REGIONAL CANCER ANALYSIS: RESOURCE ALLOCATION USING MACHINE LEARNING

Optimizing healthcare resources through advanced data techniques

11/14/2025

GLENN BLAKE
DHVANI SOLANKI
ALEXANDRE TUGIRUMUBANO

Background / Business Case

Methodology

Findings

Issues / Surprises

Recommendations

BACKGROUND & BUSINESS CASE

BUSINESS CASE

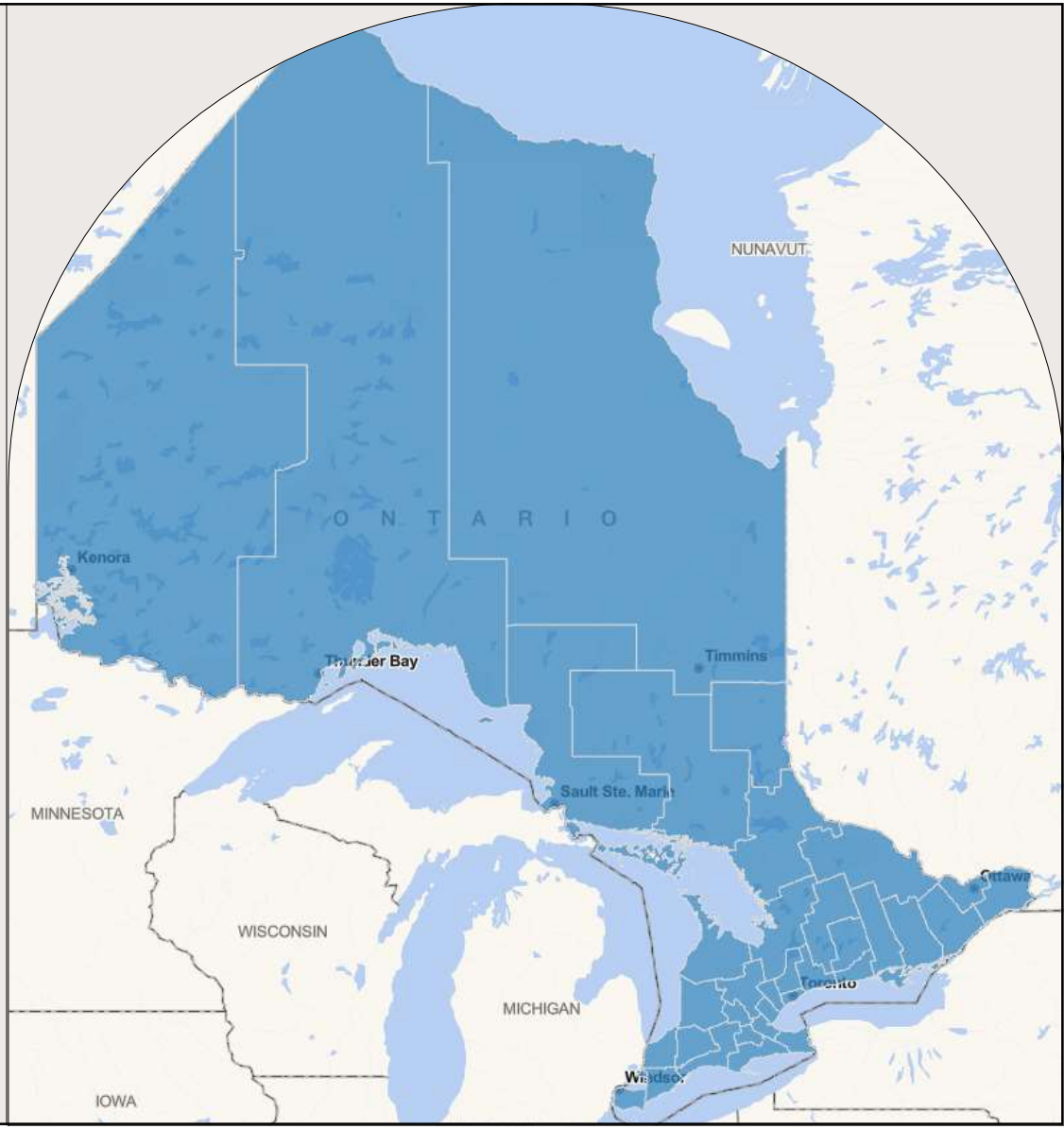
The Ministry of Health:
Has a finite budget to allocate toward resources for cancer support

Public Health Units:
Ontario is divided into 34, each addressing local health needs with tailored services and programs. They have limited human resources to dedicate to cancer programming

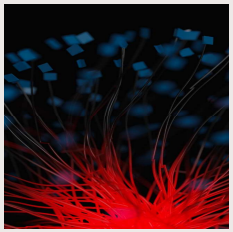
Cancer Burden Variation
Incidence and mortality vary significantly across Ontario's regions, reflecting uneven health outcomes.

Social Determinants:
Vary across the province and influence cancer types, outcomes and resource needs in PHUs.

Resource Allocation Strategy:
How would you advise government health officials to allocate resources to prioritize interventions in areas with greatest cancer burden and needs with consideration for regional Social Determinants of Health circumstances?

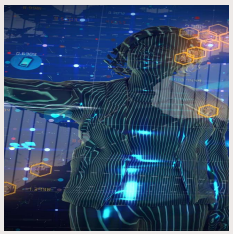


HOW CAN CANCER DATA AND SOCIAL DETERMINANTS GUIDE RESOURCE ALLOCATION?



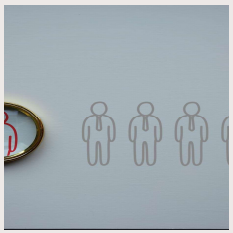
Integrating Cancer and Social Data

Combining cancer incidence and mortality rates with social determinants information (such as income and housing) can reveal regional health disparities.



Machine Learning for Pattern Identification

Machine learning models can be used identify patterns and predictors of cancer burden and social vulnerability across the province.

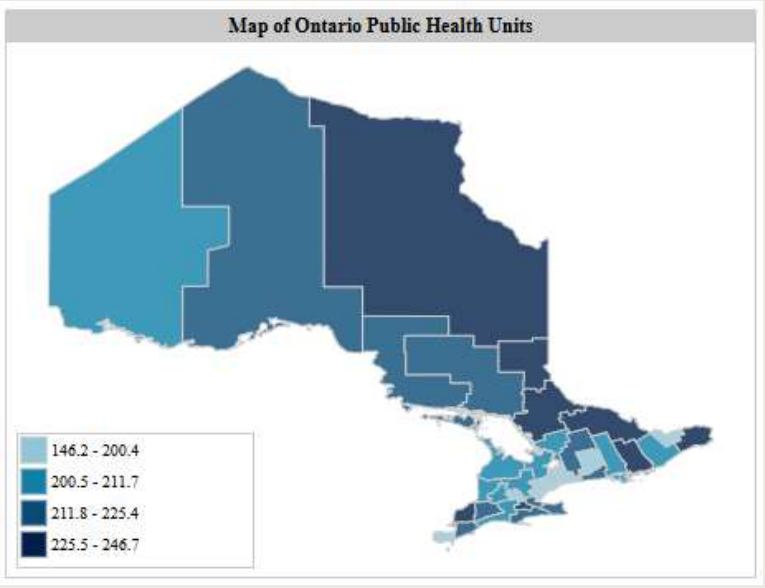


Targeted Resource Allocation

Insights garnered from identified patterns can guide strategic, equity-focused resource allocation to improve cancer outcomes and reduce disparities.

PUBLIC HEALTH ONTARIO DATA SETS

Mortality From all Cancers
Age-standardized rate (both sexes) 2015. Public Health Ontario
https://www.publichealthontario.ca/en/Data-and-Analysis/Chronic-Disease/Cancer-Mortality



Cancer Incidence Data

This data details new cancer cases reported across various public health units in Ontario, essential for tracking disease trends.

Cancer Mortality Data

Tracks deaths caused by cancer, providing critical insight into cancer outcomes and mortality patterns across the province.

Social Determinants of Health

Captures 2016 and 2021 Census Data census-based data for 34 Public Health Units, 14 Local Health Integration Networks (LHIN), and other geographic comparators in Ontario.

Data Integration for Insights

Combining these datasets enables nuanced understanding of cancer burden influenced by social and economic conditions.

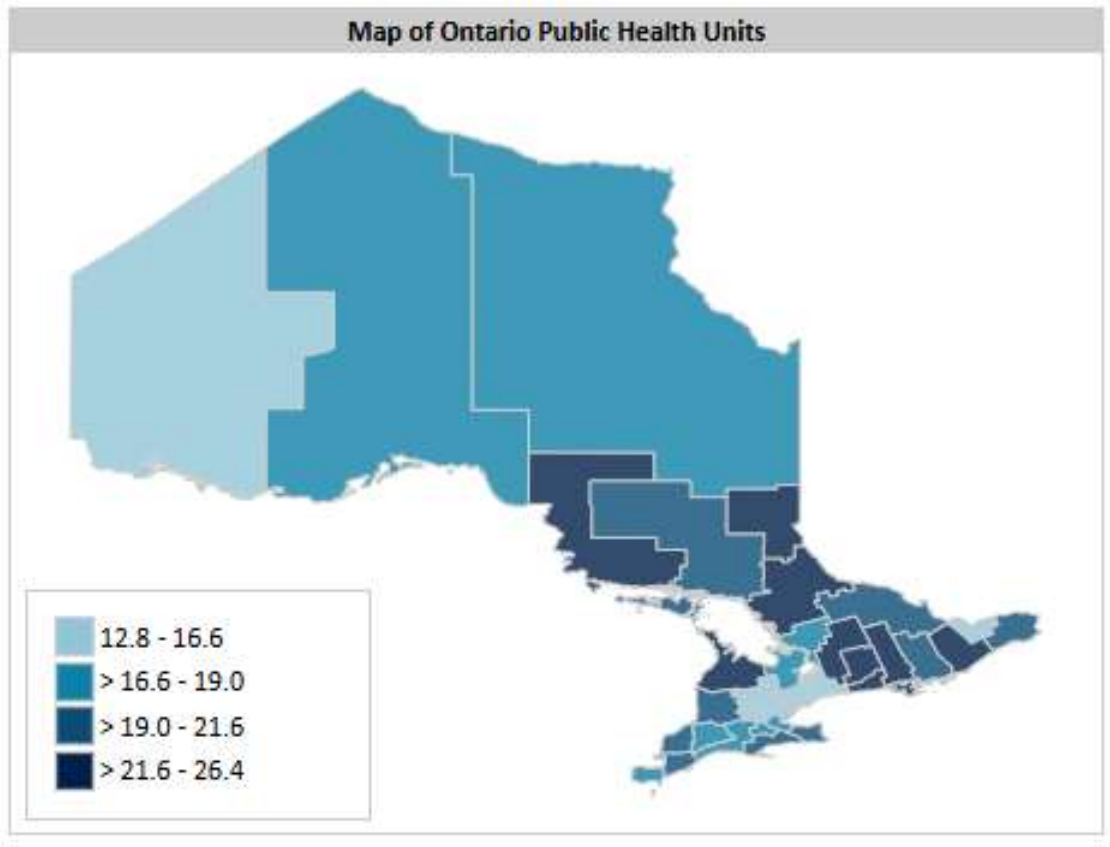
SOCIAL DETERMINANTS INDICATORS

Social Determinants of Health Provincial Snapshot

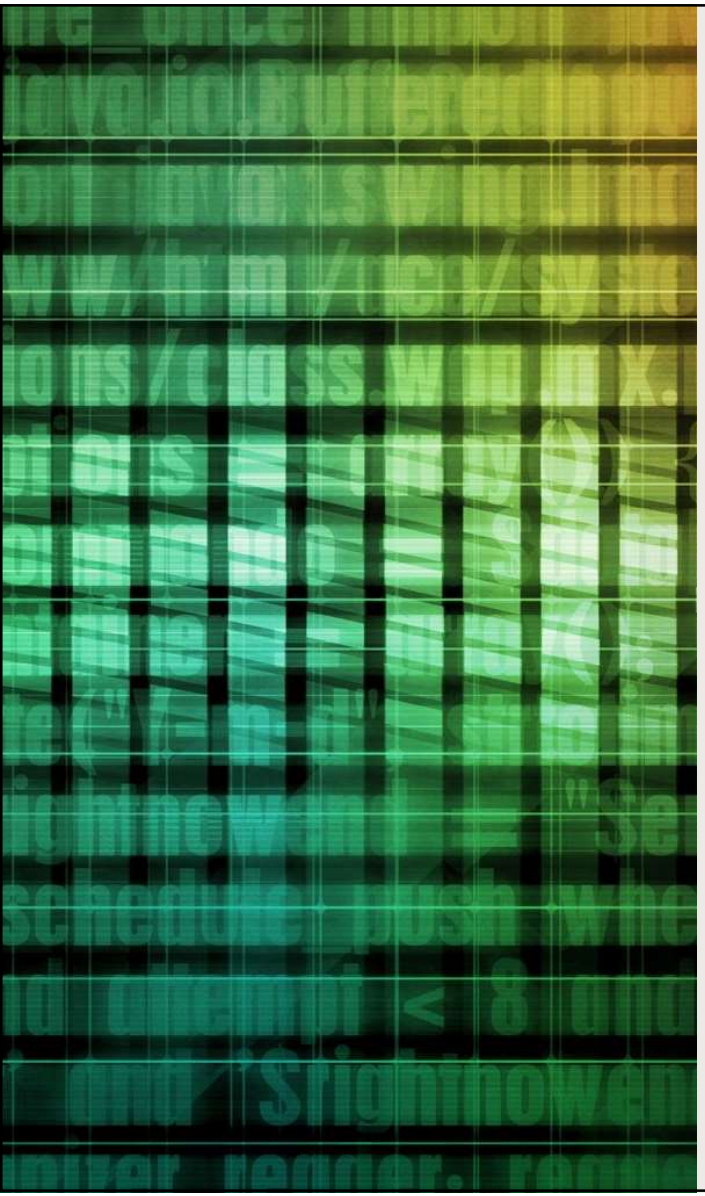
Social Determinants Indicators includes income, education, employment, and housing data to understand social factors impacting health outcomes in Ontario.

% Senior Population
% Of the population who cannot speak English or French
% Immigrant population
% Recent immigrant population
% Visible minority population
% Of the population living in low income
% Lone parent households
% Of the population who are employed
% Of the labour force population who are unemployed
% Of the population without a high school diploma
% Of households spending more than 30% of their income on shelter costs

% Senior Population (both sexes) 2016.
Public Health Ontario
https://www.publichealthontario.ca/en/Data-and-Analysis/Health-Equity/sdoh



METHODOLOGY



MACHINE LEARNING APPROACH AND DATA INTEGRATION

Data Cleaning and Preprocessing:

- Ensuring data consistency and accuracy by cleaning and preprocessing cancer and social determinants datasets. Merge three datasets by year and geography.

- PHU names were cleaned so everything lined up properly.

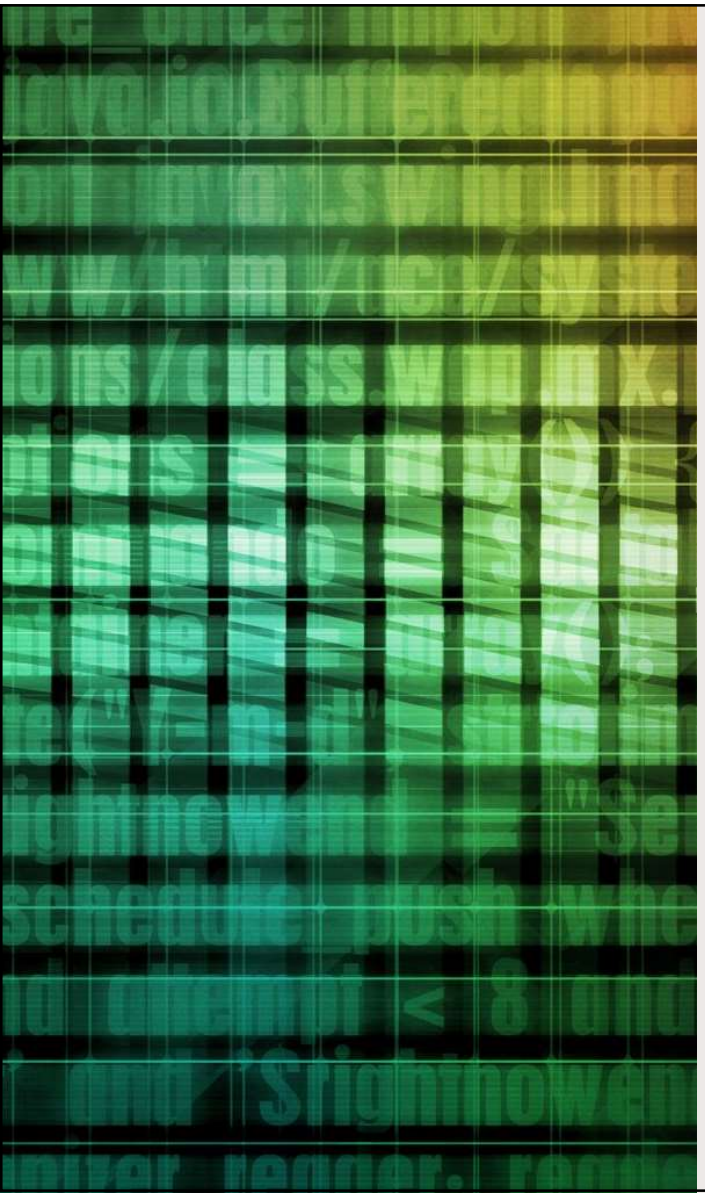
- For each PHU, we calculated the average incidence and mortality values, kept only the 2016 SDOH data, and removed columns that were non-numeric, mostly empty, or had no variation.

- All remaining features were scaled so the models could learn consistently.

Model Building:

- For both cancer incidence and mortality, we split the data into training and testing sets (80/20).
- We then trained the following models:

- **Linear Regression**
- **LassoCV**
- **XGBoost Regressor**
- **LightGBM Regressor**



MACHINE LEARNING APPROACH AND DATA INTEGRATION

Evaluating Performance:

- We measured each model using the same three metrics: **R^2**, **MAE**, and **RMSE**.

This allowed us to compare how well each approach predicted PHU-level outcomes.

Interpretation & Explainability:

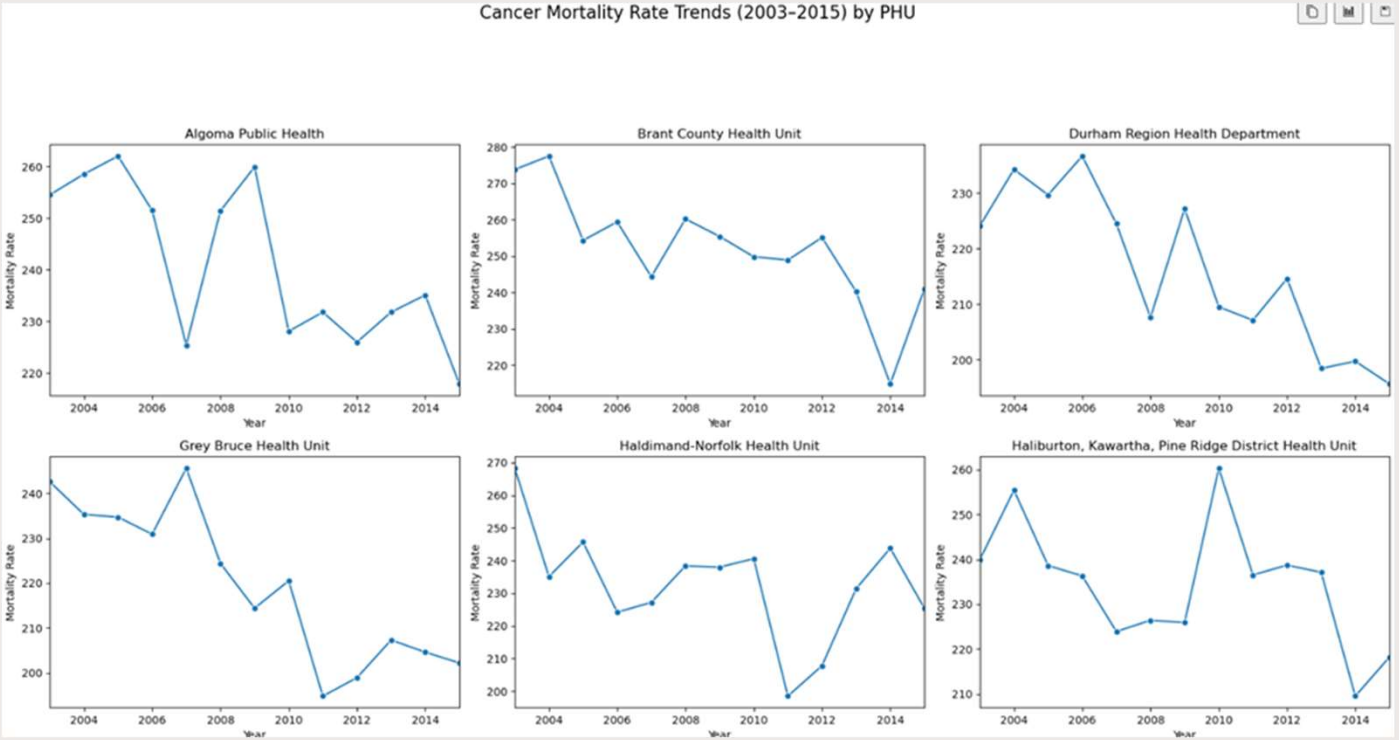
- Reviewed **Lasso coefficients** to identify the most influential SDOH predictors.
- Examined **feature importance** from tree-based models for nonlinear patterns
- Used **SHAP values** to understand how individual SDOH features push predictions up or down

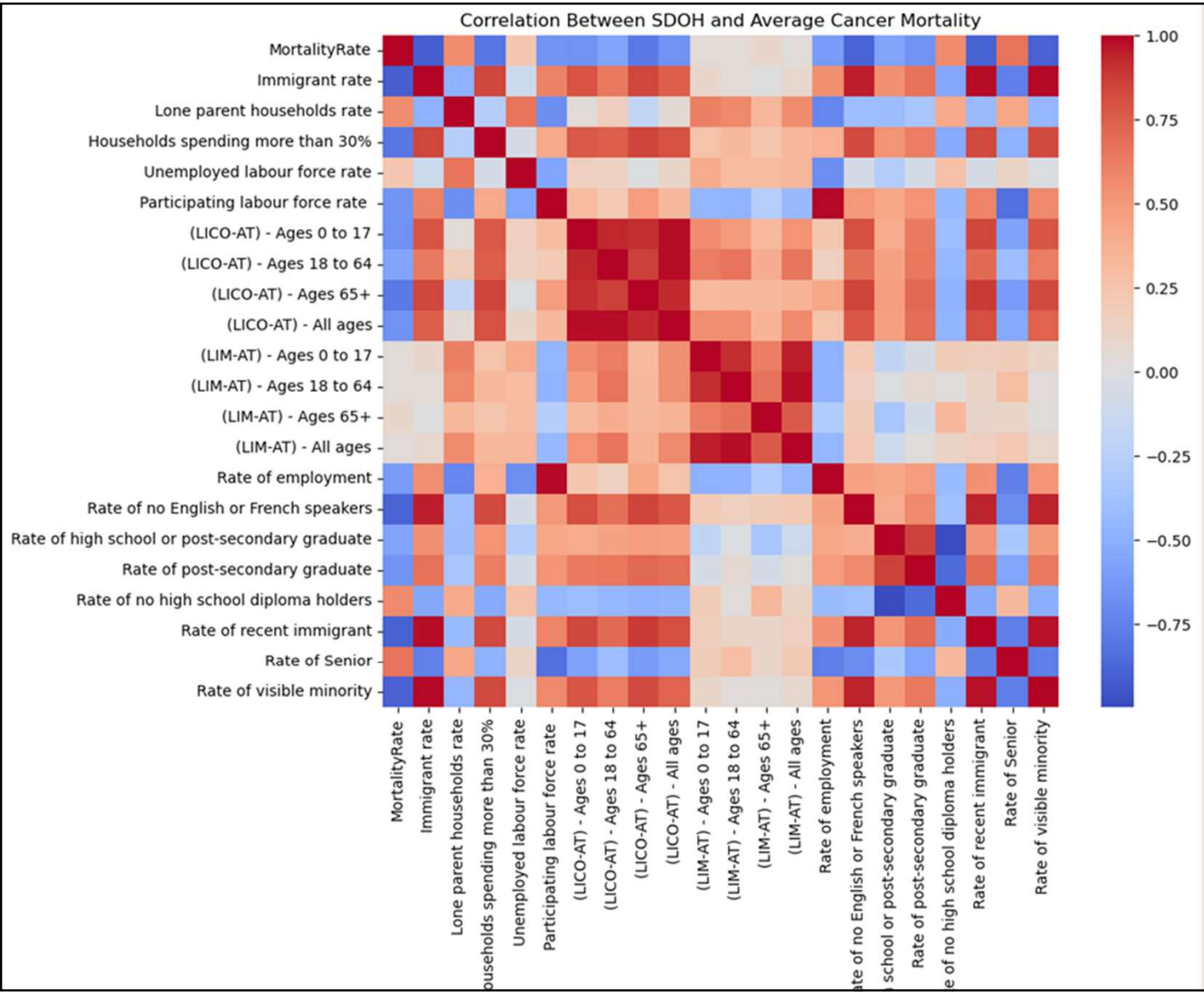
Final Outputs:

- All trained models and the scaler were saved for future reuse.
- Generated a summary file of actual vs. predicted values for visualization and reporting.
- Produced visualizations for feature importance and SHAP interpretation to support findings

FINDINGS

MORTALITY TRENDS OVER TIME (SELECTED PUBLIC HEALTH UNITS)





FINDINGS

Correlation of 2016 SDOH with the averaged mortality rate (2003-2015)

CANCER INCIDENCE – SHAP INTERPRETATION

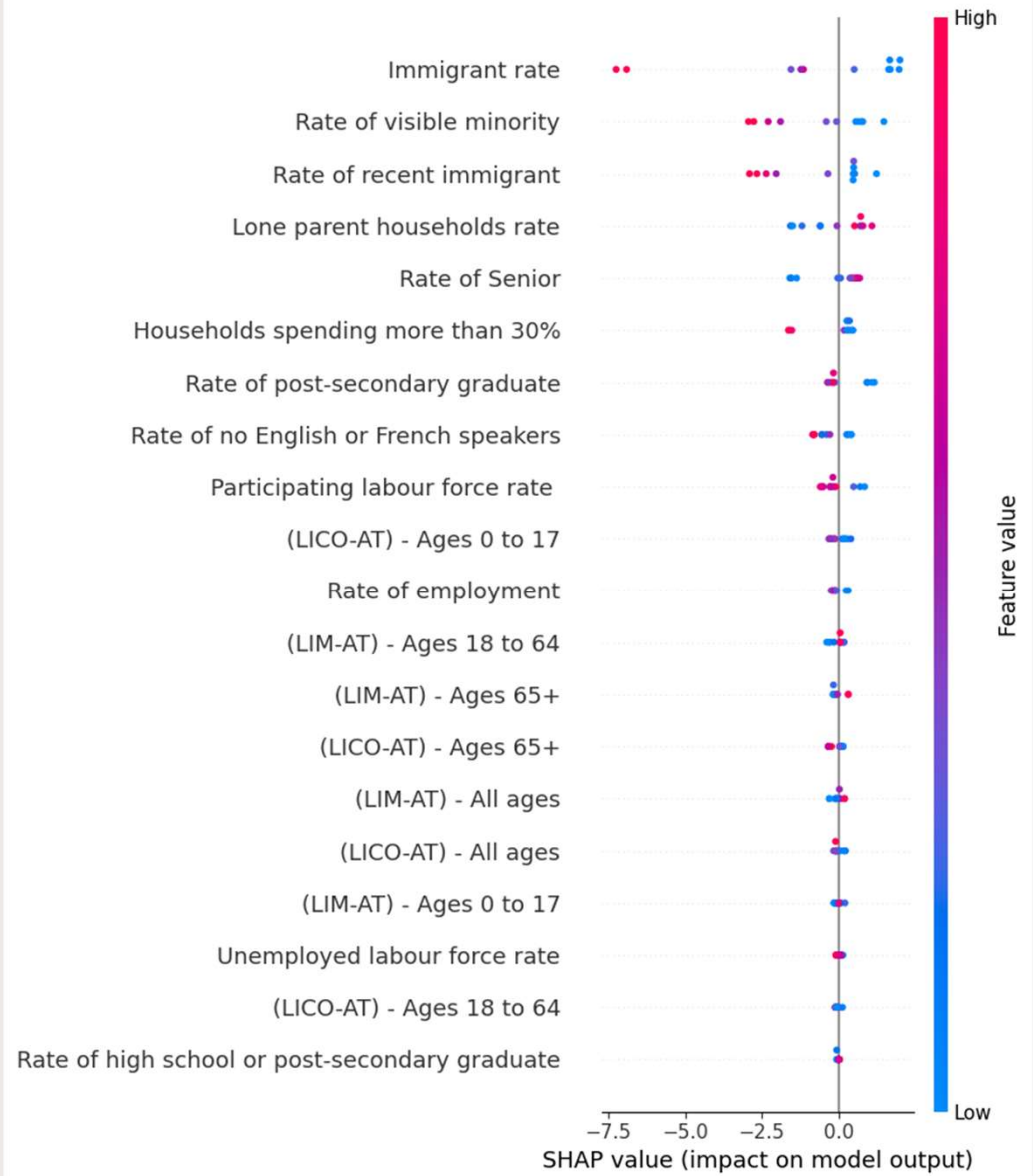
Age and demographic structure are the strongest drivers of incidence predictions.

Higher senior population rates significantly shift incidence outcomes.

Immigrant-related features (immigrant rate, visible minority rate, recent immigrants, language barriers) show major influence.

Economic variables (LIM/LICO, unemployment, education) have moderate to low impact on incidence.

The model suggests population composition is more predictive of incidence than socioeconomic hardship.



CANCER MORTALITY – SHAP INTERPRETATION

Immigration-related factors (immigrant rate, visible minority rate, recent immigrants) dominate mortality predictions.

Lone-parent households and housing cost burden contribute significantly, indicating social vulnerability effects.

Senior population rate remains an important mortality predictor.

Low-income indicators appear but have less influence than demographic and social structure variables.

The model indicates **mortality is more sensitive to community vulnerability** and social determinants than incidence.

FINDINGS

Cancer Incidence

Age was the primary determinant associated with cancer incidence

Cancer Mortality

Immigrant rate was the primary determinant associated with cancer mortality

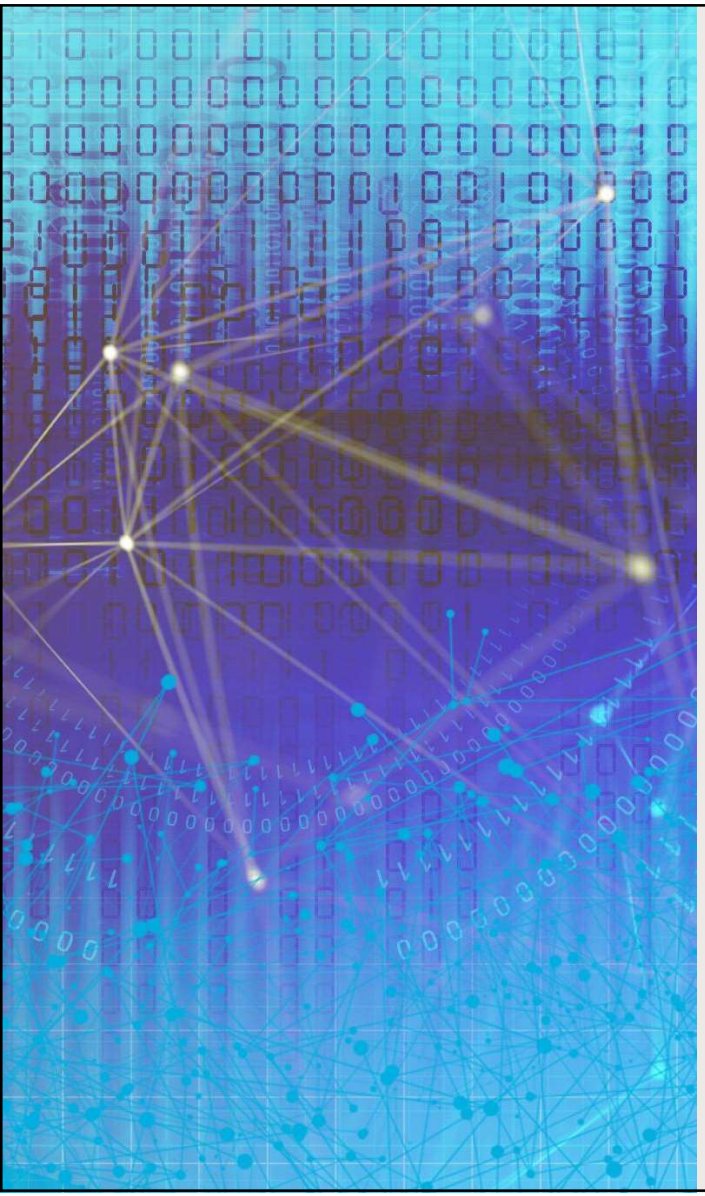
Suggestions

Ensure PHUs are providing appropriate programs and resources among communities with high immigration populations

Next Steps (if we had more time)

- *Conduct a region-by-region analysis of population and immigration rates*
- *Identify contributing factors to cancer mortality among immigrant populations*
- *Conduct an analysis of cancer-related resources specifically targeting immigrant populations*

ISSUES & CHALLENGES



DATA ALIGNMENT AND GEOGRAPHIC INCONSISTENCIES

Temporal Data Misalignment

Mismatched years between cancer data and social determinants hindered accurate correlation efforts.

Geographic Boundary Issues

Ontario Marginalization Dataset excluded due to incompatible geographic boundaries with PHU regions.

PHU Merging Adjustments

Merging of two PHUs required dataset adjustments to accurately reflect new configurations.

Need for Standardized Data

Standardized, interoperable datasets are essential for improving analysis accuracy and resource allocation.

REFERENCES

REFERENCES

Cancer Incidence Snapshot
Ontario Agency for Health Protection and Promotion (Public Health Ontario). Snapshots: cancer incidence snapshot: [Internet]. Toronto, ON: King's Printer for Ontario; c2024[cited 2025 Nov 14]. Available from: https://www.publichealthontario.ca/en/Data-and-Analysis/Chronic-Disease/Cancer-Incidence

Cancer Mortality Snapshot
Ontario Agency for Health Protection and Promotion (Public Health Ontario). Snapshots: cancer mortality snapshot: [Internet]. Toronto, ON: King's Printer for Ontario; c2024[cited 2025 Nov 14]. Available from: https://www.publichealthontario.ca/en/Data-and-Analysis/Chronic-Disease/Cancer-Mortality

Social Determinants of Health (SDOH) Snapshot

Ontario Agency for Health Protection and Promotion (Public Health Ontario). Snapshots: social determinants of health snapshot: [Internet]. Toronto, ON: King's Printer for Ontario; c2024[modified 2024 Feb: cited 2025 Nov 14]. Available from: https://www.publichealthontario.ca/en/Data-and-Analysis/Health-Equity/sdoh