Niki Jafari (nikijafari@cmail.carleton.ca)
School of Public Policy and Administration

Yujing Yang (yujingyang@cmail.carleton.ca)
Sprott School of Business

Mahitha Sangem (mahithasangem@cmail.carleton.ca)
Department of Systems and Computer Engineering

Carleton University

PREDICTING WATERMAIN BREAKS IN THE CITY OF OTTAWA

INTRODUCTION

Watermains refer to the pipes that distribute water throughout Ottawa. When a watermain break occurs, there is potential for a significant loss of water and associated service interruptions. This has financial and environmental ramifications, as well as prevents the City from delivering safe drinking water to the citizens of Ottawa. This research aims to identify the factors that contribute to watermain breaks, to predict where the City of Ottawa will need to conduct condition assessments.

METHODOLOGY

The datasets were cleaned using SQL. A statistical analysis was performed, including regressions, hypothesis tests like Chi-squared test to determine the relationship between features. The analysis was conducted on 20 years of data provided by the City of Ottawa, focusing on pipe material, pipe diameter, and the time of watermain breaks. A machine learning model was trained to forecast the frost depth based on temperature. Data visualization of the historical repair record using Tableau was performed.

ACKNOWLEDGEMENTS

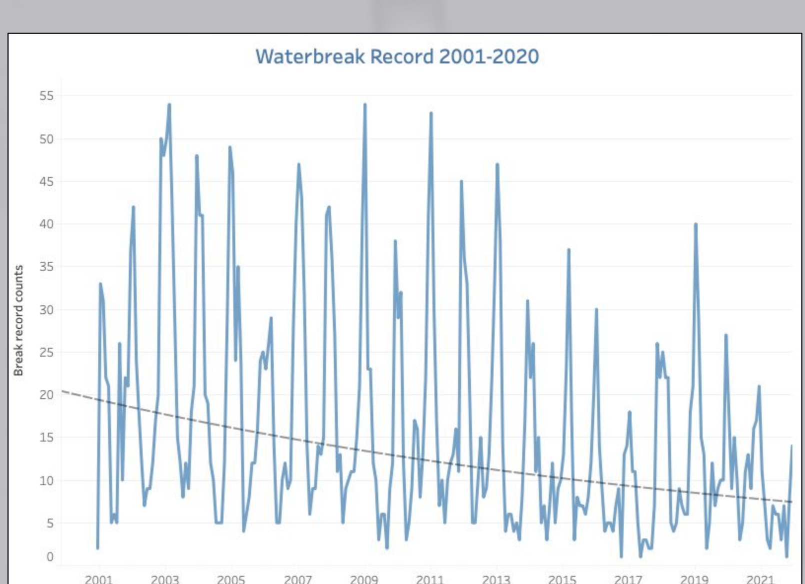
We would like to thank our DATA5000 Professor Olga Baysal for her guidance, as well as the City of Ottawa for sharing their data and ambitions with us.

DATASETS AND LIMITATIONS

This research uses two datasets provided by the City of Ottawa.First dataset consists of data from 2001 to 2022 of the watermain repairs that have occurred in Ottawa, with information regarding the time and location of the watermain breaks, and the pipe material and diameter of the watermain pipes. Second dataset is a frost model from 2011 to 2022 with records of frost depth and temperature in Ottawa (for the months of November to March). The limitation of the datasets is that it lacks variables such as the age of the watermain pipes, depth of installation, and the temperatures of soil and water, which would contribute to a more accurate prediction of watermain breaks. With additional variables, including data on watermains that did not break, models can be trained to predict the occurrence of watermain breaks in future.

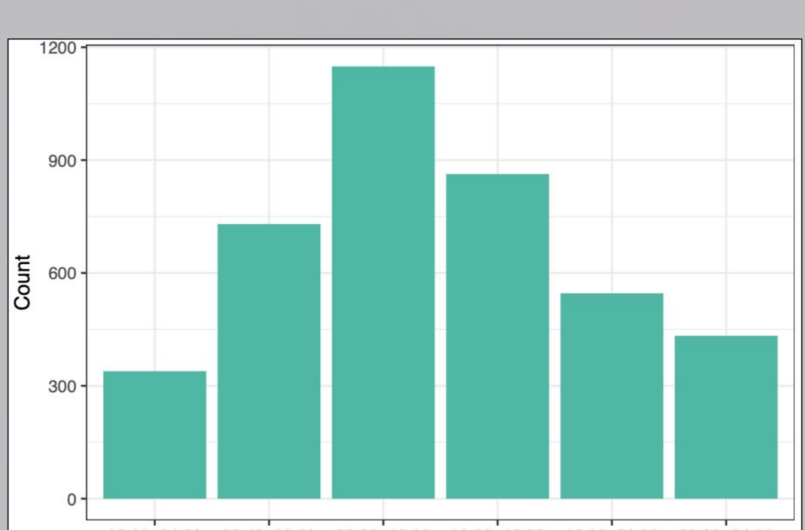
RESULTS

Watermain Break Record



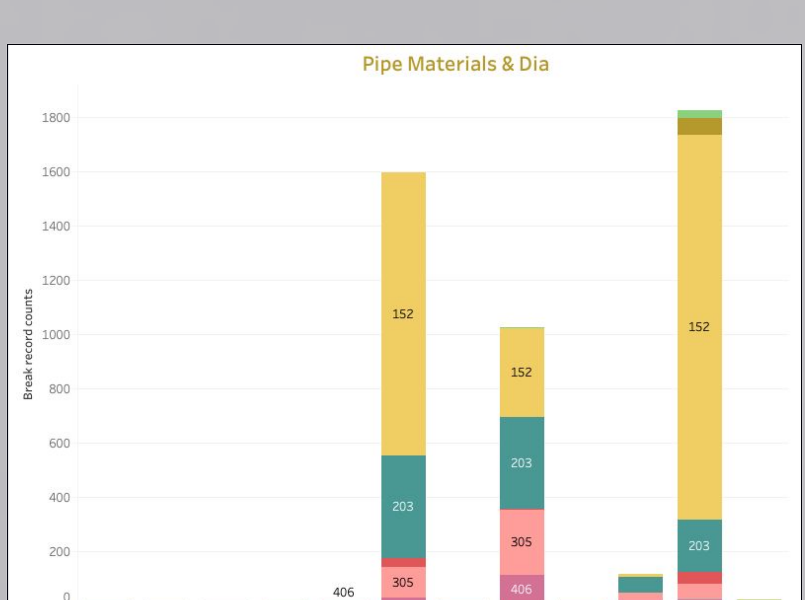
The watermain break record from 2001 to 2020 indicates a slightly downward facing trend line, demonstrating that watermain breaks are decreasing overall.

Time Analysis

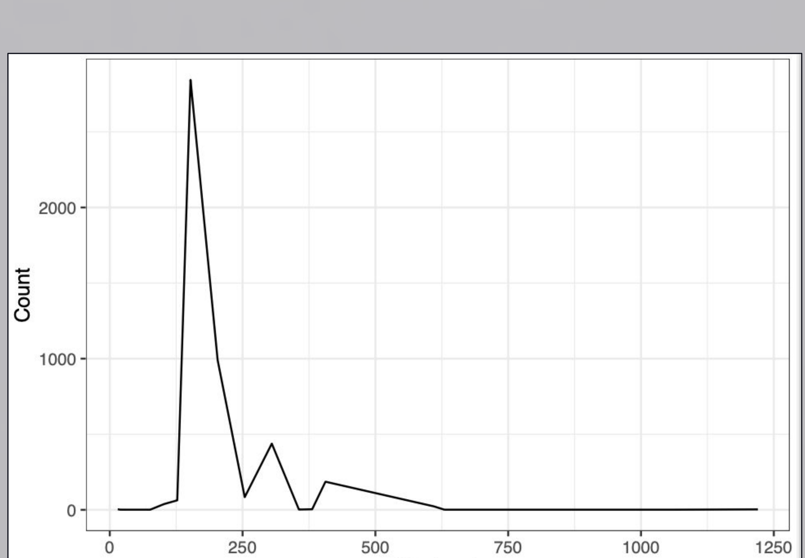


Our analysis demonstrates that there is a significant relationship between watermain breaks and the time of day. The 8:00 to 12:00 time period has the most occurrences of watermain breaks, followed by the 12:00 to 16:00 period. A Chi-squared test has been performed that confimed this hypothesis.This may be related to the periods when they are more frequently used.

Pipe Material and Diameter Analysis



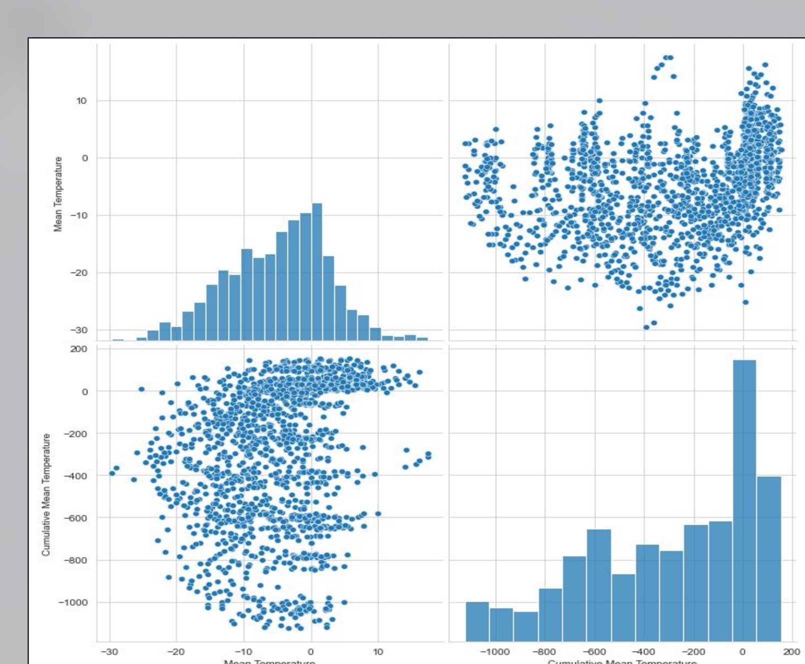
When examining pipe material, our analysis indicates that CI, DI, and UCI have the highest break rate (or that they are the most frequently used material). Moreover, the 152 mm, 203 mm, and 305 mm pipe diameters are the most frequently used diameters. Among these three diameters, the widest pipe diameter has the lowest break rate.



The frequency of watermain breaks is the highest when the pipe diameter is 152 mm, followed by 203mm. A regression analysis was performed that confirmed the relationship. Despite this, it was discovered that the relationship between watermain breaks and pipe thickness was not significant. Therefore, a smaller pipe diameter is not correlated with an increased number of watermain breaks.

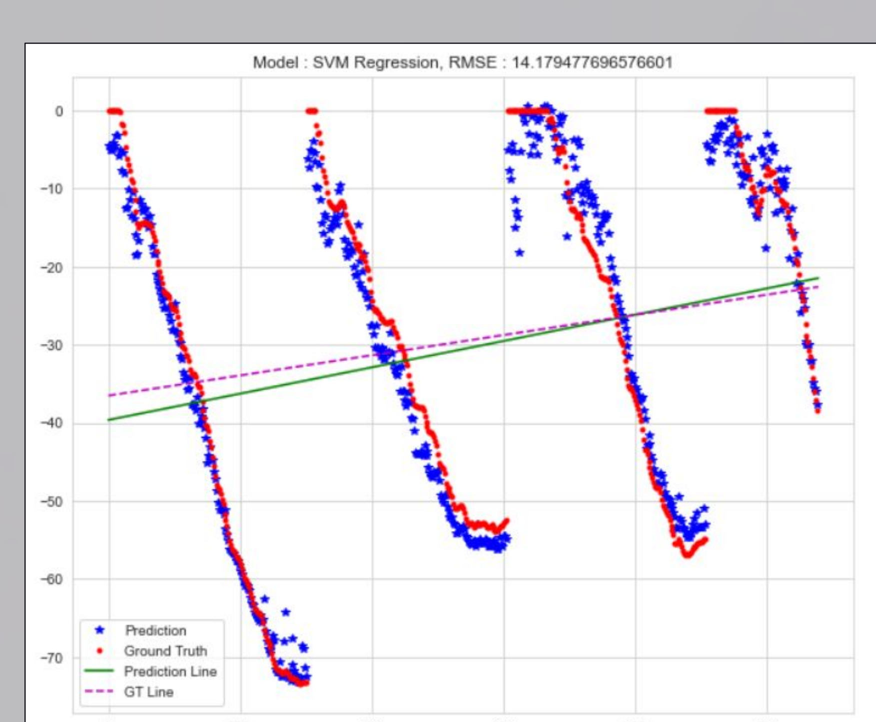
Frost Depth Analysis and Prediction

- The data consists of temperature and frost depth from 2011 to 2022, from November to March.
- Analysis on the features “Mean Temperature” and “Cumulative Mean Temperature” was conducted for the prediction of frost depth, after data framing and analyzing the correlation of features.
- The training and testing data have been divided as follows: November 2011 to March 2018 – Training. November 2018 to January 2022 – Testing (to test the prediction of the model with actual values).
- Several regression algorithms have been implemented for modeling to predict the frost depth based on the temperature change data.



Their performance metrics are as follows:

- SVM Regressor is optimal as the errors are minimal, and the R2-score (how well the regression model fits the actual/observed data) is the maximum comparatively.
- The graph for SVM Regressor prediction. Ground truth (of testing data) is closest to the prediction for SVM Regressor.



	Regressor	R2-score
0	Linear Regression	0.96
1	KNN Regression	0.97
2	SVM Regression	0.97
3	Decision Tree Regression	0.96
4	Random Forest Regression	0.97
5	GradientBoostingRegressor	0.97

	Regressor	Mean Absolute Error	RMSE
0	Linear Regression	3.35	18.11
1	KNN Regression	3.22	15.56
2	SVM Regression	3.02	14.18
3	Decision Tree Regression	3.45	21.67
4	Random Forest Regression	3.15	15.97
5	GradientBoostingRegressor	3.13	14.71

CONCLUSION

Identifying pipes that are more likely to be at-risk of breaking based on our analysis will help prevent watermain breaks in Ottawa. Predicting watermain breaks through our analysis will enable the City of Ottawa to allocate funds to maintain, rehabilitate, and replace at-risk pipes. The City of Ottawa can also use cameras and other tools to closely monitor at-risk pipes and avoid major failures through the condition assessment program.