**Predicting Structural Conditions of Sewers based on Related Pipe Properties**

Springboard Data Science Intensive Capstone Project

# Objectives

Understand patterns of structural conditions of sewers in the Greater Toronto Region and develop a recommendation system to identify pipe properties that would increase the likelihood of poor structural conditions

# Data Provided

* 4 years of sewer structural condition data and their respective repair recommendations (2,362,347 m of sewer surveyed)
* Pipe Properties: age, length, size, depth below ground, sewer use (sanitary, combined, stormwater), location type (highway, residential, airport, building, parking lot, railway, wood, yards)
* Structural Conditions: NASSCO PACP pipe structural grading (1-5, 5 being failed)

# Previous Studies Findings

* Structural condition of the pipes and various pipe properties: age, length, size, depth, material, shape and surrounding soil type
  + Methods
    - Used simple linear regression, one-way analysis of variance (ANOVA) and Student’s T-test
  + Findings
    - Correlations by statistical analysis does not prove a causation
    - Simple linear regression, Pearson’s coefficient: structural condition and age, length, size, and depth are 0.29, 0.10, -0.10, and -0.12, respectively
    - One-way ANOVA of pipe materials, asbestos cement is found to be the best pipe material
    - Pipe shape did not have a statistical significant effect on the structural condition
    - Pipes in clay soil performed better than pipes in sandy soils
* Correlation analysis between soil type and sewer conditions in the City of Toronto
  + Methods:
    - Hypothesis test used to select the sewer condition codes that are representative of the severe sewer conditions (structural, infiltration, and joint issues)
    - Spatial analysis (weighted sum method in ArcGIS) of statistical analysis results (simplified soil data from borehole logs and sewer conditions)
  + Findings
    - Correlation exist between some soil types and sewer conditions
    - Structural issues are more likely to be observed in fill, shale, and bedrock, and less likely in silt; infiltration issues are more likely to be observed in sand, fill, and bedrock, and less likely in till and silt; joint issues are more likely to be observed in sand and shale
    - Results only indicated correlation, but causation could not be proved