Water Access Analysis

Using Machine Learning to predict water stress in Canadian river basins.

by: Xavier Toledo
water.xavierdatapro.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 help mitigate adverse effects.

Vision

Using machine learning, we could predict the long term effects of different chemicals polluting our water supply

By targeting different affected sectors, business stakeholders will learn about potential opportunities

And, as more models are trained and new solutions are created, more individuals could benefit

Impact

Corporations could optimize their consumption by identifying and replacing water intensive processes

Non-governmental organizations could allocate resources to mitigate human impact (e.g. prepare for water shortages)

Federal and Provincial agencies could plan infrastructure projects more efficiently. For instance, they could plan and build water treatment plants faster

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 ANALYIS 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

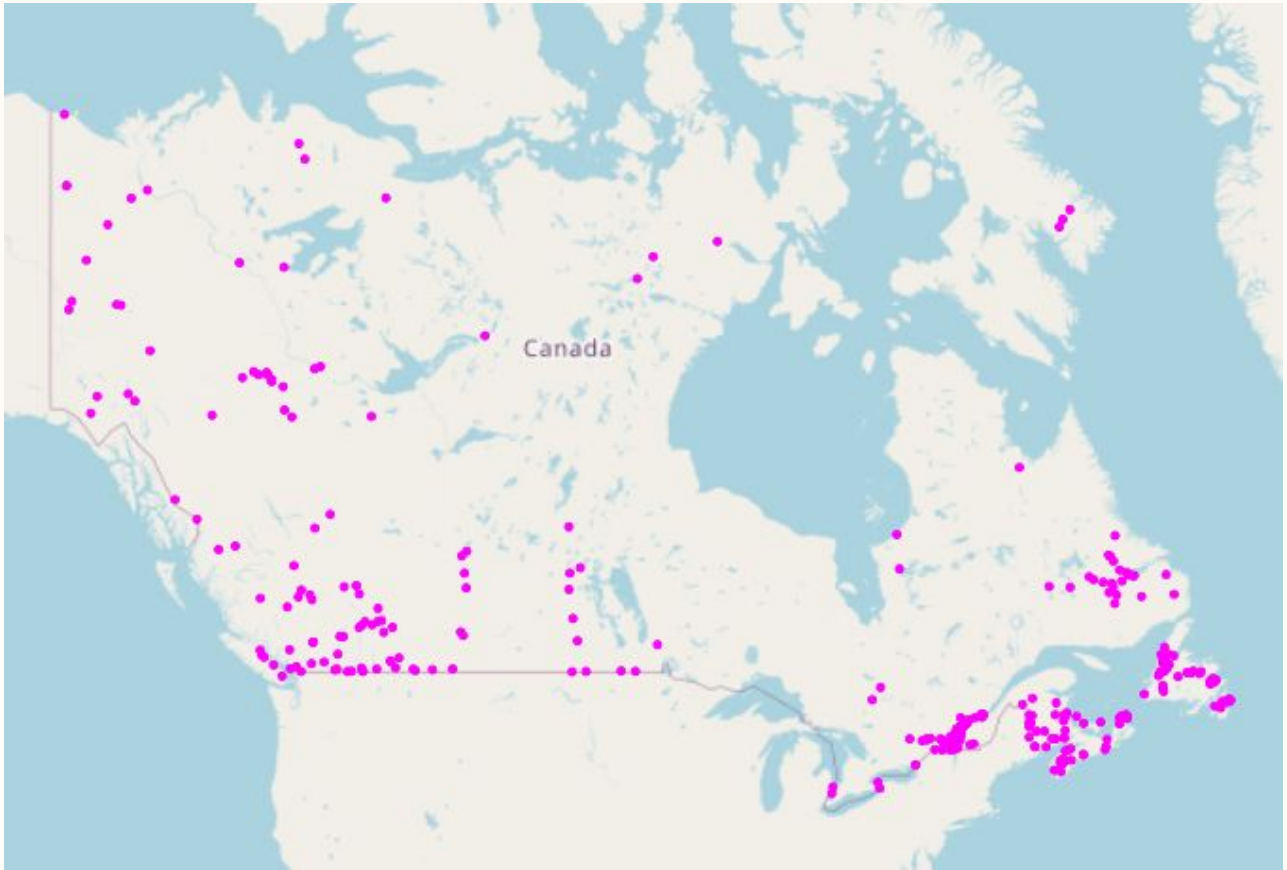
Progression

Exploratory Data Analysis

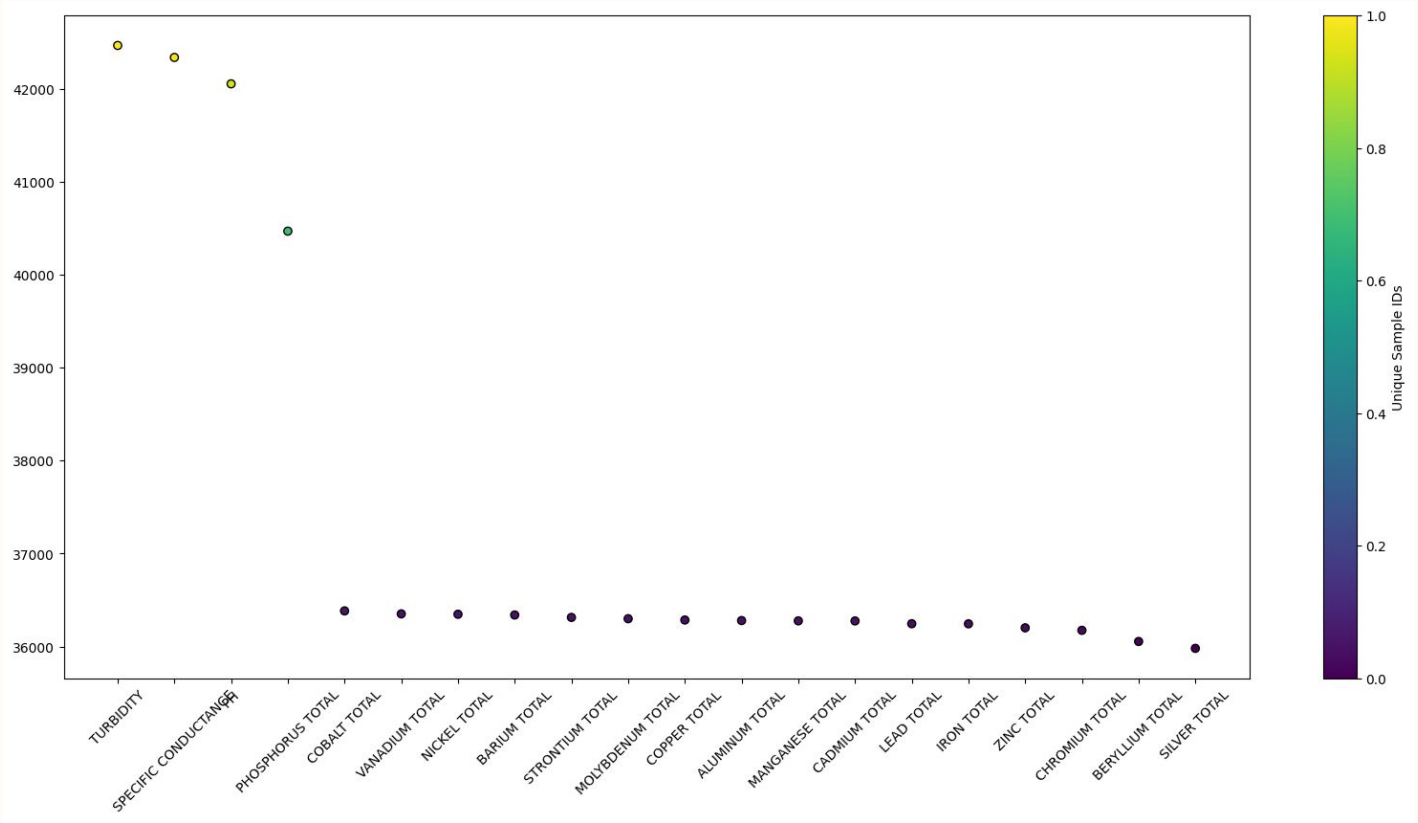
Basic Statistical Analysis

Baseline Modeling

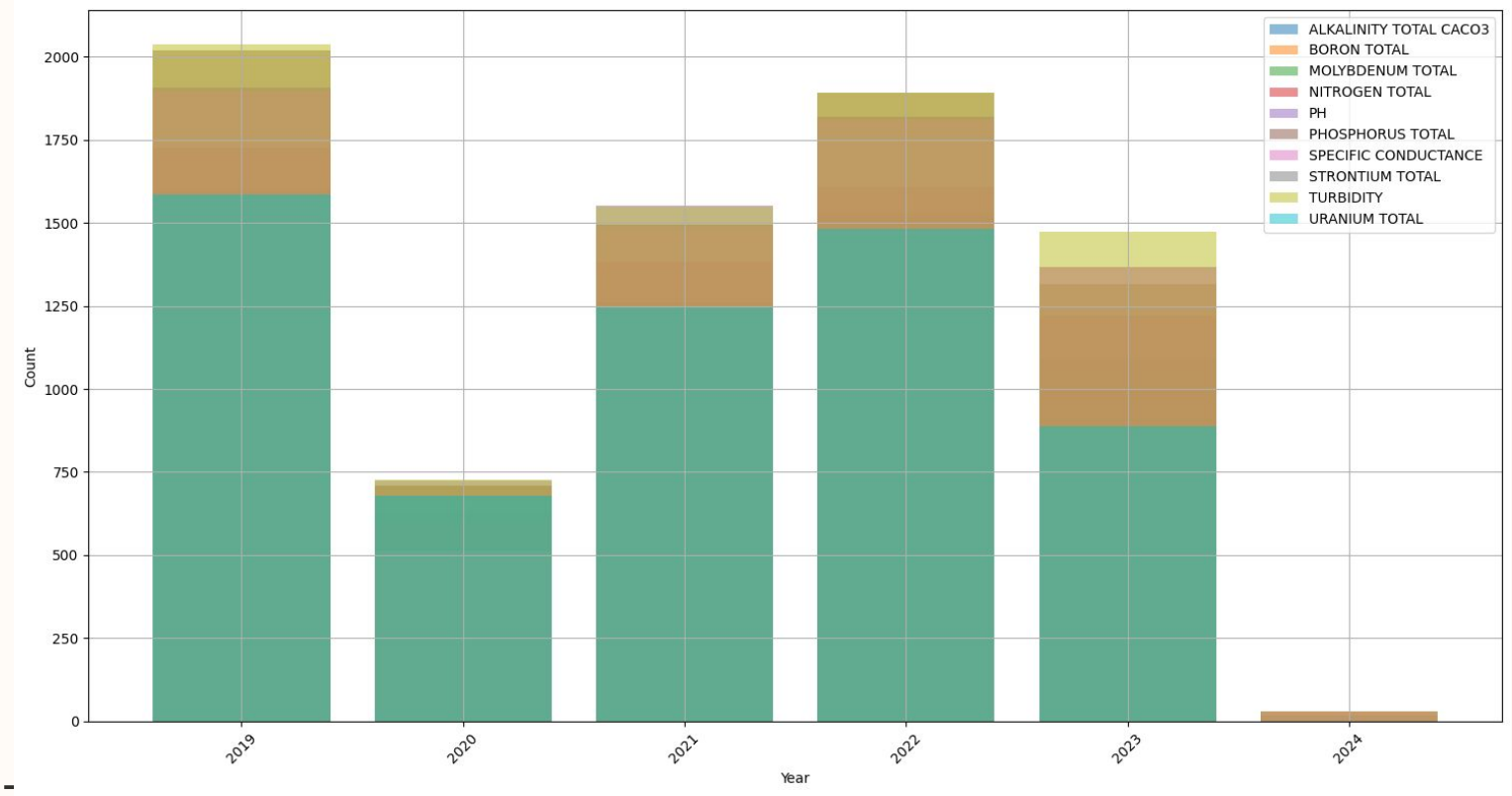
Water Monitoring Sites



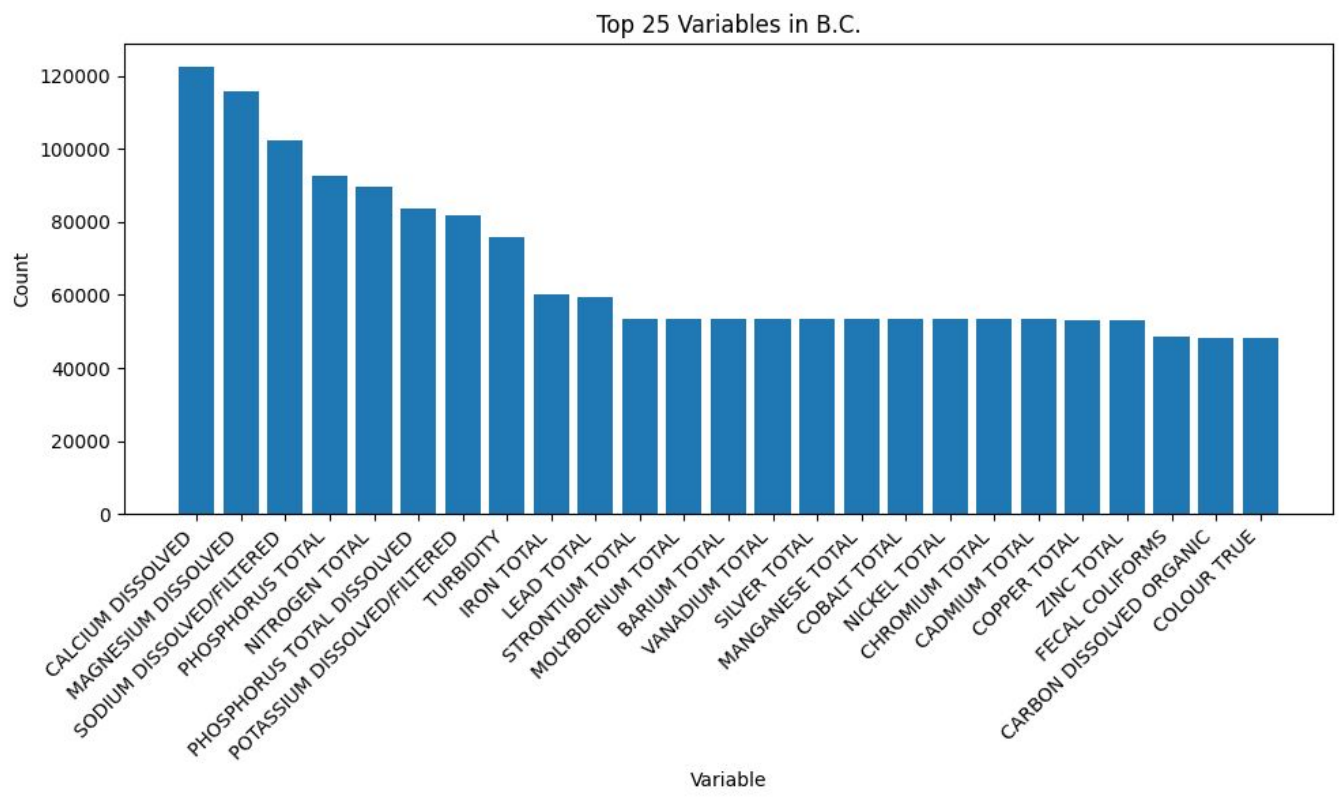
Top 20 pollutants



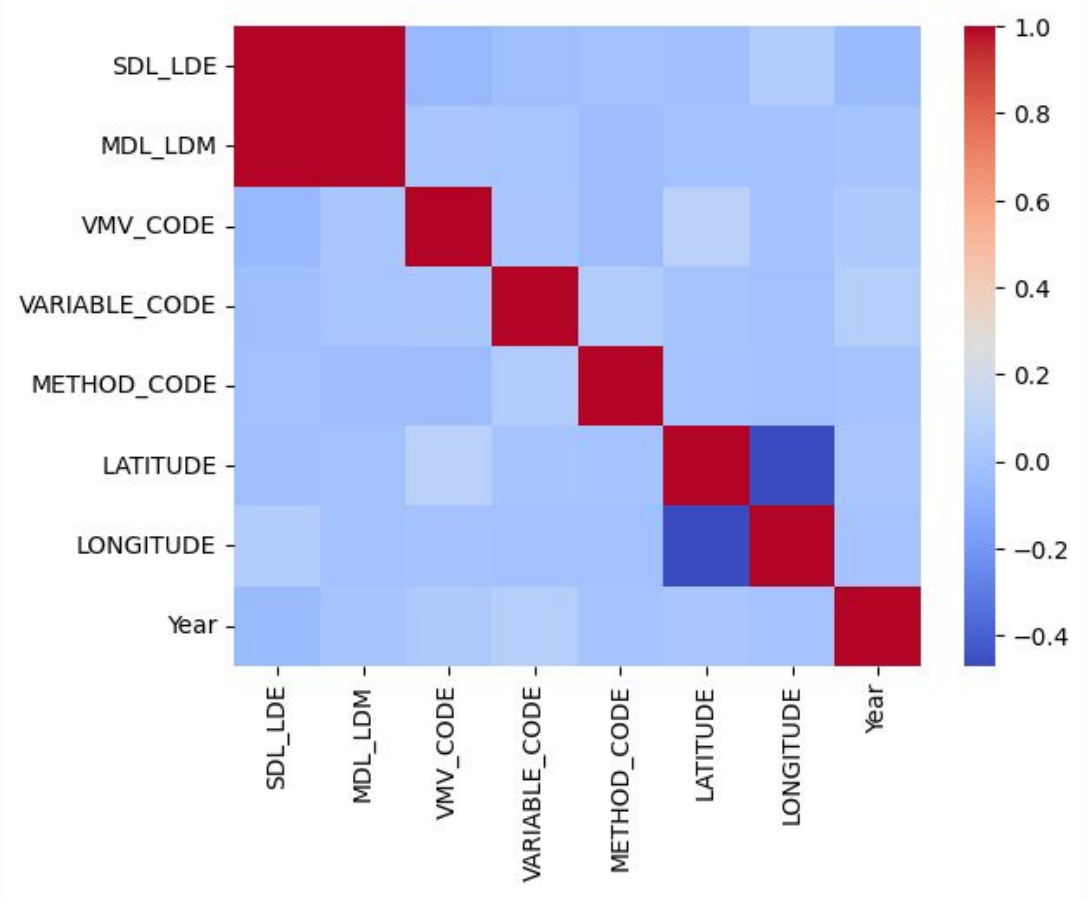
Top 10 pollutants since 2019



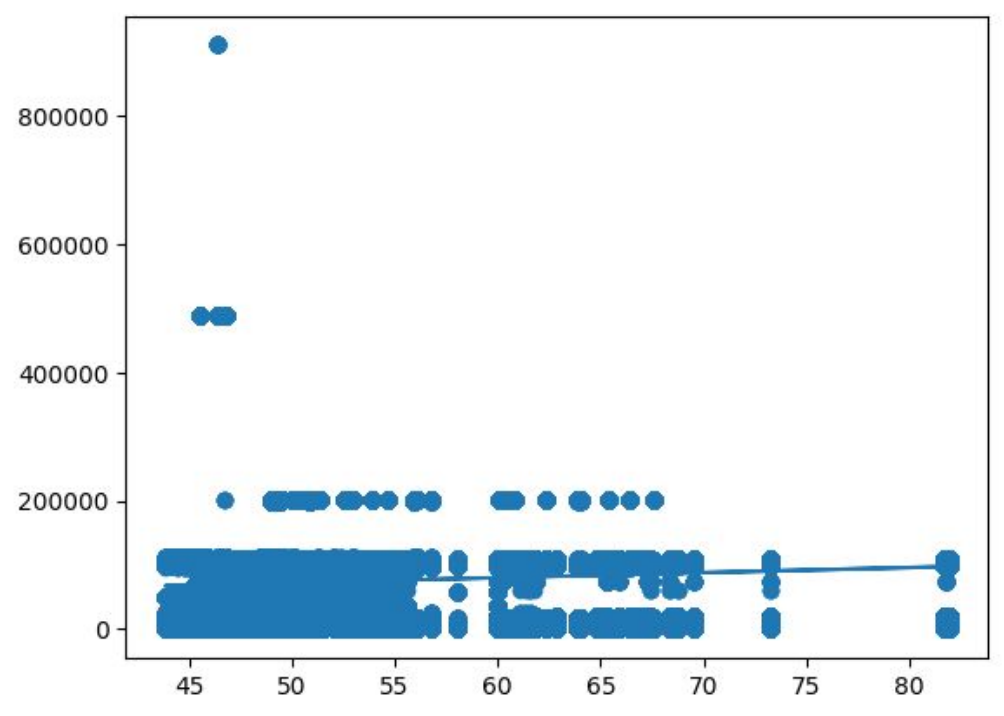
Top 25 pollutants in BC



Heatmap (LATITUDE & VMV_CODE)



Prediction
(weak relationship)



Next Steps

Feature engineering definition

ML Pipeline (Random Forest, AdaBoost)

Model comparison and interpretation

Product demo (Dashboard, Web App)

Questions?