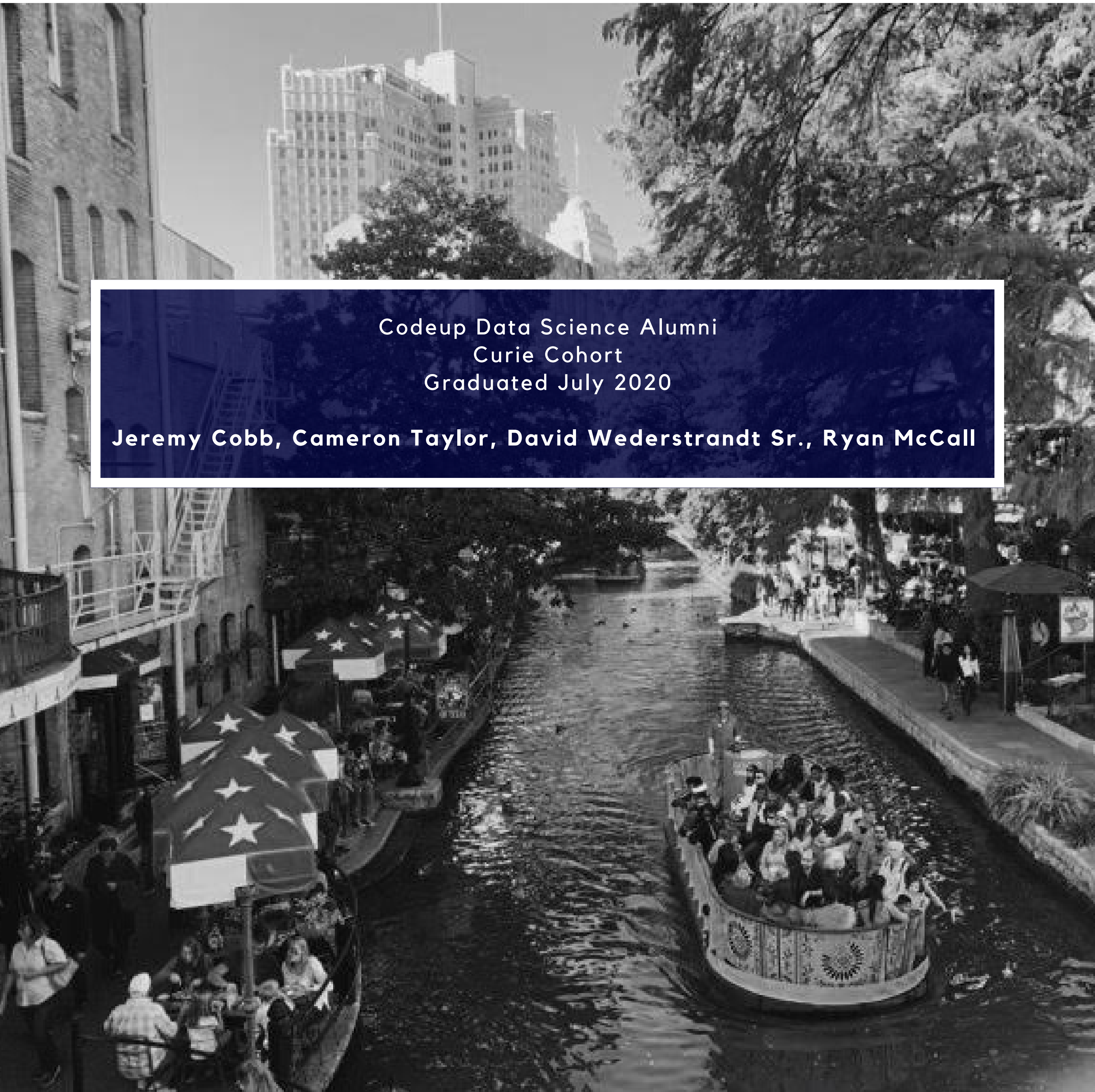
WE CAME, WE SAWS, WE CONQUERED

PREDICTING SANITARY SEWER OVERFLOW EVENTS IN THE CITY OF SAN ANTONIO



Codeup Data Science Alumni
Curie Cohort
Graduated July 2020

Jeremy Cobb, Cameron Taylor, David Wederstrandt Sr., Ryan McCall

INTRODUCTION

San Antonio is currently spending over $300 Million dollars a year in operations and maintenance as well as an extra $200 million in repairs and renewal of their sewer system annually. With our machine learning model, we've created a means to predict what the primary root causes of sewer related damages are and identify what scenarios drive those issues into fruition. We want to use our model to help San Antonio Water System (SAWS) prevent sewer related damages before they occur, as well as help them quickly diagnose these issues when they do occur.

BACKGROUND

Our dataset consists of over 3,000 sanitary sewer overflow (SSO) events from 2009-2019, that we obtained from the city of San Antonio. We joined that with historical weather data that we pulled from the National Oceanic and Atmospheric Administration's API captured from San Antonio's International Airport to create a framework of weather conditions on the dates these SSO events occurred.

EXPLORATION & HYPOTHESIS TESTING

We started by asking if rainfall, temperature, or age of pipe was a driver of certain root causes of SSO events. We ran a t-test to see if the amount of rainfall was significantly related to certain root causes. We found that it was statistically significant for lift stations and rain events. Through exploration, we found that when it rains heavily (>= 10" of daily precipitation) there is a greater chance of lift station failures and failures caused by rain events.

We then used t-tests to see if temperature was significantly related. We did not find any significance in temperature and through exploration discovered this is due to even distribution of temperature for all root causes.

Finally, we ran t-tests to see if the age of the pipe was a driver of certain root causes. We found that age was significant in several cases. First, between 15 and 20 years old, sewers are less likely to fail due to debris, rain, or grease, but there is an increased chance of lift station failure. We also found that structural failures are more likely to occur during the first year of installation or after the 20 year mark.

MACHINE LEARNING

Since determining root causes of failure for SSO events is a classification problem, we began by creating a baseline that used the most common cause, structural failure, as its prediction of every event's root cause. Since structural failure occurred 39% of the time, that model had a 39% accuracy. We then tested out several different models including Decision Tree Classification, Random Forest Classification, and K-Nearest Neighbors. We then picked the best model based on its performance on the test data and found the best model to be the Decision Tree Classifier. This model offered a 41% improvement over the accuracy of the baseline.

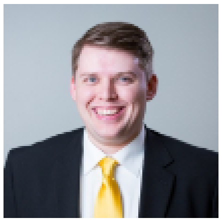
We then used the yes or no questions our decision tree classification machine learning model came up with to create a checklist that determines what events lead up to each SSO event happening. By using this checklist on sewers that have not experienced an SSO event yet, we can prioritize maintenance and upkeep of individual at-risk sewers to help prevent SSO events from occurring, thereby saving our city potentially millions of dollars of fixes and cleanup per year.

RECOMMENDATIONS & CONCLUSION

Moving forward with our project we want to obtain more data- first: using data for every sewer (even the ones that haven't experienced SSO events yet), we could predict future SSO events happening to specific sewers; and second: using location data we could increase the accuracy of our model even further by joining extra data on ZIP codes. We also want to develop a playbook that uses the model output to prioritize maintenance by severity. We believe our model can be used in a multitude of different ways to benefit SAWS, and we are excited about the future of our work and how it will benefit the city of San Antonio.

CLICK HERE TO SEE THE PROJECT REPO:

REACH US

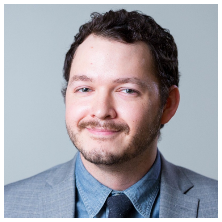


RYAN MCCALL

in/ryan-mccall

/RyanMcCall

Codeup Alumni Page

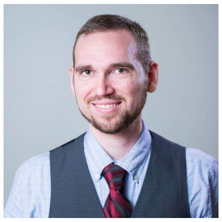


JEREMY COBB

in/jeremybcobb

/jivemachine

Codeup Alumni Page

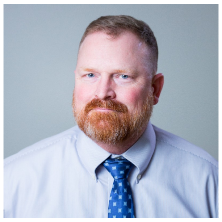


CAMERON TAYLOR

in/camerongtaylor

/CameronGTaylor

Codeup Alumni Page



DAVID WEDERSTRANDT SR.

in/davidwederstrandtsr

/davidwederstrandtsr

Codeup Alumni Page