Exploratory Data Analysis on Vehicle Insurance Fraud

Jeremiah Olabisi Johnson

2022-11-24

## 0.1 Introduction

Exploratory Data Analysis will be used in this project to examine the Insurance Fraud Dataset and give the business insight (EDA). Any action taken to manipulate the insurance system is referred to as insurance fraud. It happens when a claimant, in this case, drivers or passengers, tries to get something for which they are not eligible. Hard insurance fraud is when someone purposefully fabricates an accident; soft insurance fraud is when someone makes a legitimate insurance claim but fabricates a portion of it. The goal of this research is to investigate the fraud datasets and offer insights.

## 0.2 Solution Summary

The steps taking

1. Data cleaning process
2. Feature engineering to aid process
3. Gain insights through visualization
4. Conclusions

## 0.3 Data Cleaning Proccess

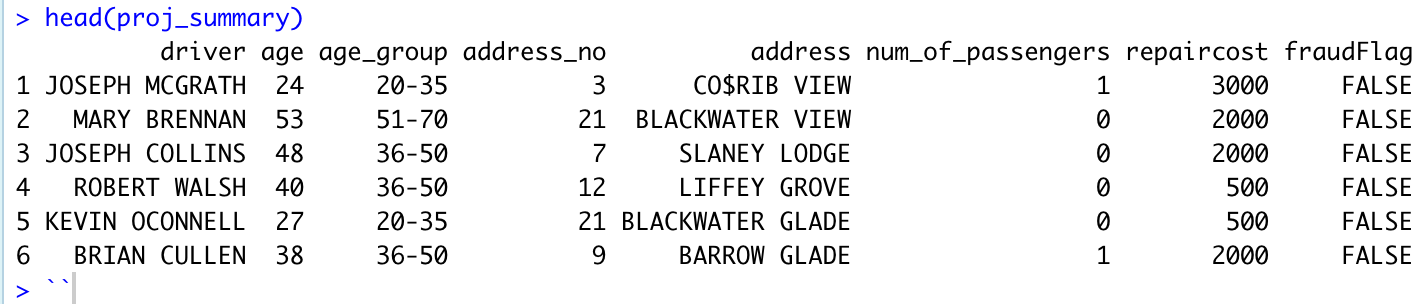
The data were cleaned using the following procedure before being subjected to an exploratory analysis.

1. To show the relationship between a driver’s age and certain behavioral characteristics associated with insurance fraud, I formed an age group.
2. To get a separate column for the address name and number, I used the str splt() technique.
3. To reduce repair costs and create a consistent column, I replaced and eliminated any unnecessary characters.

Then, the data that have been cleansed and are prepared for exploratory analysis are chosen.

## 0.4 Feature Enginnering

Feature engineering is employed to speed up the exploratory data analysis procedure. For exploratory data analysis, Feature engineering is the process of turning raw data into features that more accurately depict the underlying issue. In this instance, the number of passengers present is 1 and the number of passengers absent is 0. This is done to enhance exploratory visualization. A portion of the cleaned dataset used for the exploratory study is displayed in the image below.



## 0.5 Exploratory Analysis

The following stage is to obtain some understanding of the nature of auto insurance fraud in relation to the behavior of those who commit it.

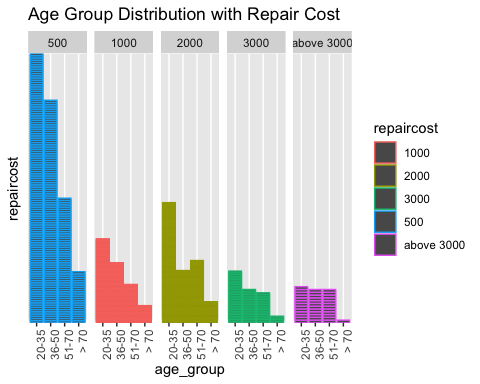


Figure 1: Age Group Distribution with RepairCost

According to the graph above, the category of 500 repair costs experiences the most fraud. It also happens to be the least expensive in terms of repairs, which implies that some scams take place when the repairs cost is more than 3000. The graph shows that the age group between 20 and 35 is the one most frequently involved in fraud across all repair costs. In contrast, across all repair cost categories, the fraud rate for age groups above 70 is comparatively low.

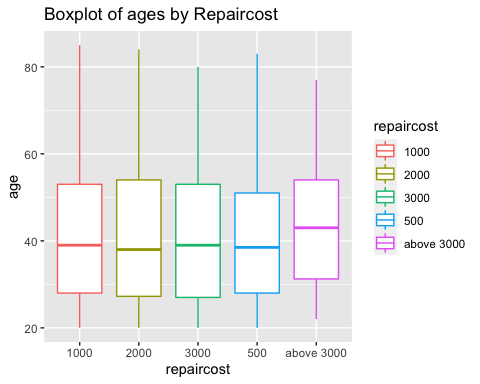


Figure 2: highligts the variation in repaircost

The boxplot shows that the median age of repair cost claims over 3,000 is higher. Given that the type of car would affect the amount requested for repairs, this suggests that the cost of repairs would be more in this situation than 3,000 for a luxury car. The fact that their repaircost claim is over 3,000 may further indicate that age groups over 70 drive expensive vehicles.

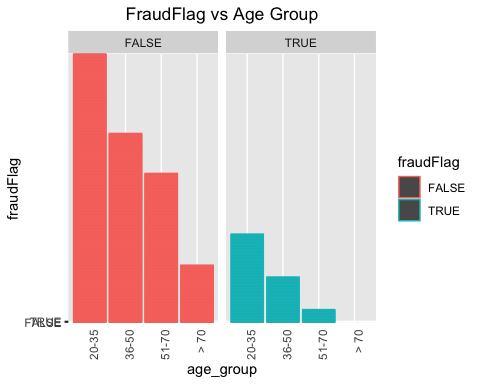


Figure 3: Highlights the impact of age group in Fraud Insurance

While the age group over 70 is involved in the least amount of fraud, the 20 to 35 age group had the highest number of vehicle insurance fraud.



Figure 4: Proportion of number of Passengers involved in Fraud

According to the graph above, the fraud rate was higher when there were more passengers This indicates that the proportion of fraud increased when there were 2 passengers compared to just 1 or none. A fraud incidence was reported in 60% of the circumstances involving two travelers. Additionally, in more than 80% of cases involving only one passenger, there was no fraud discovered.

## 0.6 Conclusion

The following are the conclusions reached after the exploratory analysis:

* The average repair cost claim for fraudulent auto insurance is 500.
* The majority of vehicle insurance fraudsters are in the 20–35 age bracket.
* Older individuals drive high-end vehicles
* The likelihood of auto insurance fraud increases with the number of passengers.