Data Analytics Challenge

"Roadway fatalities and the fatality rate declined consistently for 30 years, but progress has stalled over the past decade and went in the wrong direction in 2020 and 2021." – *Dept. of Transportation*

"At 55 miles per hour, sending or reading a text is like driving the length of a football field with your eyes closed." – *CDC*

Learning Objectives

- Understand and summarize the accident dataset.
- Analyze traffic fatalities, explain your choice of algorithms and technologies.
- Develop recommendations (backed up by your analysis) for Arizona traffic safety backed up by your analysis.

Task Overview

You are tasked with providing recommendations for the state of Arizona to improve traffic safety and reduce fatalities. To help you provide recommendations, you are given the following datasets relating to traffic accidents (for Arizona and the nation as a whole) resulting in fatalities. The data was collected by the National Highway Traffic Safety Administration (NHTSA) and spans the years 2016-2020 (data up to 2020 was available at the time of preparing this assignment; new data for 2021 has been released since and is available from the NHTSA website). While Arizona is your focus, you are given data for all states as a matter of reference and understand why certain states have better safety records.

- **Accidents**: crash characteristics and environmental conditions at the time of the crash (one record per crash).
- **Vehicles**: information describing the motor vehicles and the drivers involved in the crash. There is one record per vehicle.
- **Persons**: information describing all people involved in the crash including motorists (i.e., drivers and passengers of motor vehicles in transport) and non-motorists (e.g., pedestrians, cyclists, and occupants of motor vehicles not in transport).
- **Pedestrians and Bicyclists**: information about crashes between motor vehicles and pedestrians, people on personal conveyances, and bicyclists.
- **Driving Impairment**: information about physical impairments of drivers of motor vehicles. There is one record per impairment and there is at least one record for each driver.
- **Distracted Driving**: information about driver distractions. Each vehicle driver has at least one record in this file.

During 2020 there were roughly 5.25 million accidents reported nationwide which led to 1,593,390 injuries and 35,766 fatalities (about 1,000 injuries and 17 fatalities per 100K licensed drivers, respectively). A study released in 2014 estimated that accidents cost our society over $800 billion a year when we factor in workplace and personal losses. According to the IIHS Arizona's fatality rate during 2021 of 16.2 deaths per 100K population was 25.6% higher than the national average of 12.9. Similarly, Arizona had 1.60 deaths for every 100 million vehicle miles traveled which was 16.79% higher than the national average of 1.37. Nationally, about 19% of the fatalities

involved pedestrians or bicyclists, while in Arizona that metric was about 25%. Between 2011 and 2021, Arizona saw a 33% increase in total vehicles, a 17% increase in crashes (from 103,958 to 121,345) and a 41% increase in fatal crashes (from 756 to 1063).

1. *Task 1*: You start with a dataset and the meaning of each column. To better understand the contents of fields you should carefully study the metadata and the contents of each file (e.g., to understand the range of values in each column). Pre-processing the data can help you, e.g., converting the records into Excel or relational database tables (you'll want to understand the primary/foreign key relationships before you join tables and run queries). You can also gather additional data from the NHTSA FARS database from previous years or for other states.
2. *Task 2*: Analyze the data to determine what you consider to be the most significant are the "risk factors" associated with accident fatalities and provide a summary of your analysis. Along with database queries, you are free to use other statistical techniques and visualization tools such as Excel, Tableau, or PowerBI to produce word clouds, charts, graphs, relationship or geospatial maps, etc.
    a. Describe the process of your analysis including the techniques and tools you used.
    b. Justify why you chose the analysis methods you selected.
3. *Task 3*: Provide recommendations to support traffic safety in Arizona. Your recommendations must be supported by your analysis. Note: you can also look at what others (states or countries) have done to successfully improve traffic safety.
4. *Task 4*: Prepare an executive summary of your project.

Note: you are free to use other statistical techniques and visualization tools such as Excel, Tableau, or PowerBI to produce tables, charts, word clouds, diagrams, geospatial maps, etc.

Deliverables

You have two deliverables.

1. A short PowerPoint presentation. Your presentation should have only **4 content slides to address Tasks 1-4 above**. Also include team member information in your presentation.
    - The following material does not count as "content" slides (i.e., does not count towards the 4-slide limit): a cover slide, team member information, outline, references, or other administrative material.
2. A report submitted in Microsoft Word document format. DO NOT use other file formats (e.g., Apple Pages, PDF, Open Office, etc.). Submissions will be checked by TurnItIn.com for originality. The report should be structured as follows.
    - Page 1 – Two sections
        - Section 1: The names of the members of your team
        - Section 2: A one-paragraph executive summary
    - Page 2 – One section (12-point font)
        - A summary of the dataset used for your analysis (Task 1 above).
    - Page 3 – Two sections (12-point font)
        - Task 2a and 2b: Provide an analysis of the risk factors. Include your methodology (process of analysis and tools/techniques you used).
        - Task 3: Provide your recommendations for improving traffic safety and reducing fatalities in Arizona. Include how you arrived at your conclusions.
    - Page 4 – Visualizations. Provide appropriate visualizations of the analysis of your dataset, such as graphing, machine learning and statistical techniques, etc.
    - Page 5 – Appropriately cited references of any external sources used in your analysis (across all sections). Use either the APA or MLA citation style.

Instructions on Accessing the Accident Fatalities Dataset

2. You are given a dataset with Accident Fatalities taken from NHTSA (2016 – 2020). You have zip files containing CSV data (national data) for each year (there are 5 zip files). Arizona data is separately included (extracted from the national data) in its own zip files. For a list of columns in each table and a brief explanation, see the file (on D2L) *Appendix A: Data Tables and Columns*. Also review the NHTSA analytical user manual which has detailed information about each table (on D2L)
3. The data can be downloaded from https://arizona.box.com/s/63ivkvf1qdxjegqbigin0gkqseg4dx7a
   The dataset is made available by NHTSA and more details can be found at https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars

Submission Guidelines

*This is a **group assignment**. You should collaborate as a group and submit ONE deliverable per group.*

Please submit your presentation as a Microsoft PowerPoint document on D2L, under the "MIS Career Management" course > MIS Orientation > Data Analytics Challenge. DO NOT use other file formats (e.g., Apple Keynote, PDF, or Open Office). Submissions will be checked by TurnItIn.com for originality.

Submit your report by the specified date and time on D2L. To submit your report on D2L, please use the following steps.

- Click on the assignment submission-box title.
- Click on the **Add a File** button.
- Find the file you would like to upload. If you are using the most current version of your internet browser, you can drag and drop the file into the target area of the **Add a File** window. You can also use the **Upload** button to browse for and select the file you would like to add.
- Once you have selected your file(s), click the **Add** button in the bottom, left corner of the window.
- Click on the **Submit** button to complete the file submission.
- Upon successful assignment submission, you will receive a confirmation email.
- Verify that the file was submitted by returning to assignment and make sure that you don't see a 0 in the "submissions" column for the assignment. If there is, then you did not successfully submit the file.

For further details on submitting assignments to a D2L submit, please refer to the following website for additional information: https://help.d2l.arizona.edu/content/students-assignmentssubmit

Please note: If you have a problem or question, contact us promptly at wtn@arizona.edu or currim@arizona.edu before the due date.

References

Centers for Disease Control and Prevention. (2022, April 26). Distracted driving. Centers for Disease Control and Prevention. https://www.cdc.gov/transportationsafety/Distracted_Driving/

Joan Lowry, (2014, May 29). Traffic accidents in the U.S. cost $871 billion a year, federal study finds. https://www.pbs.org/newshour/nation/motor-vehicle-crashes-u-s-cost-871-billion-year-federal-study-finds

National Center for Statistics and Analysis. (2022, October). Traffic safety facts 2020: A compilation of motor vehicle crash data (Report No. DOT HS 813 375). NHTSA. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813375

US Department of Transportation, (2023, February 2). Our nation's roadway safety crisis. ArcGIS StoryMaps. https://storymaps.arcgis.com/stories/9e0e6b7397734c1387172bbc0001f29b

Merissa Yellman and Erin Sauber-Schatz. Motor Vehicle Crash Deaths — United States and 28 Other High-Income Countries, 2015 and 2019. MMWR Morbidity and Mortality Weekly Report 2022; 71:837–843. DOI: http://dx.doi.org/10.15585/mmwr.mm7126a1