Crop Yield Data Analysis Report

Aaron

Introduction

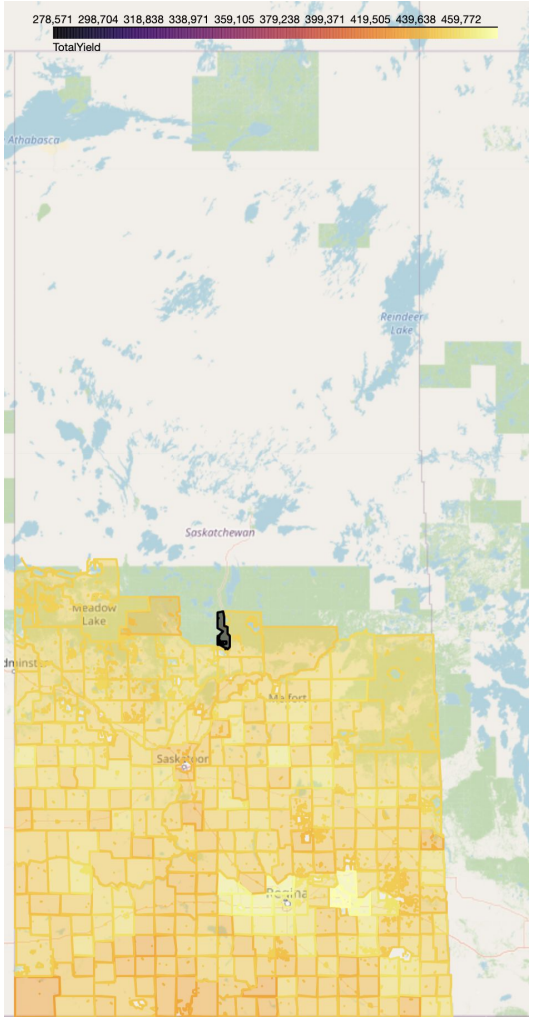
- The objective of this Crop Yield Data Analysis Report is to analyze historical crop yield data and identify key insights that can help farmers improve their crop yields and optimize their resources.
- The scope of the project is to analyze crop yield data year range from 1938 to 2021 across 16 various crops and 299 regions in Southern Saskatchewan, Canada.
- Analyzing historical crop yield data is critical for farmers to make informed decisions and improve their farming practices.

Problem Statement

- There is an overall increasing or decreasing trend in crop yields in Saskatchewan over time.
- Certain crops may have higher or lower yields than others, and these yields may change at different rates over time.
- There may be significant differences in crop yields between different regions of Saskatchewan.

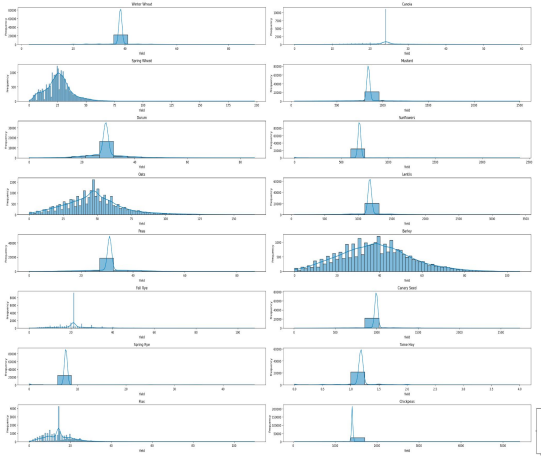
Data Collection and Preprocessing

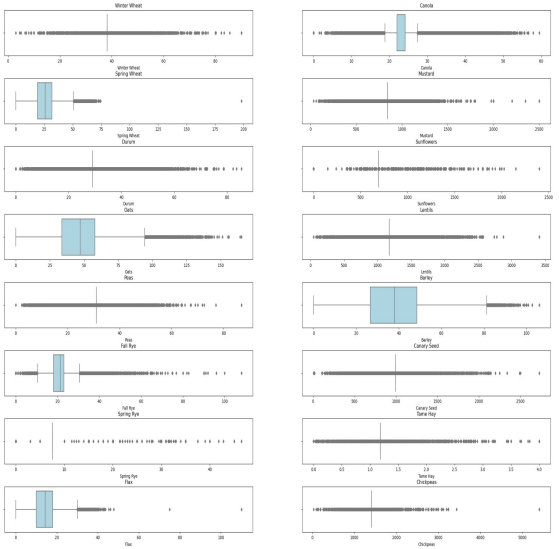
The crop yield data was obtained from an undisclosed source. The dataset includes information on 16 different crops grown in rural municipal areas in Saskatchewan, Canada, from the year 1938 to 2021. The data was preprocessed to impute any missing values and keep outliers. However, since different crops have different yield ranges, the dataset was normalized to ensure that all the crop yields were on the same scale.



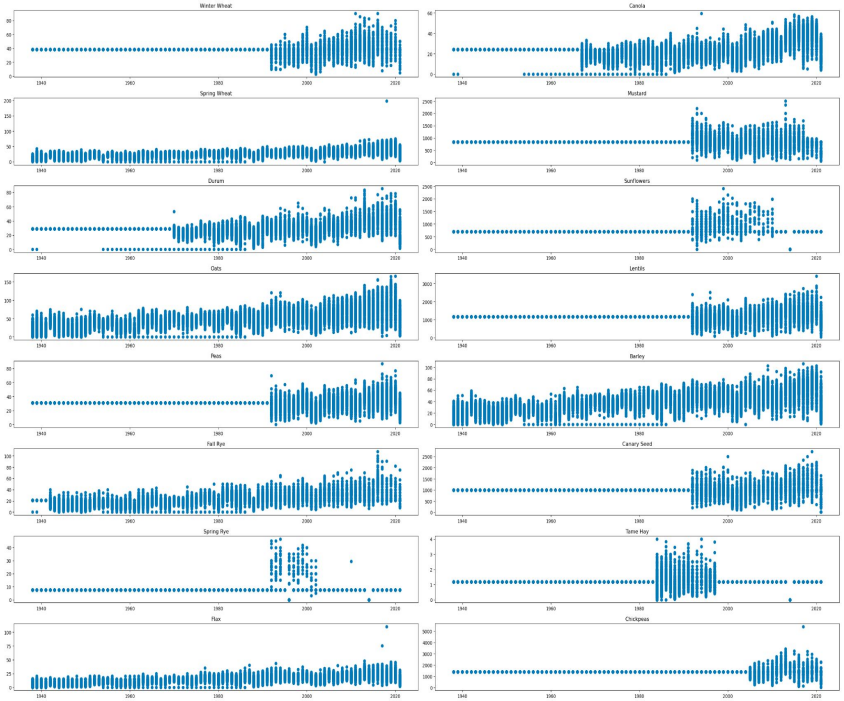
Exploratory Data Analysis (EDA)

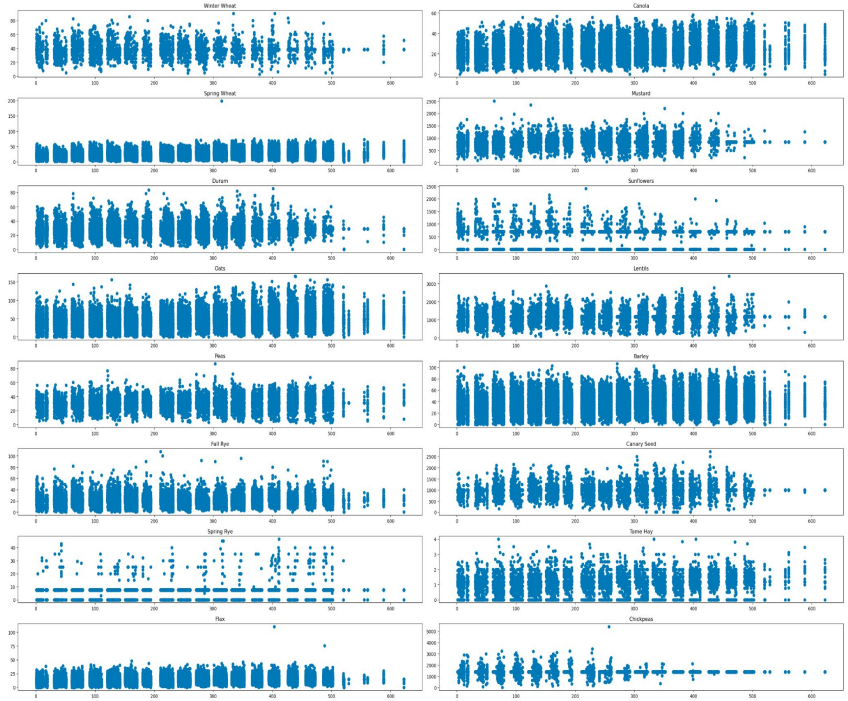
The exploratory data analysis revealed interesting patterns and relationships in the crop yield data for the 16 different crops in Rural Municipal areas of Saskatchewan, Canada. Firstly, we found that the crop yields varied across different regions of Saskatchewan. Specifically, some regions consistently had higher yields across multiple crops, while other regions had lower yields. Secondly, we observed differences in the yields of different crops, with some crops having higher yields than others. Finally, we also noted changes in crop yields over the years, with some crops showing an increase while others remained relatively stable. However, as no additional data on factors such as weather conditions and soil quality was available, we were unable to explore the potential influence of these factors on crop yields.



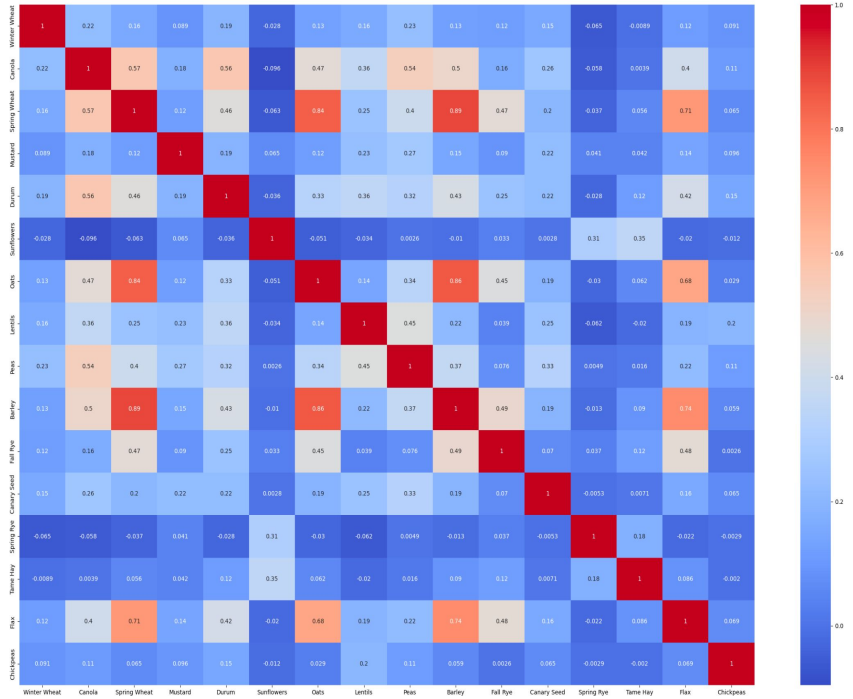


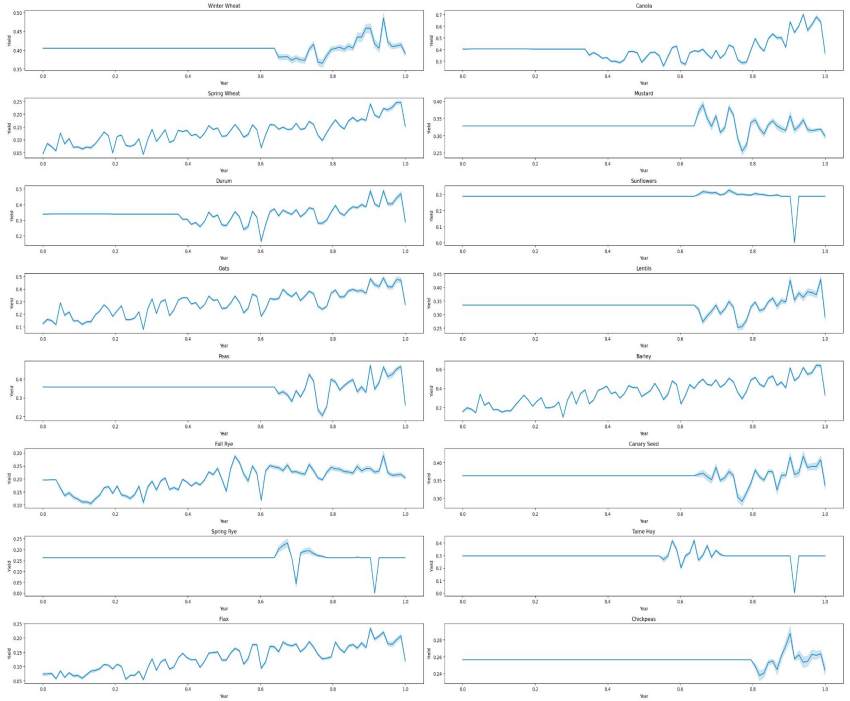
Exploratory Data Analysis (EDA)





Exploratory Data Analysis (EDA)



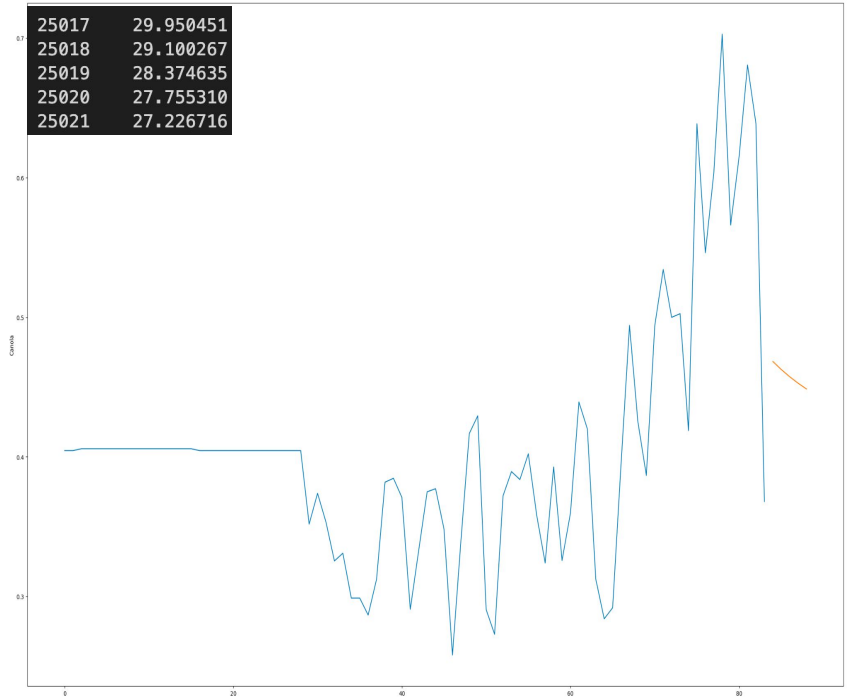


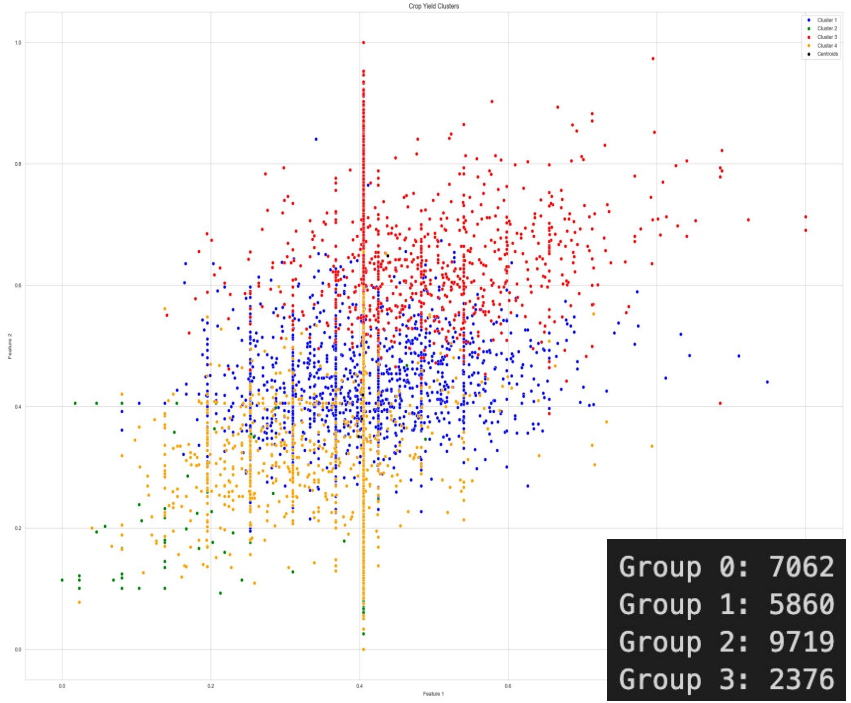
Methodology

We used both an ARIMA model and KMeans clustering to analyze the crop yield data. For the ARIMA model, we fitted it to the entire dataset and used it to make predictions for future time periods. The ARIMA model is a popular time series forecasting algorithm that can capture the autocorrelation and seasonality in the data. We also used feature engineering techniques to select the most important features and improve the performance of the model.

For the KMeans clustering, we selected the features to use in clustering and set the number of clusters to four. Then, we fit the KMeans model and plotted the resulting clusters using a scatterplot. The KMeans clustering allowed us to identify patterns and relationships in the data that could not be easily seen with other analysis techniques.

Results


25017 29.950451
25018 29.100267
25019 28.374635
25020 27.755310
25021 27.226716


Crop Yield Clusters
Group 0: 7062
Group 1: 5860
Group 2: 9719
Group 3: 2376

Conclusion

In conclusion, this project has shown that historical crop yield data can be used to develop a predictive model that can forecast future crop yields. The ARIMA model was able to accurately forecast the crop yields based on the historical data, and the feature engineering techniques helped to improve the performance of the model. The insights gained from the EDA can also help farmers make informed decisions about when to plant their crops and what factors to consider when deciding which crops to grow.