Kyle Absten

Juan Gutierrez

Ramo Tucakovic

Ranjita Kaur

### 

### 

### **Final Project**

### 

### 

### 

### 

### 

### 

### **Introduction**

In this final project, we got our hands with a hefty dataset from Stanford University's OpenPolicing project, specifically targeting traffic stop data from Nashville, Tennessee. Our goal was to build a model that can predict the outcome of these stops: either a warning or something more serious, like a ticket or arrest. This project was undertaken to create a data-driven solution to predict the outcome of a traffic stop.

### **Selection of Data**

As far as the libraries for our data, we used pandas for data manipulation, along with seaborn and matplotlib for visuals. The source of the dataset is the U.S. Census Bureau. Characteristics of data were things like race, age, sex. There was feature engineering used in the data we pulled from the data related to racial disparities in police stops across the United States. After gathering what we needed from there, we dove into the dataset. It is a big one, with around three million records to explore and utilize. Initially, we were looking at forty-two different features in the dataset, which range from boolean and numeric to object types. However, not everything here we found was gold. The mission was to filter out the noise and focus on what really matters for our prediction model.

### **Methods**

A closer look at the features told us that not all of them were helpful in predicting outcomes. Some, like the raw row numbers or the exact coordinates of the stops, were not going to be useful for our purposes. So, we made the call to drop them. It is all about streamlining the dataset to keep only the features that could actually influence the outcome of a traffic stop.

### **Results**

With the dataset cleaned up, we were leaning towards using a decision tree model. Why? Because our data is mostly categorical, and decision trees are pretty good at handling that type of data. The result was a trained model with our cleaned dataset, it learned to accurately predict whether a traffic stop will end in just a warning or escalate to something more.

### **Discussion**

The plan was to get our hands dirty with model training and validation. We taught our decision tree model to understand patterns in the data that might indicate the outcome of a stop. It is a mix of coding, testing, and tweaking to get it right. All team members were involved in answering the research question.

In essence, this project was about applying what we know about data science and machine learning to make sense of real-world issues. By predicting the outcomes of traffic stops, we're not just flexing our technical muscles; we're also contributing to a larger conversation about law enforcement practices.

### **Summary**

Our study found predictions on, given all the documented information of a traffic stop, the specified decision an officer will make. It's a cool intersection of tech skills and social impact, which is pretty much what we're all about as we wrap up our final year in college. The answer implies officer behavior and can help drivers make better decision and drive more carefully. This matters because a large part of the population drives and information like this can help them learn more related to the law or for their own safety. The perspectives for future research are about how drivers can do better and how officers can do better, as well.

References

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Ramachandran, V., Phillips, C., & Goel, S. (2017, June 18). *A large-scale analysis of racial disparities in police stops across the United States*. arXiv.org. https://arxiv.org/abs/1706.05678

U.S. Census Bureau. “Quickfacts: Nashville-Davidson Metropolitan Government (Balance), Tennessee.” *QuickFacts*, www.census.gov/quickfacts/fact/table/nashvilledavidsonmetropolitangovernmentbalancetennessee/PST045222. Accessed 24 Feb. 2024.