CS 2704 : Data Analytics using Python

Nicolas Porter
Student ID: 3269520

Cedrick TAHMO
Student ID: 3727271

Project Proposal

April 4, 2025

0.1 Dataset chosen

1. StatCan: Population estimates, quarterly : 1980-2025
2. StatCan: Estimated areas, yield, production, average farm price and total farm value of principal field crops, in metric and imperial units: 1980-2025

0.2 Github Repository for project

GitHub Repo (currently private)

0.3 Hypothesis

The increasing population in Canada correlates with a lower annual crop yield due to reduced farmland.

0.4 Expected Outcome

The expected outcome of this hypothesis is that as population growth increases, there will be a greater demand for agricultural land, water, and resources, potentially leading to lower crop yields. Factors such as urban expansion will reduce farmland, increase resource competition, and environmental degradation could contribute to this trend. However, advances in agricultural technology, improved farming practices, and government policies may counterbalance these effects, maintaining or even increasing yields despite population growth. The analysis of the provided datasets should reveal whether there is a clear negative correlation between population growth and crop yields.

0.5 Plan for Testing Hypothesis

To analyze whether population growth negatively impacts crop yields, we need to:

- **Retrieve the Data** - Download the datasets from the provided StatCan links.
- **Clean & Process the Data** - Extract relevant variables:
  - Population estimates (independent variable).

  - Crop yields (dependent variable).

- **Perform Statistical Analysis** - Use correlation tests to check the relationship.
- **Compute correlation statistics** - Determine if the correlation is statistically significant with the help of a Regression Coefficient and R-Squared analysis.
- **Visual Trends** - Create graphs to observe trends over time.
- **Other** - Other statistics may be used as the project progresses.