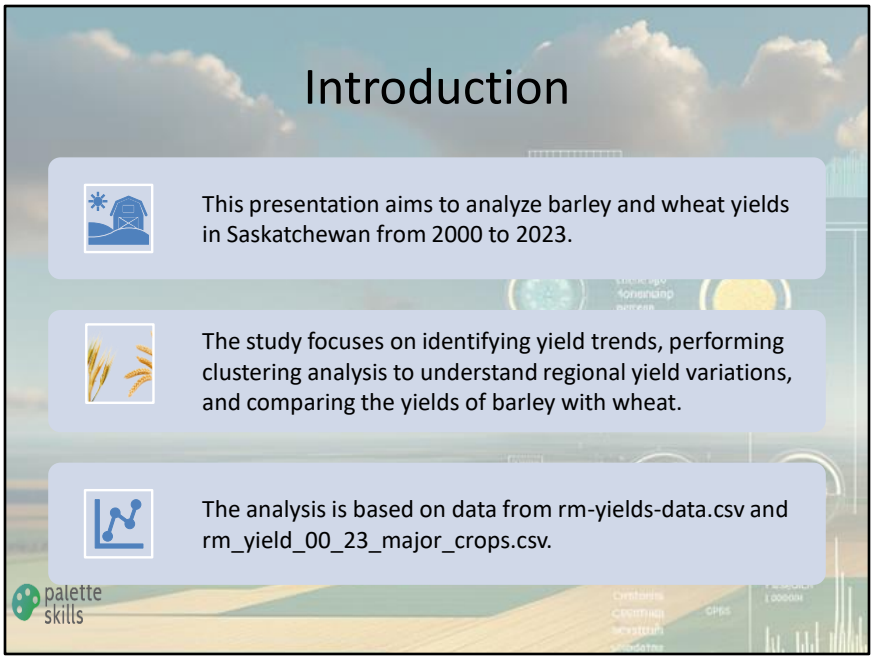
Analysis of Barley and Wheat Yields in Saskatchewan (2000-2023)

A Comprehensive Study on Yield Trends, Clustering, and Comparative Analysis

Presented by
Muhammad Zeeshan Gondal

palette skills



In this analysis, we aim to investigate the trends, distribution, and clustering of barley yields in Saskatchewan over the period from 2000 to 2023. We use two primary datasets: one containing yield data aggregated by Rural Municipality (RM) and another with comprehensive annual yield data. The techniques employed include outlier detection, trend analysis, K-Means clustering, and comparative analysis with wheat yields. By understanding these factors, we can derive insights into the performance and variability of barley yields over time and across different regions.

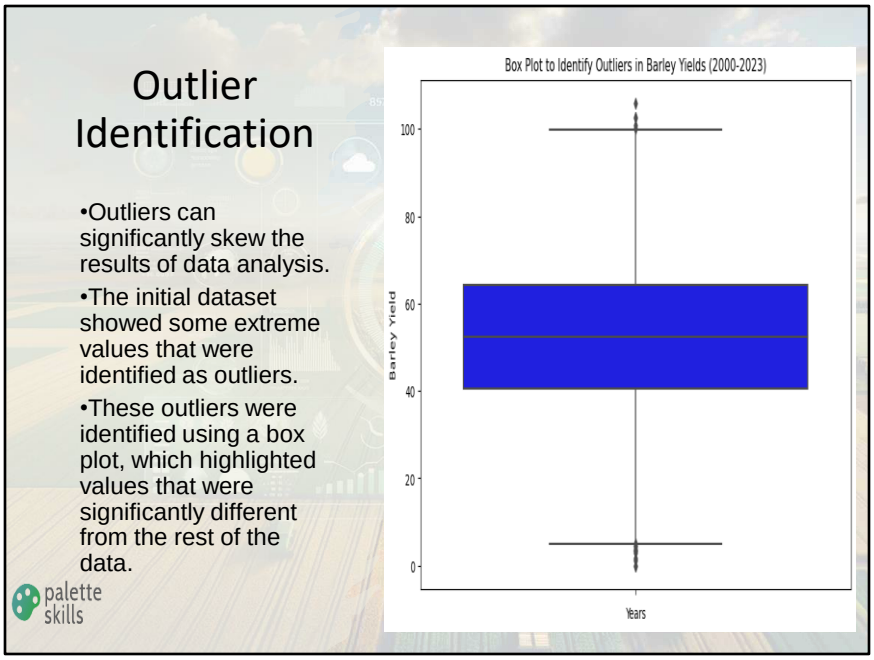
Overview of Datasets

•RM-level Yield Data (2000-2023)

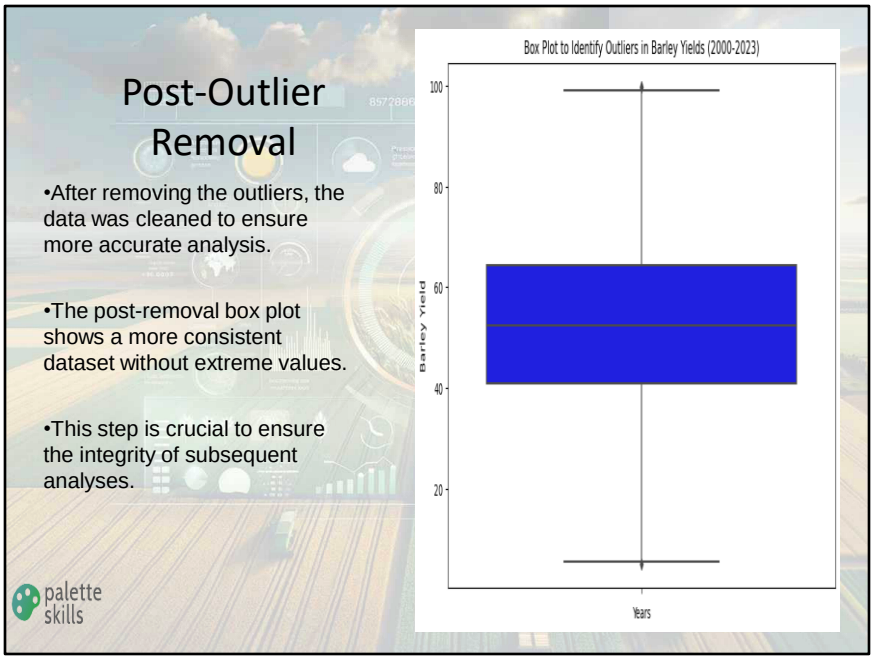
•Comprehensive Annual Yield Data

•GIS Data for Saskatchewan RMs

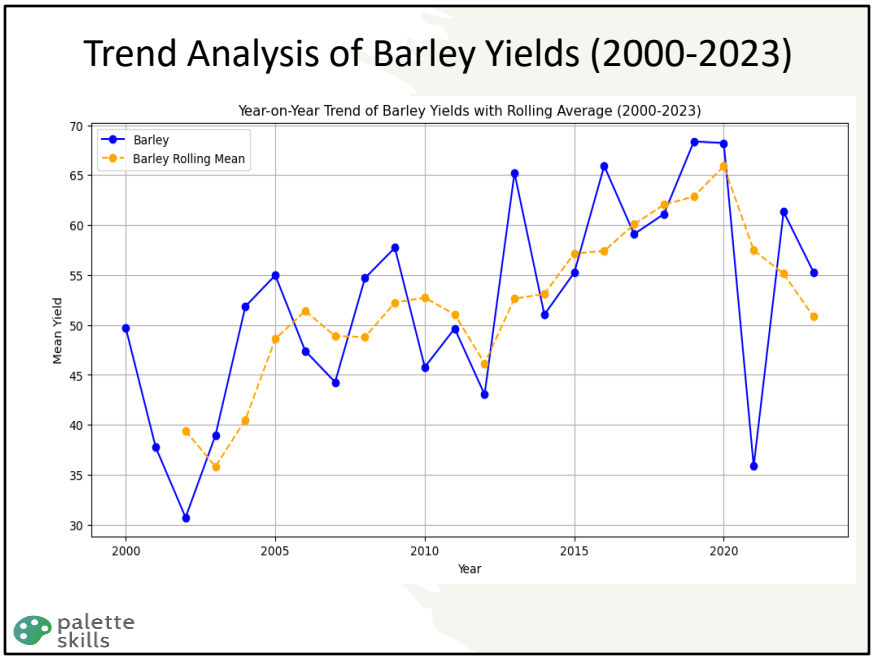
We used RM-level yield data for barley and wheat from 2000 to 2023. This dataset provides annual yield information for various crops across different RMs. We also utilized GIS data to map and visualize the spatial distribution of yields. The integration of yield data with spatial information allows for a more comprehensive analysis of agricultural productivity across the region.



We started by identifying outliers in the barley yield data using a box plot. Outliers can skew the analysis and lead to incorrect conclusions. A box plot is an effective visualization for this purpose as it highlights the spread and central tendency of the data while clearly marking outliers. The initial box plot revealed extreme values that needed to be addressed to ensure the accuracy of our analysis.



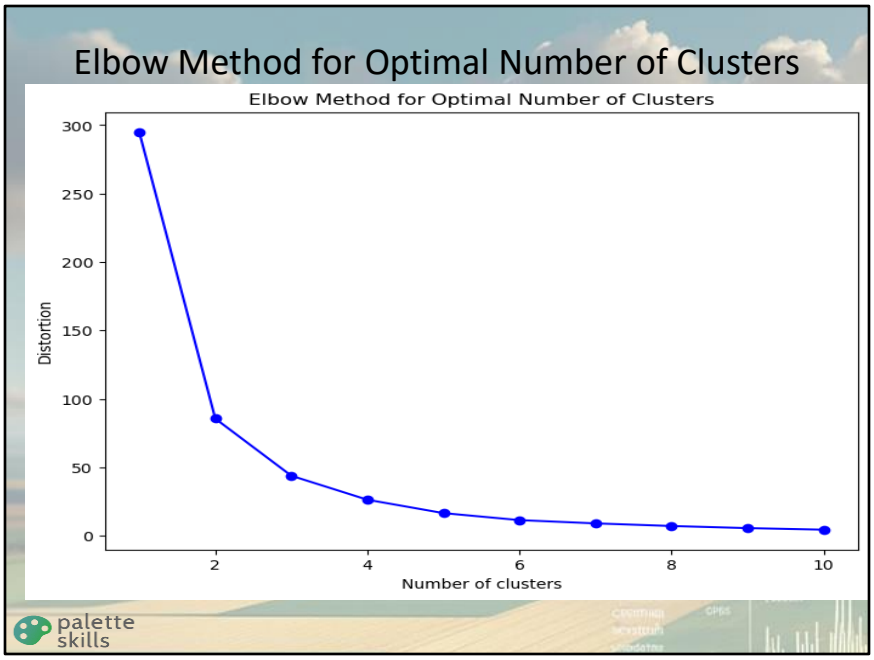
After identifying outliers, we removed them to clean the data. The post-removal box plot shows a more consistent dataset without extreme values. This step is crucial to ensure the integrity of subsequent analyses. By removing outliers, we reduce the potential for skewed results and improve the reliability of our findings. The cleaned data provides a solid foundation for further analysis.



Trend Analysis of Barley Yields (2000-2023)

- The trend analysis reveals how barley yields have changed over the years.
- The line plot shows the annual average yields of barley, with significant fluctuations observed in certain years.
- The rolling average helps to smooth out short-term fluctuations and highlight longer-term trends. For example, the yields generally increased until 2016, followed by a decline.

To understand the overall trend in barley yields, we calculated the annual average yields and plotted them over time. A rolling mean was also calculated to smooth short-term fluctuations and highlight longer-term trends. This helps in identifying whether the yields have generally increased, decreased, or remained stable over the years. The trend analysis provides valuable insights into the performance and variability of barley yields over time.



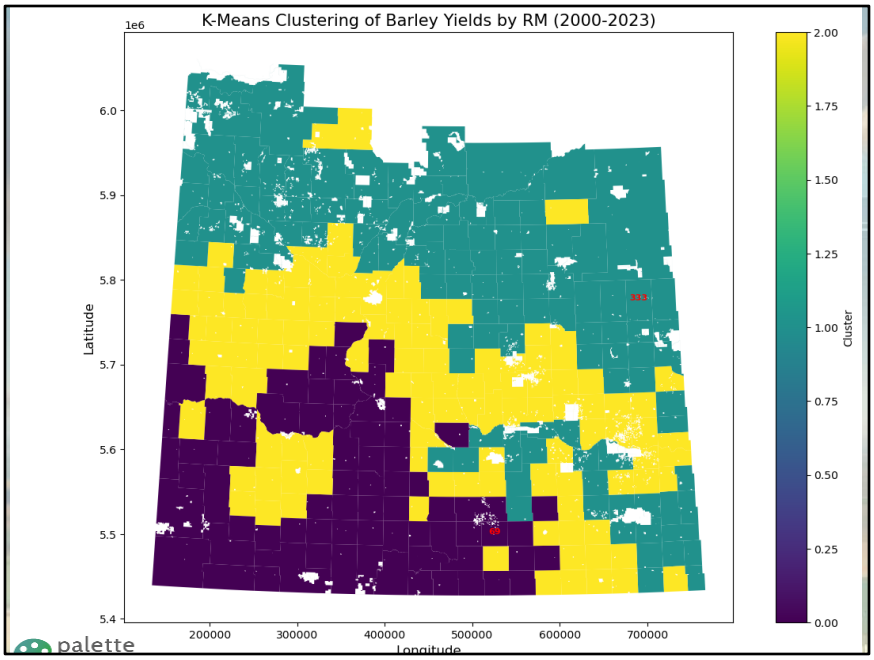
Slide Points:

•Purpose of Elbow Method

•Distortion vs. Number of Clusters

•Identifying the Optimal Number of Clusters

Explanation: The Elbow Method is used to determine the optimal number of clusters for K-Means clustering. By plotting the distortion (sum of squared distances from each point to its assigned cluster center) against the number of clusters, we can identify the "elbow point" where the distortion decreases abruptly. This point indicates the optimal number of clusters to use. The elbow point helps to balance the trade-off between the number of clusters and the compactness of the clusters.



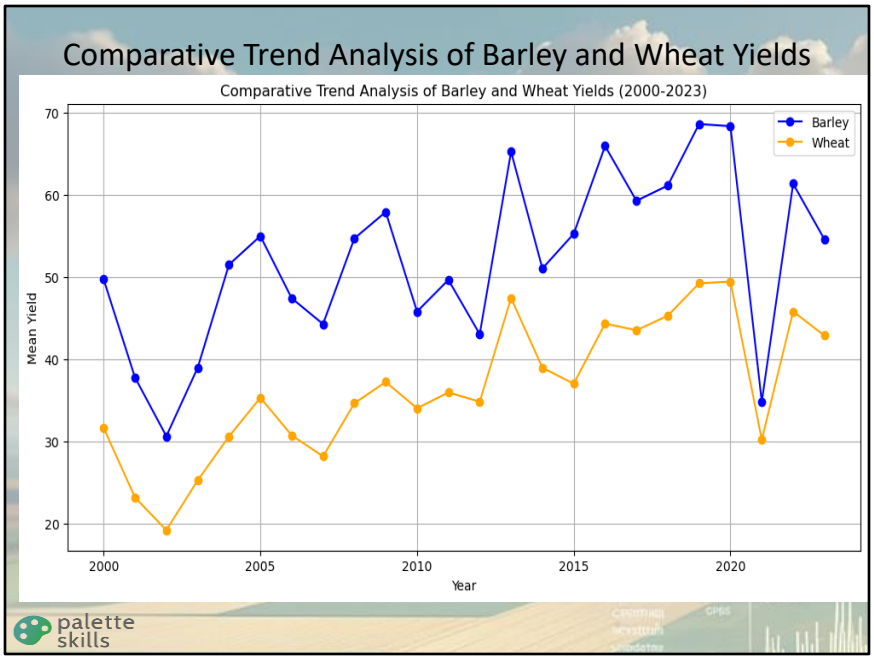
Slide Points:

•Performing K-Means Clustering

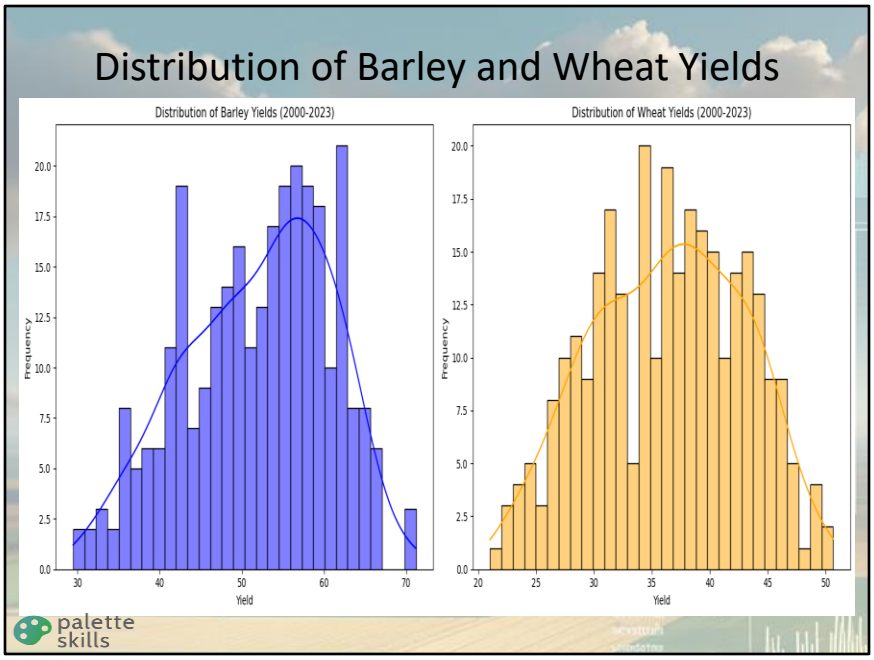
•Visualizing Clusters on the Map

•Significant Regions Identified

Explanation: Using the optimal number of clusters determined by the Elbow Method, we performed K-Means clustering on the barley yield data. The clusters were then visualized on a map of Saskatchewan RMs, highlighting regions with different yield characteristics. Annotations were added to mark the regions with the highest and lowest yields. This analysis helps in identifying regions with similar yield patterns and understanding the spatial distribution of barley yields.



Slide 8: Comparative Trend Analysis of Barley and Wheat Yields
Title: Comparative Trend Analysis of Barley and Wheat Yields (2000-2023)
Slide Points:
•Yearly Average Yields of Barley and Wheat
•Comparison of Trends
•Insights from Comparison
Explanation: We compared the yearly average yields of barley and wheat to understand the relative performance and trends of these two crops. This comparative analysis helps in identifying any significant differences or similarities in the yield patterns of barley and wheat over the years. By analyzing the trends of both crops, we can gain insights into their relative performance and the factors influencing their yields.



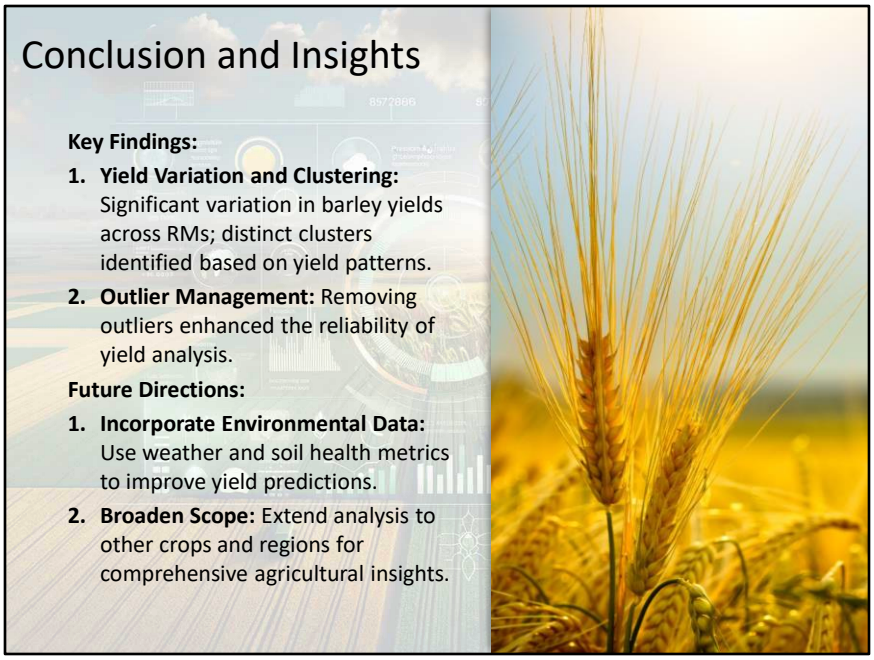
Slide Points:

•Histogram of Barley Yields

•Histogram of Wheat Yields

•Insights from Distribution Comparison

Explanation: We visualized the distribution of barley and wheat yields using histograms. This helps in understanding the frequency and spread of yields for both crops. The comparison provides insights into the variability and central tendency of yields, highlighting any differences in the yield distributions. Histograms are useful for showing the distribution of data and identifying patterns and anomalies.



Key Findings:

•Barley yields have shown significant variation across different RMs and years.

•Clustering analysis revealed distinct groups of RMs with similar yield patterns.

•Comparative analysis with wheat showed different yield trends and distributions.

•Identification and removal of outliers improved the reliability of the analysis.

Future Directions:

•Further research could incorporate additional variables such as weather data and soil health metrics to enhance yield predictions.

•Extending the analysis to other crops and regions could provide broader agricultural insights.

•Investigation into the impact of specific agricultural practices and policies on yield performance.