KANSAS WHEAT CROP YIELD PREDICTION

Leveraging Machine Learning for Better Agricultural Decisions : Predict, Prevent, Prosper

Presentation By :
DENISH TRADA

Overview:

- Kansas is a major wheat-producing region.
- Yield variability impacts farmers, agribusiness, and policymakers.

Problem Statement:

- Predicting wheat yield is challenging due to complex weather, soil, and management interactions.

Opportunity:

- A data-driven approach can optimize resource use and improve decision-making.

THE BIG IDEA

Integrative Strategy:

- Combine historical wheat yield data with 2010–2020 climate records.
- Use EDA to reveal patterns and anomalies.

Analytical Techniques:

- Regression, clustering, and PCA for robust predictions.
- Interactive visualizations (e.g., Plotly, GeoPandas) to engage stakeholders.

Goal:

- Develop actionable insights to drive better resource allocation and risk mitigation.

THE IMPACT

GOAL

Improve yield prediction accuracy by at least 15%

ECONOMIC BENEFITS

- Better yield predictions can save costs by reducing resource waste.
- Increased forecast accuracy may stabilize incomes and markets

ENVIRONMENTAL GAINS

- Optimized use of water and fertilizers reduces environmental footprint.

POLICY SUPPORT

- Data-driven insights support targeted policymaking and efficient planning.

DATA OVERVIEW

DATASET 1: USDA KANSAS WHEAT YIELD DATA

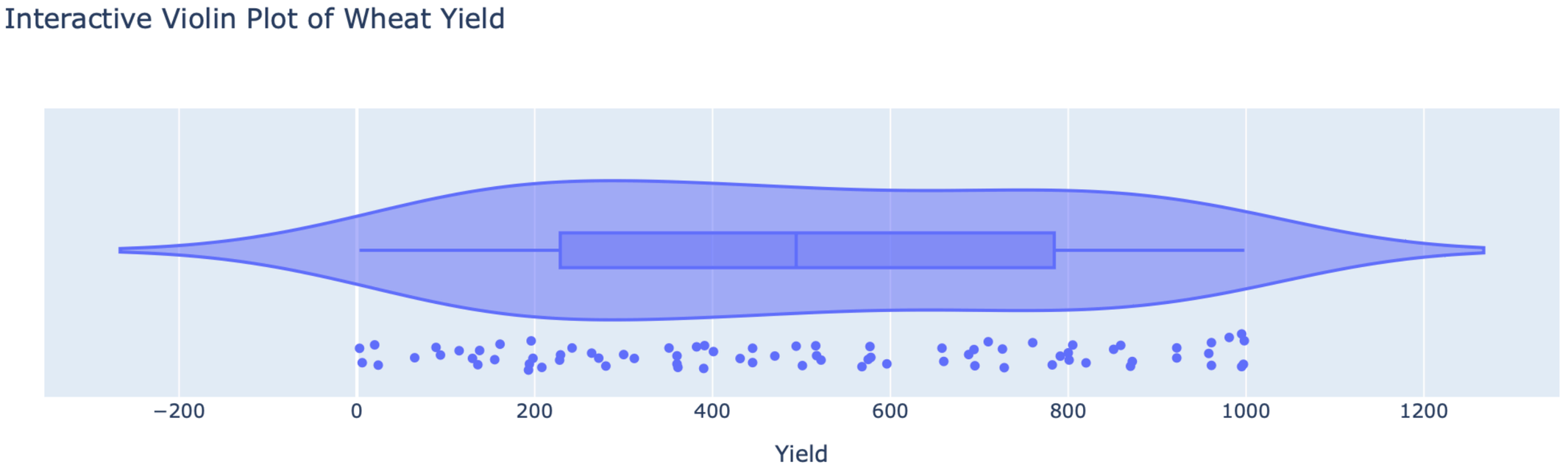
1,127 county-level records (2010–2020) with yield (Bu/Acre).

SUPPLEMENTARY PEST/ECONOMIC FACTORS

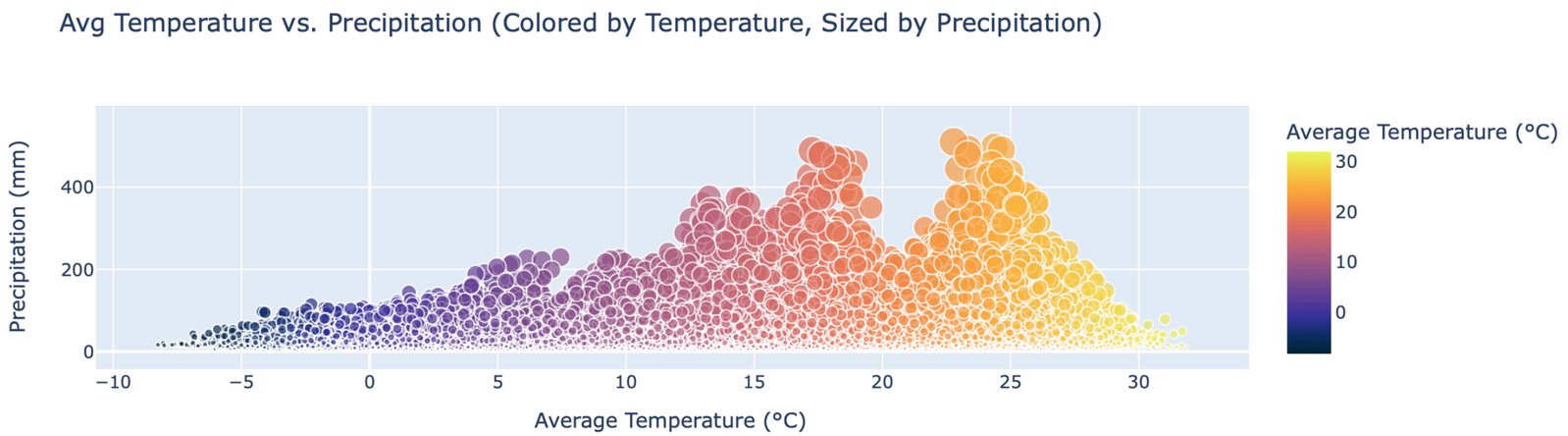
Multi-year climate records (temperature, precipitation, summer rainfall).

EDA - BOX PLOT OF THE WHEAT YIELD DATA

- **Central Tendency:** The median yield gives a clear indication of the central value.
- **Data Spread:** The interquartile range (IQR) shows the variability within the middle 50% of yields.
- **Outliers:** Points outside the whiskers flag potential anomalies that may require further investigation.
- **Skewness:** Any asymmetry in the box plot suggests possible skewness in the data distribution.

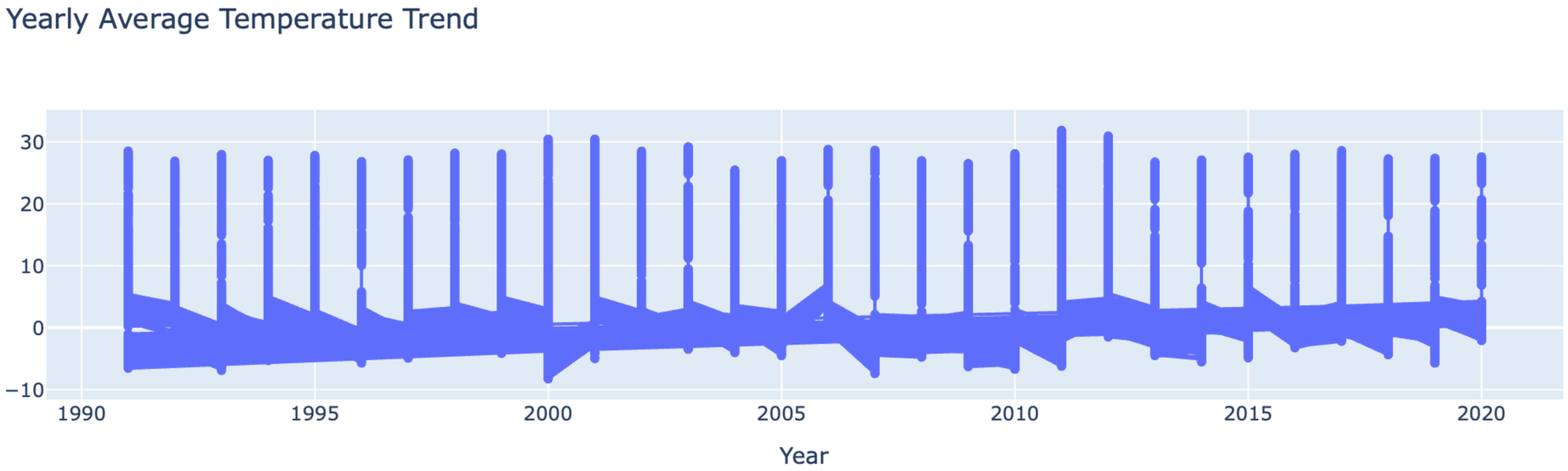


SCATTER PLOT: TEMPERATURE VS. PRECIPITATION

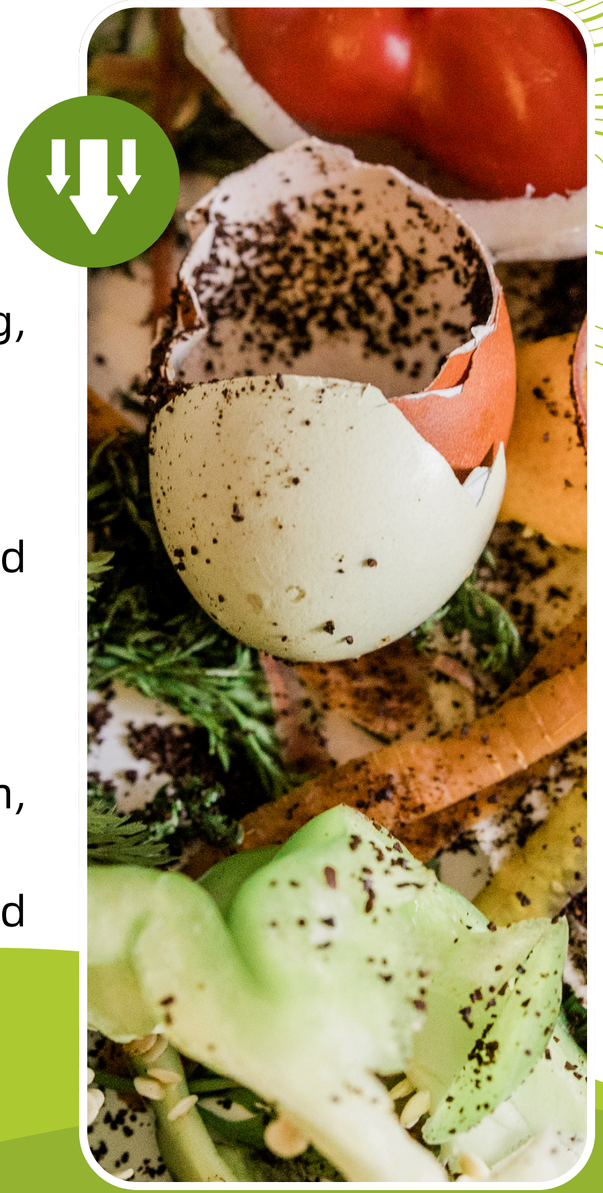


- This plot illustrates strong positive correlations between summer precipitation and yield while also indicating non-linear interactions between temperature and yield.
- Such insights are crucial for selecting appropriate model features and interaction terms.

TIME-SERIES LINE CHART FOR CLIMATE VARIABLES



NEXT STEPS



Immediate Actions:

- Enhance data cleaning (outlier handling, normalization).
- Improve merging strategies for data integrity.

Feature Engineering:

- Create temporal features (year, season) and interaction terms (temperature × precipitation).
- Introduce lag variables for temporal dependencies.

Baseline Modeling:

- Develop initial models (e.g., linear regression, decision trees).
- Evaluate and iterate toward more advanced techniques.

THANK YOU

Sustainable agriculture is key to a healthier, more resilient planet.