

Analyzing Federal Prosecutor Misconduct

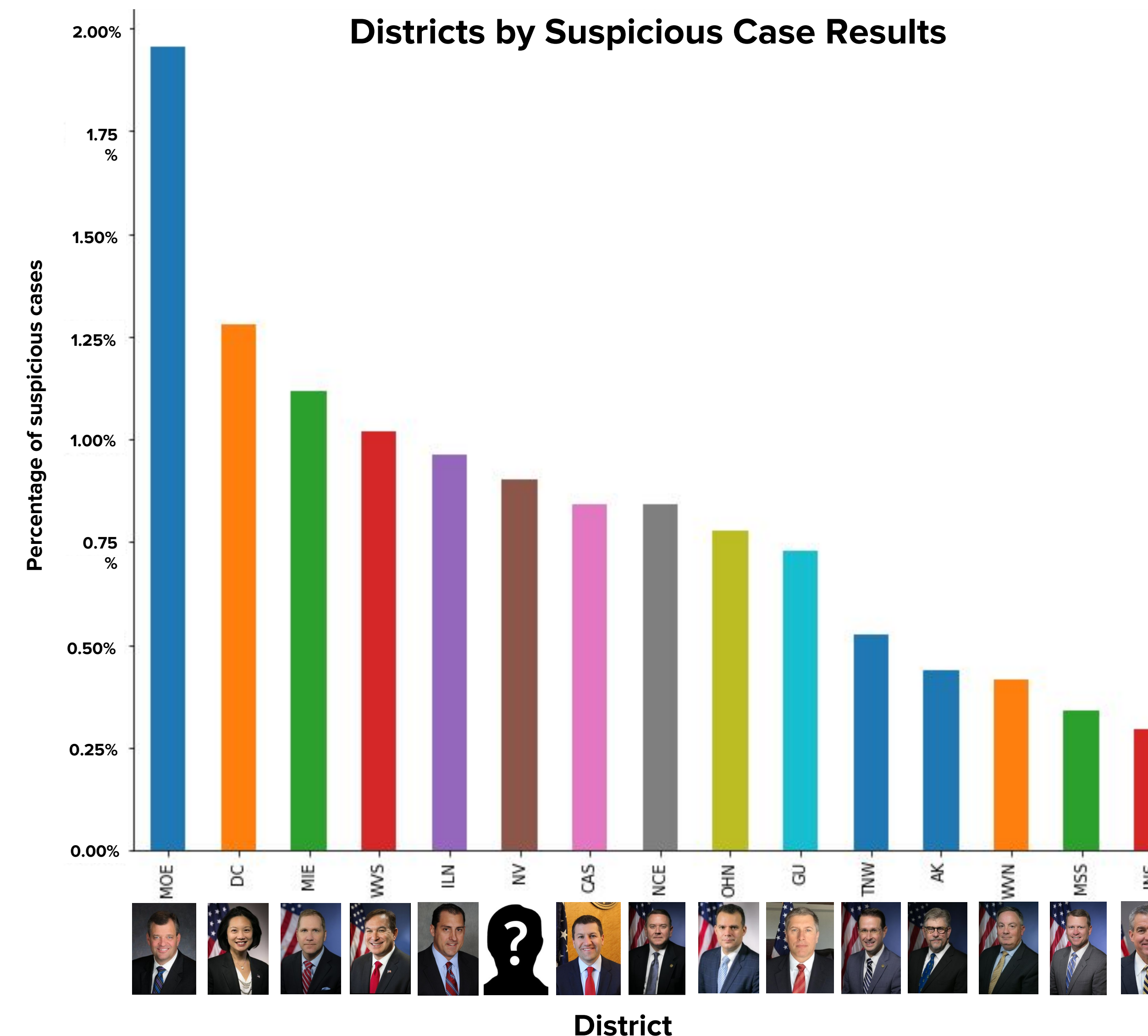
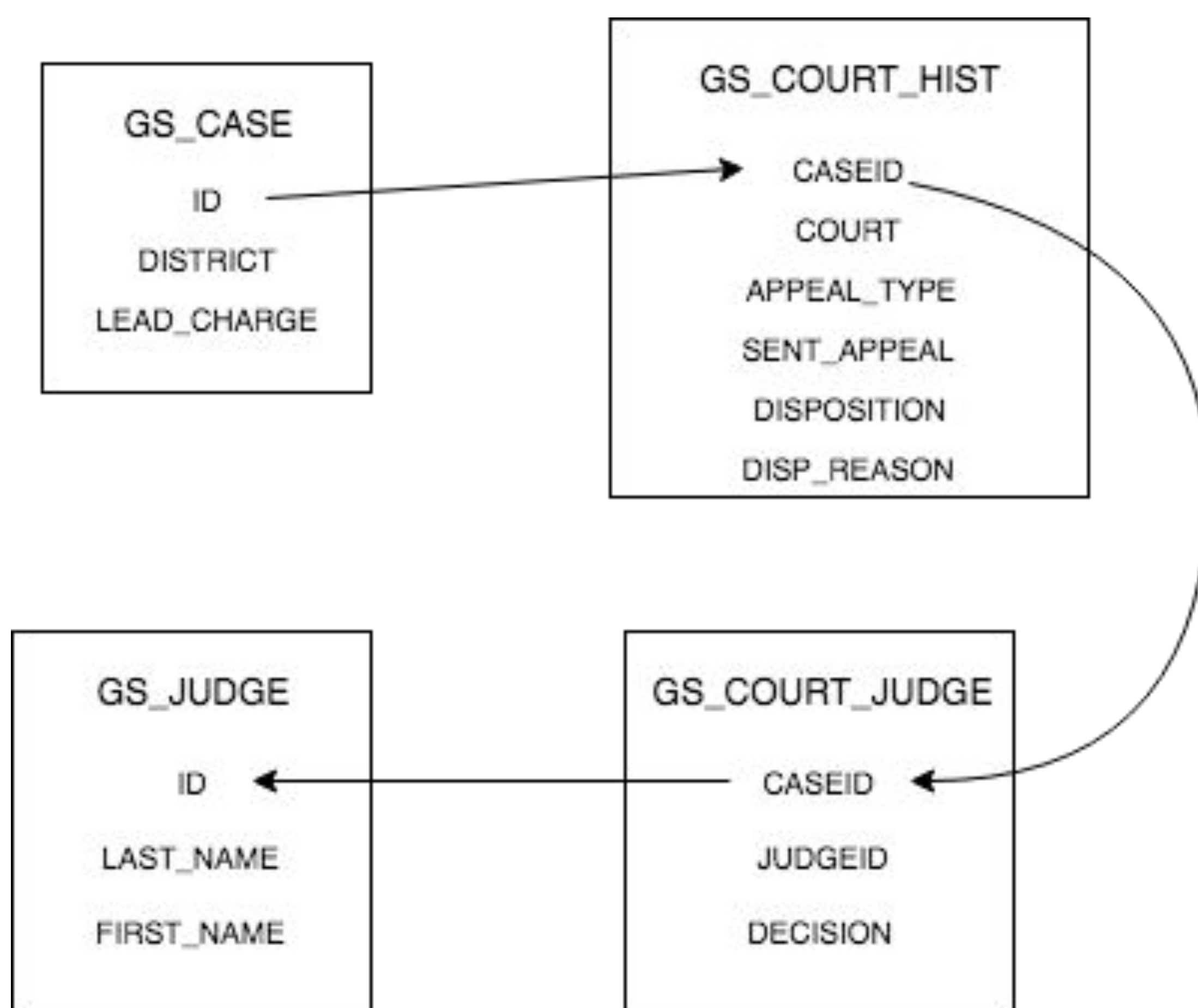
Justin Ingwersen | Hao Chen | Ben Gaudiosi

Introduction

Reporters have heard rumors from their colleagues and the opinions of judicial experts of an epidemic of prosecutorial misconduct. However, the data to back this up is obfuscated behind a complex government dump of files that is difficult for reporters to analyze. This dataset is called the Legal Information Office Network System, or LIONS. In order to figure out the most likely cases that have experienced misconduct, we dug deep into the dataset and devised classifiers to better predict where the epidemic occurs.

Dataset

- Parsed over 100 large (some >5Gb) text files
- Created a SQL database which stored data from text files
- Used pandas dataframes to organize 4 specific tables
 - Dropped insignificant features
 - Applied one hot encoding for categorical features
 - Inner merged the dataframes together
 - Split the dataset based on if the target feature is determined or not for classification purposes

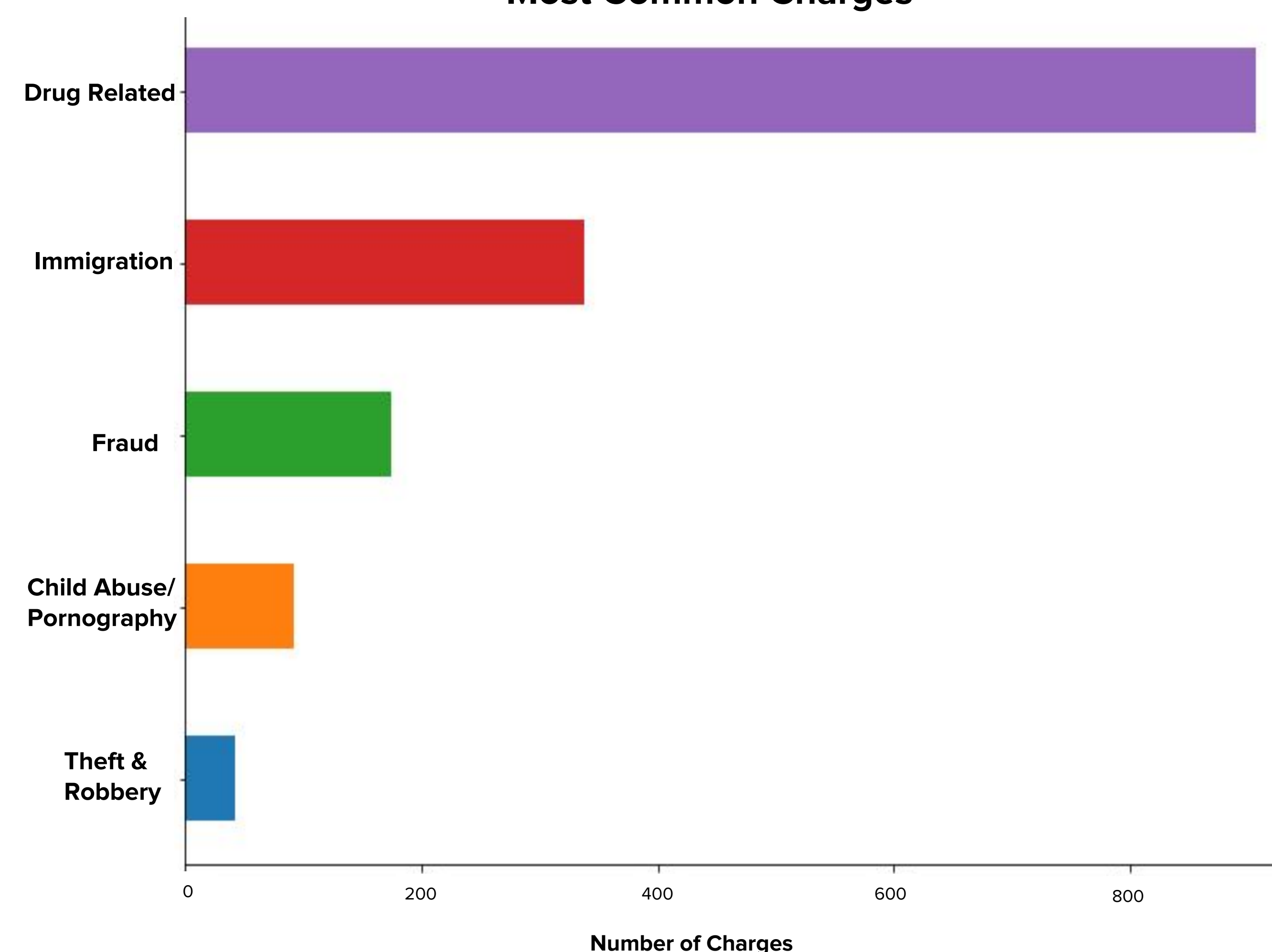


Analysis & Results

- Applied two classification models to find misconduct cases with no features directly pointing to misconduct
- Found the top 15 districts that have the highest percentage of prosecutorial misconducts
- Found the top 20 judges who are related to most prosecutorial misconduct cases
- Found that the top 5 charge categories most prone to suspicious results were drug related, immigration, fraud, child abuse/pornography, and theft/robbery

Our results were unfortunately limited by a severe lack of data. Much of the data in LIONS is redacted. In addition, key fields were missing from the database, including the one that listed the lead prosecutor in individual cases.

Most Common Charges



Conclusion & Future Work

We believe our analysis provides valuable insight into federal prosecutions in the United States and highlights districts that are potentially abusing their authority. However, there is plenty room for improvement. Much of the data from LIONS database is redacted, which severely limited our analysis to only older cases. In addition, though we may have districts where there are prosecutorial issues, it's hard to pinpoint exactly who is responsible, as there are many attorneys working at the offices of each district. Hopefully, our analysis will allow reporters to more efficiently investigate potential prosecutorial misconduct.