Data Science Presentation

Abstract

- Analyzing San Francisco Fire Department Data to find interesting observations to improve efficiency
- Exploring Large Dataset using Various Data Science Techniques
- Exploratory, Statistical, Correlation analysis, and machine learning analysis of the dataset
- Comprehensive research to provide some suggestions and insight into making the Fire Department Calls more efficient

Introduction and Hypothesis

- All fire Department calls from 2001 to present
- 2.5GBs large
- 6 million observations
- 46 starting variables, made an additional 12
- I love SF, and have an interest in medicine
- Interested in a few variables:
  - what effects Response Times
  - Does number of alarms affect FD time
  - What variables are most correlated
  - Finding unique and interesting findings

Title

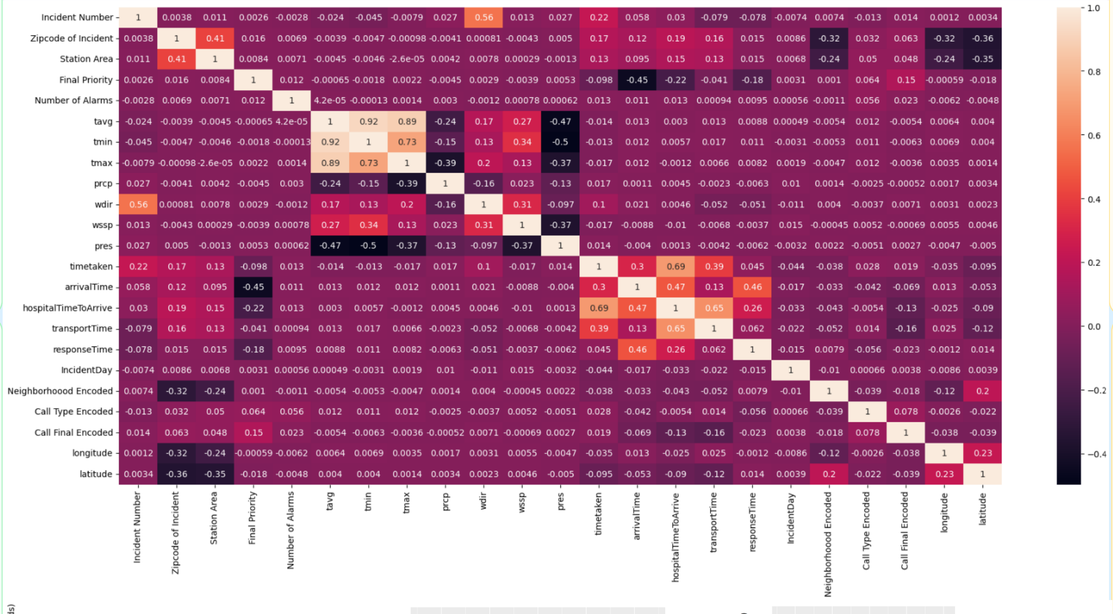
Fire Department Research Project

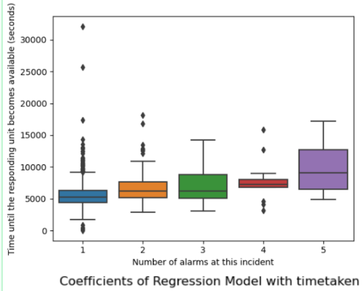
Sean Malloy

COMP-261: Data Science

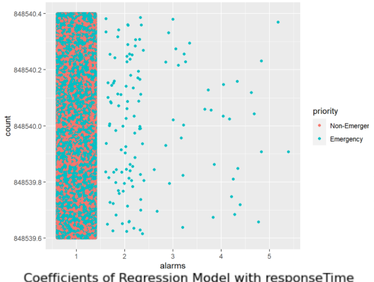
Methods

- Software: R, Python, Jupyter, Tableau, Spark, ScikitLearn, Pandas
- Hardware: Intel Core I7 and 16GBs of RAM
- Machine Learning Methods: Logistic Regression, Linear Regression, Decision Tree Classifiers
- Correlation Testing: Pearson and Spearman Correlation Testing



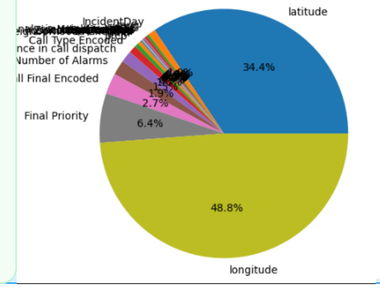


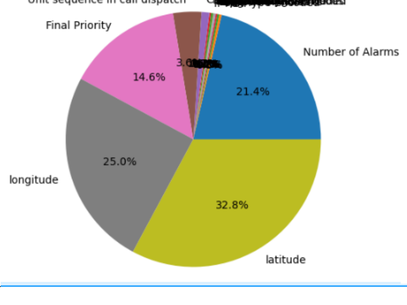
Coefficients of Regression Model with timetaken

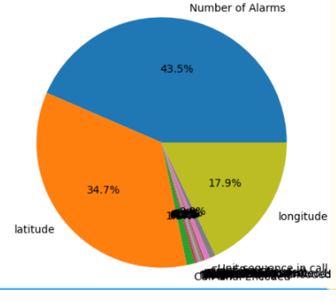


Coefficients of Regression Model with responseTime

Coefficients of Regression Model with transportTime







Results

- Correlation between zipcode of an incident, and the times it takes to respond and handle a incident
- As number of alarms increases at a incident, the more urgent and longer a incident will take
- Location, Number of alarms, and final priority seem to make the largest effects in the machine learning models
- Times are increasing since 2020
- The most common call type is Medical
- Outdoor fires tend to take a lot more time than other types of fires
- Thursdays seems to be less efficient

Conclusions

- Seems that some areas need more resources to be efficient, and some might be underserved or more difficult to address in general
- There is a disturbing trend that over the years since 2020 that times to react to an incident is increasing
- Since medical responses are the most common, it is important to add more resources in ambulances
- More resources could be used to detect outdoor fires, since the response times to react to a outdoor fire is much higher than other types of fires, and needs to be optimized