Mitigating Crop Yield Risk in Saskatchewan

Insightful Risk Management for Agriculture

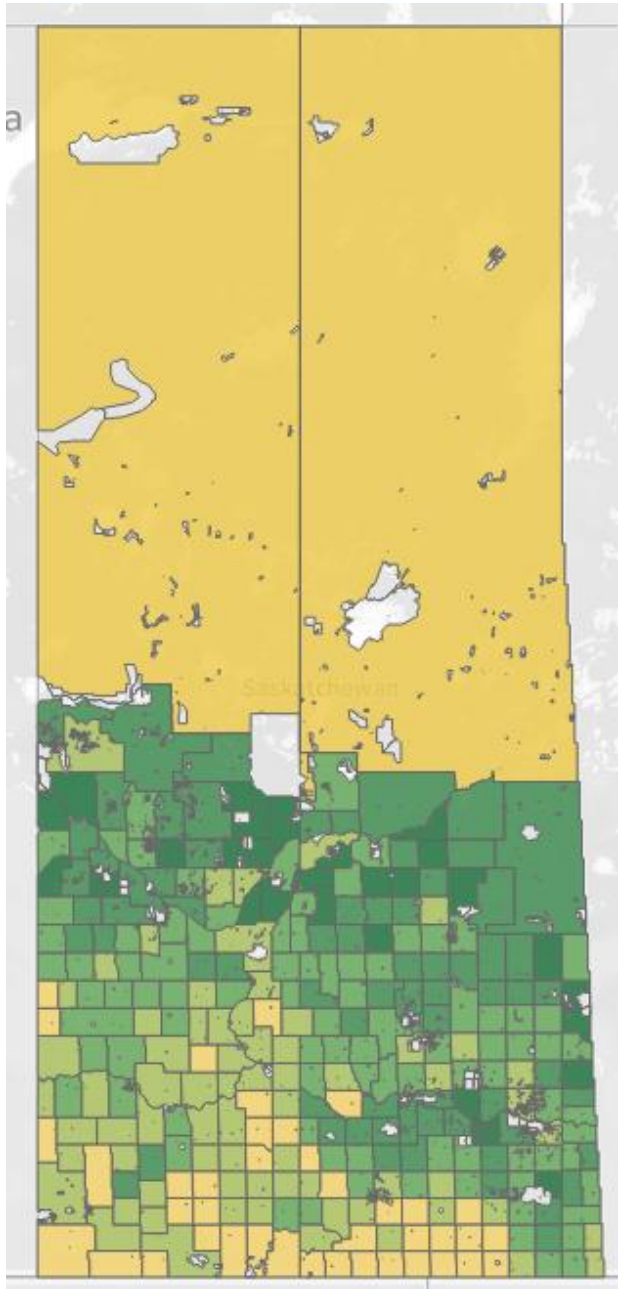
Insightful Risk Management for Agriculture

Objective:

An analysis that leverages data to provide improved risk assessment for the insurance and financing industry in Saskatchewan.

Scope:

- 296 Rural Municipalities.
- 15 Crop Types: Winter Wheat, Canola, Spring Wheat, Mustard, Durum, Oats, Lentils, Peas, Barley, Fall Rye, Canary Seed, Flax.
- Yield data from 1940 to 2021.

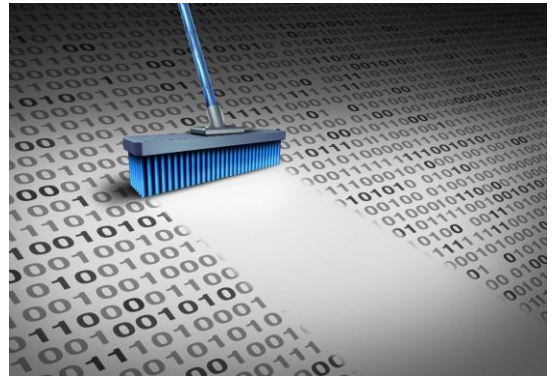


Problem Statement

- Insurance companies to set appropriate premiums.
- More effective pricing by insurance and financing companies, resulting in fairer premiums and interest rates.
- Better management of risk and protecting the company's bottom line.

Data Collection & Preprocessing

- Data Source: Government of Saskatchewan



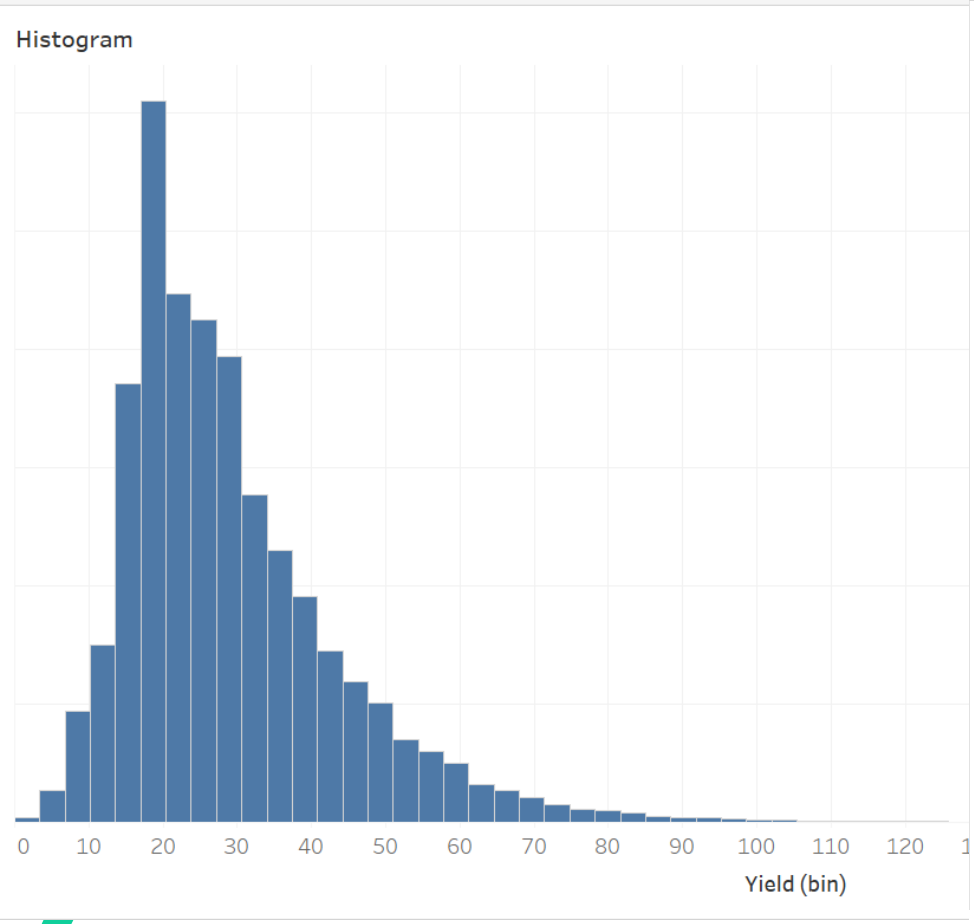
- Data Cleaning: Below Data was excluded from the analysis

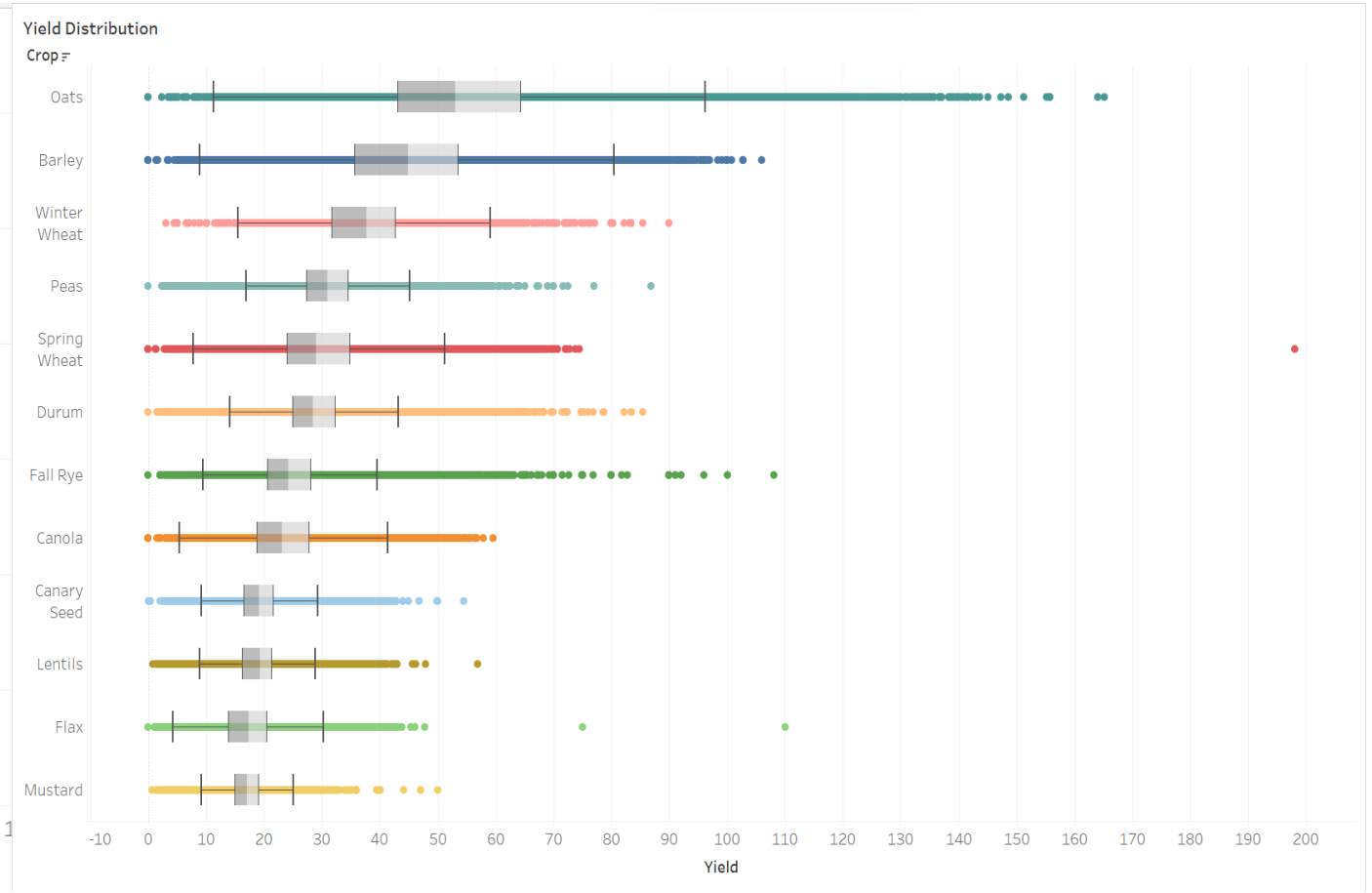
1. Before 1965.
2. RM number 521.
3. The crops (Tame Hay, Spring Ray, Sunflower and Chickpeas).
4. Replaced the remaining Null values for every RM with its mean.

- Data transformation: Pounds to Bushels.

Explanatory Data Analysis

- Data Distribution

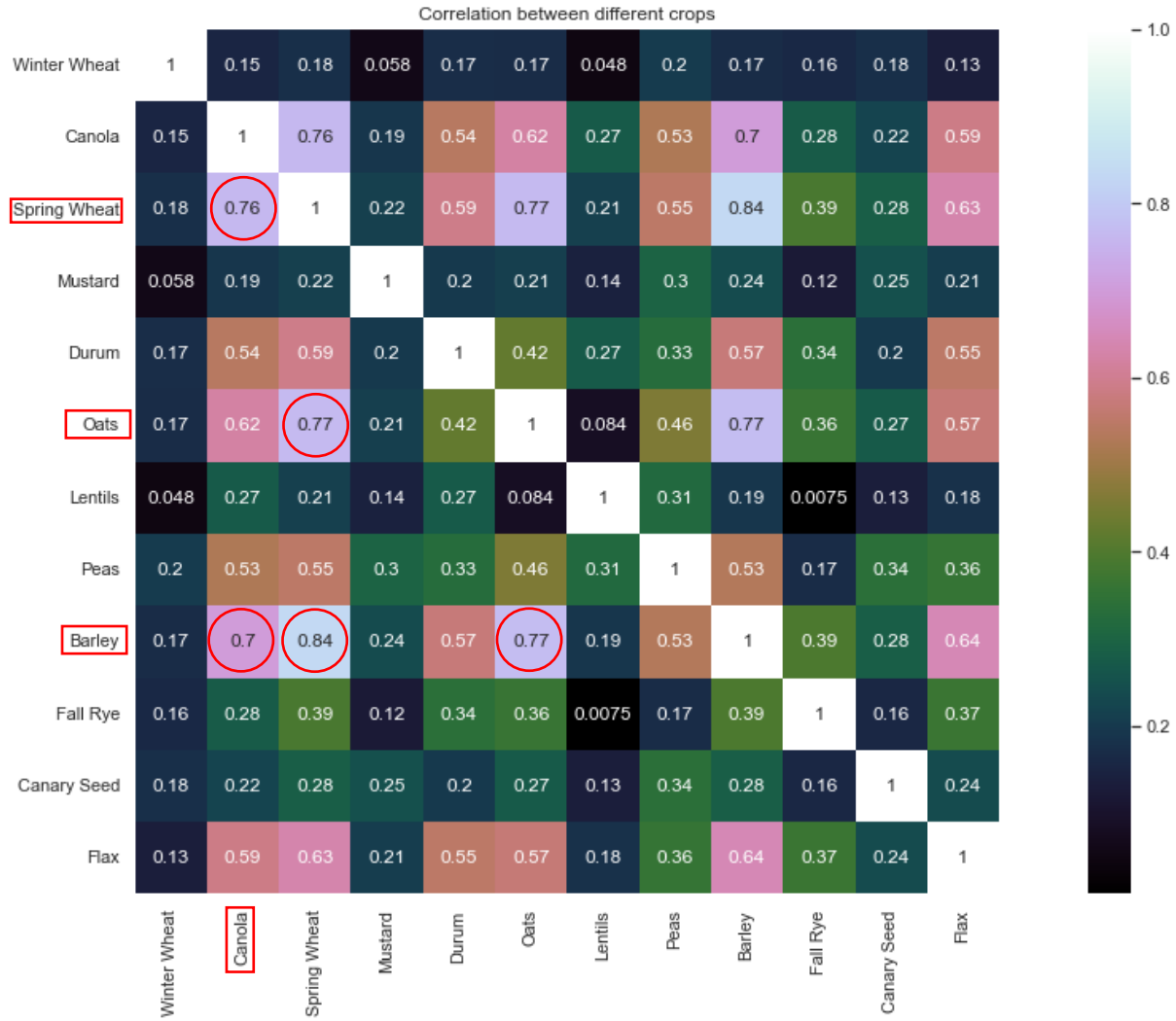




Explanatory Data Analysis

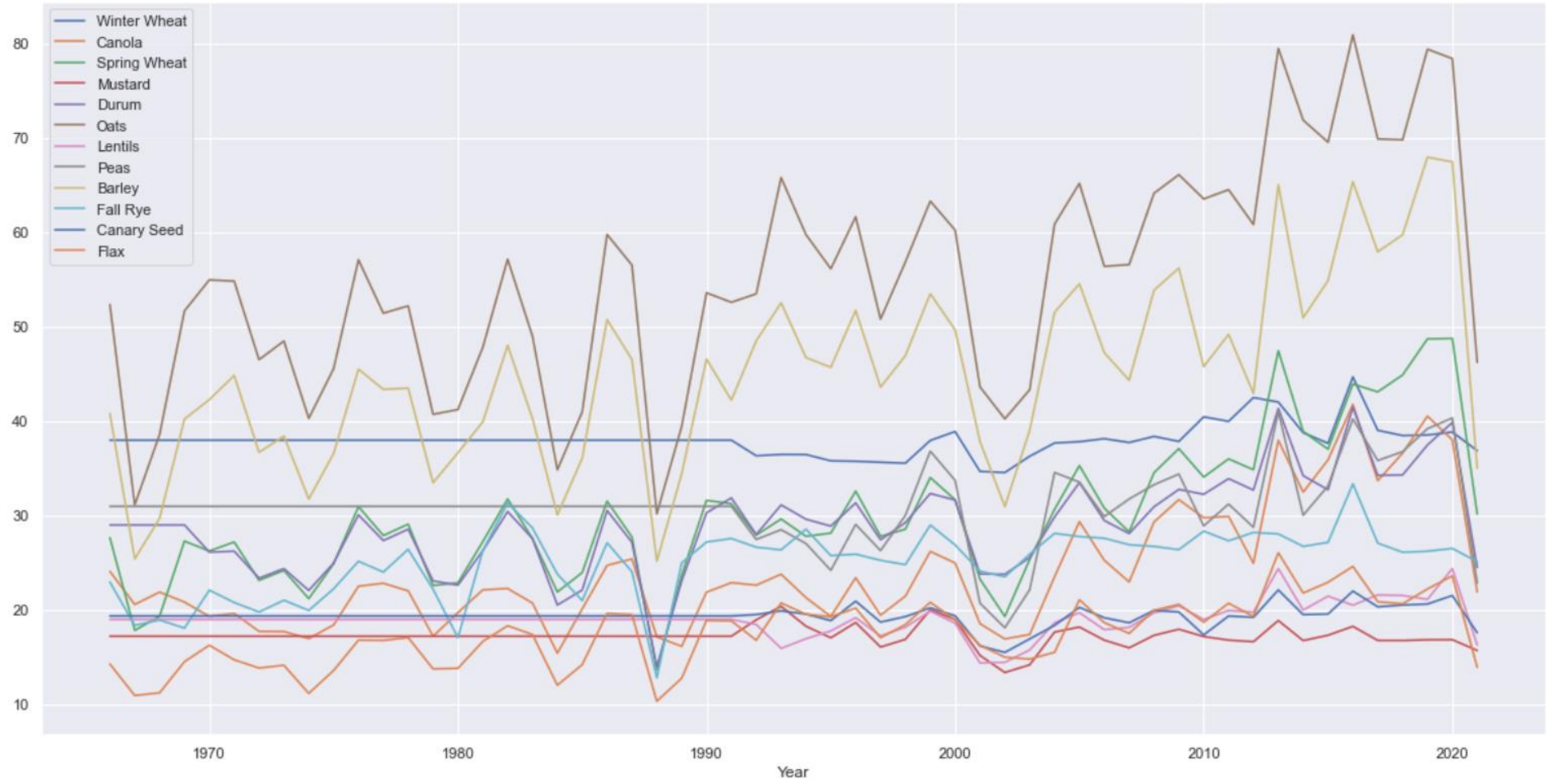
- Correlation Analysis:

There is strong positive Correlation between (Oats, Barley, Spring Wheat and Canola).



Explanatory Data Analysis

- **Time Series Analysis**



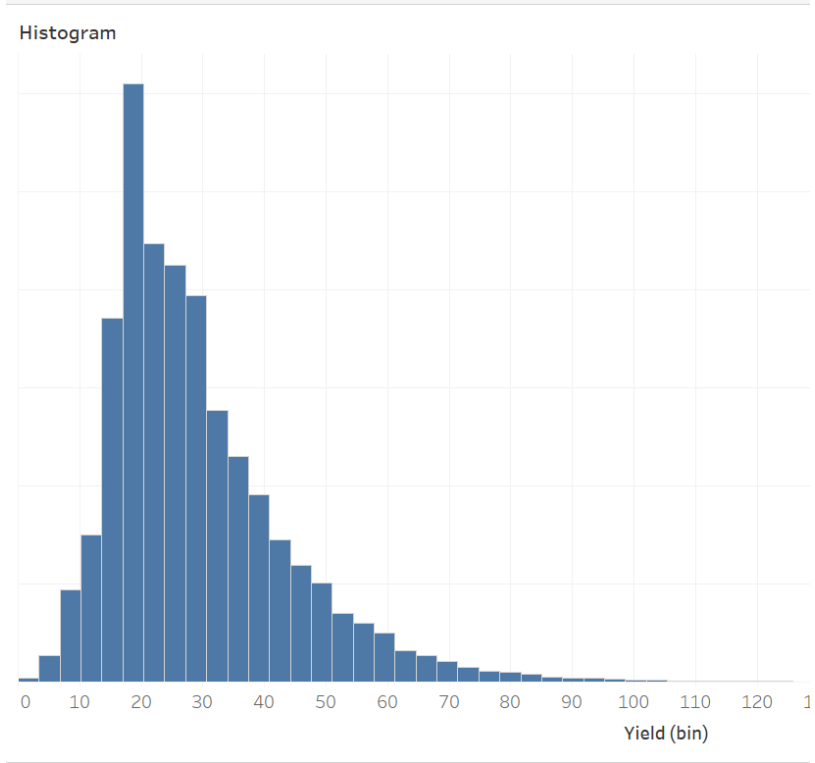
Explanatory Data Analysis

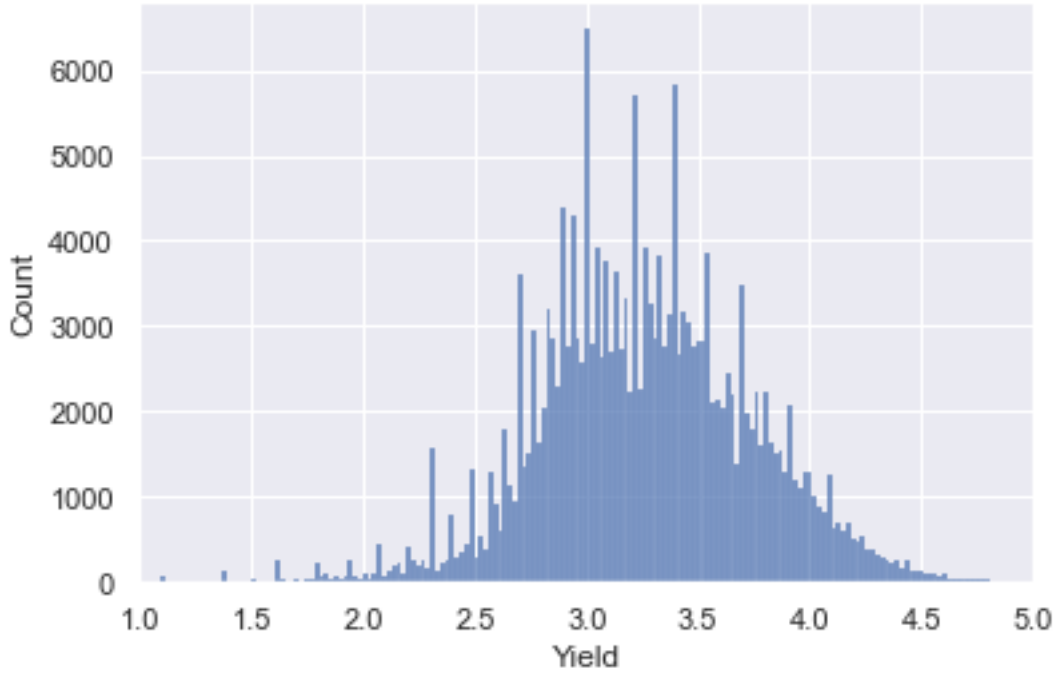
- **Autocorrelation**

Methodology

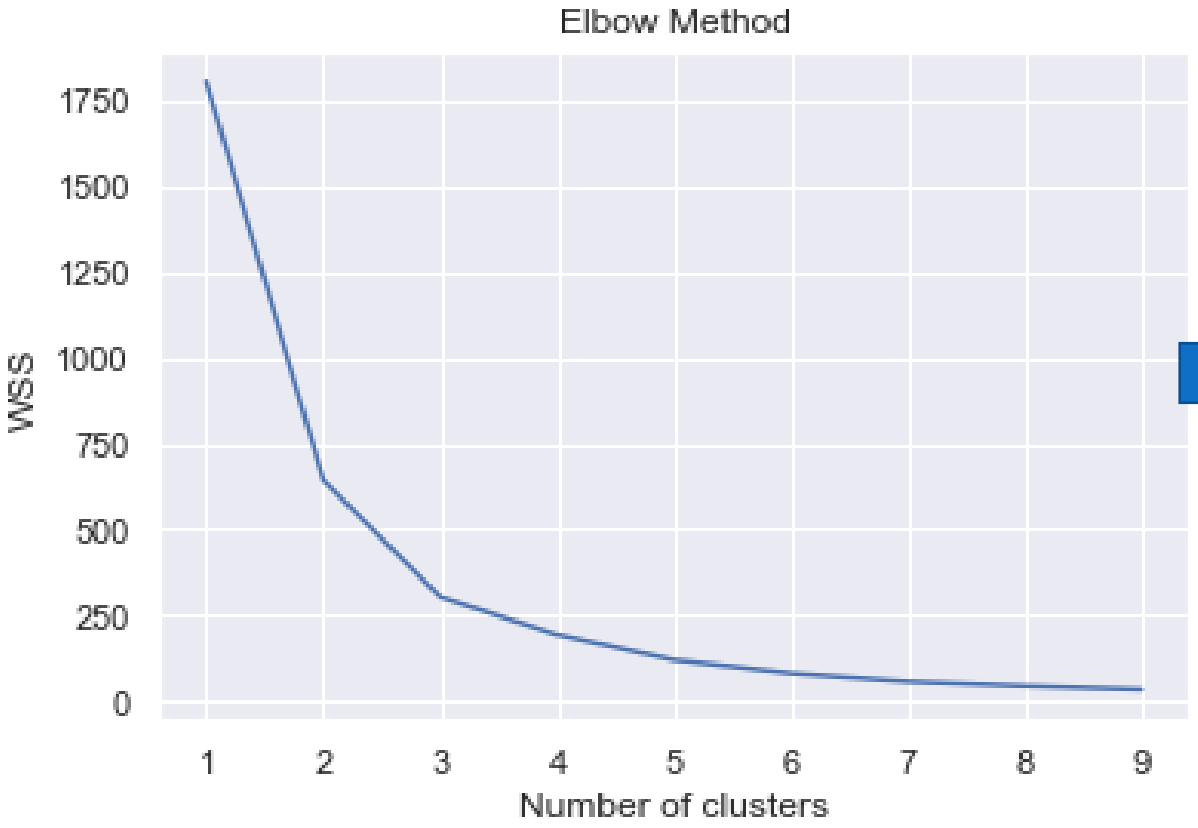
- For each crop type in the dataset, I calculated the yield mean as a representation of the expected value.
- To better represent the variability of the yield data as a risk factor, I calculated the coefficient of variation (CV) for each crop type. The CV allows for a comparison of the variability between different Rural Municipalities (RM) even if they have different yield means.
- Using the CV values, I identified the highest risk areas based on yield variability.
- To group similar crops, I conducted unsupervised machine learning using clustering techniques on the yield data for each RM.
- I used supervised machine learning to predict the yield for each crop type in each RM based on historical yield data, weather patterns, and other relevant factor.

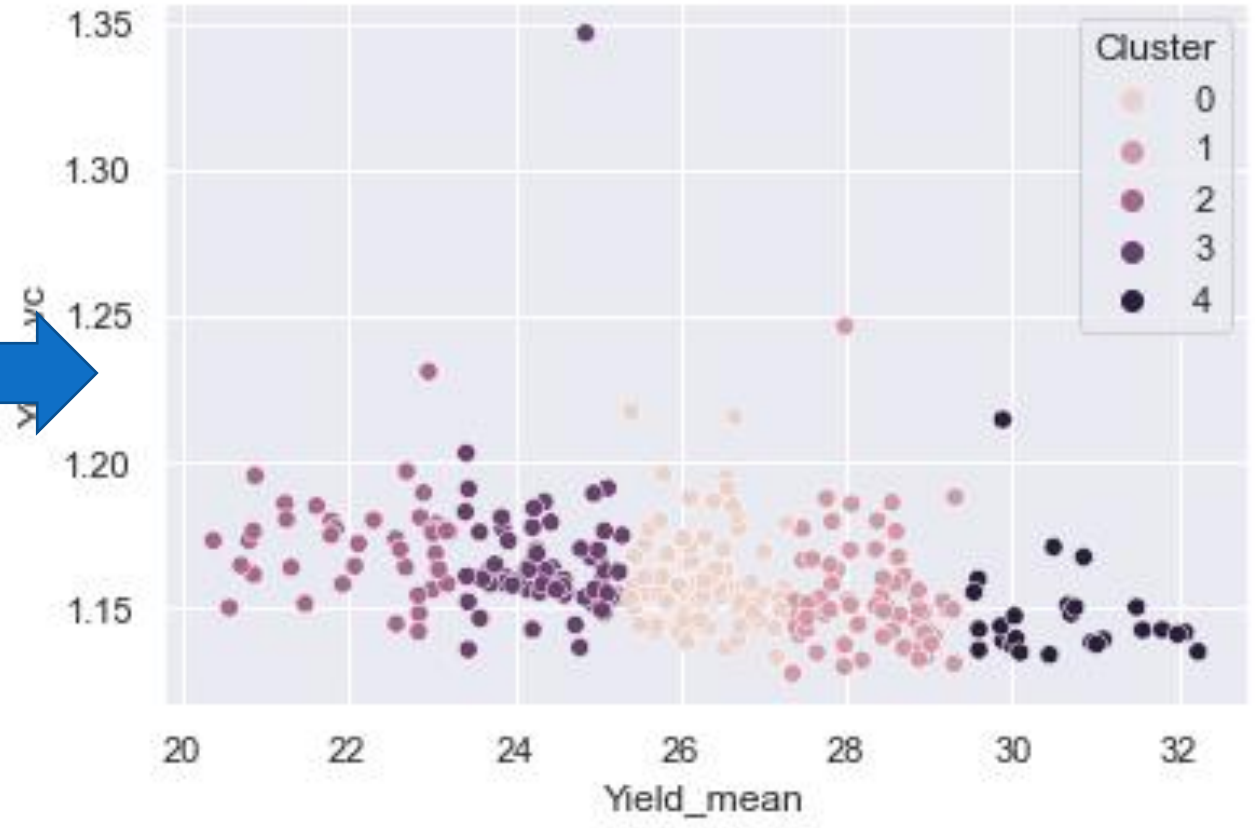
Un-Supervised Machine Learning : K-Means Clustering



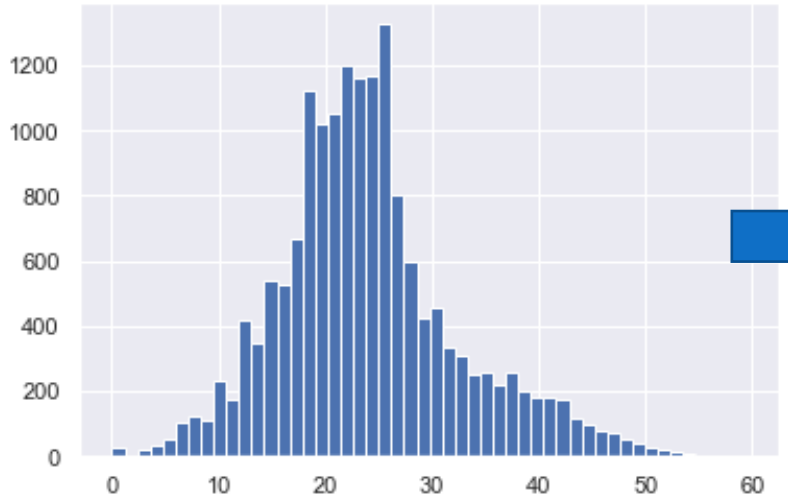


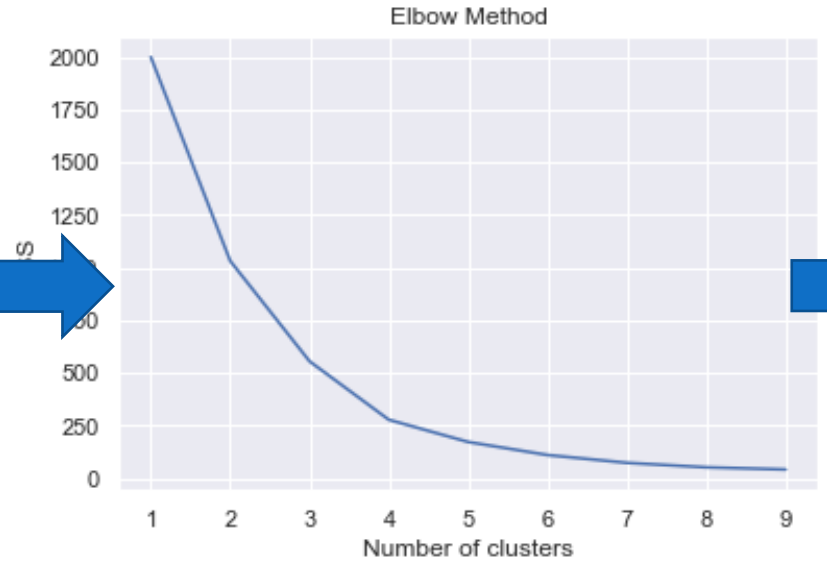
Un-Supervised Machine Learning : K-Means Clustering

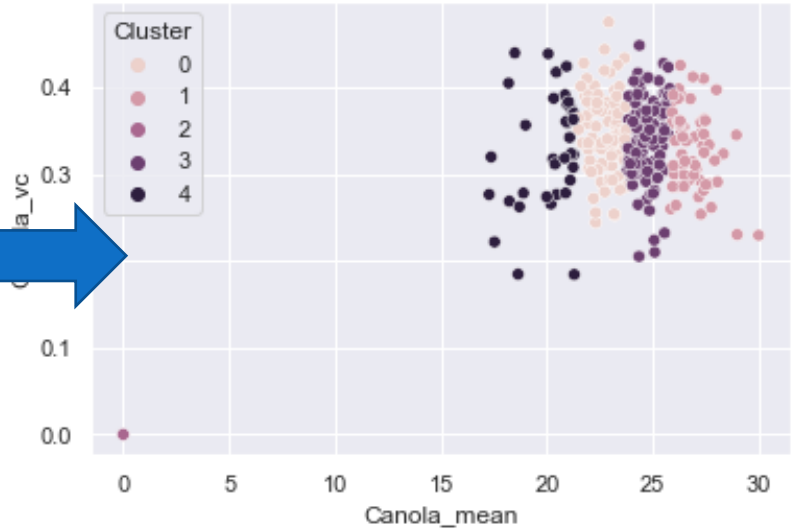


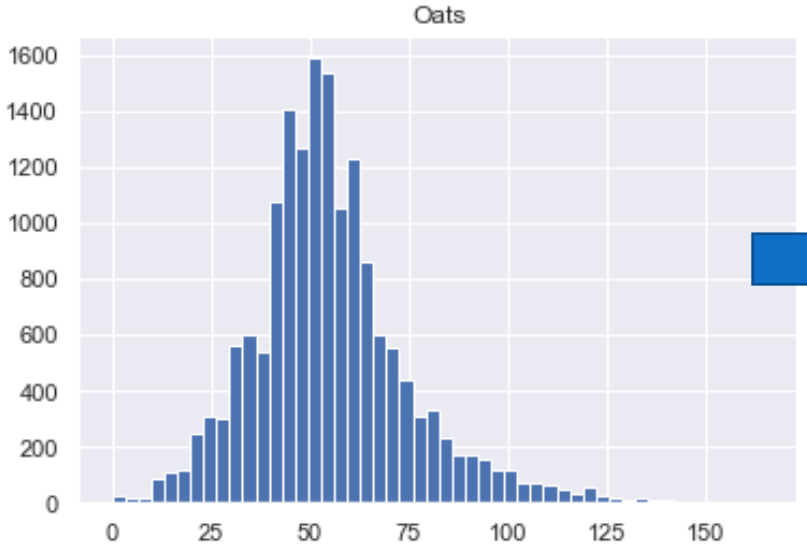


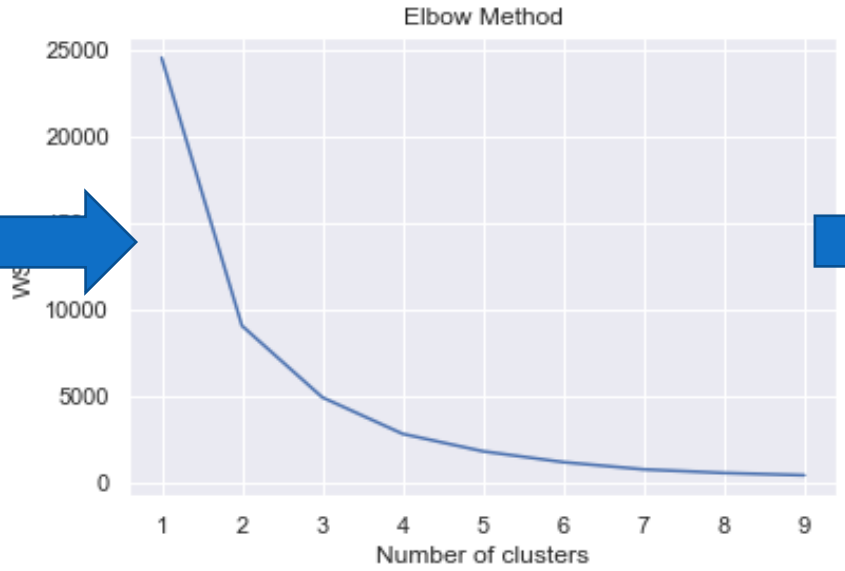
Un-Supervised Machine Learning : K-Means Clustering

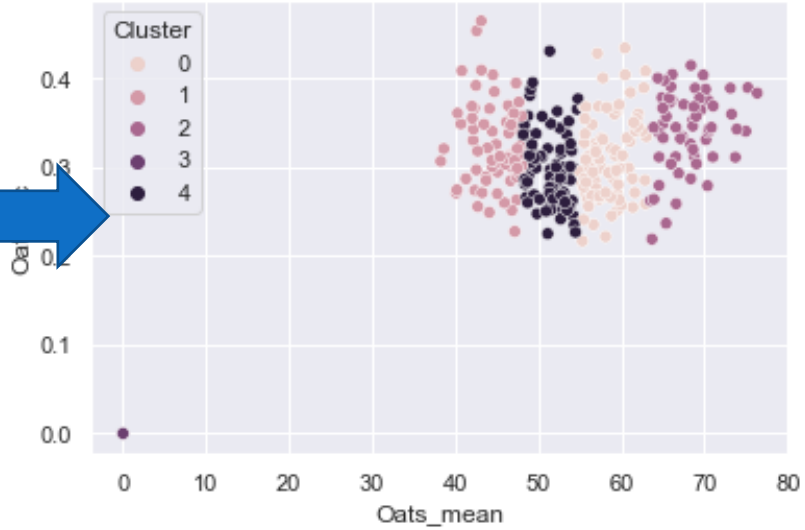




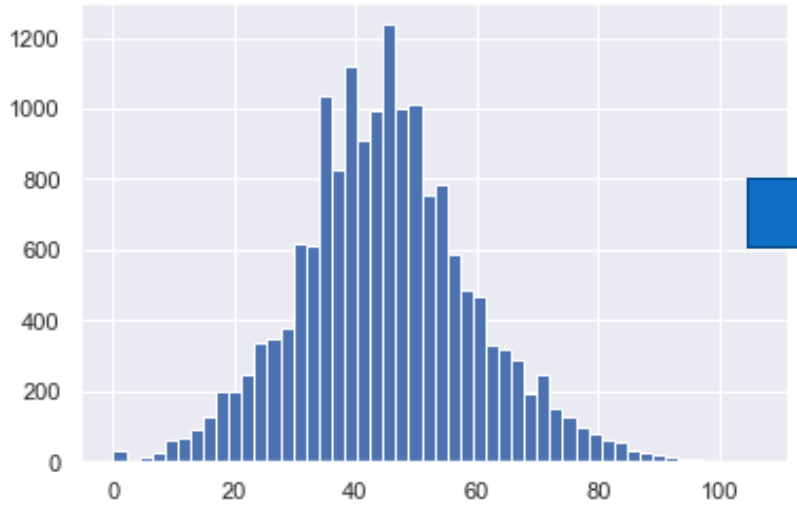


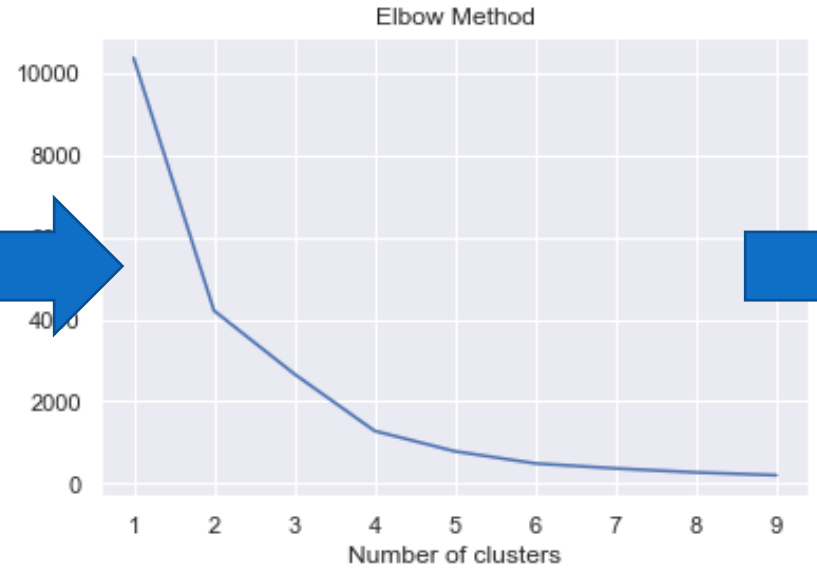


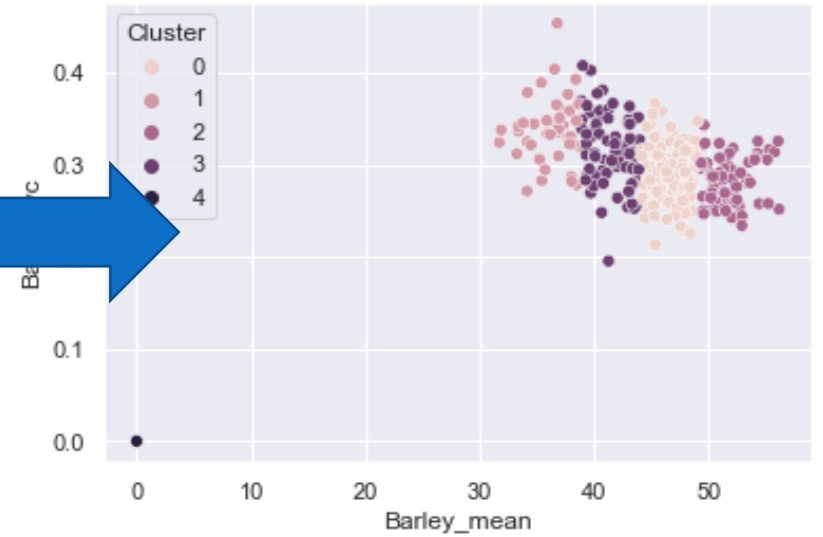


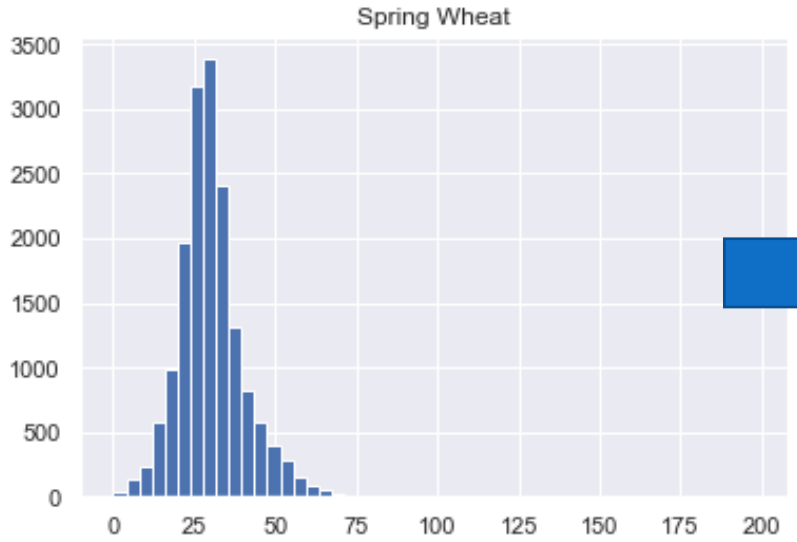


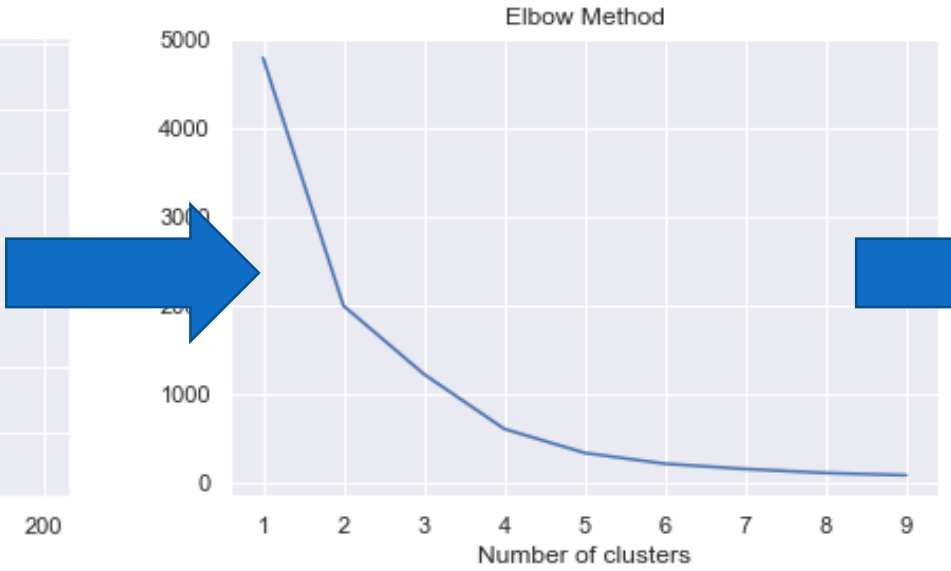
Un-Supervised Machine Learning : K-Means Clustering

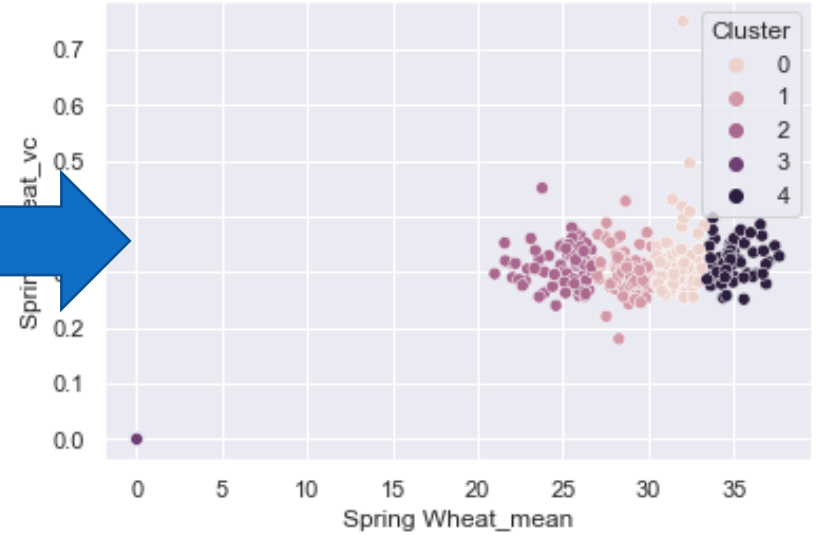






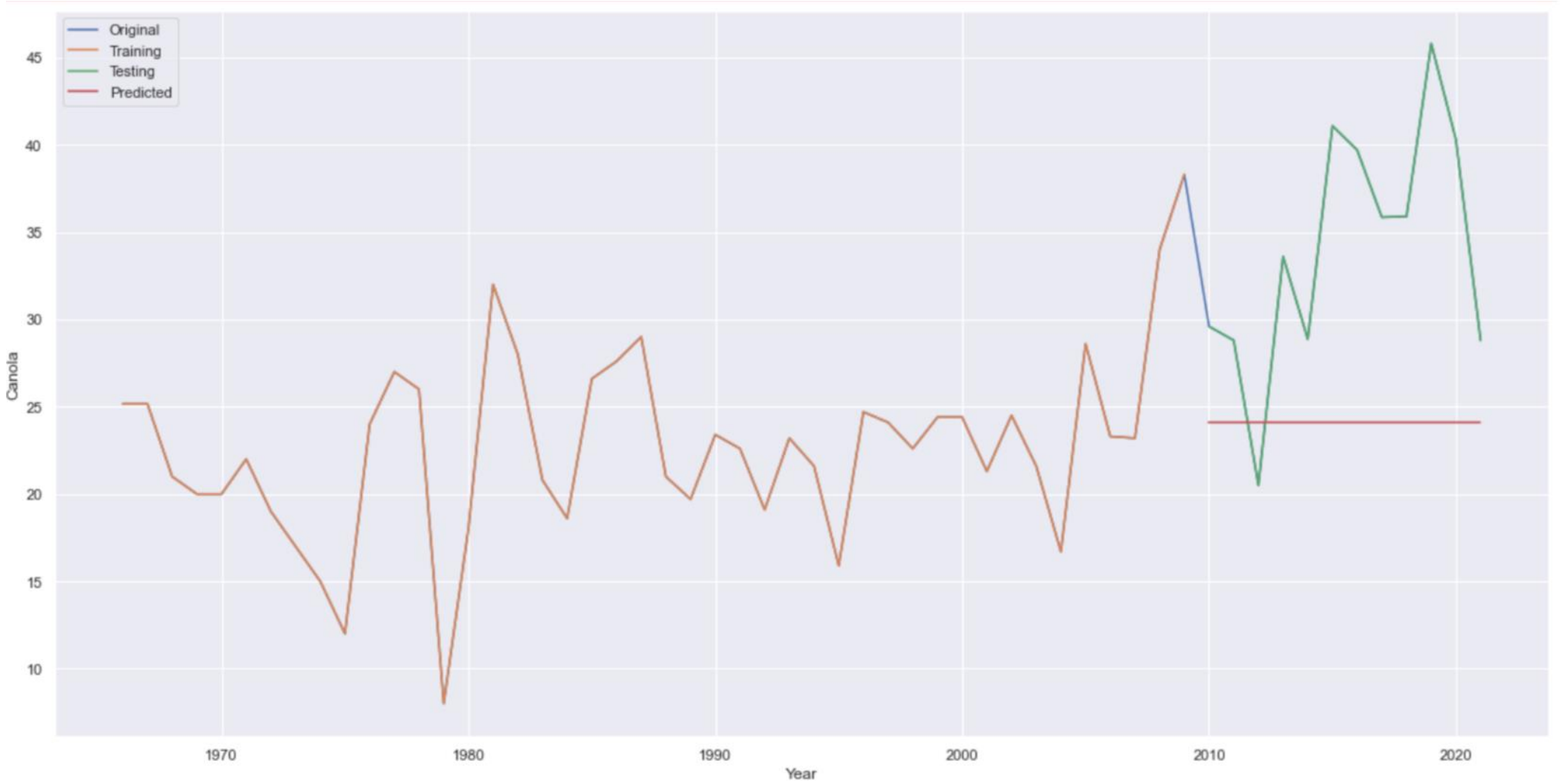






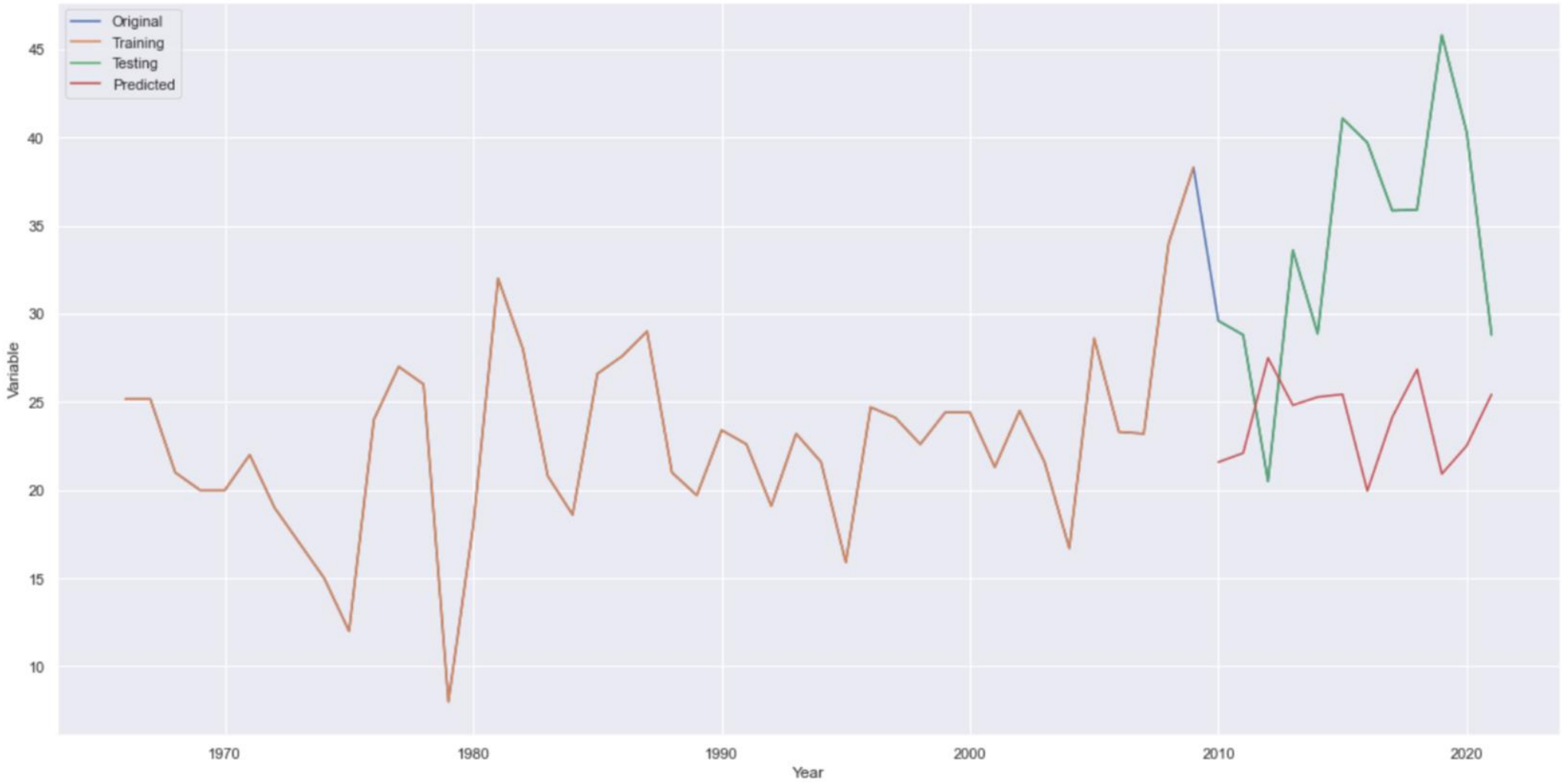
Supervised Machine Learning : Auto_Regression (Canola, RM#213)

MAE = 10.6



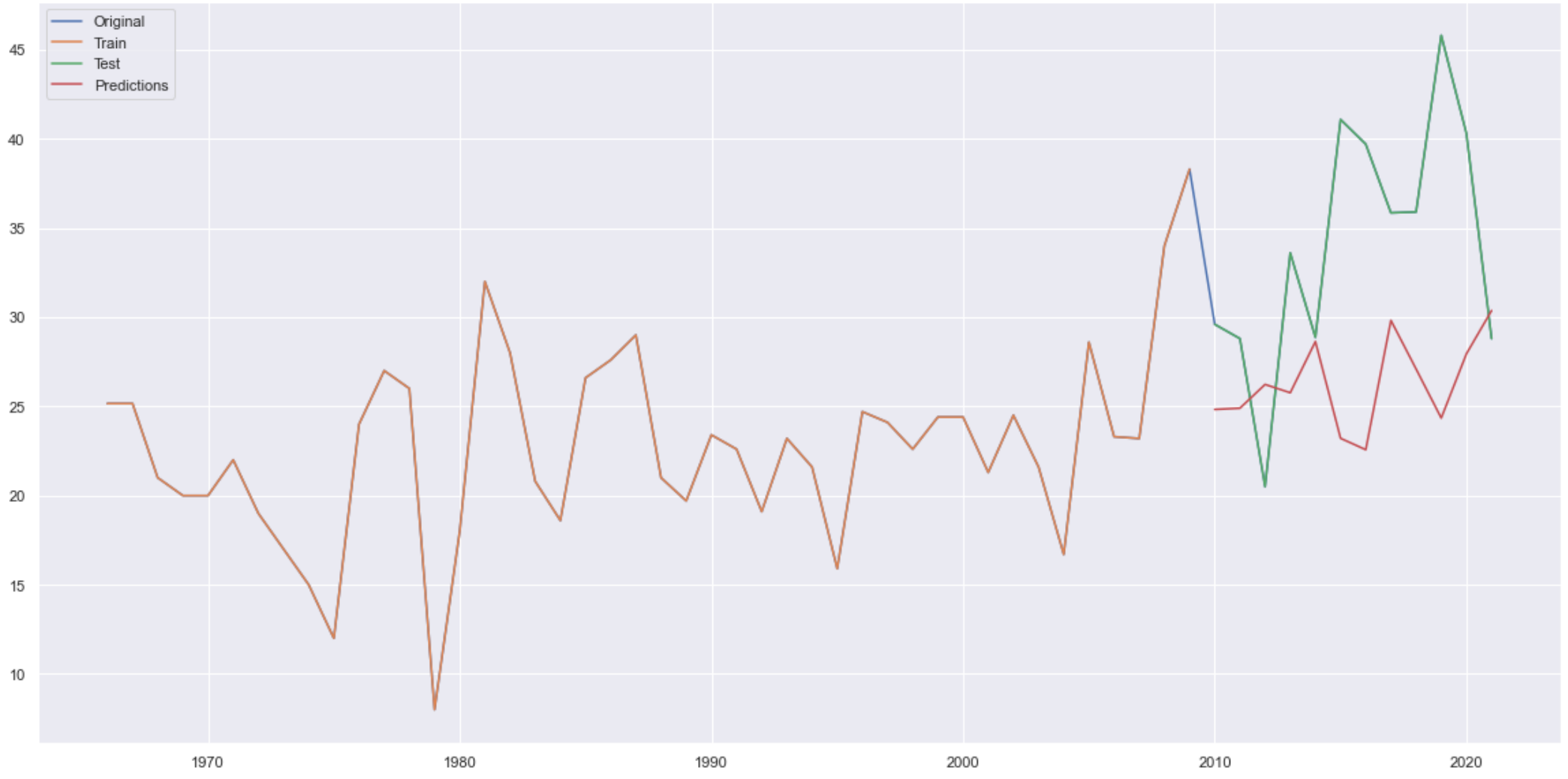
Supervised Machine Learning : ARIMA (Canola, RM#213)

MAE = 11.5



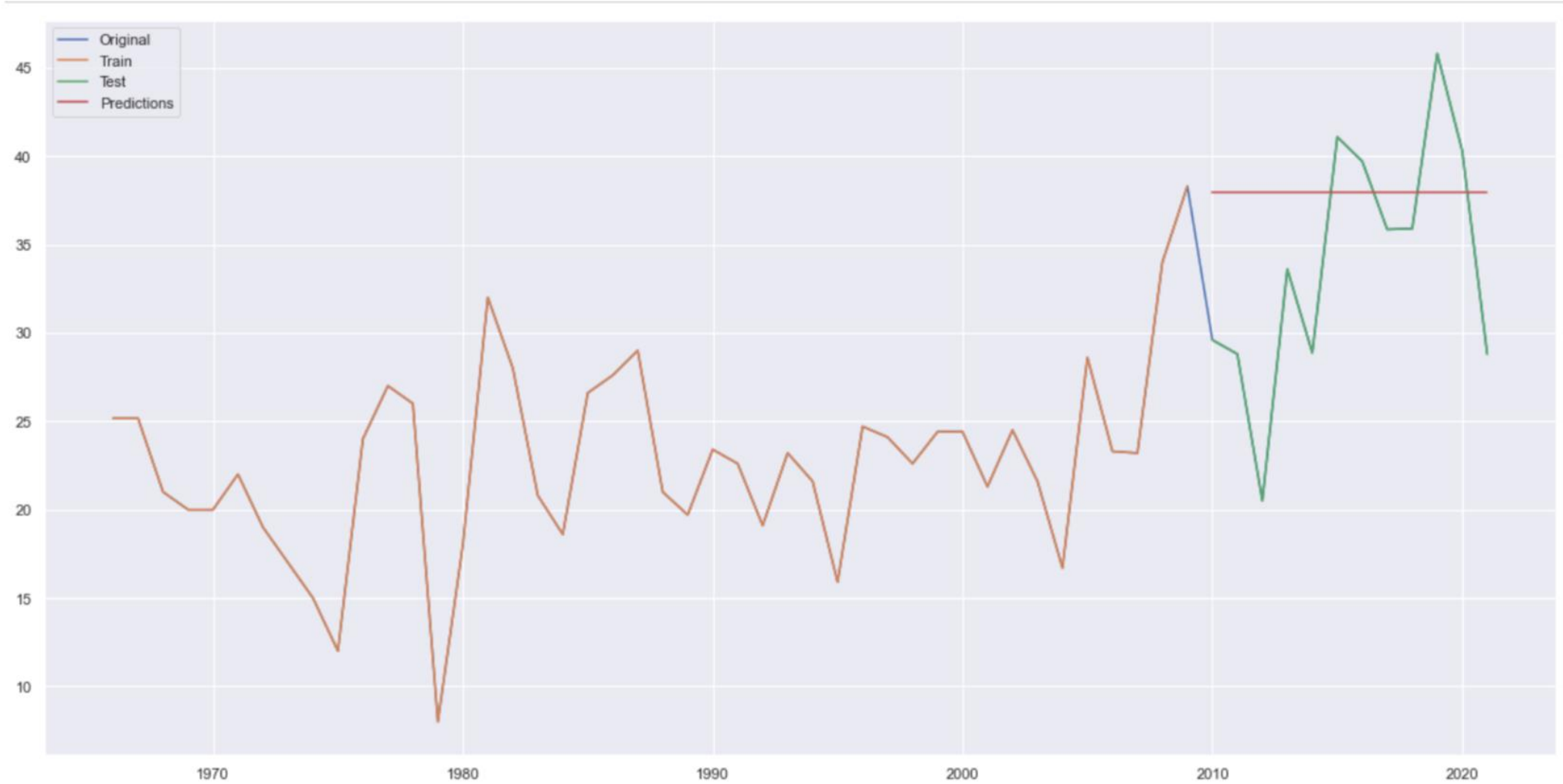
Supervised Machine Learning : Exponential Smoothing (Canola, RM#213)

MAE = 8.9



Supervised Machine Learning : Gradient Boosting Regressor (Canola, RM#213)

MAE = 6.4



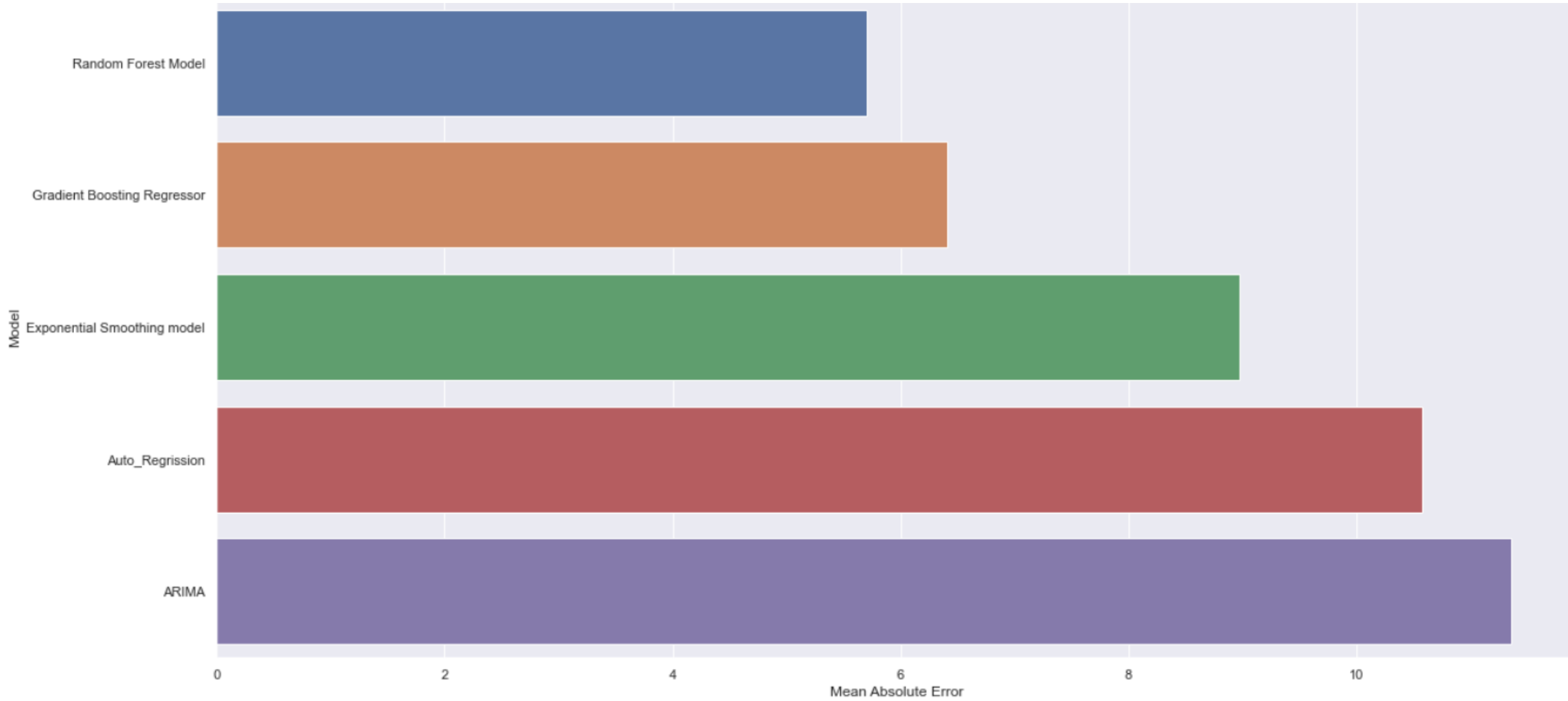
Supervised Machine Learning : Random Forest (Canola, RM#213)

MAE = 5.7

Supervised Machine Learning : Models Comparison (Canola, RM#213)

Percentage of Error

18%



Results

- Interactive dashboard for analyzing yield data in Saskatchewan's rural municipalities
- Visualizations:
    - Clustering scatter plot to identify groups of similar RMs based on yield and coefficient of variation
    - Spatial map with color-coding of yield groups
    - Time series chart for each crop showing historical patterns
    - Box-and-whisker plot for each crop to compare yield distributions and identify outliers
- Dashboard is interactive, allowing users to filter data based on specific criteria
- Provides comprehensive view of yield data for various crops in Saskatchewan's rural municipalities.

Conclusions

- After identifying the yield rank for rural municipalities and the coefficient of variance as an indicator of yield consistency (Risk).
- 1) identification of clusters of RMs with similar yield and variability,
- 2) spatial patterns of high and low yields.
- 3) trends in the data.

- I would recommend the following:

1. Develop mitigation strategies.
2. Develop insurance products.
3. Develop financing products.

Future Work

- Improving the prediction model by including more factors such as:

1. Temperature Data.
2. Precipitation Data.
3. Soil types and quality.
4. Crop rotation history.
5. Pest and disease history.
6. Crops management practices (fertilizers, pesticides, planting and tillage)