Executive Summary

The purpose of the project is to find the major factors for fatal, injury and car damage due to car crash in Maryland State. SHSP (Strategic Highway Safety plan, 2016) reported that in the past five years on average 502 people died, 3702 people injured and 84, 955 cars were damaged. In order to address the problem, three datasets (car crash, driver info, and vehicle info) which were recorded at the time of accident were taken (DATA, GOV, 2016). The datasets were merged using MS-SQL techniques, cleaned and prepared in using excel and SAS EM for the best outcome. Tableau was used for Data visualization. Nine models were built in using SAS EM software with a binary target variable (report Type: fatal/injury = 1, car damage = 0) and 31 predictors. out of all, Logistic with backward Regression was selected with ROC value (0.81), Sensitivity Accuracy rate (0.67), minimum overfitting (0.001), less complex (using chi square and p- value <0.001) and ease of implementation (identifying important variables easily). In addition, the result selected the top 5 variables which affects fatal/injury car crash. These are Airbag deployment performance (not deployed at all and deployed in other direction), Alcohol use (refused or positive test), using belt (unbuckled), Car Collision type (single car crash, same direction), the driver condition (being abnormal). The

result helps SHSP, DMV, stakeholders, insurance company, car manufacturers, and partnerships to allocate budget for education and enforcement to act on the result in order to achieve the goal of minimizing fatal/injury car crash in Maryland.

Project Scope

The scope of this project is finding appropriate, reliable, robust and more features dataset related with car crash in Maryland State. Next, cleaning, Preparing and summarizing statistical analysis of fatal/injury and car will be done. Then developing and comparing different models by using SAS EM like stepwise logistic regression, SVM, Random forest, Decision model, Ensemble will be the third task in this project. Finally, selecting model with high performance rate of with a minimum accuracy rate of 85% will be the last task of the project. In addition, data visualization will be done to show the most significant factors of fatal/ injury accident by county, reasons, date of birth, time, safety equipment (Belt and Air bag) and other relevant factors to provide easy and understandable information for stakeholders.

The project is unique in providing insights for various business areas (MHSO, MVA, Insurance companies, Emergency Health facilities, car manufactures, society) since it merges three robust datasets (car crash, driver info, vehicle info) with many attributes). Moreover, after searching up previously made studies in related to accident in Maryland, it has been found out that this project will be the first one to build predictive model on accident dataset in Maryland.

Importance of the Project

The result from this project answers question which is important for Maryland Highway Safety Office to prepare a budget request list and allocate funds properly to reduce accident. For example, which city, and which road has the highest fatalities/injuries? and what is the most frequent reason for those accidents? These insights help MHSO to plan

and take action with priority intervention. Emergency Rooms or Urgent Care Health Centers in Maryland can also utilize the results to allocate the number of emergency rooms based on the frequency of injuries in each city and assigning professionals based on the type of collision.

Questions, like what percentage of young adults are involved in car crashes which cities are exposed to the highest amount of car crashes help life and car insurance companies to set the minimum cost to cover liability. The voice of Maryland Consumer Right Coalition reported that most drivers, even those with perfect record, pay more than $500 a year for the legally required minimum liability coverage. Today in Maryland, location triumphs driving skills. The Maryland Insurance Administration (MIA) allows insurance companies

to use geographic areas, sex, age, marital status, education and credit histories to set rates (MCRC, 2019).

Problem Description

According to the report of the Maryland Motor Vehicle Administration (MMVA), the number of fatal accidents showed an increasing trend after 2014. 557 people died in Maryland traffic crashes during 2017 whereas a total of 522 people was killed in Maryland car accidents in 2016 and 521 people were killed in 2015. About 30% of the police-reported car accidents resulted in physical injuries to drivers or passengers (Miller, 2016). Fatal accidents are serious tragedies, taking away the life of a loved one unexpectedly, and the consequence go beyond one's imagination with respect to the emotional pain and financial burdens that strain day to day activity. For example, children who lost their

parents due to accidents are susceptible to mental illnesses, such as depression, which affect their education, physical and emotional well-being. Moreover, those children are unable to get enough financial support for food, school materials, medical and living situations. All these negative consequences expose children to be ineffective and unproductive in their school performance and prohibit them from achieving their goal. Furthermore, the community and society also share the burden of these children in providing their basic needs.

Second, Car crash injuries put a person in a long-term physical and psychological disability status that affect their daily life and can complicate basic survival skills. Even if the injury is minor, most car accident injuries leave some sort of stain, resulting in long-term consequences that are difficult to cope with. In 2007, 7357 teenage drivers in Maryland were

involved in crashes with injuries. Even if the injury is short-term, it results in high medical bills, preventing a person from being able to work and take care of family and other financial obligations. All these outcomes limit a someone's ability to accomplish pre-determined goals in life.

Property damage due to car crashes is another challenge in life which invites an unplanned expense due to the fulfilling the need to buy a new car or fix damage one. The person who is involved in accident is expected to deal with the insurance company. Specifically, if the person is found to be at fault for the accident, the monthly insurance premiums increase and the driving record will have a permanent mark detailing the accident, which could have a negative impact to be acceptable in certain jobs which requires the stability of the person. For

example, transportation, security jobs, or occupations that consider driving record as one of the criteria

for job offering will be able to see the record and likely cross that individual off the list of potential hires. In general, car crashes affect the life of an individual untimely or impose a disability

Business Understanding

The Maryland Strategic Highway Safety Plan (SHSP) is a statewide, coordinated, comprehensive, traffic safety plan to reduce highway fatalities and serious injuries on all public streets and highways. It establishes overall goals and objectives, as well as strategies within each of six key areas. Despite increases in vehicle miles traveled, reported traffic crashes declined in Maryland, dropping to a historic low

of 96,392 in 2009. Between 2005 and 2009, the number of fatalities decreased by 10.4 percent, overall injuries decreased by 14.4 percent, and serious injuries declined by 39.9 percent. To continue this positive trend, Maryland updated the plan in 2010 under the direction of the SHSP Management and Implementation Teams (MSHS, 2016).

Safety is a State priority. Maryland joined the American Association of State Highway and Transportation Officials (AASHTO), other national organizations and several other states in adopting a Toward Zero Deaths fatality goal to reduce motor vehicle-related fatalities and injuries by one-half by 2030. Each emphasis area also adopted measurable fatality and injury objectives in line with these goals (MSHS, 2016)

Organization

The Maryland Strategic safety plan works for the reduction of fatality and injury. SHSP Maryland Department of Transportation's Highway Safety Office (MHSO) and SHSP Area Teams developed the 2016–2020 SHSP to implement safety strategies during the next five years.

The Maryland Highway Safety office (MHSO) works under the Maryland Motor Vehicle Administration (MMVA). Administrators provide overall leadership for the state's highway safety program. The MHSO is housed within the MVA, with direct supervision provided by the MVA's Chief Deputy Administrator. The mission is dedicated to saving lives and preventing injuries by reducing motor vehicle crashes through the administration of a comprehensive network of traffic safety programs. The vision of MHSO is Moving Maryland

Toward Zero Deaths since death is not an acceptable consequence of driving (MDT, 2018). Maryland maintains the Toward Zero Deaths (TZD) approach with a philosophy of MHSO's "Maryland's Toward Zero Deaths Campaign; Every life count." by developing interim targets to reduce fatalities by at least 50 percent in the next two decades (from 592 in 2008 to 296 in 2030) (MMVA, 2018).

Stakeholders

Stakeholder groups includes federal, state and local government agencies, nongovernmental organizations, regional authorities, and individual advocates who are participated in the development of the SHSP. Each Team which includes regional and local agencies held at least two facilitated discussions to identify, develop, and finalize

strategies for the 2016-2020 SHSP. Each team participated with difficult decisions regarding how to cover the essentials of transportation safety while remaining strategic and focused on the most vital needs (SHSP, 2016).

SHSP Stakeholders are service provider organizations in improving road way safety in Maryland through implementing the strategies and action steps related to Engineering, Enforcement, Education and Emergency Medical Services. SHSP has a plan to develop new partnerships for the success of roadway safely. Achieving these goals will require a sustained and steadfast commitment from State and local agencies and key safety partners “Safety leaders convened a wide range of stakeholder groups to develop the new plan, and they participated in a series of meetings to confirm the final list of emphasis areas, develop strategies, and begin

working on action steps to meet the new performance targets (MSHS, 2016)

Define Business Area

The business area of this project is wide and deep that surly touch the life of everyone, family, society, companies and organization. SHSP is government organization who plan to achieve the goal of reducing fatal/injury. Data driven planning will help to allocate budget for each county by reason of accident. For example, educating aggressive drivers, using belt, construction of roads, changing rules and so on. bring success. MSHS is working the implementation with

stakeholders and partners. Maryland's Motor Vehicle Administration issue driver licenses by being dedicated to the safety of every driver and occupants traveling on Maryland's roadways. MMVA can update their requirement to issue driver license based on summary statistics of this project. Next, SHSP and MHSO plan and implement the safety of reducing accidents and achieve the goal of Toward Zero Deaths. Every year MHSO allocate budget to reduce accident and the information from this project helps MHSO to prioritize the issue to be addressed and to allocate the money according to the seriousness of the accident.

Car and life Insurance companies are benefited from the result to have a knowledge of the most recurring accident type, places, reasons in Maryland state which helps insurance companies to set up their policies with information. For

example, area with the most prevalent accident, the sell price of insurance is higher in order to compensate the coverage expense at the time when their customer is found at fault. Insurance companies provided grants to intervene the most prevalent type of accident. For example, in 2016, the statewide seat belt usage rate for drivers and front seat passengers was 90.8 percent, which is a two percent decrease from 2015. In 2016 the MHSO received a $15,000 grant from State Farm Insurance to educate teens concerning seat belt usage and prevent distracted, impaired, and aggressive driving (MHSO, 2016).

Urgent and Emergency care get data driven information from this project to plan schedules of emergency room and professional to save the life of injured person. Health care facility provides more emergency room and urgent care

professionals around areas where more car crash was experienced.

Car manufactures are another beneficiary to use the result of this project if they need the information of car crash and Air bag deployment. A lot of fatal/injury accident happened when the airbag fails to deploy at all or improperly. This helps to review their performance and change their plan to increase their business.

Business Objectives and Business success criteria

- Reduce fatalities
- Reduce property damage
- Reduce insurance rates
- Reduce medical bills
- Increase stability and healthy lifestyles

SHSP success criteria is measured by the goal SHSP set up in 2020. Reduce the number of distracted driving related fatalities on all roads in Maryland from the five-year average (2004 to 2008) of 333 to 177 or fewer by December 31, 202. Another success criterion is Reducing the number of distracted driving related to serious injuries on all roads in

Maryland from the five-year average (2004-2008) of 4134 to 1826 or fewer by December 31, 2020 (SHSP, 2016)

.

Describe the Gap

Previously Maryland SHSP describes a program of strategies to minimize or eliminate traffic-related fatalities and serious injuries on 6 Priority strategies that can significantly reduce roadway fatalities and serious injuries in the six identified SHSP. These are Aggressive Driving, Occupant Protection, Distracted Driving, Highway Infrastructure, Impaired Driving, Pedestrians and Bicyclist

The gap this project will fill is to show that Airbag deployment and seat belts are major factors for fatal/ injury accidents.

Specifically, the life of children is taken away by car crash with related to belt using. For example, in Maryland CNN revealed that five children killed in a single car crash in Feb 2019 because of not properly restrained. Police said all five kids were ejected and killed, however the driver of the vehicle and the passenger in the front seat were both wearing seat belts and survived (CNN, 2019). The point is that belting safe life even in the worst car accident. Therefore, the result from this project open the eyes of stakeholders to act or tighten the use of seat belt. In addition, car manufactures can evaluate themselves with the effectiveness of Airbag deployment of their product. Last but not least, to deliver a working model to predict a model which can help SHSP to forecast accident which helps the organization to take a proactive measure before the accident happen.

Background

Traffic accidents are one of the major reasons contributing to fatalities, injuries and property damage in Maryland. According to the Maryland Statewide Crash Summary, on average, one person was killed every 19 hours, 121 people were injured each day (5 injuries every hour), and 268 police-reported traffic crashes occurred daily (MHSP, 2017). Out of all regions in Maryland, 85% of state's annual crashes were accounted for in the Baltimore and Washington metropolitan regions. In 2013, more than 20,000 crashes occurred in the city of Baltimore alone and 37.8 crashes per mile per year on the I-83 highway Baltimore. That statistic is

very high compared to other Maryland intestates, with a rate of 15.7 (Brando, 2019) crashes per mile per year. With respect to age, accidents among young adult drivers in the age interval of 21 to 29 represent more than one in every five drivers. These groups constitute to 20% of Maryland Crashes, 23% of injuries and 22% of deaths. The first Judicial Highway safety conference was held in 2015 after the increasing of fatal accident. The event was designed to provide evidence-based training and education to assist judges in recognizing how the adjudication of traffic cases impact highway safety.

The goal of MHSO is to reduce the amount of yearly car crashes by implementing statewide traffic safety programs based on the crash data and summary statistics. The MHSO is tasked with the effective and efficient administration of a comprehensive, statewide traffic safety program utilizing

federal funds to reduce traffic crashes and resulting injuries and deaths on Maryland's roads (MHSP, 2017).

Proposed Project

In Maryland the most common reason for fatalities or injuries happen because of lack of using a seat belt and failure or improper deployment of air bags at the time of the accident.

Key Performance indicators

Building a logistic model with an accuracy level of 90 % and obtain highest sensitivity accuracy with a minimum percentage of 85% which is the main goal of catching fatal accident.

Expected Result

The expectation for this project belt usage, air bag deployment and the condition of the person before getting involved in the accident will be the most significant factors. These three variables will be the best predictor variables.

Project Milestones

1. Selecting scopes: 02/17/2019
2. Selecting the dataset: 03/25/2019
3. Cleaning the dataset: 03/10/2019
4. Exploring the dataset: 03/11/2019
5. Testing /evaluating: 03/25/2019
6. Developing the model: 03/30/2019

Comparison of Models

Various models will be developed using SAS EM such as Decision Tree, Logistic Regression (Back ward and Step wise), SVM, Random forest and the comparison of the model will be done by different indicators such as accuracy rate, sensitivity and specificity and ROC. The one with the best accuracy will be selected.

Dataset Description

In this project three dataset have been used and all data sets were taken from DATA.GOV. Please see Dataset Description at Appendix 1

High-Level Data Diagram

The initial exploration in the below bar graph, worth by variables, showed that the first five most important variables for the occurrence fatal/ injury accident are Airbag-Deployed, Person- condition, Accident time, movement code, Alcohol Test, safety- equipment, Damage code, collision. On the contrary, speed limit, weather condition, road type have little contribution. Variable Cluster Node was used see the correlation between interval variables (Distance, latitude, Longitude and speed limit). As it can be seen from the below matrix none of them have strong correlation.

Data Preparation/ Cleaning /Transformation

Three dataset have been used to prepare the merged dataset and all data sets were taken from DATA.GOV.

Crash Dataset

The first dataset is Accident Crashes in Maryland for 2018. The dataset has been taken from the Gov. Data (Maryland Statewide Vehicle crashes: CY 2016). The dataset has 26,974 rows and 37 attributes. Crash dataset is telling about the scenario of the accident, for example, what time, where, the road, the county, brightness of the day, weather condition, Junction code, speed limit, collision type and so on.

Person's information Dataset

The second dataset is Person's information which is about the people who were involved in the accident. The total number of rows were 61,032 rows and 26 attributes. In this dataset some of the attributes are sex of the person,

Alcohol _ Test, Drug_ test, Equip_ wearing (e.g., belts, anchor), Date of Birth, License State-code, and so on. All the data set and description of the datasets are found at the appendix (Appendix 1)

Vehicle Information Dataset

The third dataset is vehicle information with 49,138 rows and 27 variables. In the dataset some of the important variables are the damaged parts model, make, year, Body type, Vehicle ID and so on and some other irrelevant variables such as plate number which was removed as it doesn't help for the analysis.

Population size of Maryland

Population size of Maryland by county was taken from United Census to calculate the number of accidents per 1000 population.