Predicting Watermain Breaks in the City of Ottawa

Yujing Yang
Niki Jafari
Mahitha Sangem

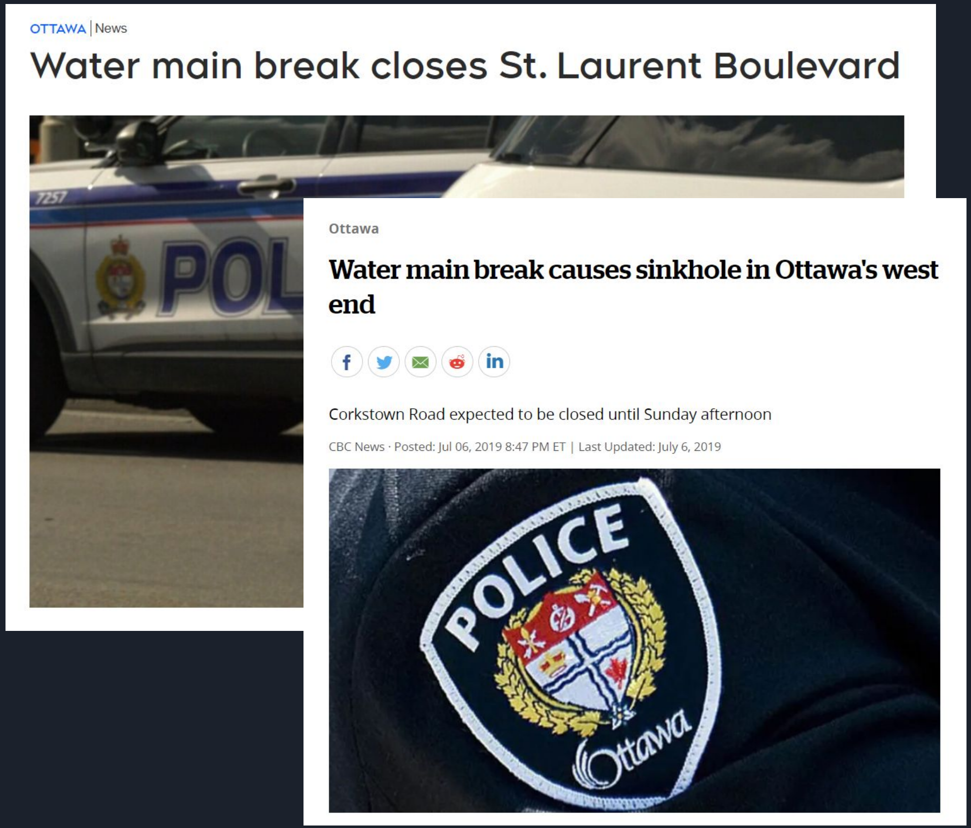
DATA 5000 - Professor Olga Baysal

Introduction

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.

Research Questions:

- What factors contribute to watermain breaks in the City of Ottawa?
- How can we predict watermain breaks?



Dataset

- Dataset 1: Data from 2001 - 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.

- Dataset 2: Frost model from 2011 - 2022 of the frost depth and temperature in Ottawa (for the months from November - March).

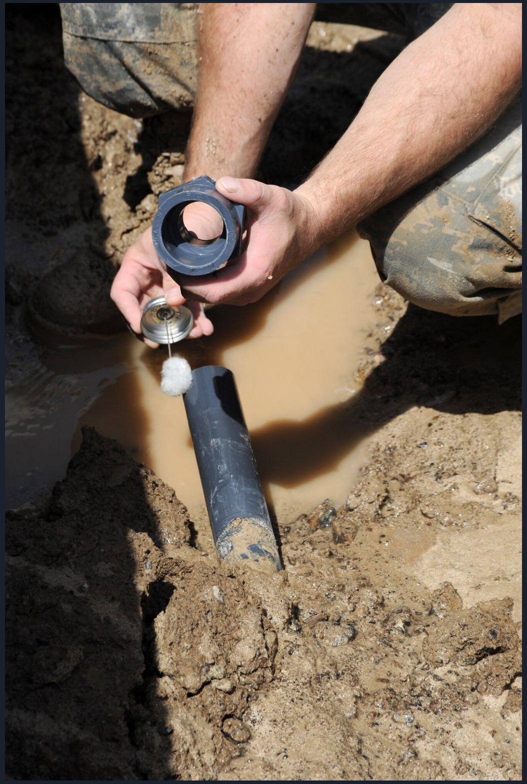
Methodology

Data visualization - Tableau.

Statistical analysis - regressions, hypothesis tests like Chi-squared test focusing on:

- Pipe Material
- Pipe Diameter
- Time of Watermain Break

Machine learning model - forecast the frost depth based on temperature.

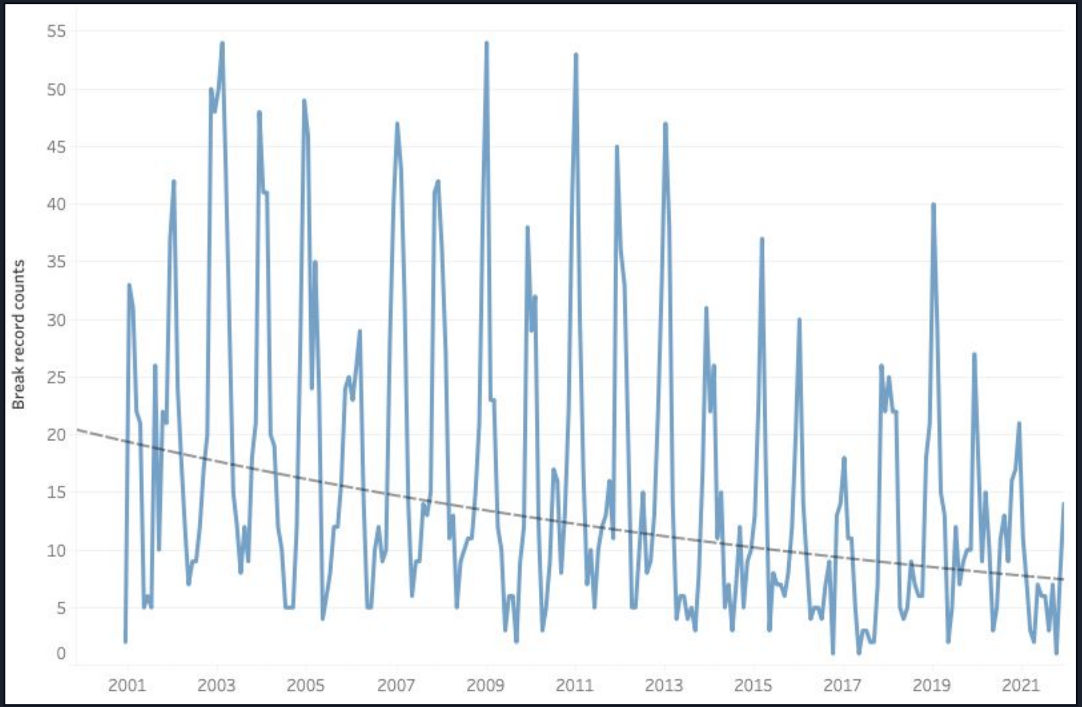


Anticipated Results

We predicted that:

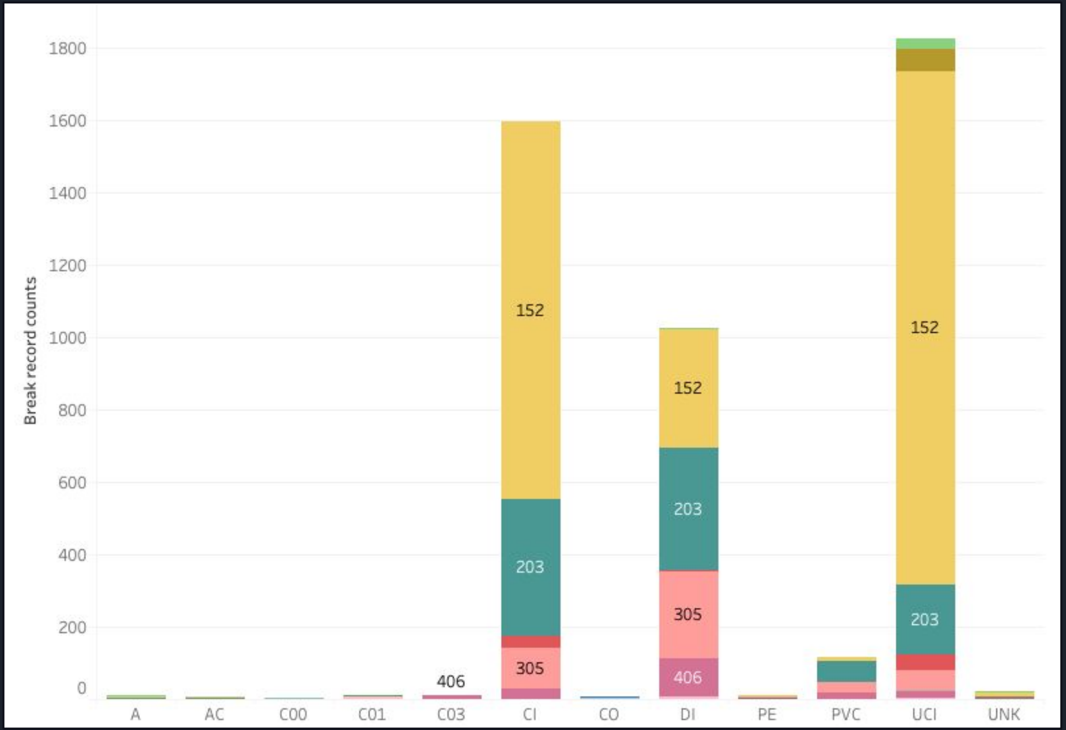
- The relationship between the pipe material and watermain breaks is highly correlated, as well as the pipe diameter.
- There is a certain time of day that watermain breaks will occur more frequently than others, likely in the mornings.
- A frost model to predict the frost depth based on the temperature change on data will be provided.

Watermain Break Record (2001 - 2020)



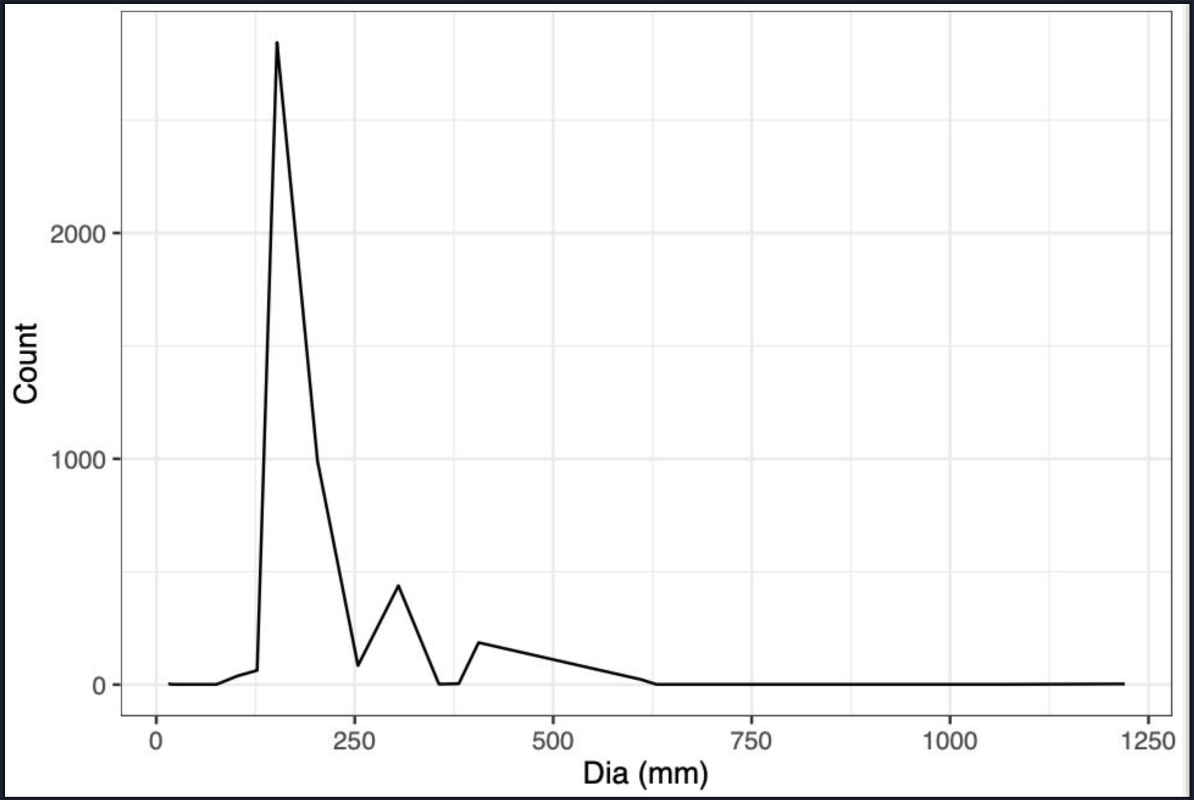
- The line graph indicates a slightly downward facing trend line
- This demonstrates that watermain breaks are decreasing overall.

Pipe Materials and Diameter



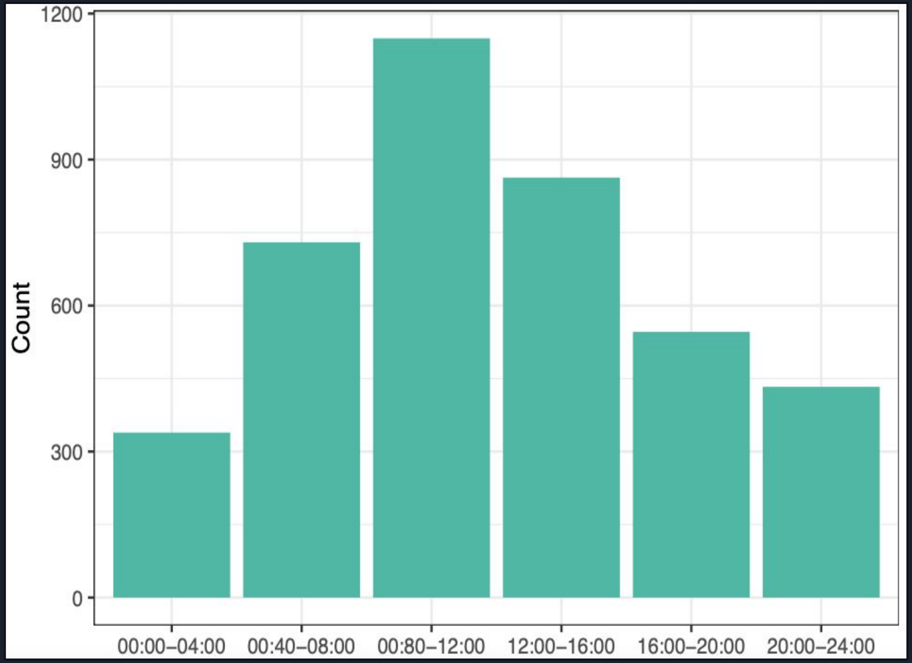
- Our analysis indicates that CI, DI, and UCI have the highest break rates.
- The 152 mm, 203 mm, and 305 mm pipe diameters are the most frequently used diameters.
- Among these three diameters, the widest pipe diameter has the lowest break rate.

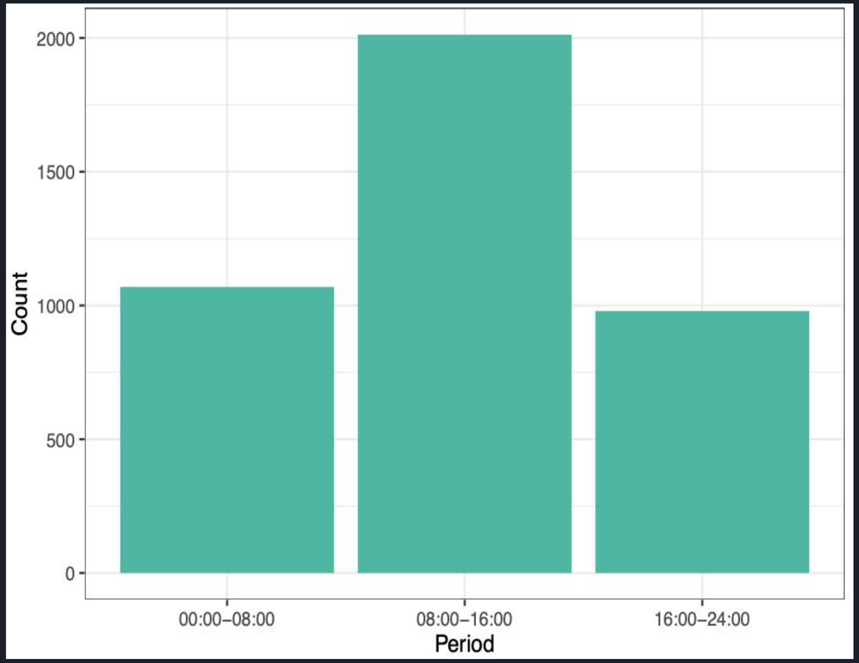
Watermain Pipe Diameter



- The frequency of watermain breaks is the highest when the pipe diameter is 152 mm, followed by 203mm.
- A regression analysis confirmed the relationship.
- The relationship between watermain breaks and pipe thickness was not significant.
- A smaller pipe diameter is not correlated with an increased number of watermain breaks.

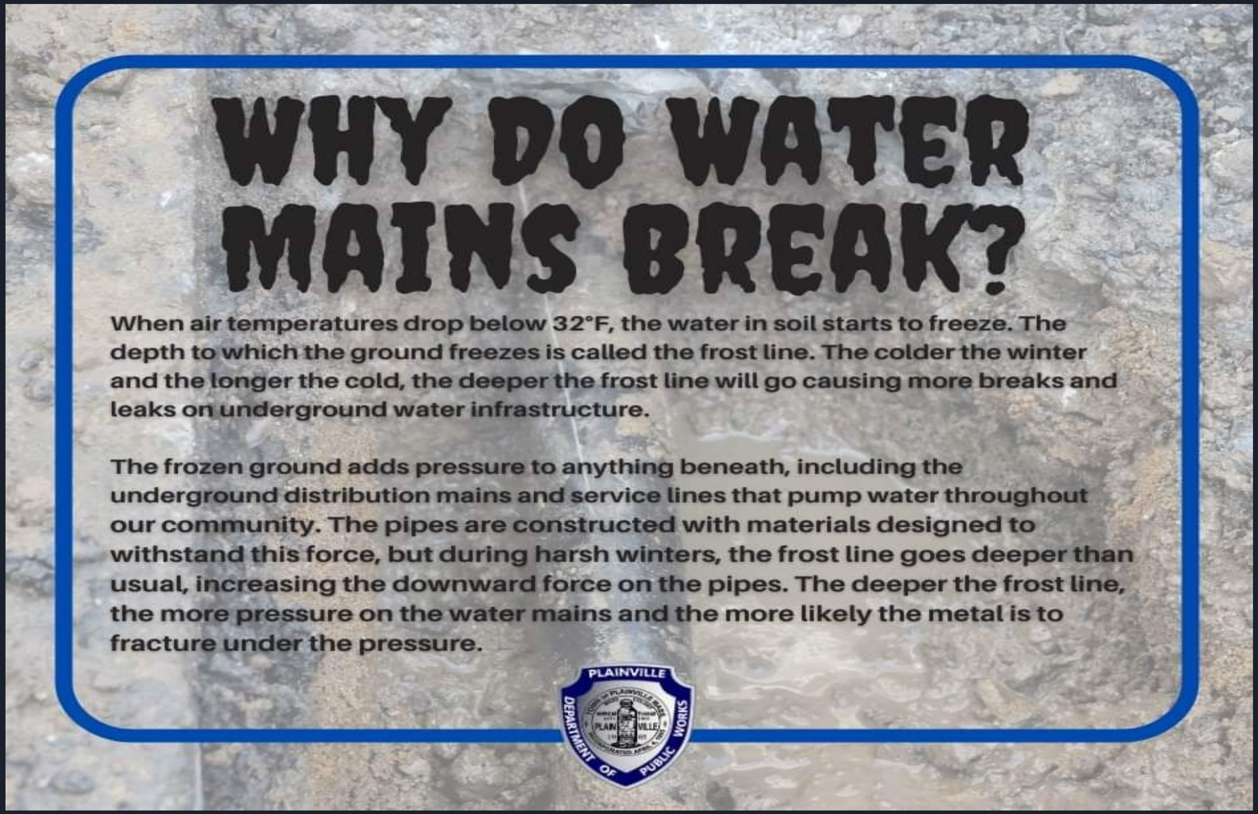
Time Analysis



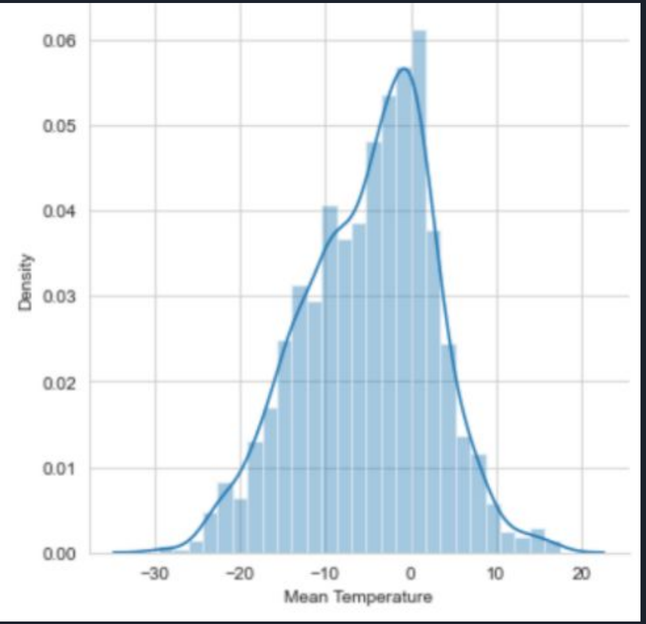


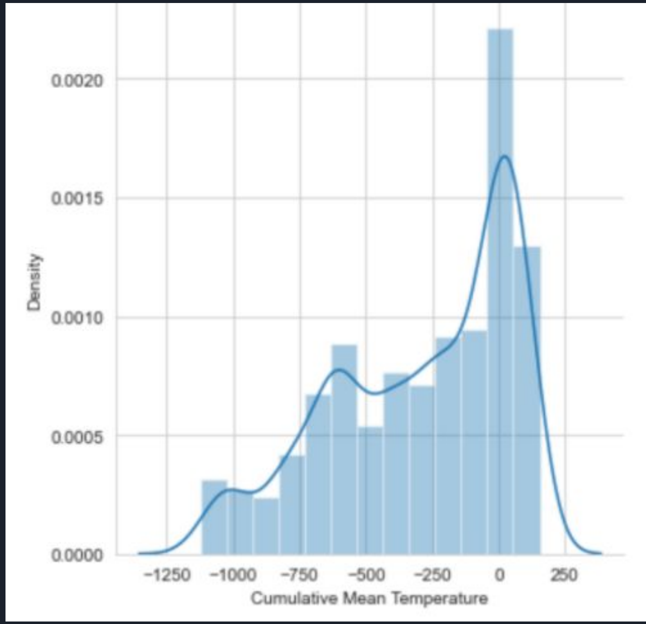
- Significant relationship found between watermain breaks and the time of day.
- The 8:00 to 12:00 time period has the most occurrences of watermain breaks, followed by the 12:00 to 16:00 period.
- A Chi-squared test has been performed that confirmed this hypothesis.
- This may be related to the periods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

Frost Affecting Watermains



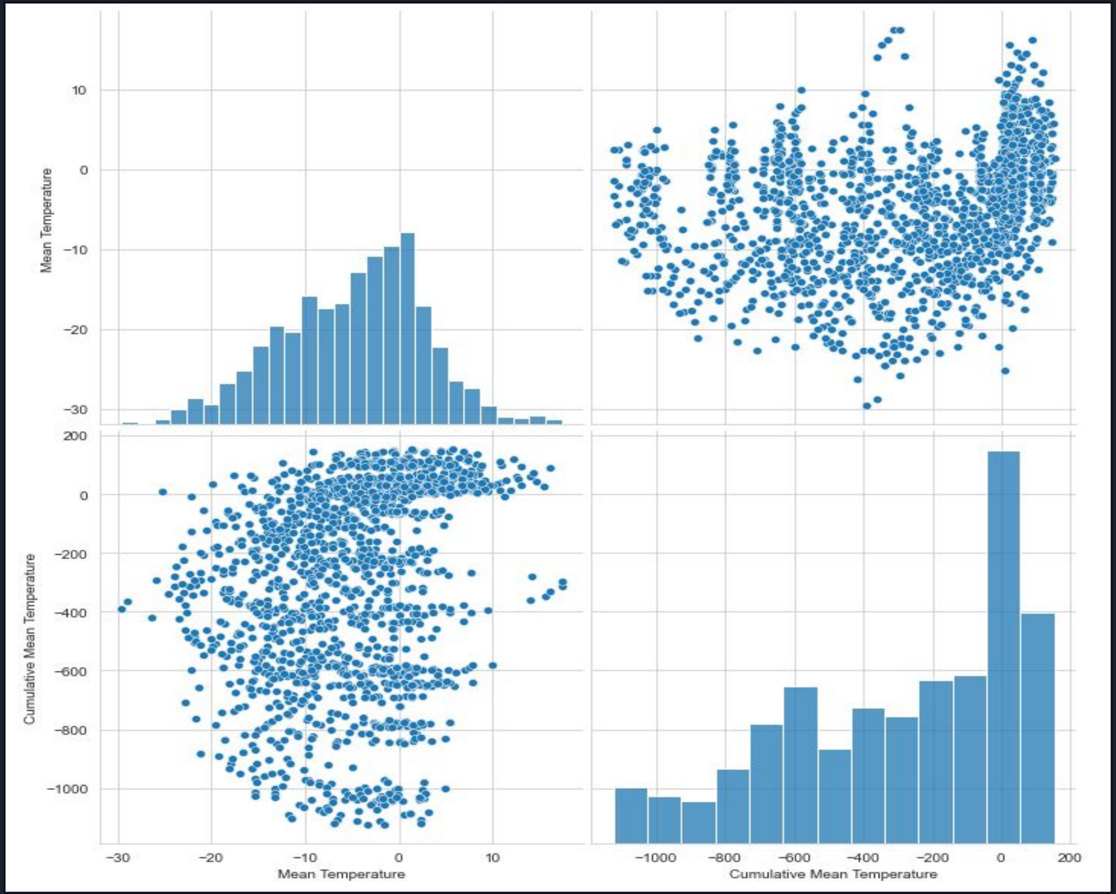
Frost Depth Prediction and Analysis (I)





- Data: temperature and frost depth from 2011 to 2022, from November to March.
- Data framing and analyzing the correlation of features.
- Analysis on the features “Mean Temperature” and “Cumulative Mean Temperature”

Frost Depth Prediction and Analysis (II)



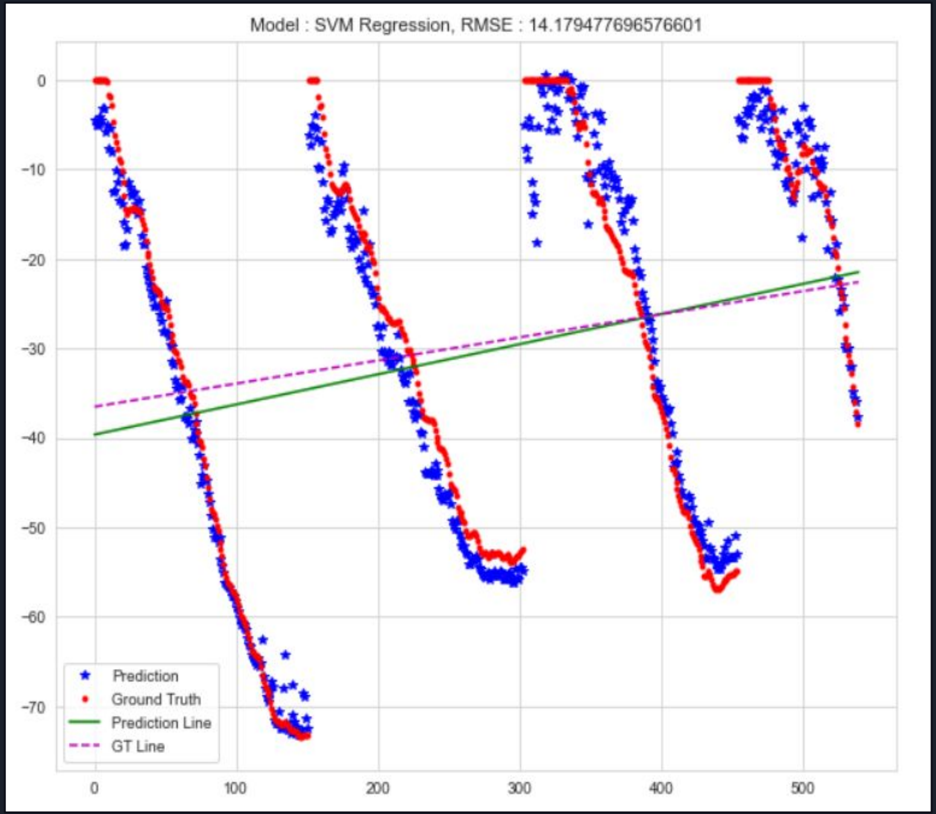
- The training and testing data have been divided as follows:

  - November 2011 to March 2018 – Training.
  - November 2018 to January 2022 – Testing
- Scaling was performed.
- Several regression algorithms have been implemented for modeling to predict the frost depth based on the temperature change data.

Frost Depth Prediction and Analysis (III)

	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



Their performance metrics:

- SVM Regressor (Support Vector Machine) is optimal as the errors are minimal.
- R2-score is the maximum comparatively.
- SVM Regressor- Ground truth (of testing data) is closest to the prediction

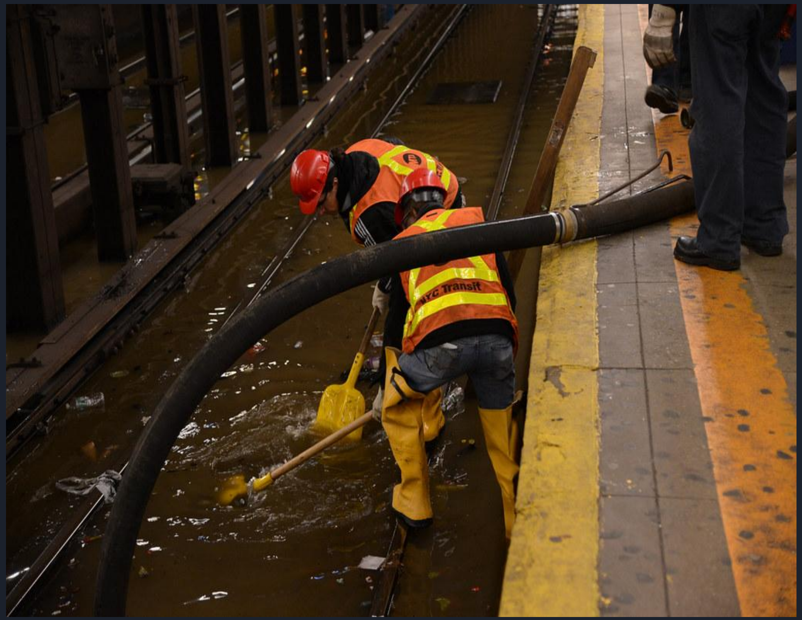
Implications

- This project is essential because it is associated with Ottawa citizens' quality of life.
- Watermain breaks lead to service interruptions, traffic disruptions, and and environmental and financial costs

Conclusion

Identifying pipes that are more likely to be at-risk of breaking based on our analysis will help prevent watermain breaks in Ottawa.

The City of Ottawa will be able to maintain, rehabilitate, and replace at-risk pipes.



Thank you for your attention :)

Team Contributions

Yujing Yang - Statistical Analysis of Factors Contributing to Watermain Breaks

Niki Jafari - Poster, Presentation, Background Research

Mahitha Sangem - Frost Model Prediction and Analysis