



# Water access analysis

Using Machine Learning to predict water stress.  
[water.xavieruxchamp.dev](http://water.xavieruxchamp.dev)

# Overview

- Water is becoming scarce in many communities across Canada. The current situation warrants a detailed analysis of the available data
- Problem statement: pollutants present on river basins across Canada are affecting drinking water supplies and our food security
- Opportunity: the right machine learning solution could become a cleantech startup

# Vision

- Using machine learning, we could predict the long term effects of different stressors contributing to the depletion of our water supply
- By targeting different segments of society, the solutions become increasingly attractive to business stakeholders
- And, as more models are trained and new solutions become available, more individuals could benefit

# Impact

- Corporations could optimize their consumption
  - Water intensive processes could be replaced
- Non-profits could allocate resources to mitigate human impact
  - NGO's could better prepare for drinking water shortages
- Governments could plan infrastructure problems more efficiently
  - Water treatment or desalination plants could be planned and built sooner

# Data Dictionary

Variable	Column Name	Description
SAMPLE DATE AND TIME	DATE_TIME_HEURE	SAMPLE DATE AND TIME IN DD/MM/YYYY HH:MM FORMAT USING LOCAL TIME ZONE
DATUM	DATUM	DATUM USED FOR LATITUDE AND LONGITUDE
MEASUREMENT FLAG	FLAG_MARQUEUR	FLAG APPLICABLE TO MEASUREMENT VALUE (<-LESS THAN, T-TRACE VALUE REPORTED BELOW DETECTION LIMIT BY LABORATORY)
LATITUDE	LATITUDE	LATITUDE IN DECIMAL DEGREES FOR THE NOTED DATUM
LONGITUDE	LONGITUDE	LONGITUDE IN DECIMAL DEGREES FOR THE NOTED DATUM
METHOD DETECTION LIMIT	MDL_LDM	METHOD DETECTION LIMIT REPORTED BY THE LABORATORY FOR THE NOTED METHOD
METHOD CODE	METHOD_CODE	NUMERIC ANALYTICAL METHOD CODE FROM THE NATIONAL VMV DICTIONARY
SHORT METHOD TITLE	METHOD_TITLE	SHORT METHOD TITLE FROM THE NATIONAL VMV DICTIONARY. FULL METHOD DESCRIPTIONS AVAILABLE UPON REQUEST AT EC.MSQEINFORMATION-
OCEAN DRAINAGE AREA	OCEANDA	OCEAN DRAINAGE AREA THAT MONITORING SITE IS LOCATED IN
PEARSE DRAINAGE AREA	PEARSEDA	PEARSE DRAINAGE AREA THAT MONITORING SITE IS LOCATED IN
PROVINCE/TERRITORY	PROV_TERR	PROVINCE OR TERRITORY THAT MONITORING SITE IS LOCATED IN
SAMPLE IDENTIFIER	SAMPLE_ID_ÉCHANTILLON	UNIQUE SAMPLE IDENTIFIER FROM MONITORING PROGRAM
SAMPLE DETECTION LIMIT	SDL_LDE	SAMPLE DETECTION LIMIT REPORTED BY THE LABORATORY FOR THE SPECIFIC ANALYSIS OF THE REPORTED MEASUREMENT. ELEVATED SAMPLE DETECTION LIMIT ABOVE METHOD DETECTION LIMITS MAY INDICATE DILUTION, MATRIX INTERFERENCE OR OTHER ISSUES WITH THE
SITE DESCRIPTION	SITE_DESC	SITE DESCRIPTION THAT MAY INCLUDE SPECIFIC SAMPLING LOCATION, RATIONALE FOR MONITORING LOCATION (E.G. REFERENCE, UPSTREAM/DOWNSTREAM, ETC.), AND LOCAL
SITE NAME	SITE_NAME	FULL SITE NAME FROM MONITORING PROGRAM
SITE NUMBER	SITE_NO	UNIQUE SITE NUMBER/IDENTIFICATION FROM MONITORING PROGRAM
SITE TYPE	SITE_TYPE	SAMPLE SITE TYPE (E.G. RIVER, LAKE, ETC.)
VALIDATION STATUS	STATUS_STATUT	LEVEL OF VALIDATION (P-PROVISIONAL, V-VALIDATED)
UNIT NAME	UNIT_NAME	FULL MEASUREMENT UNIT NAME FROM NATIONAL VMV DICTIONARY
UNIT	UNIT_UNITÉ	MEASUREMENT UNIT ABBREVIATION FROM NATIONAL VMV DICTIONARY
MEASUREMENT VALUE	VALUE_VALEUR	MEASUREMENT VALUE OR CONCENTRATION
VARIABLE NAME	VARIABLE	VARIABLE NAME FROM NATIONAL VMV DICTIONARY
VARIABLE CODE	VARIABLE_CODE	NUMERIC VARIABLE CODE FROM NATIONAL VMV DICTIONARY
VARIABLE TYPE	VARIABLE_TYPE	VARIABLE TYPE FROM NATIONAL VMV DICTIONARY
VMV CODE	VMV_CODE	NUMERIC VALID METHOD VARIABLE (VMV) CODE FROM NATIONAL VMV DICTIONARY

Source: Government of Canada's National Long-term Water Quality Monitoring Data

# Progression

- Data collection, preliminary visualization, and initial Exploratory Data Analysis (EDA)
- Data limitations could hinder machine learning work. Namely, the datasets are focused on chemicals
- Data quality could become apparent upon closer review of sources (*i.e. sample data collected at multiple sites by multiple parties and devices*)

# Next Steps

- EDA, statistical analysis, and modeling
- Feature engineering definition, and supplemental data gathering
- Machine learning
- Neural networks, generative AI, and monetization