality rate considering safety measures while driving a car.

Phase 1:

UNDERSTANDING THE DATA

Two Phases of Execution

The project will be executed in two phases, each phase focusing on specific data points.

In Phase I:

The added effectiveness of combining both safety measures on Fatality rate. Here 2 safety measures are Airbags and Seat Belts.

Team will plot Injury Severity information along with fatality count with Safety measures information and compare bar plots to understand the relationship between the impact of collision on driver/passenger and safety measures used.

Team will then apply a data mining algorithm to analyze the effect of safety measures on Survival rate.

Team will make validate the research question whether people who use both Airbag and Seat belts as safety measures have more chance to survive Car crashes.

Phase 2:

Deep Diving

In Phase 2

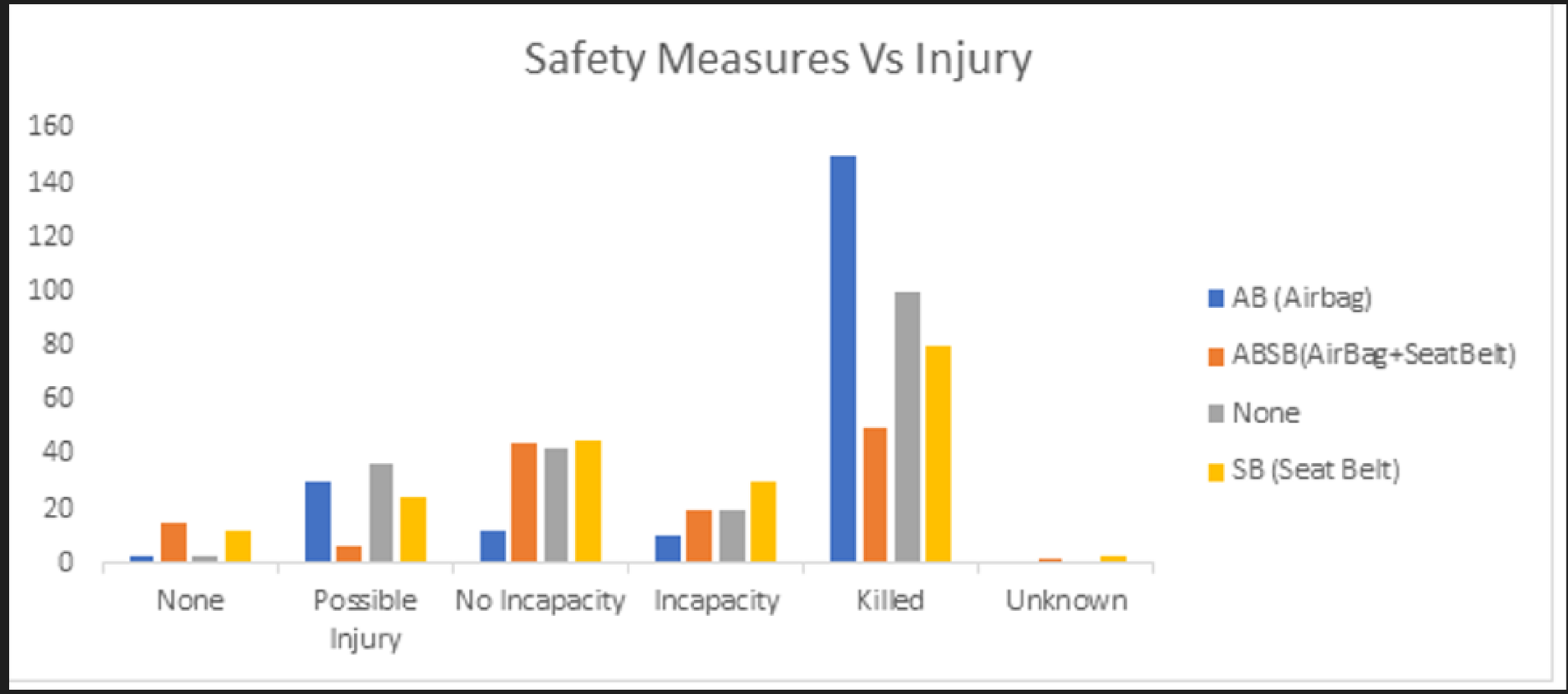
Team further investigates to identify the effect of Vehicle Speed for the scenarios where individuals use both Safety measures- Seat-belt and Airbags both.

In other words, we explore if vehicle speed affects the Fatality rate given the drivers use both the Safety Measures (Seat-belt, Airbags).

Data Overview

- Data Source - Kaggle
- 26217 observations
- 15 variables
- Target Variable - "Dead"

Initial Visualization



Data Pre-processing

1. No Missing Values
2. No Noisy Data
3. New "Safety_Measure" attribute

Safety_Measure values

Only "Seatbelts" protecting occupants "SB"

Only "Airbag" protecting occupants "AB"

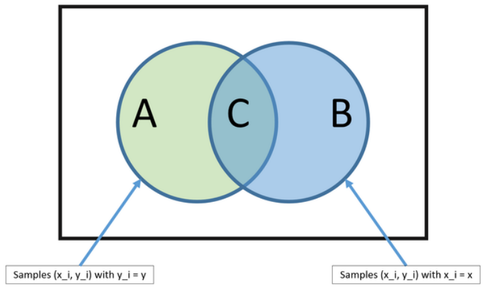
"Seatbelts" and "Airbag" protecting occupants "ABSB"

No Safety_Measures "None"

Comparison between different techniques

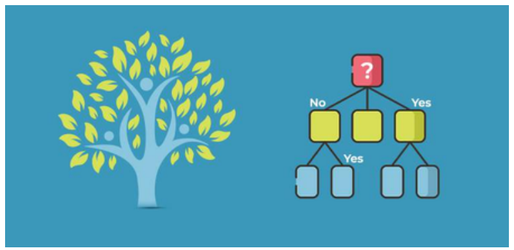
1

Naive Bayes
71.20%



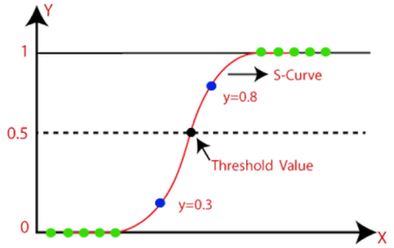
2

CART
71.80%



3

Logistic Regression
73.30%

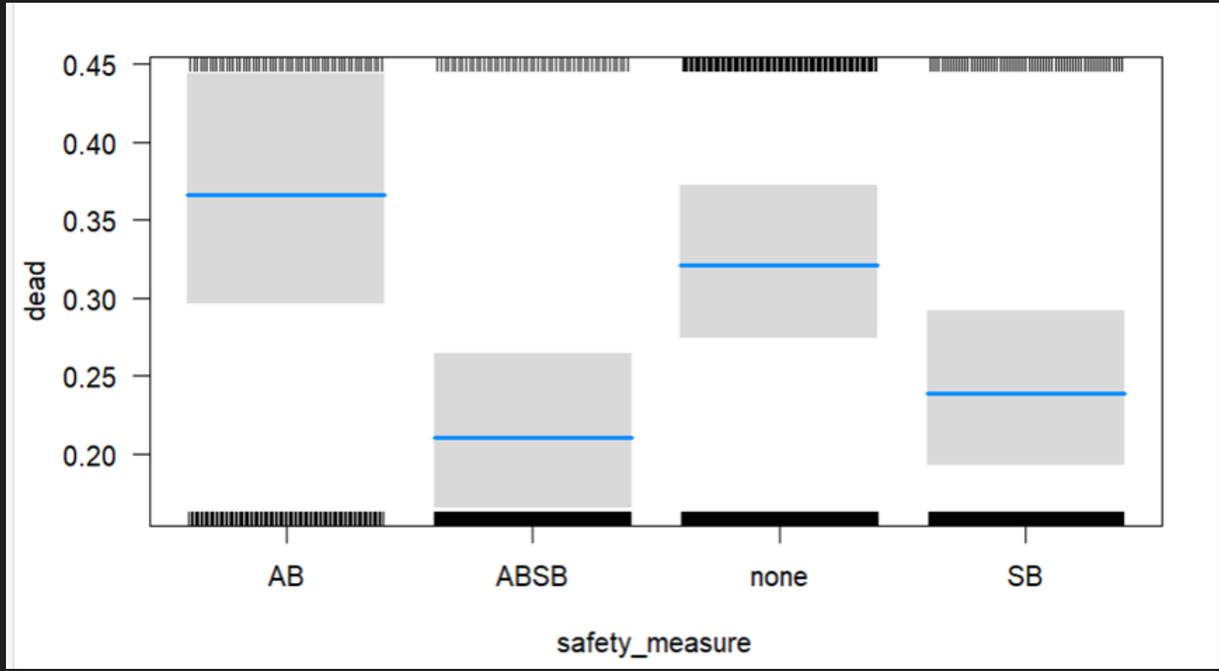


Objective 1

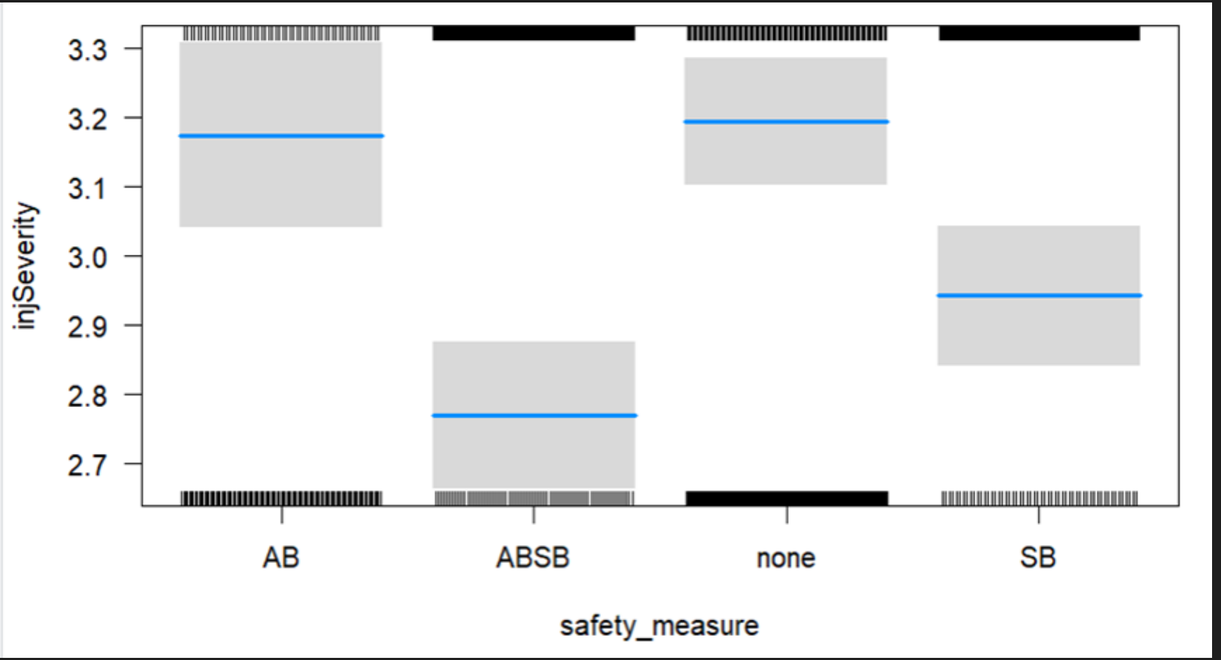
```
Call:
glm(formula = dead ~ safety_measure, family = binomial, data = train_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.9554  -0.8796  -0.7381   1.4169   1.7664

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        -0.5474     0.1636  -3.347 0.000817 ***
safety_measureABSB -0.7768     0.2240  -3.467 0.000526 ***
```

1. Binary Logistic Regression
2. Data Split - 70:30
3. 5% significance level
4. Research Question

Safety_Measure Vs Fatality Ratee



Safety_Measure Vs Injury Rate

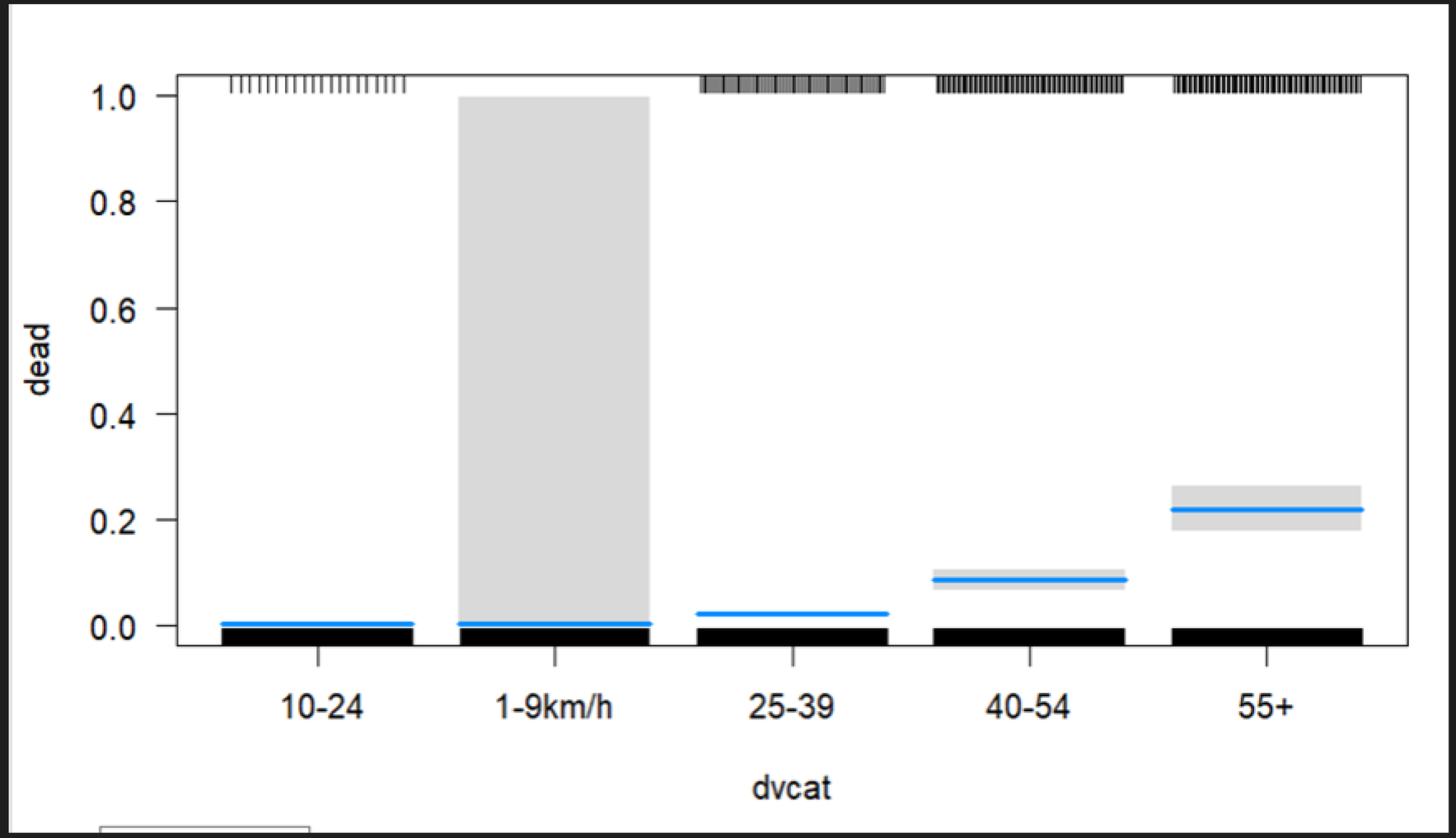


Objective 2

```
Call:
glm(formula = dead ~ dvcat, family = binomial, data = train_data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.6989  -0.2000  -0.0791  -0.0791   3.3963

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -5.7644     0.2240 -25.739  < 2e-16 ***
dvcat1-9km/h -12.8017   326.1320  -0.039    0.969
dvcat25-39     1.8623     0.2586   7.201 5.98e-13 ***
dvcat40-54     3.3594     0.2540  13.227  < 2e-16 ***
dvcat55+       4.4792     0.2579  17.367  < 2e-16 ***
---
```

1. Binary Logistic Regression
2. Data Split - 70:30
3. 5% significance level
4. Research Question

Speed_Limit Vs Fatality Ratee



Future Scope

1 Key scope 1

Extend scope of project by analyzing more data attributes:-

- manufacturing year of vehicle
- drivers age
- whether an old car is prone to accident or causing fatality
- which age group is recording high mortality rate

2 Key scope 2

What interesting conclusions can be drawn and whether any conflicting theories emerge from those observations

3 Key scope 3

We have also observed that mortality rate is high when only Airbags are used as safety measure. It is almost as high as when no safety measures are used. This interesting find needs further investigation with more data points gathered over the period

4 Key scope 4

We can use different data mining techniques and algorithms to cross validate like naive bayes or artificial neural network

Conclusions

1 Action Step 1

After going through the 2 phases of analysis- we can conclude that we can observe that drivers following both safety measures- air bags and seat belts, observing the lowest death rate of only 20% whereas drivers using none of them , observing a mortality rate of 35%.

2 Action Step 2

Also, for drivers who are using both safety measures but driving at a high speed - observing a mortality rate of 25% whereas if they are driving with moderate speed (<40) , the mortality rate is very less as 1-2%.

3 Action Step 3

This data and graphs can be very much useful to drive the campaign.


Thank you!