**IMPACT OF CLIMATIC EXTREMES ON SOYBEAN AND CORN PRODUCTION IN ONTARIO**

**Name: Sundas Ahmed**

**Student Number: 500910967**

**Data: 20th October 2023**

**CIND820: Big Data Analytics Project**

**Introduction:**

Climate change is the most pressing issue at this time. Although the 21st dawned with new advancements, it also brought devastating impacts of changing climate. Climate change is a global phenomenon that has affected countries across the globe. Various catastrophic events have happened in the 21st century such as Hurricane Katrina in 2005 which killed about 1,800 people [(Worst Natural Disasters Of The 21st Century, 2011)](https://allthatsinteresting.com/worst-natural-disasters), Myanmar Cyclone in 2008 affected 2.4 million people [(Blair & Maclean, 2023)](https://www.reuters.com/business/environment/turkey-quake-other-major-natural-disasters-this-century-2023-02-09/), Cyclone Nargis in 2008 resulting in 100 thousand deaths [(Worst Natural Disasters Of The 21st Century, 2011)](https://allthatsinteresting.com/worst-natural-disasters), Indonesia Quake/Tsunami in 2018 killing more than 4,300 people [(Blair & Maclean, 2023)](https://www.reuters.com/business/environment/turkey-quake-other-major-natural-disasters-this-century-2023-02-09/). These events throughout the globe are the result of climatic changes.

Furthermore, these climatic changes didn’t only increase the catastrophic events but also adversely impact agricultural crops across the world. The agricultural production is highly dependent on the climate. Without adequate climatic conditions, it is becoming difficult day by day to achieve crop yields. Extreme weather conditions have affected crop yields in global north. There is considerable evidence that shows that the climate is changing throughout Canada. The seasonal temperatures throughout Canada have increased, causing warming temperatures in the winter and northern and southern regions. These changes have increased the frequency of extreme events such as wildfires, storms, and tornadoes. The arctic glaciers have melted to a great extent in the past decade due to the increased temperature.

The increase in temperature, levels of precipitation, and the increased number of extreme events have become more challenging for the agricultural sector. These changes can impact different regions across Canada across seasons. Some regions have opportunities to grow crops due to the warm weather, while on the other hand, some regions might face challenges in production due to the rising water level, heat waves, and strong winds that can affect the soil.

In Canada, springs are wet, summers are hot, and winters are mild. Due to the variability in the climatic conditions, this makes it difficult for the farmers to seed and harvest the crops accordingly. In Ontario, warm springs can extend the growing season and wet springs can delay the seeding of the crops due to the deficiency in soil nutrients [(Climate change impacts on agriculture, 2020)](https://agriculture.canada.ca/en/environment/climate-change/climate-scenarios-agriculture). The warm summer can increase water stress in various plants because of evapotranspiration and higher atmospheric CO2 [(Climate change impacts on agriculture, 2020)](https://agriculture.canada.ca/en/environment/climate-change/climate-scenarios-agriculture). There are longer growing seasons for warmer weather crops in Ontario, such as corn and soybeans.

In this paper, the effect of climatic changes on corn and soybeans will be analyzed in Ontario from 1991 to 2020. The historical patterns will be analyzed to understand the effect of different level of temperature and precipitation on the important grains, corn, and soybeans. Since there are many factors that can affect the yield of crops and grains in the province of Ontario, however, selected factors will be used in this study to analyze the effect on the production of the grains.

On the policy aspect, it is essential to understand the effect of various factors on the production of crops and grains. This study will analyze the effect of temperature and precipitation on crop yield, which will assist policymakers in creating such policies that would aid farmers in adopting new methods and techniques for greater yield in Ontario. There is an important point to keep in mind that there could be variations in the impact of the factors on the crop yield based on the geographical location. The climatic conditions of a particular locality can affect the levels of temperature and precipitation in a region.

**GitHub Link to this Study:**

<https://github.com/sundasahmed/Capstone-Project.git>

**Literature Review:**

Due to the increased number of extreme events in the 21st century and food scarcity, the study of climate change and crops has become crucial for researcher, policy makers and farmers. The effect of climate change on the crops has attained significant attention. There are numerous research papers can be found on this subject and have motivated many scientists and researchers to explore the applications of data science specifically machine learning models.

[He et al. (2018)](https://www.sciencedirect.com/science/article/abs/pii/S0308521X16305716?via%3Dihub) predicted the impacts of soil temperature, precipitation rates and N2O concentrations on crop yields and GHG emissions, using a denitrification-decomposition (DNNC) model to measure the impacts of climate change. It was based on the winter wheat, maize, and soybean under conventional and no tillage practices in Ontario from 2071 to 2100.

In this study, [Pearson et al. (2008)](https://www.sciencedirect.com/science/article/abs/pii/S1364815208000285?via%3Dihub) uses crop growth index, Growth, to calculate the growth for over 40 years in eastern Canada, to measure the regional growth and variability. This also highlights the impacts of climate change on the variability of the crop yields. This study elaborated median integrated seasonal growth index (MIGI) to describe the growth variability. This results from MIGI were compared to regional yield data to resolve policy related issues such as the impacts of climate change.

In this study, [Bhattarai et al. (2017)](https://www.sciencedirect.com/science/article/abs/pii/S0308521X16309179?via%3Dihub) investigated corn and soybean yields in the Raccoon watershed, in US. The projected data used was from 2015-2099 to implement Environmental Policy Ingetrated Climate (EPIC) model to understand the impact of climate change and Atmosphere-Ocean General Circulation Models (AOGCMs) to reflect on the GHG emissions. This study found that crop yield would decline at the RCP 8.5 rate which corresponds to high carbon emissions. If the scientists would not contain the emissions, it would have negative effects on the climate change and thus, on the agricultural industry.

[Mhawej et al. (2022)](https://globalchange.mit.edu/sites/default/files/MITJPSPGC_Rpt362.pdf) research developed SEBALIGEE (v2) in which machine learning RF algorithm was used to calculate the heat flux instead of hot/cold pixel approach. Major crops were analyzed such as corn, soybean, and wheat from 2013-2021 including seasonal and annual dataset along with climate variables such as air temperature, dewpoint temperature, surface net solar radiation and wind speed.

[Sun et al. (2021)](https://ui.adsabs.harvard.edu/abs/2021AGUFMGC32B..06S/abstract) research investigates the impacts of climate change on farmer’s income and agricultural production using Agent-Based Simulation model. This study uses corn, wheat, and soybean from 2016-2050 in Ontario along with climate variables such as average temperature and precipitation. Other variables were used such as crop prices and operation costs data to calibrate farm operations. Based on the historical time-series data, the authors implemented supervised machine learning algorithm to forecast climate prices. Through this study, it was concluded that climate changes cause drastic fluctuations on the farming within 10% for corn, wheat, and soybean along with a decline in the profitability.

[Bootsma et al. (2005)](https://cdnsciencepub.com/doi/pdf/10.4141/S04-025) explores the climatic changes on the corn, soybean, and barley in the Atlantic region of Canada for the 2040 to 2069 period. Various statistical techniques such as regression and correlation were used to understand the relation between the crops and the climate. It was concluded through General Circulation Model (GCM) that the average yields can be improved for corn and soybeans and lower average yield for barley with high water deficit area and with water surpluses by the direct effect of CO2 emissions.

[Yu et al. (2022)](https://www.emerald.com/insight/content/doi/10.1108/IJCCSM-11-2020-0124/full/html) investigates the effects of price and climate variables on the rice acreage in China using Multivariate Adaptive Regression (MARS) model from 1992 to 2017. Climatic variables such as minimum and maximum temperature and precipitation were used. It was concluded that there is a strong relation between the climate and the production and price of rice.

[Abbasi et al. (2021)](https://www.sciencedirect.com/science/article/abs/pii/S1537511021002397?via%3Dihub) explores the influence of CO2 emissions from agricultural soils of maize and soy using machine learning. There are six models used in this study such as Support Vector Machine (SVM), Random Forest (RF), [least absolute shrinkage and selection operator](https://www.sciencedirect.com/topics/engineering/least-absolute-shrinkage-and-selection-operator) (LASSO), feed-forward [neural network](https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/neural-network) (FNN), radial basis function [neural network](https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/neural-network) (RBFNN), and [extreme learning machine](https://www.sciencedirect.com/topics/engineering/extreme-learning-machine) (ELM). The selected variables were soil moisture, [soil temperature](https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/soil-temperature), [soil organic matter](https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/soil-organic-matter), soil total carbon, soil total nitrogen, air temperature, solar radiation and pan evaporation from 1989 to 2008. This study found that random forest (RF) was best to predict the CO2 emissions from inorganic fertilizer.

**Overview of the Data:**

The data utilized for this project was sourced from Statistics Canada and Climate Change Knowledge Portal which is disseminated by the World Bank.

For this study, I will emphasize on High Temperature, Low Temperature, Precipitation, Soybean production and Corn production. The variables are collected from different settings and sites. Temperature and Precipitation was collected from Climate Change Knowledge Portal and crop productions are selected from Statistics Canada. These datasets are collected for Ontario based on the annual reporting from 1991 to 2020. Based on the particular years, the variables ware amalgamated from different files into one dataset.

To mitigate prediction bias, the variables will be analyzed based on the latest 30 years as there has been drastic changes in the climate in the recent years.

1. **Maximum Temperature:**

The temperature is a primary factor that affects the rate of plant growth. Extreme temperatures can affect the plant productivity. Pollination is a phenological stage in the development of a plant where adequate temperature is necessary. High temperatures affect the variability of the pollen, fertilization and grain formation during the reproduction stage of a plant [(Hatfield & Pruegar, 2015)](https://www.sciencedirect.com/science/article/pii/S2212094715300116#bib15). The yield growth for corn and soybean would increase with temperatures up to 29°C to 32°C and then would decrease on temperature beyond this range [(Hatfield & Pruegar, 2015)](https://www.sciencedirect.com/science/article/pii/S2212094715300116#bib15). The average maximum temperature is selected for the 30 years in Celsius degree.

1. **Minimum Temperature:**

Low temperature also affects greatly on the production process of the plants. Photosynthesis and respiration will begin to increase when the temperatures rise, causing plants to grow. This is called optimal growth. The optimal temperature for corn ranges from 25° C and 33° C [(Curtis, 2022)](https://goodineverygrain.ca/2020/07/07/ontario-grains-need-a-drink/). Soybean optimal growth is around 29°C [(Curtis, 2022)](https://goodineverygrain.ca/2020/07/07/ontario-grains-need-a-drink/). The average minimum temperature is selected for the 30 years in Celsius degree.

1. **Precipitation:**

Ontario’s climate is very humid, the annual precipitation ranges from 660 to 1,020 mm. There are plentiful water supplies available [(McDonald, 2021)](https://ontariograinfarmer.ca/2020/10/01/water-management-3/). However, there have been times in the recent years when the crops have suffered drought such as in 2016 that had drastic effect on crop production and farmers [(McDonald, 2021)](https://ontariograinfarmer.ca/2020/10/01/water-management-3/). The aggregated precipitation is selected in millimeter (mm) for the 30 years.

1. **Soybean Production:**

Due to the increased production of soybean in Ontario, the production in Canada has increased from 2.69 million metric tons (Mmt) in 2007 to 7.72 (Mmt) in 2017 [(Qian et al., 2023)](https://cdnsciencepub.com/doi/full/10.1139/cjps-2022-0233). The soybean data is collected in harvested area in hectares and production in tons in order to calculate the yearly yield for 30 years.

1. **Corn Production:**

Corn is a hugely grown crop in Ontario. Over the recent decades, 2004–2015, corn acreage averaged 769,000 ha with an average yield of 9.53 t/ha [(Bagg et al.,2017)](https://www.ontario.ca/files/2022-10/omafra-agronomy-guide-for-field-crops-en-2022-10-13.pdf) . The industrial and feed has been using corn for 45% and 55% respectively in Ontario [(Bagg et al.,2017)](https://www.ontario.ca/files/2022-10/omafra-agronomy-guide-for-field-crops-en-2022-10-13.pdf) . The corn data is collected in harvested area in hectares and production in tons in order to calculate the yearly yield for 30 years.

**Crop Yield:**

The research employs the data set of crops and soyabean yield. The yields for the aforementioned crops were calculated using the following formula.

Crop yield = Production metric tons / area by hectare

**Attribute Description:**

This table summarizes the attribute, its description, type, and values.

|  |  |  |  |
| --- | --- | --- | --- |
| **Attribute** | **Description/Unit** | **Type** | **Value** |
| Precipitation | Aggregated accumulated precipitation in mm | Numeric | from 60.42 to 75.59 |
| Maximum Temperature | Average maximum temperature over the aggregation period in °C | Numeric | from 4.96 to 8.55 |
| Minimum Temperature | Average minimum temperature over the aggregation period in °C | Numeric | from -4.08 to -0.56 |
| Corn | Annual corn production in metric tons and by harvested area in hectares | Numeric | from 5.99 to 10.60 (yield) |
| Soybean | Annual soybean production in metric tons and by harvested area in hectares | Numeric | from 0.28 to 0.70 (yield) |

**Research Questions:**

What is the impact of high temperature on soybean and corn crop production in Ontario?

What is the impact of low temperature on soybean and corn crop production in Ontario?

What is the impact of precipitation levels on soybean and corn crop production in Ontario?

**Approach/Methodology:**

Creating an approach for analyzing the effect of climate change on the crop production is essential for the policy makers and farmers to make informed decisions. The following steps outline a structured methodology for analyzing the data:

1. **Data Collection:** The first step was to collect historical data for the variables being analyzed to understand the relation between the selected crops and climate factor in Ontario. The source of the collected data must be reliable and accurate. The crops data was collected from Statistics Canada and climate data was collected from Climate Change Knowledge Portal.
2. **Data Preparation:** Once the data was collected, it should be prepared for analysis. This step involves removing missing values and processing in a suitable format to make ready for analysis. There were no missing values found in the dataset. The data variables from two different sources were selected based on multiple years, which then organized according to the year.
3. **Model Selection:** In this step, various models such as Decision Tree, Naive Bayes, KNN

Regression, Logistic Regression and Random Forest are some potential options for analysis for this dataset. This step is vital for the outcome as it will determine the analysis of the project.

1. **Model Implementation:** After the models are selected for the analysis, these models will be implemented on the dataset that was prepared earlier. This step could involve adjusting the parameters of the selected models to enhance their performance.
2. **Data Visualization:** In this step, Exploratory Data Analysis (EDA) was conducted to further understand the trend, patterns, and characteristics in the dataset. Various techniques can be applied such as histograms, box plots and confusion matrix to find the correlation and detect the outliers.
3. **Results Interpretation:** In this crucial step, the results were interpreted based on the findings from the model implications. Certain statistical tests can be used to determine the significance of the findings.
4. **Conclusion and Recommendations:** In the last step, the results were concluded based on the interpretations in the previous step. Certain recommendations could be made for the future research purposes and for the policy makers.

**Exploratory Data Analysis:**

**1. Summary Statistics:**

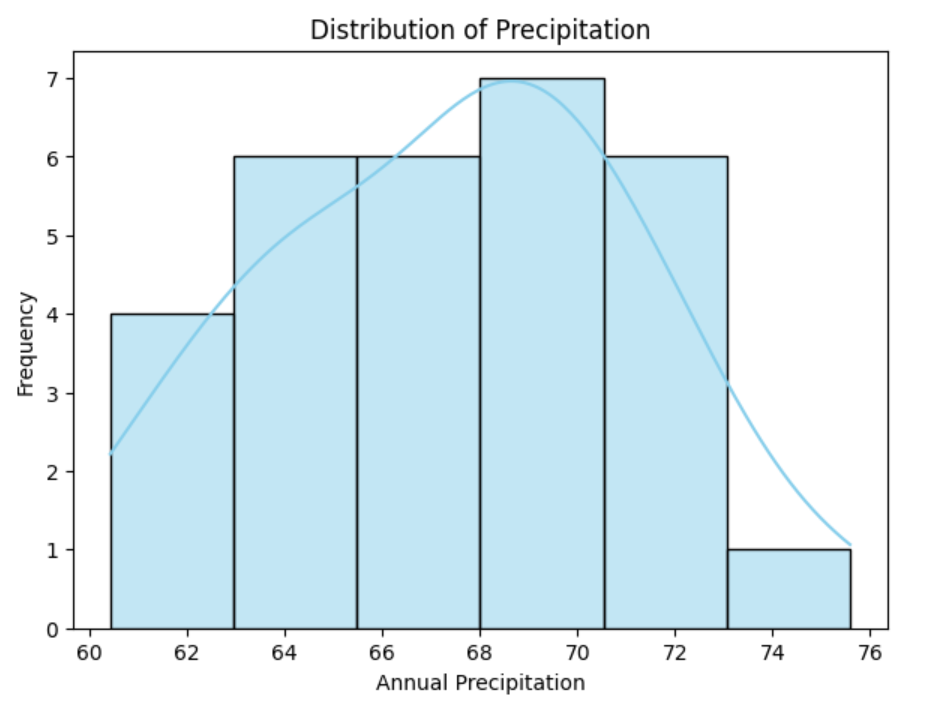
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Precipitation | Maximum Temperature | Minimum Temperature | Corn | Soybean |
| count | 30.000000 | 30.000000 | 30.000000 | 30.000000 | 30.000000 |
| mean | 67.465333 | 6.580333 | -2.699667 | 8.602932 | 0.372887 |
| std | 3.782038 | 1.04462 | 1.028254 | 1.449592 | 0.077837 |
| min | 60.420000 | 4.960000 | -4.080000 | 5.996473 | 0.289513 |
| 25% | 64.520000 | 5.735000 | -3.425000 | 7.161624 | 0.324747 |
| 50% | 67.730000 | 6.305000 | -2.965000 | 8.683487 | 0.362707 |
| 75% | 70.015000 | 7.582500 | -2.040000 | 9.910658 | 0.391111 |
| max | 75.590000 | 8.550000 | -0.56000 | 10.607699 | 0.703932 |

Summary statistics is an essential tool for that explores the measures of central tendency and measure of variation in the dataset. The above table shows the average value of the variables for the 30 years, showing average precipitation and temperatures and the average yield of crops. The maximum and minimum values of the variables can be used to understand the spread of the values. For example, the minimum yield of corn was 5.99 (metric tons per hectare) and the maximum yield was 10.60 (metric tons per hectare) in the past 30 years. The standard deviation shows how close the data is clustered around the mean value.

**2. Distribution:**

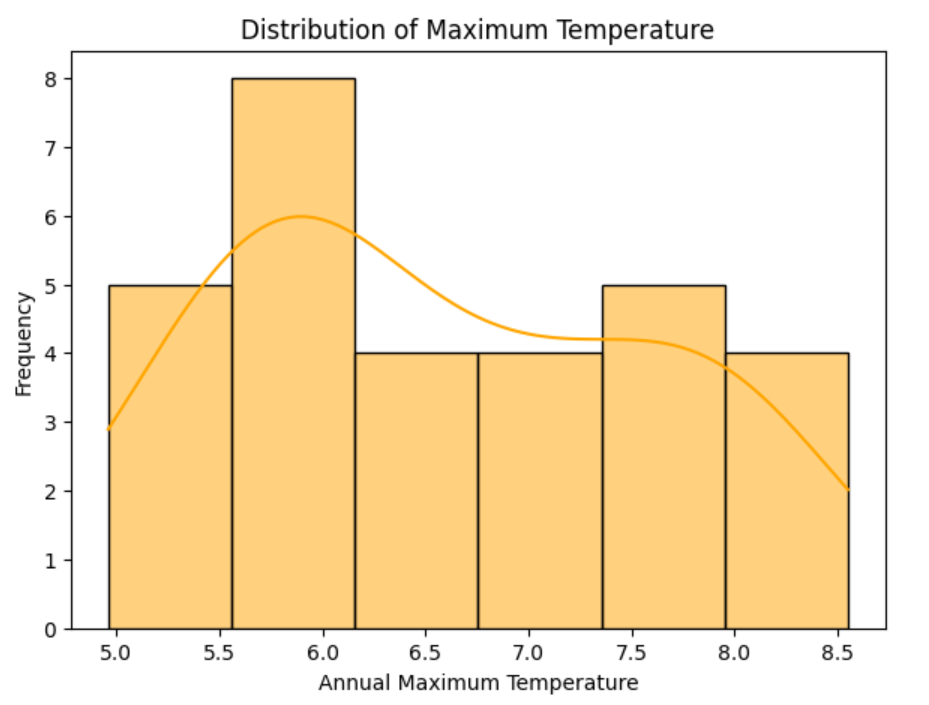
**Precipitation (mm):**

The distribution of precipitation based on the mean value of 67.5 show that the slightly left-skewed as the median value 67.73 is slightly greater than mean value. This shows that the environment of Ontario has majority of the precipitation throughout the course of 30 years around 67 mm.



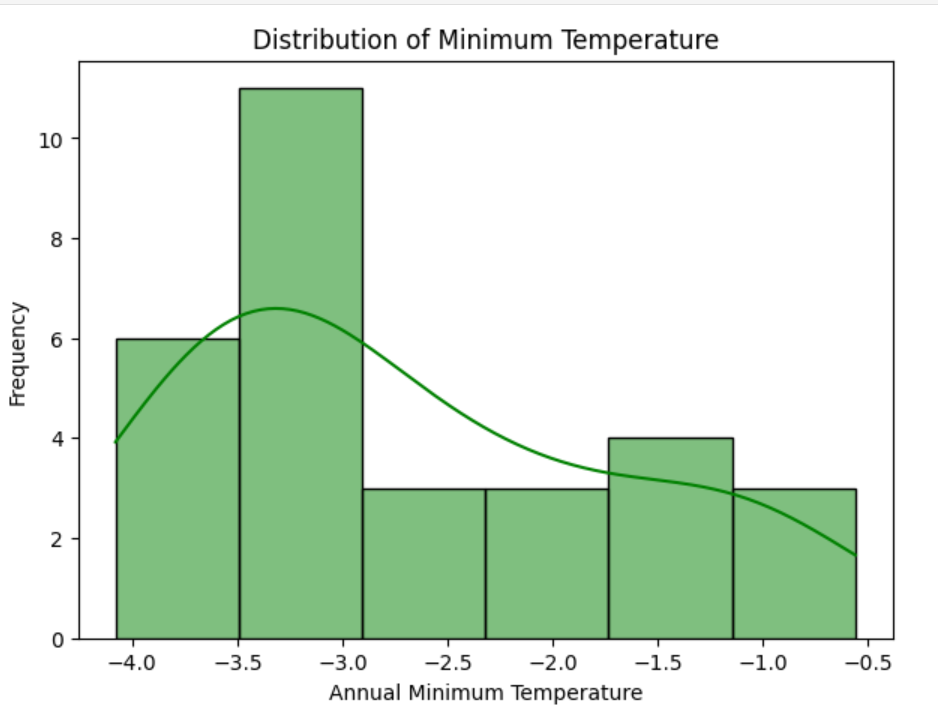
**Maximum Temperature (oC):**

The distribution of maximum temperature in Ontario over the course of 30 years based on the mean value of 6.5 oC show that the distribution is slightly right skewed as the median value 6.3 is slightly lower than mean value. This shows that the annual average maximum temperature has been around 6 oC.



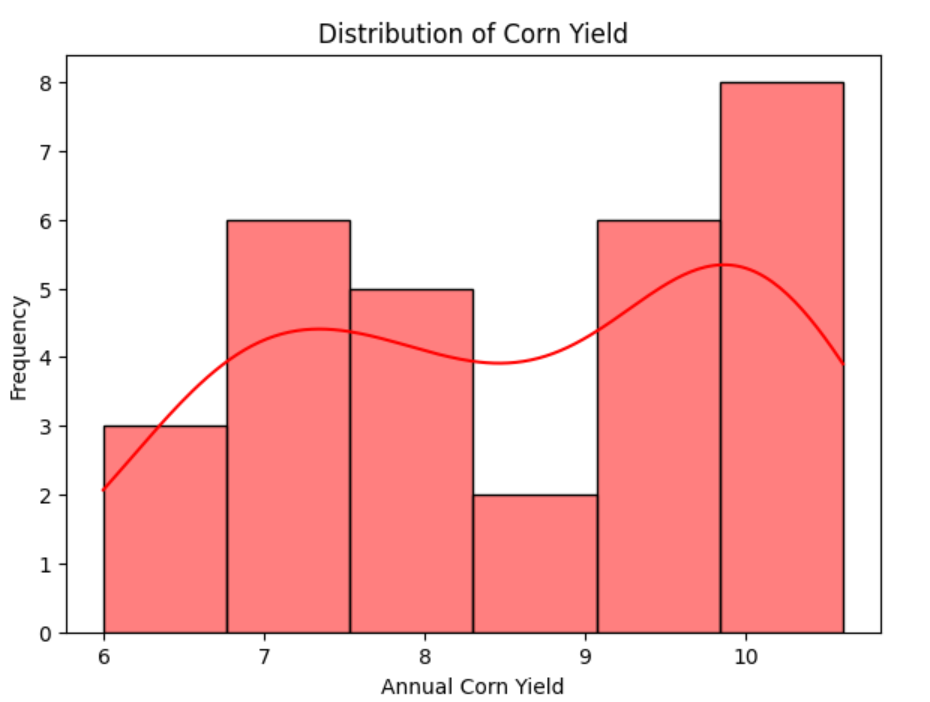
**Minimum Temperature (oC):**

The distribution of minimum temperature in Ontario over the course of 30 years based on the mean value of -2.69 oC show that the distribution is right skewed as the median value -2.96 is lower than mean value. This shows that the annual average minimum temperature has been around -2 oC.



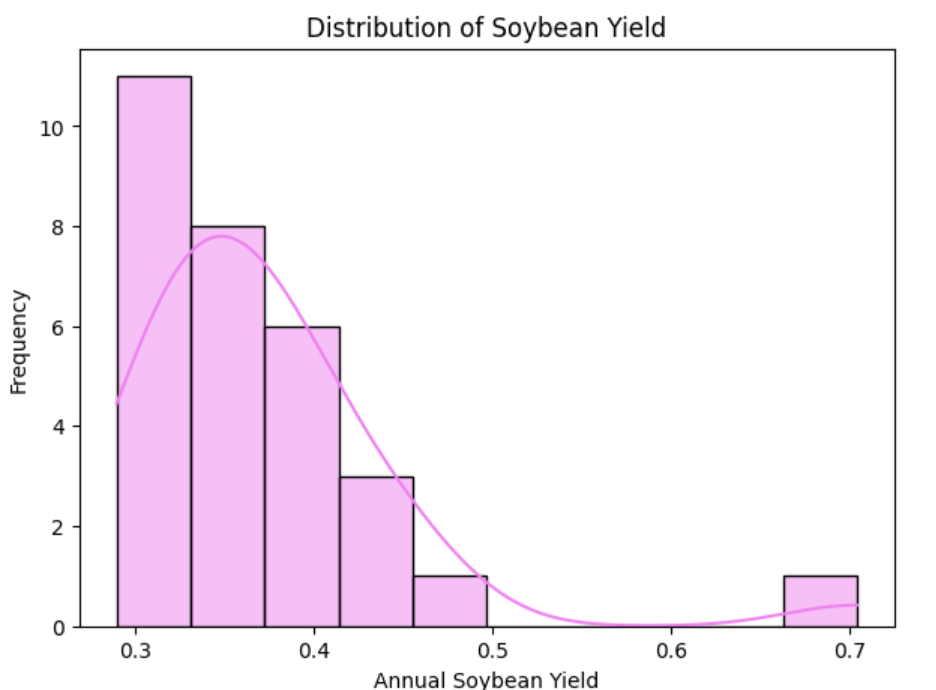
**Corn Yield (metric tons by hectare):**

The distribution of yield of corn in Ontario over the course of 30 years based on the mean value of 8.6 metric tons by hectare, shows that the distribution is slightly left skew as the median value 8.7 is slightly greater than mean value. The longest bar in histogram shows that in Ontario, there has been more production of corn grain over the pass years. However, the annual average yield of corn has been around 8 metric tons by hectare.



**Soybean Yield (metric tons by hectare):**

The distribution of yield of soybean in Ontario over the course of 30 years based on the mean value of 0.37 metric tons by hectare, shows that the distribution is right skewed as the median value 0.36 is slightly lower than mean value. The longest bar in histogram shows that in Ontario, there has been less production of soybean grain over the pass years. However, the annual average yield of soybean has been around 0.3 metric tons by hectare.

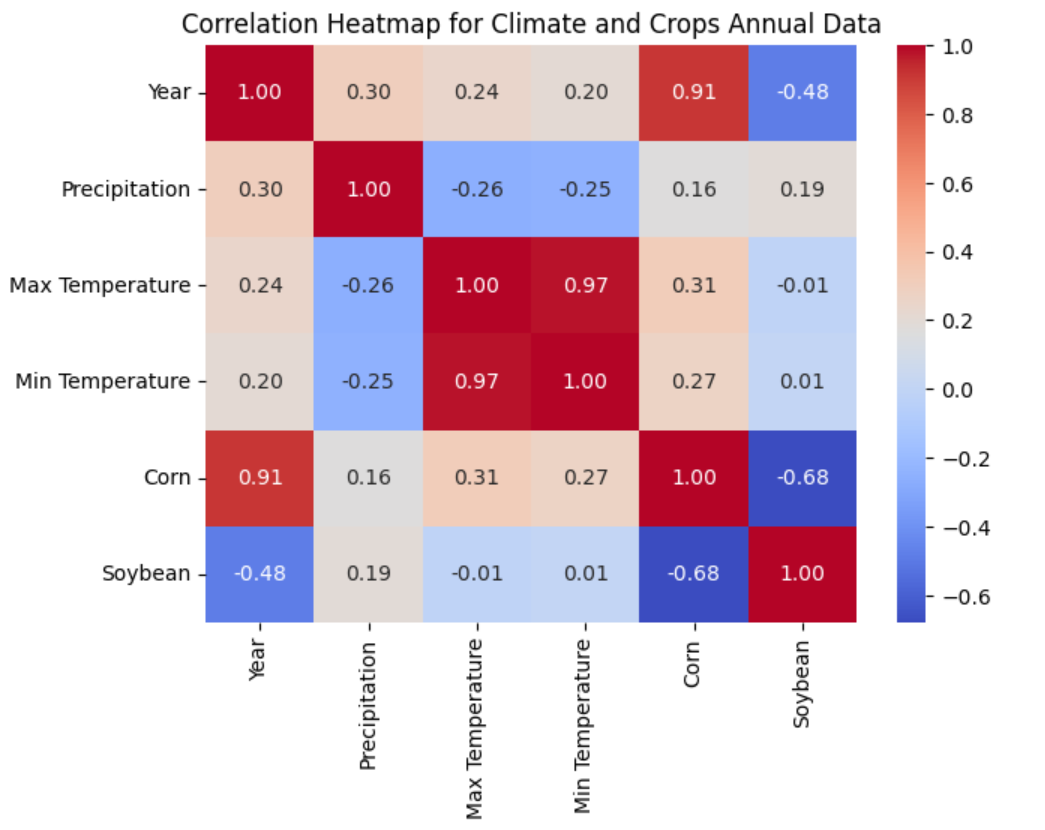


**3. Correlation:**

In the correlation heatmap, each block of cell represents the correlation coefficient between the two variables from the dataset. The precipitation, maximum temperature, minimum temperature, crop yields and the year are correlated with each other to analyze the relationship among them. The intensity of the color indicates the direction and the strength of the relationship among the variables. Red color indicates positive correlation and blue indicates negative correlation.

The below heatmap shows that the production of corn has increased with the years as it shows strong positive correlation. There is a positive correlation of corn with maximum and minimum temperature but a weak correlation with precipitation.

On the other hand, he soybean production has been reduced with the passage of time. Maximum and minimum temperature have slightly negative relationship with soybean production, but a weak relationship with precipitation.



**4. Outliers:**

Through the boxplot, outliers in the dataset can be distinctly identified. Soybean is the only variable that have an outlier as it is identified in the below boxplots. The outlier in soybean lies at 0.7. This outlier lies extremely high as compared to the rest of the data point of Soybean. As the production of the soybean has a mean 0.37 and the standard deviation value 0.077 shows that the data values are clustered closely to the mean value. The values are not widely dispersed. However, there is an outlier at 0.7, which shows that during a particular year from 1991 to 2020, there was a high production of soybean in Ontario.

A diagram of a boxplots of selected variables

Description automatically generated

A diagram with blue lines

Description automatically generated

**References:**

/ Gouvernement du Canada. (2020, January 31). *Government of Canada*. Language selection - Agriculture and Agri-Food Canada / Sélection de la langue - Agriculture et Agroalimentaire Canada. <https://agriculture.canada.ca/en/environment/climate-change/climate-scenarios-agriculture>

All That’s Interesting. (2019, June 12). *The worst natural disasters of the 21st Century*. <https://allthatsinteresting.com/worst-natural-disasters>

Bagg, J., Ball, Dr. B., Banks, S., & Baute, T. (2017). *AGRONOMY GUIDE  FOR FIELD CROPS*. Ontario. <https://www.ontario.ca/files/2022-10/omafra-agronomy-guide-for-field-crops-en-2022-10-13.pdf>

Blair, E., & Maclean, W. (2023, February 9). *Turkey earthquake and some of the worst natural disasters of this century*. Reuters. <https://www.reuters.com/business/environment/turkey-quake-other-major-natural-disasters-this-century-2023-02-09/>

Bootsma, A., Gameda, S. and McKenney, D. W. 2005. Potential impacts of climate change on corn, soybeans and barley yields in Atlantic Canada. Can. J. Plant Sci. 85: 345–357, <https://cdnsciencepub.com/doi/pdf/10.4141/S04-025>

Budong Qian, Ward Smith, Qi Jing, Yong Min Kim, Guillaume Jégo, Brian Grant, Scott Duguid, Ken Hester, and Alison Nelson. 2023. Climate conditions in the near-term, mid-term and distant future for growing soybeans in Canada. *Canadian Journal of Plant Science*. **103**(2): 161-174. <https://doi.org/10.1139/cjps-2022-0233>

Craig J. Pearson, Delia Bucknell, Gregory P. Laughlin, Modelling crop productivity and variability for policy and impacts of climate change in eastern Canada, Environmental Modelling & Software, Volume 23, Issue 12, 2008, Pages 1345-1355, ISSN 1364-8152, <https://doi.org/10.1016/j.envsoft.2008.02.008>.

Curtis, B. (2022, November 30). *Ontario grains need a drink*. Good in Every Grain. <https://goodineverygrain.ca/2020/07/07/ontario-grains-need-a-drink/>

ISSN 0308-521X, <https://doi.org/10.1016/j.agsy.2016.12.013>.

Jerry L. Hatfield, John H. Prueger, Temperature extremes: Effect on plant growth and development, Weather and Climate Extremes, Volume 10, Part A, 2015, Pages 4-10, ISSN 2212-0947, <https://doi.org/10.1016/j.wace.2015.08.001>.

McDonald, I. (2021, April 1). *Water management*. Ontario Grain Farmer. <https://ontariograinfarmer.ca/2020/10/01/water-management-3/>

Mhawej, M., Gao, X., Reilly, J. M., & Abunnasr, Y. (2022). *SEBALIGEE v2: Global Evapotranspiration Estimation Replacing .*. Cambridge: Massachusetts Institute of Technology. <https://globalchange.mit.edu/sites/default/files/MITJPSPGC_Rpt362.pdf>

Mukesh Dev Bhattarai, Silvia Secchi, Justin Schoof, Projecting corn and soybeans yields under climate change in a Corn Belt watershed, Agricultural Systems, Volume 152, 2017, Pages 90-99,

Naeem A. Abbasi, Abderrachid Hamrani, Chandra A. Madramootoo, Tiequan Zhang, Chin S. Tan, Manish K. Goyal, Modelling carbon dioxide emissions under a maize-soy rotation using machine learning, Biosystems Engineering, Volume 212, 2021, Pages 1-18, ISSN 1537-5110, <https://doi.org/10.1016/j.biosystemseng.2021.09.013>

Sun, R., Kulshreshtha, S., & Nolan, J. (2021, December). *Challenges and opportunities from farm-level impacts of climate change on agriculture based on agent-based modelling simulation*. NASA/ADS. <https://ui.adsabs.harvard.edu/abs/2021AGUFMGC32B..06S/abstract>

Wentian He, J.Y. Yang, C.F. Drury, W.N. Smith, B.B. Grant, Ping He, B. Qian, Wei Zhou, G. Hoogenboom, Estimating the impacts of climate change on crop yields and N2O emissions for conventional and no-tillage in Southwestern Ontario, Canada, Agricultural Systems, Volume 159, 2018, Pages 187-198, ISSN 0308-521X, <https://doi.org/10.1016/j.agsy.2017.01.025>.

[Yu, Y.](https://www.emerald.com/insight/search?q=Yan%20Yu), [Tian, Q.](https://www.emerald.com/insight/search?q=Qingsong%20Tian) and [Yan, F.](https://www.emerald.com/insight/search?q=Fengxian%20Yan) (2022), "Climate change and its impact on rice acreage in high-latitude regions of China: an estimation by machine learning", [*International Journal of Climate Change Strategies and Management*](https://www.emerald.com/insight/publication/issn/1756-8692), Vol. 14 No. 4, pp. 313-331. <https://doi.org/10.1108/IJCCSM-11-2020-0124>