

# Finding High Value Canola Farmland

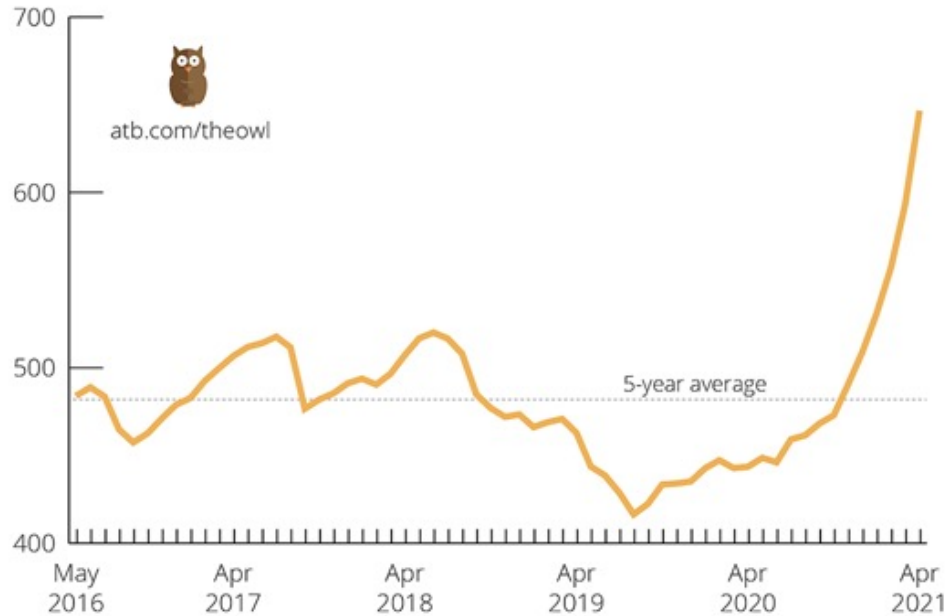
Agricultural Insights from  
Saskatchewan and Manitoba

By Clinton Boyda Oct 2023



## Canola prices, Alberta

\$ per metric tonne (includes rapeseed)



Source: Statistics Canada Table 32-10-0077-01

Graph by ATB Economics

# Agricultural Insights

## Objective:

Leverage Canola public stock price to calculate the value from yield data.

$$\text{Value} = \text{Crop Yield} \times \text{Stock Price}$$

## Scope:

- ▶ 397 Rural Municipalities (RM)
- ▶ Year 2010-2022 from SK and MB





# Problem Statement

Let's go shopping for the best Canola Valued Farmland!

- ▶ Where is the Top 10 Canola Yield RM locations?
- ▶ Where is the Top 10 Canola Valued RM locations?
- ▶ Are there any trends or discrepancies between yield and value?



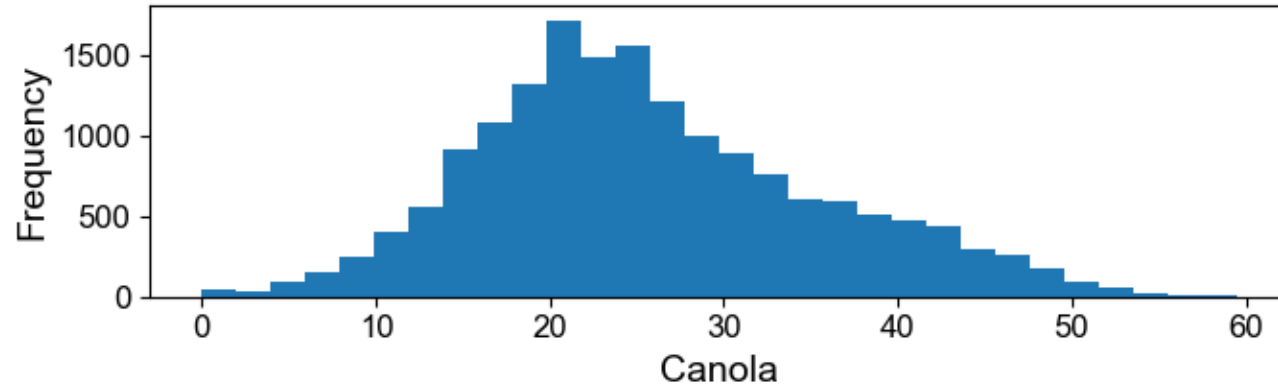
# Data Collection & Preprocessing

- ▶ Crop Yield Data for 397 RM's from Provincial Data Sources
- ▶ Geospatial Data found for 478 RMs but 4 still missing from Yield list
- ▶ Canola Future Stock Price from Investing.com limited to 2010-2022
- ▶ Transformed Yield to Bushels, Cleaned GIS RM names, Converted \$ per Bushel
- ▶ Scaled Yield and Value for GIS Comparison as % Diff

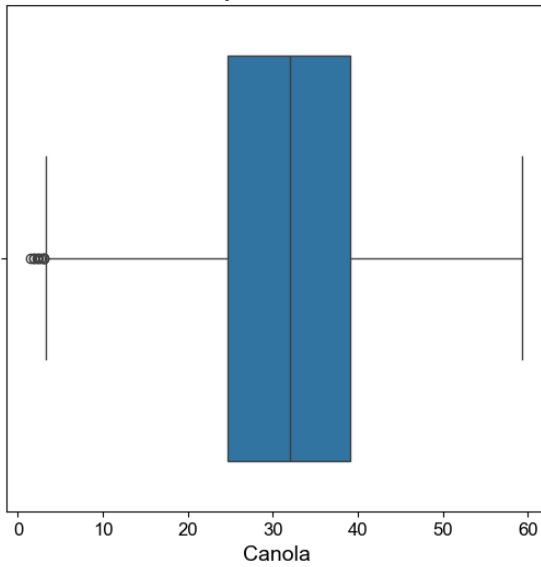
# Exploratory Data Analysis (EDA)

Quality of our Yield and Value Data

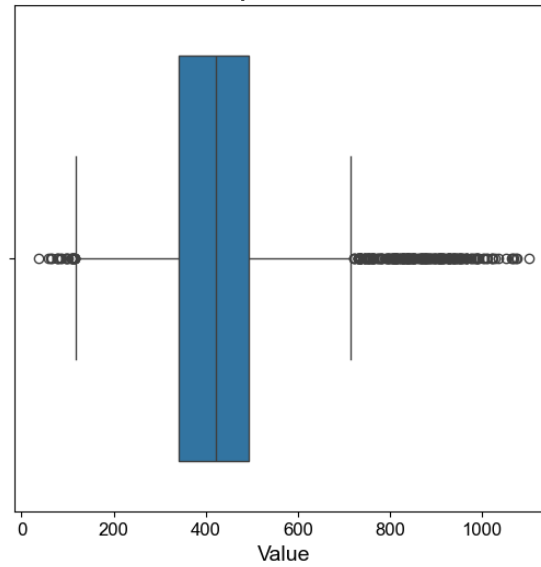
Canola Histogram



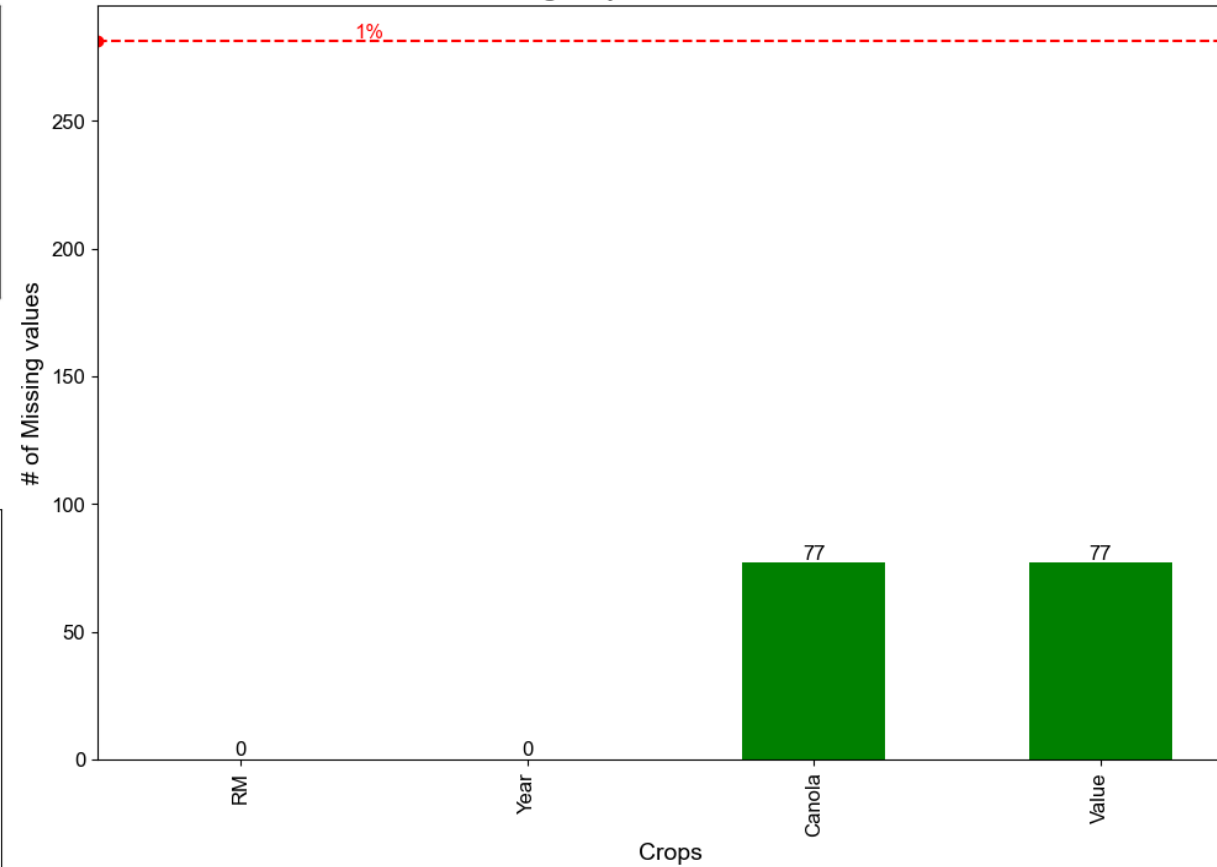
Boxplot of Canola



Boxplot of Value



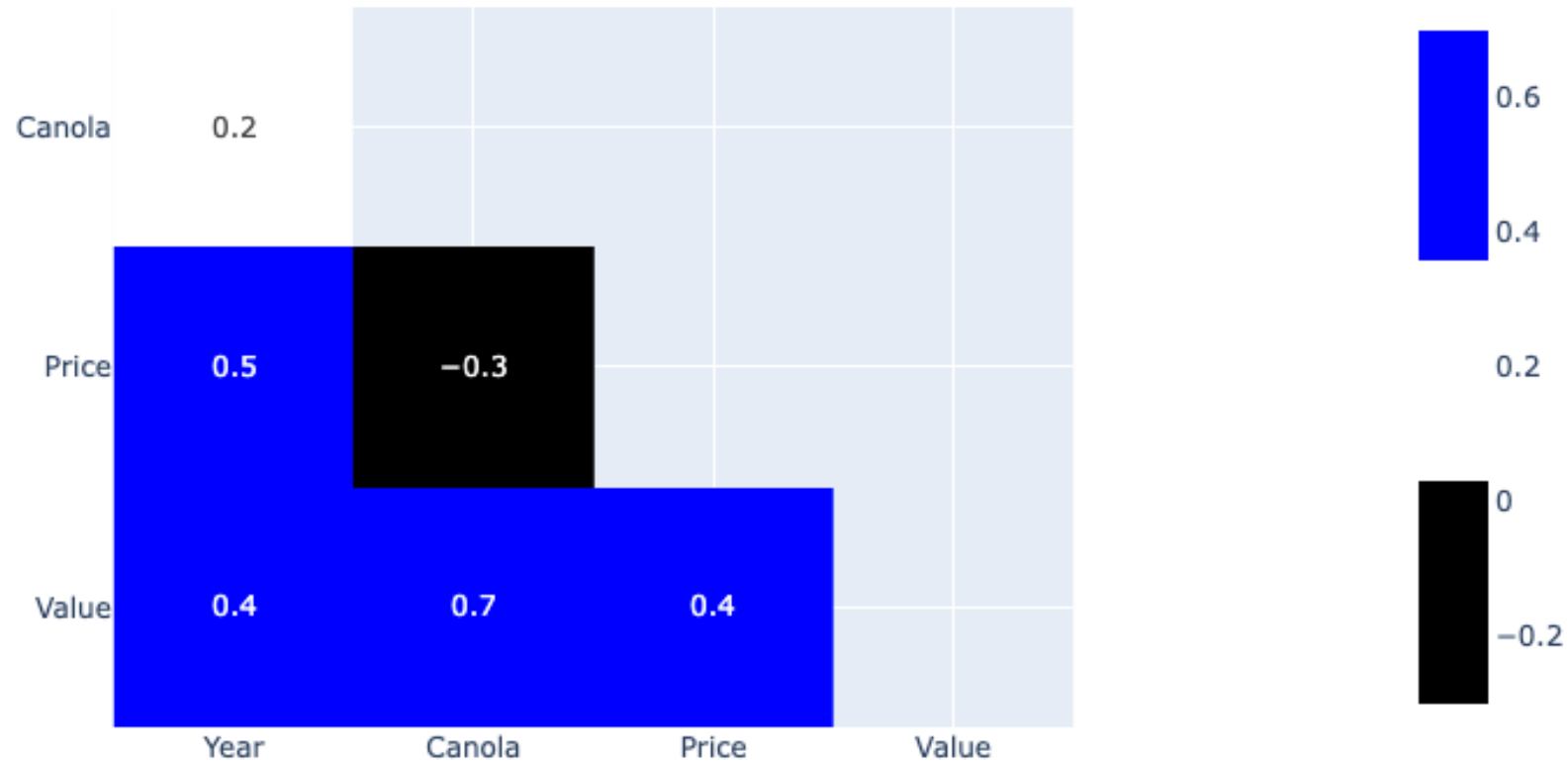
Missing Crop Values - 2010 to 2022





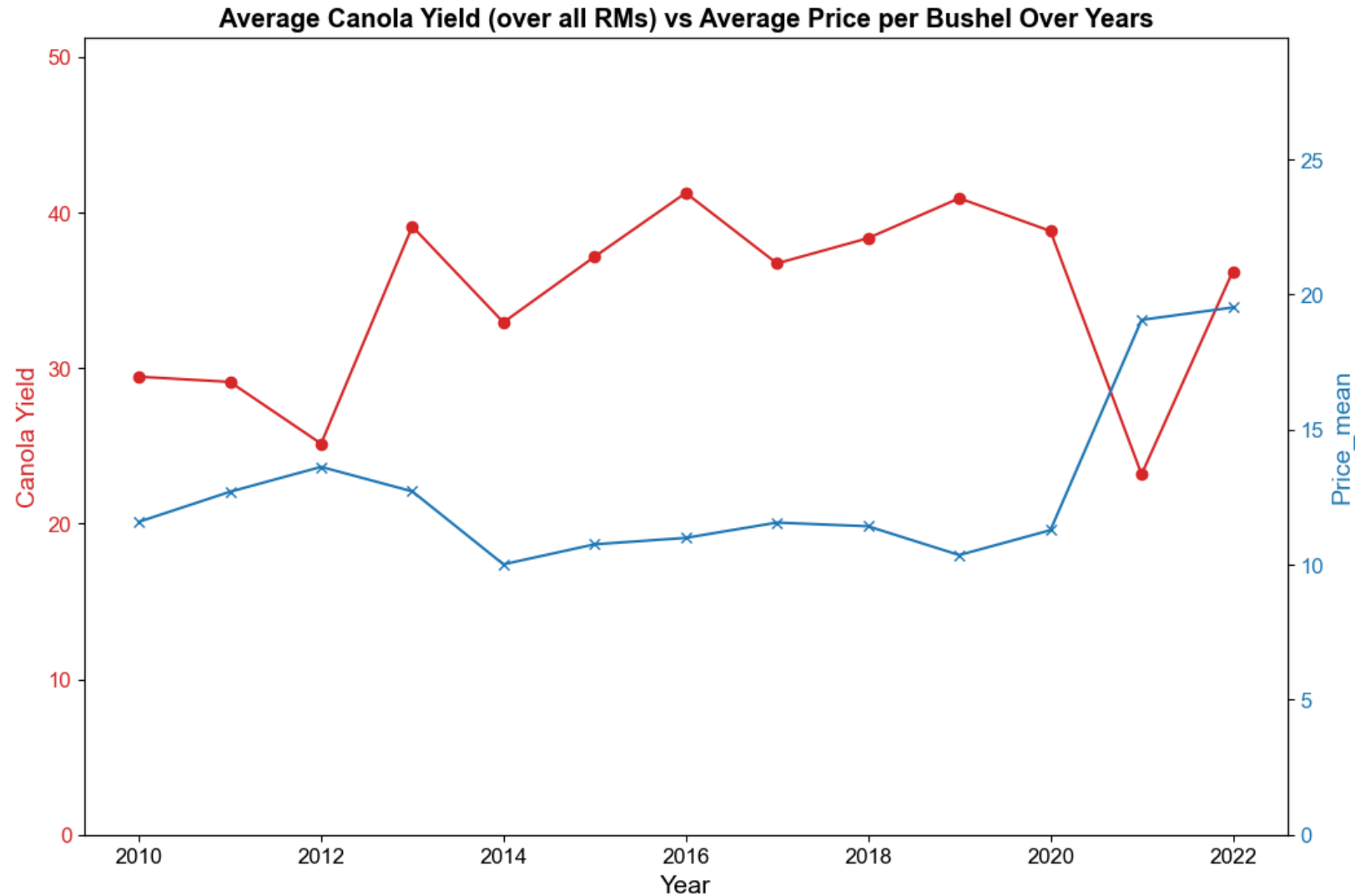
# Exploratory Data Analysis (EDA)

Correlation Matrix shows several positive correlations



# Exploratory Data Analysis (EDA)

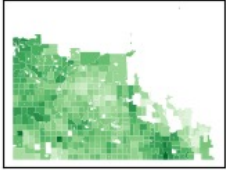
Relationship between Canola Crop Yields and Stock Price Over Time



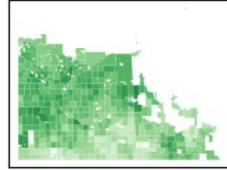
# Exploratory Data Analysis (EDA)

Explore Choropleth Maps by RM and Year for Canola Crop Yields

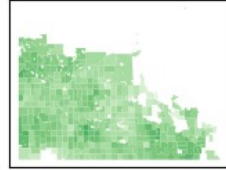
Canola Yield in 2010



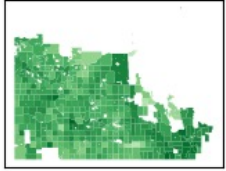
Canola Yield in 2011



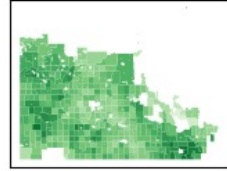
Canola Yield in 2012



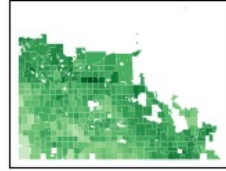
Canola Yield in 2013



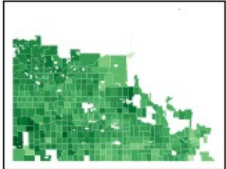
Canola Yield in 2014



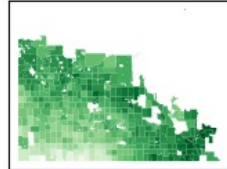
Canola Yield in 2015



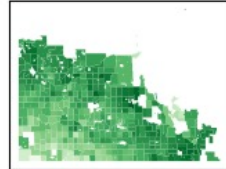
Canola Yield in 2016



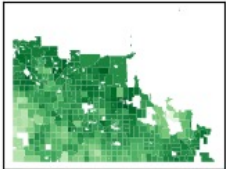
Canola Yield in 2017



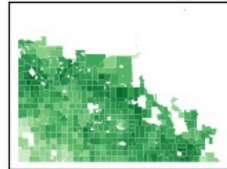
Canola Yield in 2018



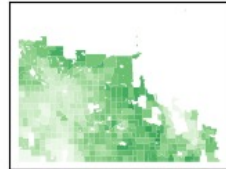
Canola Yield in 2019



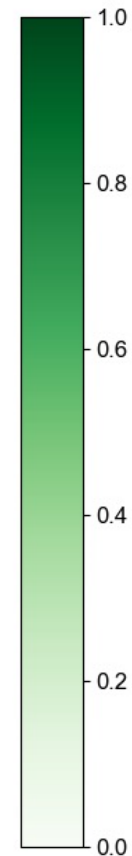
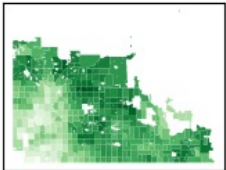
Canola Yield in 2020



Canola Yield in 2021



Canola Yield in 2022

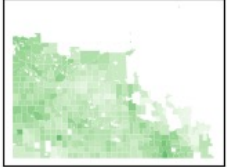




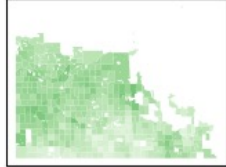
# Exploratory Data Analysis (EDA)

Explore Choropleth Maps by RM and Year for Canola Crop Values

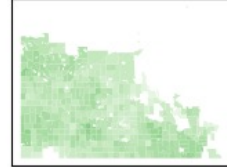
Canola Value in 2010



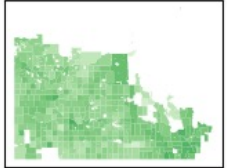
Canola Value in 2011



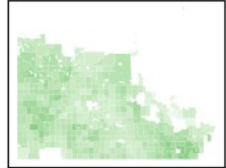
Canola Value in 2012



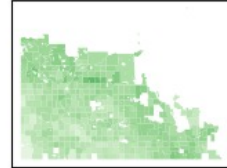
Canola Value in 2013



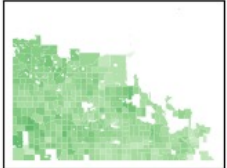
Canola Value in 2014



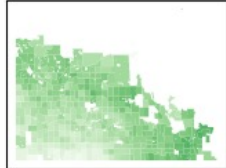
Canola Value in 2015



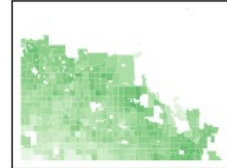
Canola Value in 2016



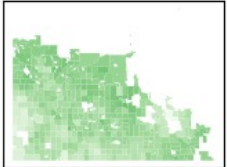
Canola Value in 2017



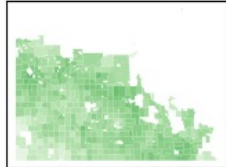
Canola Value in 2018



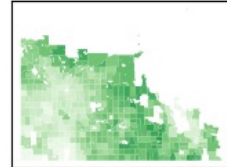
Canola Value in 2019



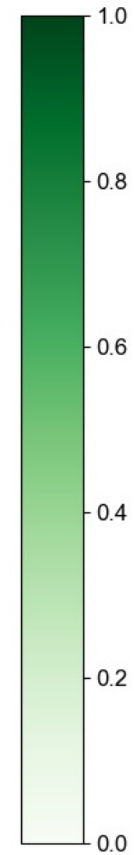
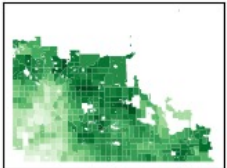
Canola Value in 2020



Canola Value in 2021



Canola Value in 2022



# Exploratory Data Analysis (EDA)

## Percentage Difference of Value vs Yield over Time Analysis

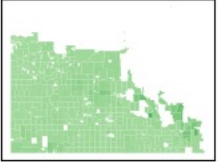
	min	max	std	mean	median	count
Year						
2010	64.7	110.6	4.999063	69.533073	68.3	384
2011	49.9	70.2	2.496016	52.665183	51.8	382
2012	39.7	49.9	1.311535	41.697668	41.4	386
2013	49.1	56.2	0.848976	50.414323	50.3	384
2014	92.7	137.3	4.116000	97.850914	97.2	383
2015	77.7	94.7	1.975660	81.105469	80.6	384
2016	73.8	90.2	1.354939	75.827105	75.7	380
2017	64.5	151.4	7.323034	68.511719	66.9	384
2018	67.0	187.9	6.481849	69.755352	68.8	383
2019	85.2	131.4	3.412380	88.137337	87.3	383
2020	69.4	85.0	1.996041	71.592167	71.0	383
2021	-2.4	-0.4	0.238938	-2.233159	-2.3	383
2022	-4.8	0.0	0.244949	-4.787500	-4.8	384

- ▶ Range: Lowest in 2021 (-2-0.4%) to Highest in 2014 (93-137%) and 2018 (67-188%) indicates significant fluctuations
- ▶ STD: Greatest variability 2014 & 2018, also Best aka highest Mean vs 21/22
- ▶ Mean~Median mostly similar suggesting not heavily skewed except '21 & '22 where yield was lower than its value
- ▶ 2017 stands out as highest STD, explore?

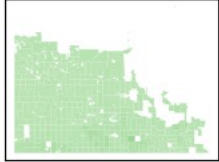
# Exploratory Data Analysis (EDA)

Percentage Difference of Value vs Yield over Time

Canola Scaled % Diff in 2010



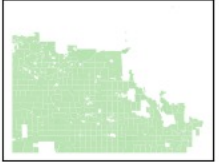
Canola Scaled % Diff in 2011



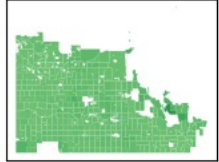
Canola Scaled % Diff in 2012



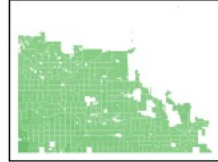
Canola Scaled % Diff in 2013



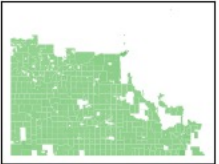
Canola Scaled % Diff in 2014



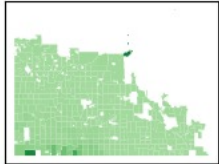
Canola Scaled % Diff in 2015



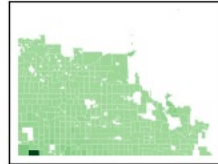
Canola Scaled % Diff in 2016



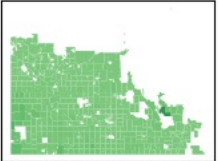
Canola Scaled % Diff in 2017



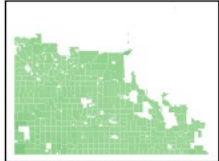
Canola Scaled % Diff in 2018



Canola Scaled % Diff in 2019



Highest STD  
Canola Scaled % Diff in 2020



Canola Scaled % Diff in 2021

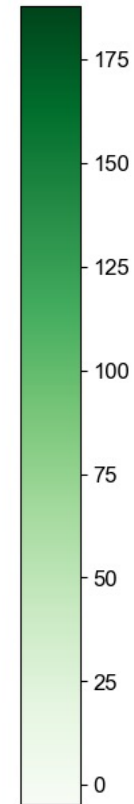


Canola Scaled % Diff in 2022



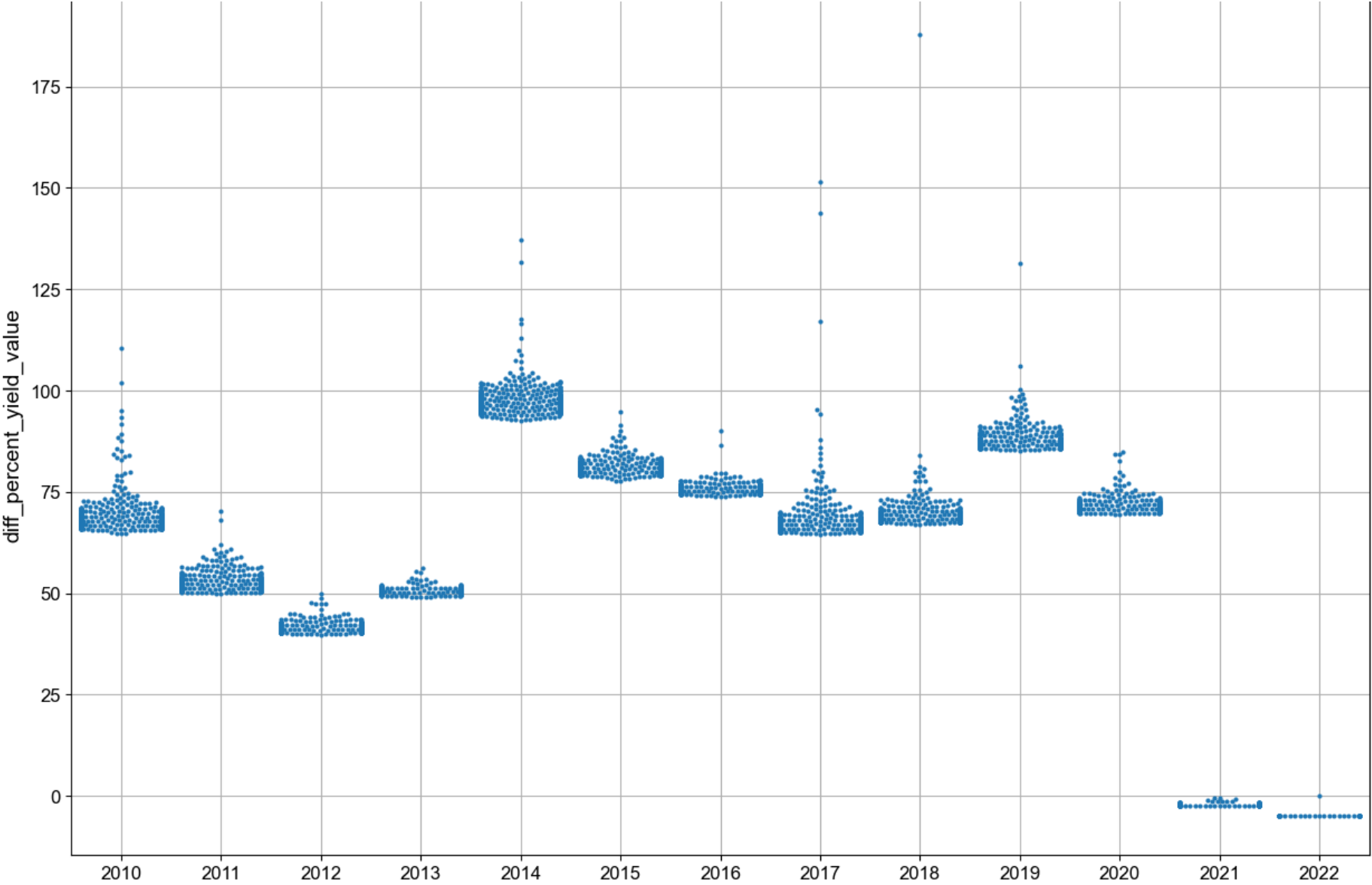
LOW

LOW



# Exploratory Data Analysis (EDA)

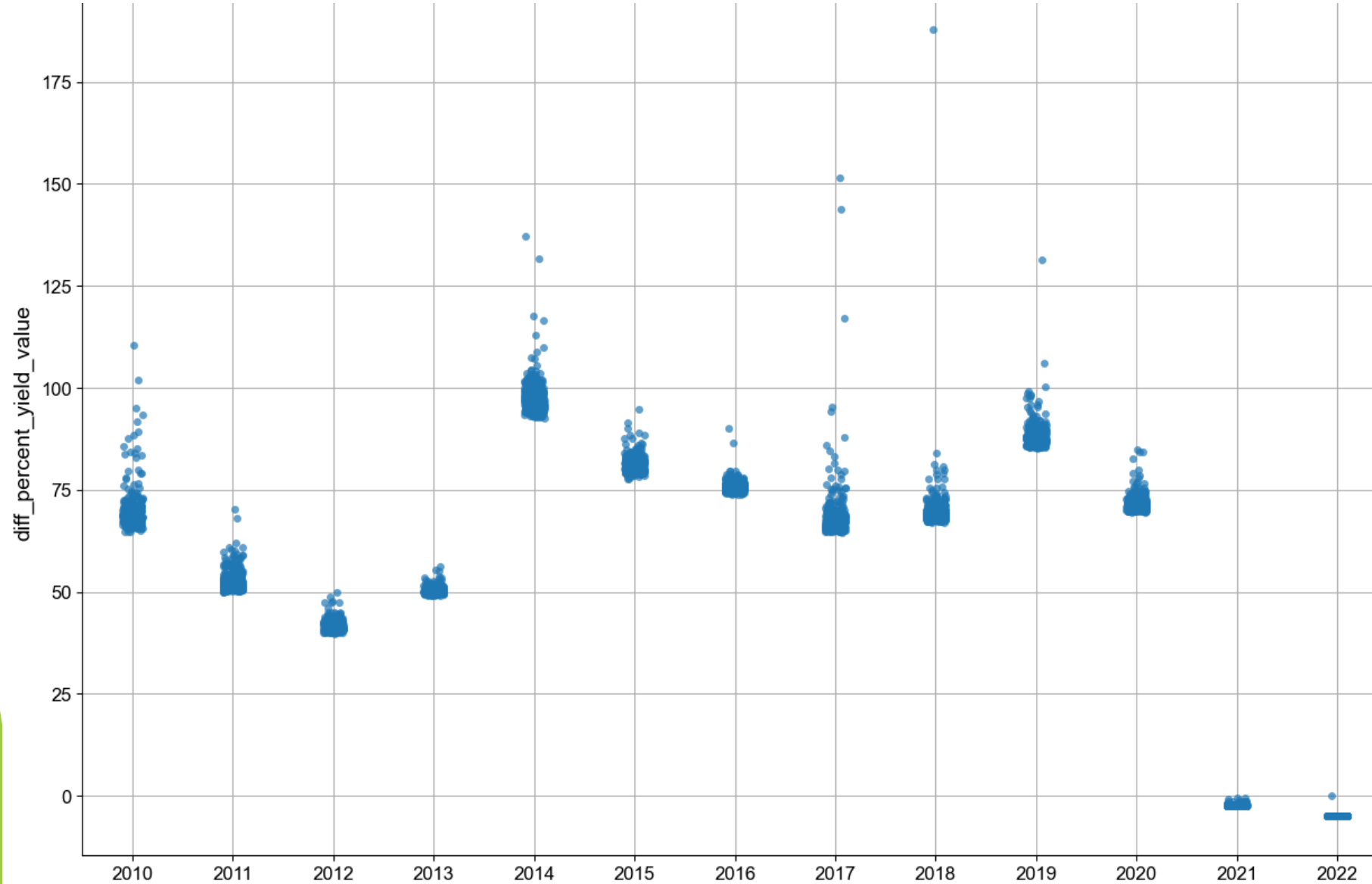
Swarm Plot of same % Difference between Value and Yield over Time





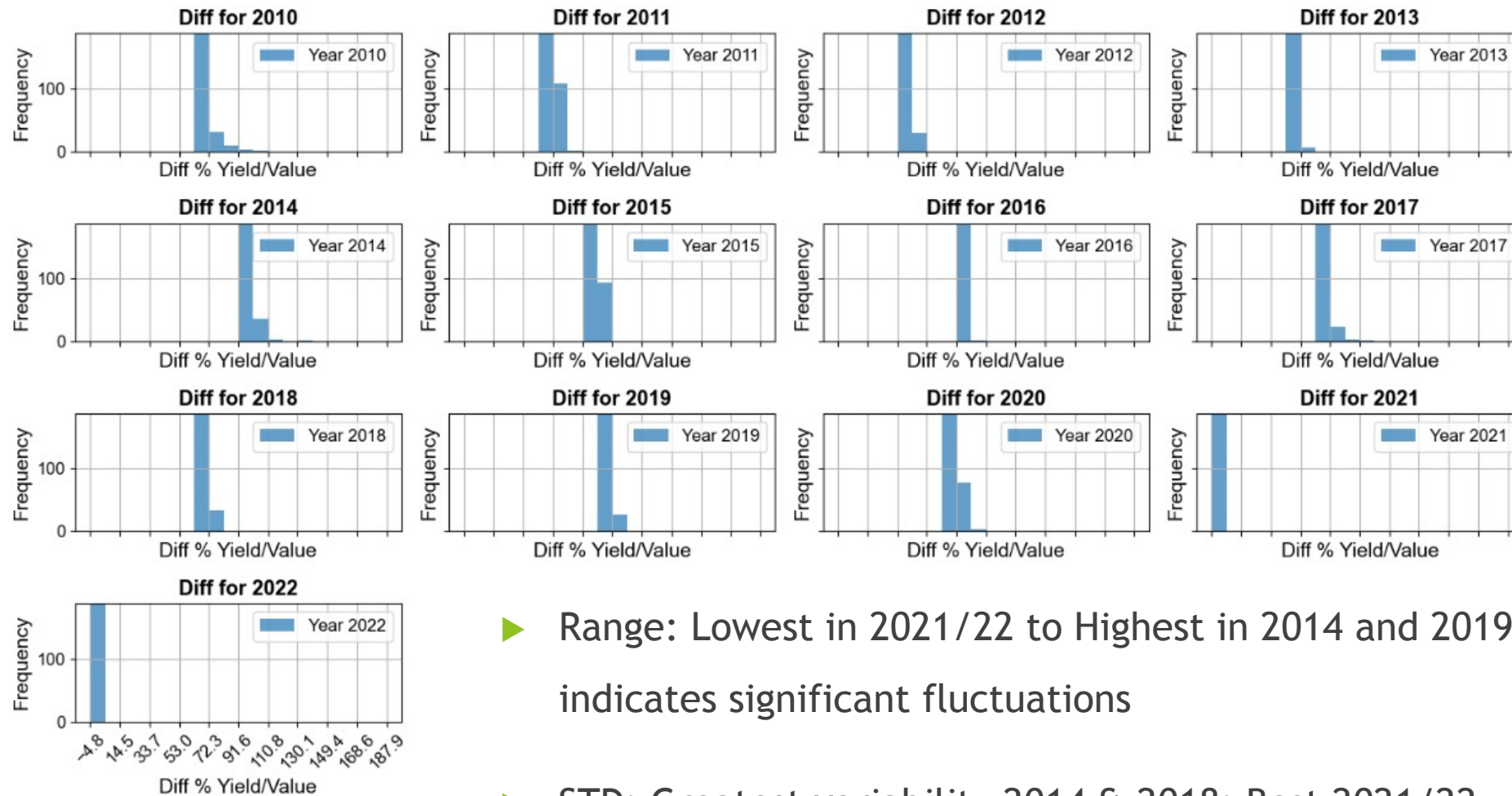
# Exploratory Data Analysis (EDA)

Swarm Plot of same % Difference between Value and Yield over Time



# Exploratory Data Analysis (EDA)

Histograms of % Difference between Value and Yield



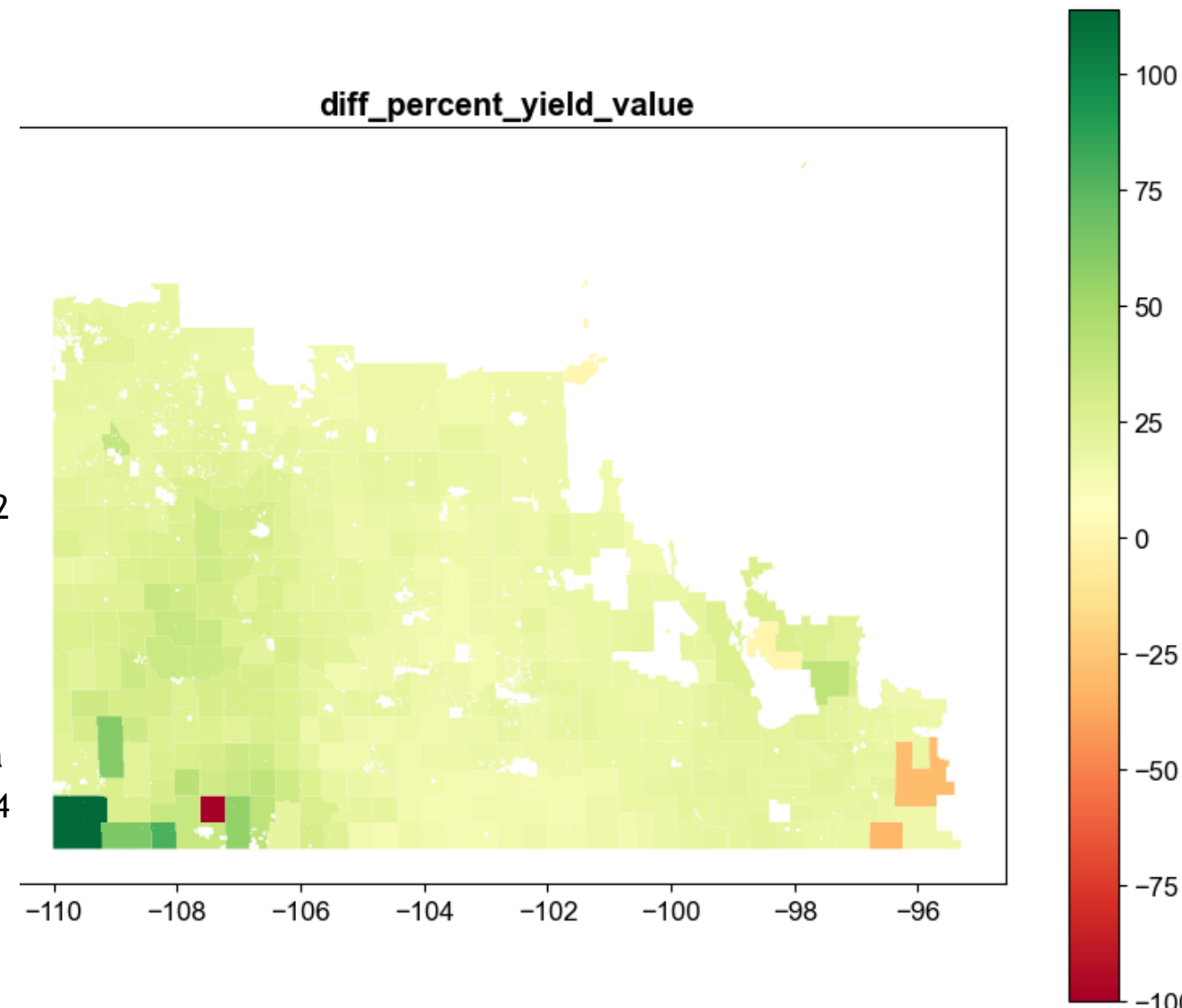
- ▶ Range: Lowest in 2021/22 to Highest in 2014 and 2019 indicates significant fluctuations
- ▶ STD: Greatest variability 2014 & 2018; Best 2021/22

# Interactive Python

Choropleth of Selectable Column / Scheme / k

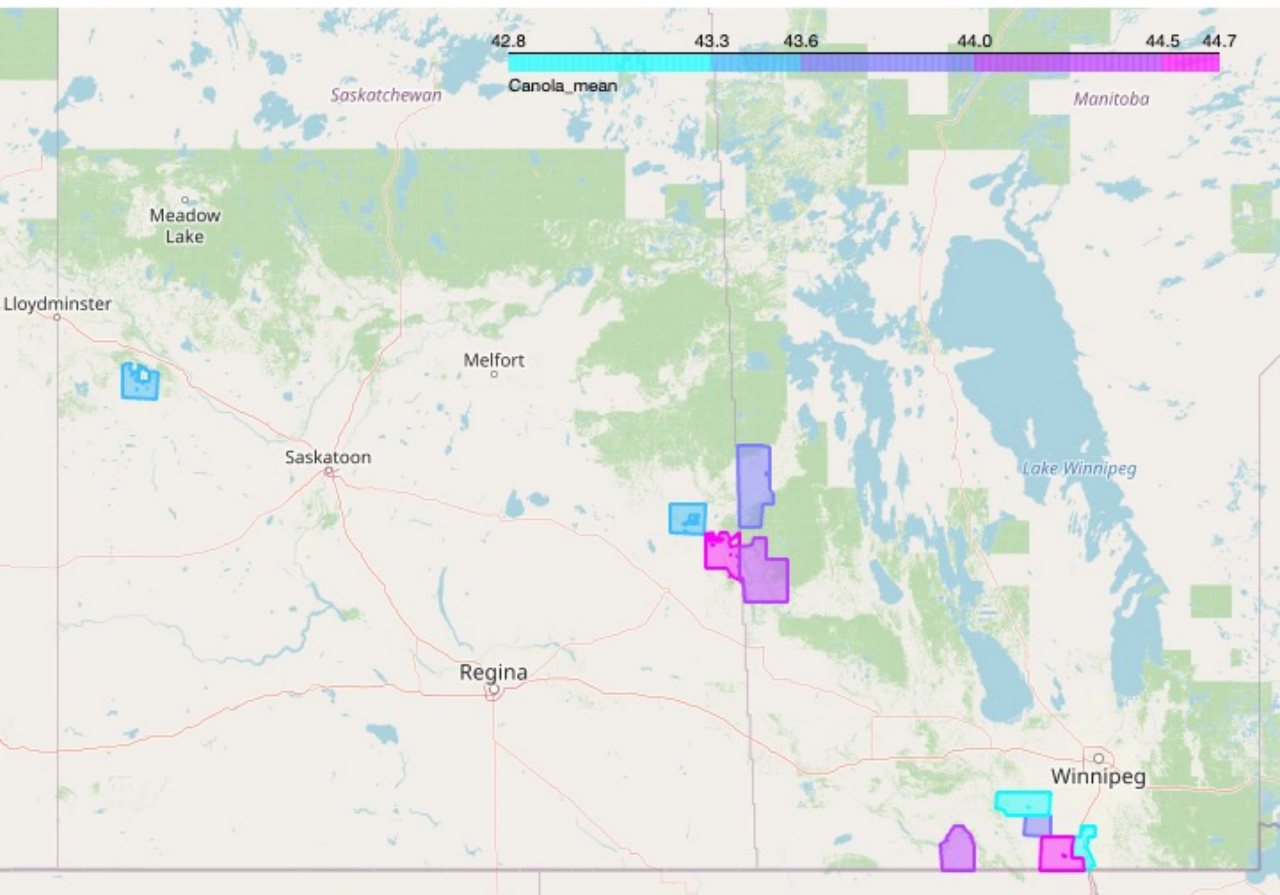
Column?	diff_percent_yield_value	▼
Width:	10	
Height:	8	
Scheme:		▼
k:	automatic	▼

- Details on 3 negative RMs (for future analysis?):
- For RM 46, the Canola yield had a mean of 16.79 with a standard deviation of 11.77, while the Value had a mean of 224.02 with a standard deviation of 113.11. The Canola was scaled to 0.00, and the Value was scaled to 0.04. The Diff %is -100.00, indicating a decrease.
- For RM REYNOLDS, the Canola yield had a mean of 36.55 with a standard deviation of 11.19, while the Value had a mean of 629.82 with a standard deviation of 313.03. The Canola was scaled to 0.71, and the Value was scaled to 1.00. The Diff %is -29.20, indicating a decrease.
- For RM STUARTBURN, the Canola yield had a mean of 29.34 with a standard deviation of 13.38, while the Value had a mean of 490.84 with a standard deviation of 377.04. The Canola was scaled to 0.45, and the Value was scaled to 0.67. The Diff %is -32.80, indicating a decrease.



# Top 10 Interactive Map

Canola RM's by YIELD



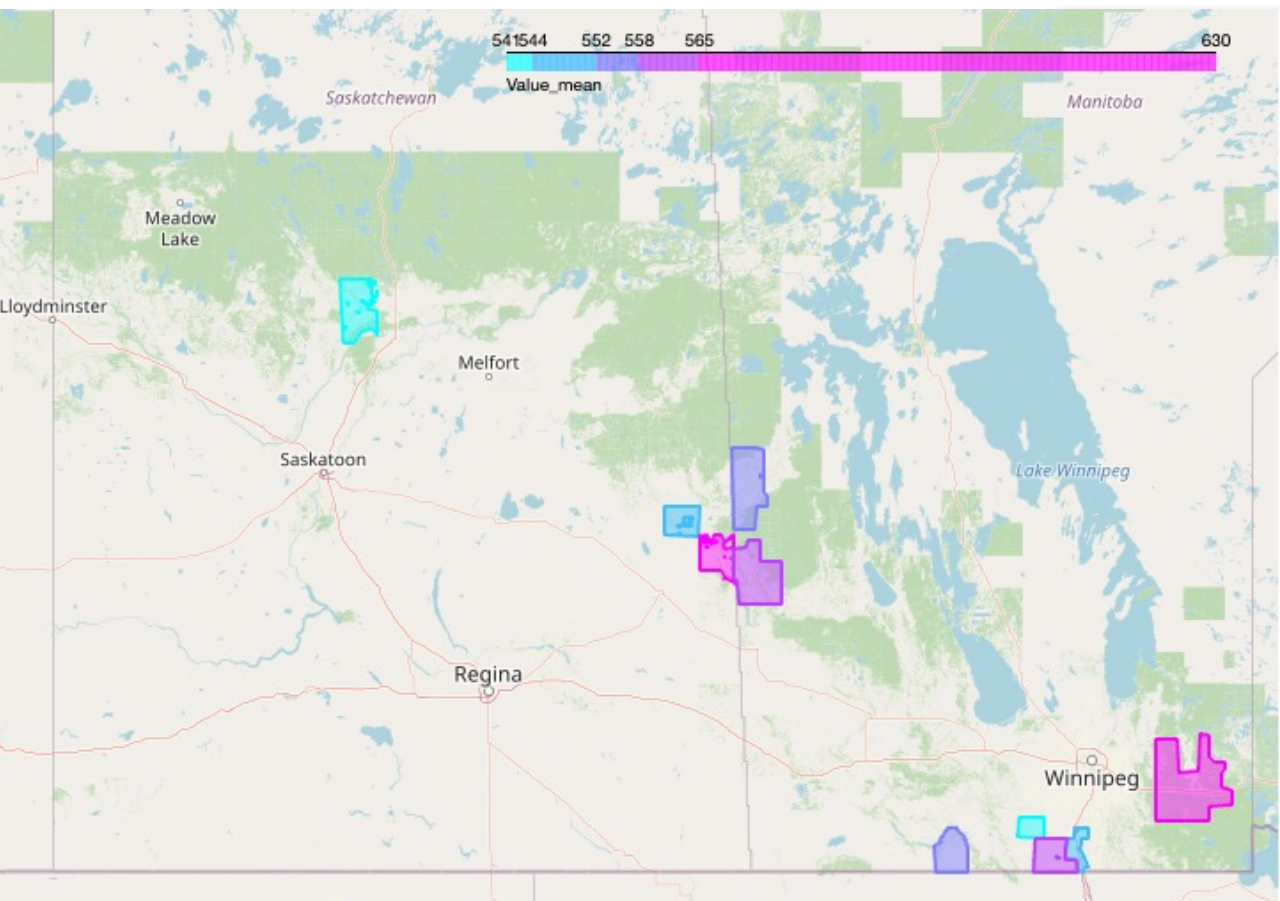
	RM	Canola_mean	Canola_std	Value_mean	Value_std	Rank
	RHINELAND	44.687	7.208	564.825	148.398	1
	271	44.572	9.911	566.433	191.948	2
	HILLSBURG-ROBLIN-SHELL RIVER	44.524	8.866	561.917	171.380	3
	LOUISE	44.277	6.704	554.761	111.715	4
	SWAN VALLEY WEST	43.854	10.263	553.334	170.970	5
	ROLAND	43.576	9.880	542.053	155.549	6
	303	43.518	9.632	550.628	177.417	7
	439	43.382	9.244	535.141	109.814	8
	MONTCALM	43.030	6.444	544.912	141.354	9
	DUFFERIN	42.760	8.629	535.205	143.565	10



# Top 10 Interactive Map

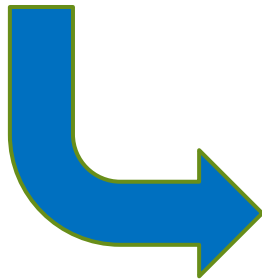
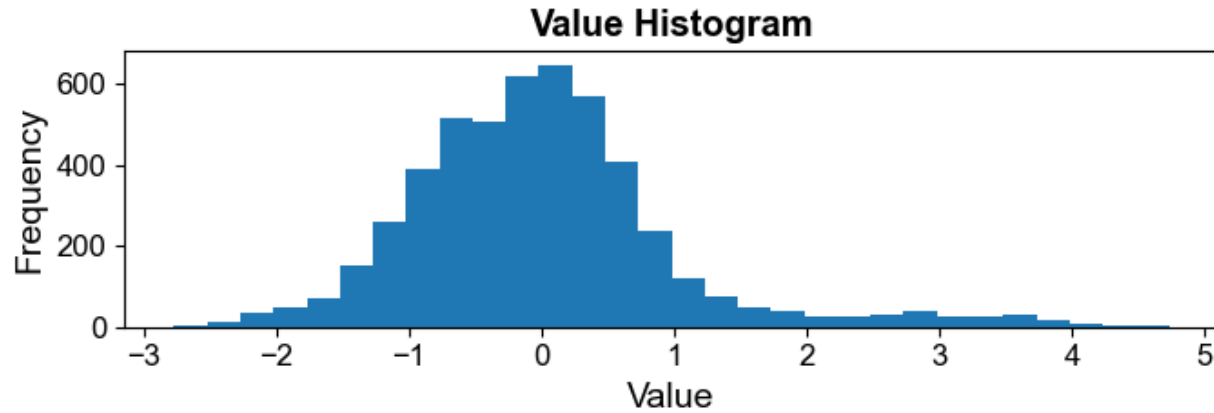
Canola RM's by VALUE

	RM	Canola_mean	Canola_std	Value_mean	Value_std	Rank
	REYNOLDS	36.552	11.190	629.820	313.026	1
	271	44.572	9.911	566.433	191.948	2
	RHINELAND	44.687	7.208	564.825	148.398	3
	HILLSBURG-ROBLIN-SHELL RIVER	44.524	8.866	561.917	171.380	4
	LOUISE	44.277	6.704	554.761	111.715	5
	SWAN VALLEY WEST	43.854	10.263	553.334	170.970	6
	303	43.518	9.632	550.628	177.417	7
	MONTCALM	43.030	6.444	544.912	141.354	8
	ROLAND	43.576	9.880	542.053	155.549	9
	493	42.515	8.314	540.992	179.664	10

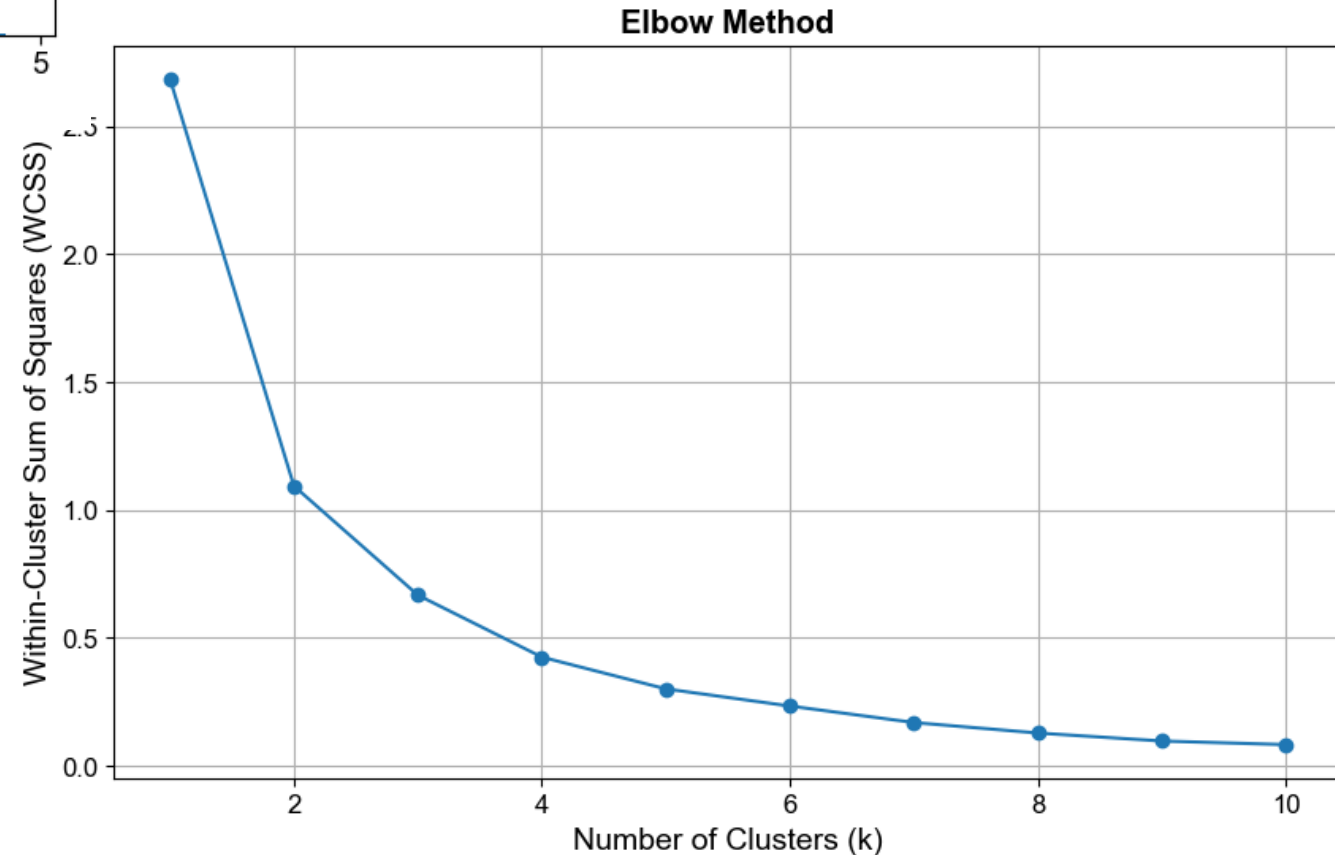


# Unsupervised Machine Learning

K-Means Clustering of Scaled Value

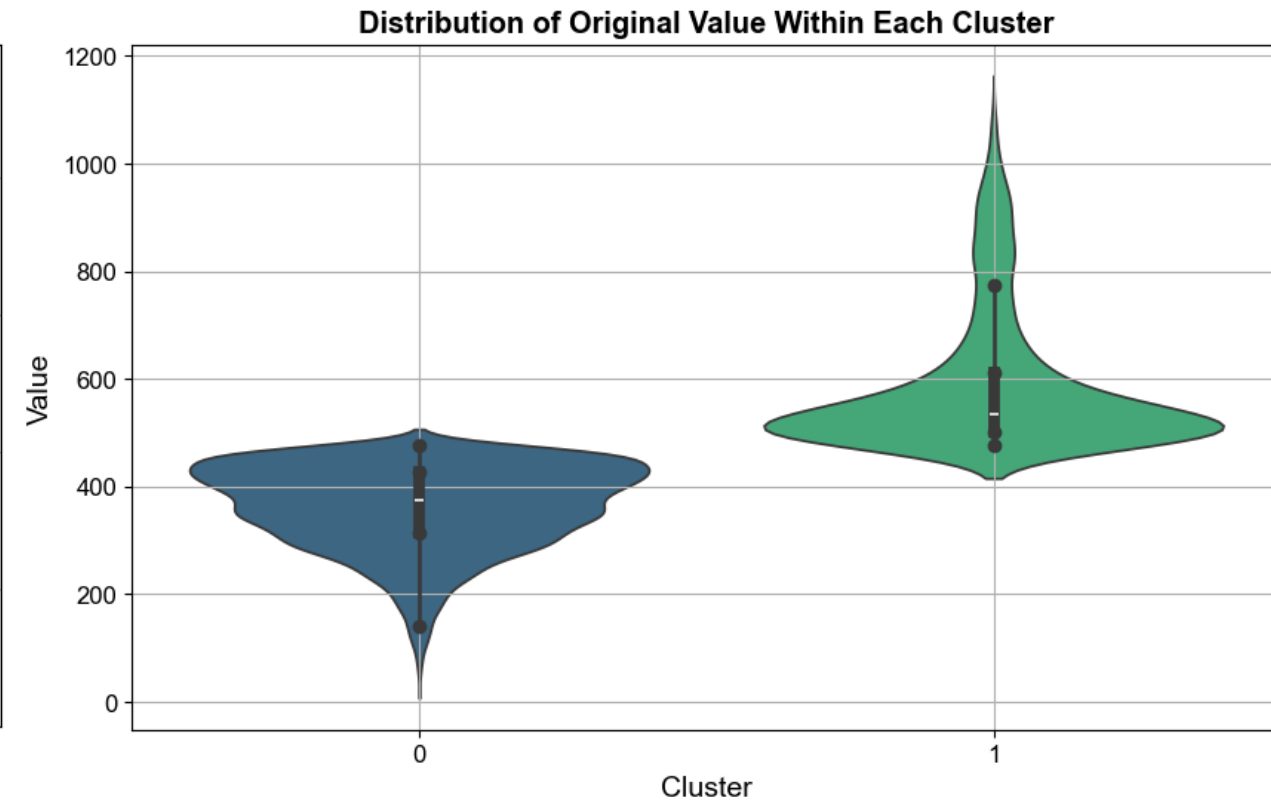
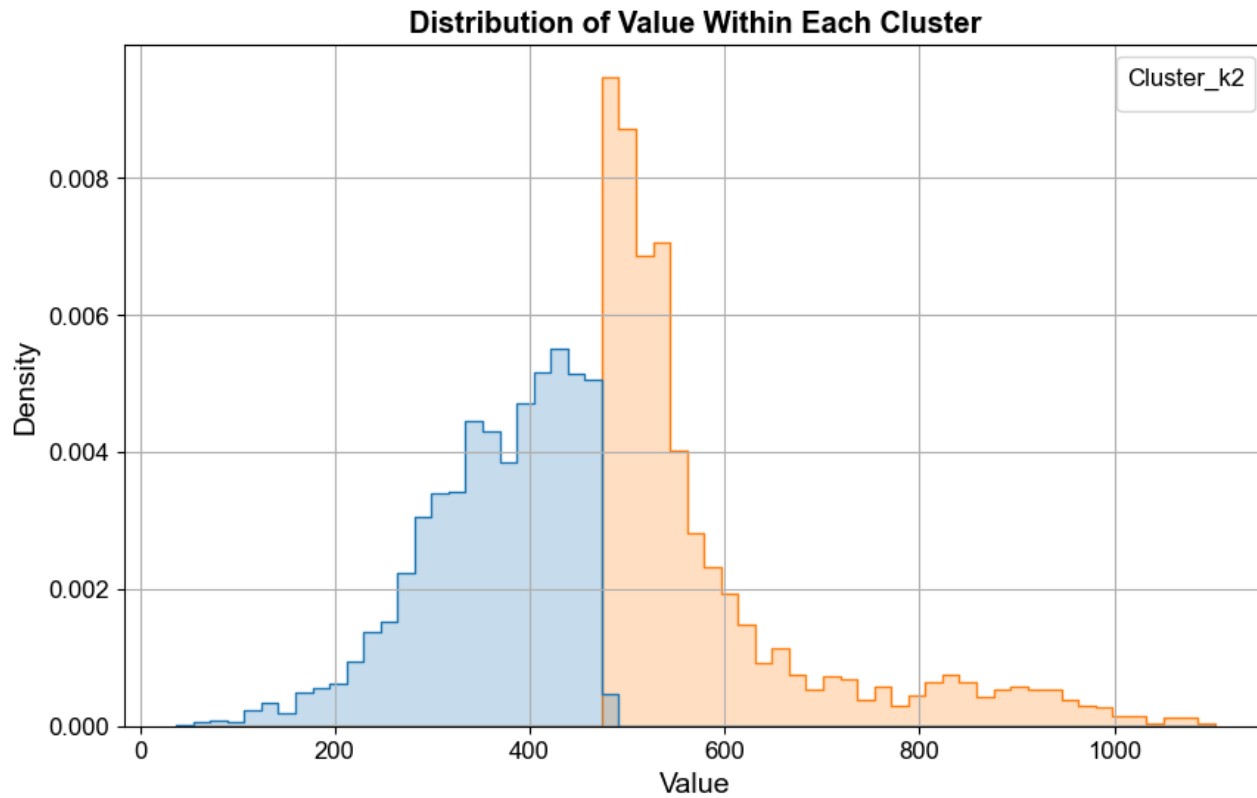


► Using Elbow Method: Selecting  $k=2$  and  $k=4$



# Unsupervised Machine Learning

K-Means Clustering with  $k=2$



# Unsupervised Machine Learning

K-Means Clustering with  $k=2$





# Unsupervised Machine Learning

## K-Means Clustering of k=2 and k=4

	k2 (0)	k2 (1)	k4 (0)	k4 (1)	k4 (2)	k4 (3)
Metrics						
Count	3472.00	1511	1219	1388	2118	258.00
Min Value	37.09	475.85	-2.774809	0.257726	-0.636971	1.80
Max Value	475.56	1102.97	-0.637957	1.786917	0.257093	4.73
Mean Value	363.19	587.21186	-1.102101	0.685314	-0.168007	2.90
Median Value	375.65	534.16	-1.000625	0.598634	-0.149166	2.86
Std Dev Value	79.34	130.702346	0.390807	0.352736	0.257825	0.69
Skewness	-0.80	1.713088	-1.27524	1.078801	-0.137471	0.29
Kurtosis	0.35	2.10047	1.435909	0.544665	-1.158503	-0.69
Silhouette Score	0.52					0.52

Overall, both k=2 and k=4 offer valuable but different perspectives on the data segmentation. The choice between them may depend on the specific needs of further analysis or domain-specific interpretation.

# Results

- ▶ Interactive Python Widget
- ▶ HTML Interactive Files for Exploration
- ▶ Visualizations:
  - ▶ Geospatial Choropleths
  - ▶ Time Series Line Charts
  - ▶ Comparative Histograms
- ▶ Provide in-depth Canola yield vs value

# Conclusions

- ▶ After Identifying the Yearly trends a Percentage Difference was used to highlight which years showed variability.
- ▶ Trends found in Canola over time
- ▶ Correlation was demonstrated between Yield and Value
- ▶ Next Steps?
  - ▶ Explore more regional specific pricing for value calculations vs public stock price.
  - ▶ Consider other correlations including real estate prices and even municipal tax rates.