

Individual Term Project

An Analysis of Traffic Trends Using Data from the National Highway Traffic Safety Administration (NHTSA)

Summer 2021

Background

The National Highway Traffic Safety Administration (NHTSA) is part of the United States Department of Transportation and whose mission is to “Save lives, prevent injuries, reduce vehicle-related crashes”. The NHTSA maintains data from the National Center for Statistics and Analysis in its Fatality Analysis Reporting System (FARS). These data are an important resource used for analyzing traffic safety trends and many forms of transportation research. For this project you are to review some of the data maintained in the FARS system, take a deep dive into one area of the traffic safety data, perform descriptive analytics using the SAP Predictive Analytics reporting and visualization tools to help understand the data better, and perform predictive analytics using some of the SAP Predictive Analytics data mining and forecasting tools. The goal is to investigate a particular area of traffic fatalities data, gain some insights from the data, then effectively communicate these findings to the reader of your final report.

Accessing and Assembling your Data

To access the FARS Data Tables go to <https://www-fars.nhtsa.dot.gov/Main/index.aspx>. There, under the **Summary** tab, you will see traffic fatalities data from 1994 to 2019. In addition, you will see tabs for **Trends, Crashes, Vehicles, People, and States**. Spend some time exploring what data is available, particularly noting the subtabs. For example, under **Trends**, you will see subtabs for **Occupants, Nonmotorists, Alcohol**, etc. Also, under the **People** tab, you will see subtabs for **Motocyclists, School Bus Related, Pedestrians, Pedalcyclists**, etc. All of these subtabs provide summary data of fatalities related to various types of traffic-related incidents. The data can be viewed by year and by state or aggregate for the USA.

[NOTE: what you are seeing at this site are various summaries of the full NHTSA dataset which can be accessed if needed. However, digging into the full dataset is beyond the scope of this term project. If you find you need more data and want to search the NHTSA site you are welcome to do so, but that is not required for this project].

Choose an area you find interesting and dive deep into the provided data. For example, perhaps you want to investigate more about trends related to pedalcyclists fatalities or school bus related fatalities. You can download much of this summary data into txt or xls files for various years and for all the USA or certain states. Download the data of interest and assemble it into a format suitable for use in SAP Predictive Analytics. **By now you should understand that this data wrangling activity is both time consuming and important for your later analysis. Make sure you spend adequate time assembling the data into the appropriate format for analysis.** Depending upon the question/issue you are exploring, it might be beneficial to find additional data to supplement your analysis. For example, you could probably find data about the number of cycling enthusiasts in the USA by year or the number of school-aged children in the USA by year. As needed, add this non-NHTSA data to your dataset.

Performing Descriptive Analytics

Using SAP Predictive Analytics explore the data using various charts and tables you learned in Assignment 3. You should experiment with various types of charts and visuals. What trends can you uncover from the data? What insights can you find using the visualization tools provided by SAP? What story can you tell about the data? What charts or tables work best to help you tell your story? You are required to **compose a series of at least six charts/tables** using your data that help tell your story.

Performing Predictive Analytics

Using SAP Predictive Analytics explore the data using various algorithms for data mining and forecasting as you learned in Assignments 4 and 5. What assumptions and methodologies are appropriate for examining your data? Can you make any predictions using your data? Can certain states be clustered together as having similarities in the data? Can you find any associations in the data? What insights can you uncover using these algorithms discussed in class? You are required to **run at least three different data mining or forecasting models** using SAP Predictive Analytics. You should use the visualizations provided by SAP Predictive Analytics to help explain the outcomes of your models.

Compose Your Term Report

Your Final Project Term Report should be formatted in a way that makes the report readable and understandable (include page numbers centered at the bottom of the page) **(5 points)**. Your report should consist of the following items:

- **Title Page** including a descriptive title of your analysis, your name, course name, and term **(5 points)**.
- **One-Page Executive Summary** explaining in overview form the purpose of your analysis and the main outcomes **(10 points)**.
- **Background Section** explaining the area you are studying and the questions you are investigating **(10 points)**.
- **Data Sources Section** explaining what data you used from the NHTSA site and any other sources of data **(40 points)**.
- **Descriptive Analytics Section** with screenshots of your minimum six charts/tables and an explanation of what insights you found in the data; here you should tell your story explaining your insights **(60 points)**.
- **Predictive Analytics Section** where you should explain the motivation, assumptions, and methodologies used for the minimum three data mining/forecasting models you ran, how you ran them using SAP Predictive Analytics, and screenshots of the outcomes from your models; here you should tell your story explaining your insights **(60 points)**.
- **Summary of Findings Section** where you should summarize the main points learned in the analysis **(10 points)**.