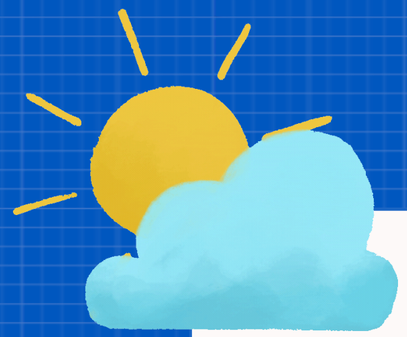


A Data Science Approach

Climate Change and Agriculture

project Aim

Many countries face challenges in producing certain crops due to factors such as climate conditions and soil problems.

These countries import these crops from other countries, so we have to study the impact of climate change on agriculture to know the best conditions for them by Identifying patterns and trends in crop yields over time.

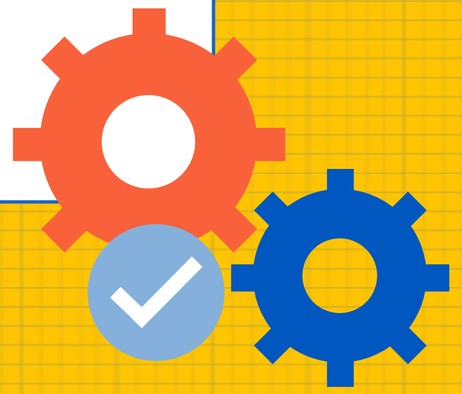
Data Source

climate_change_impact_on_agriculture_2024

Kaggle is the world’s largest data science community with powerful tools and resources to...

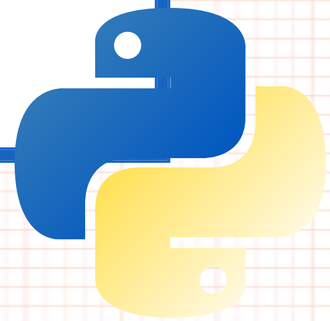
kaggle.com

Tools and Techniques



1

Python



https://www.kaggle.com/code/asmaaabdelazizradwan/climate-change-impact-on-agriculture

1
Exploratory Data Analytics (EDA)

summary of the DataFrame

```
df.info()
```

```
Data columns (total 15 columns):
 #   Column                      Non-Null Count  Dtype  
---  ------                      --------------  -----  
 0   Year                        10000 non-null  int64  
 1   Country                     10000 non-null  object 
 2   Region                      10000 non-null  object 
 3   Crop_Type                   10000 non-null  object 
 4   Average_Temperature_C       10000 non-null  float64
 5   Total_Precipitation_mm      10000 non-null  float64
 6   CO2_Emissions_MT            10000 non-null  float64
 7   Crop_Yield_MT_per_HA        10000 non-null  float64
 8   Extreme_Weather_Events      10000 non-null  int64  
 9   Irrigation_Access_%         10000 non-null  float64
 10  Pesticide_Use_KG_per_HA     10000 non-null  float64
 11  Fertilizer_Use_KG_per_HA    10000 non-null  float64
 12  Soil_Health_Index           10000 non-null  float64
 13  Adaptation_Strategies       10000 non-null  object 
 14  Economic_Impact_Million_USD 10000 non-null  float64
dtypes: float64(9), int64(2), object(4)
```

Display unique values of each categorical column

```
categorical_columns = df.select_dtypes(include=['object']).columns
numeric_columns = df.select_dtypes(include=['number']).columns

tot_len= 0
for i in df.columns:
    if i not in numeric_columns:
        print(f'\033[31m{i.upper()}\033[0m'.center(50, ' '))
        print(df[i].value_counts())
        tot_len= tot_len+ df[i].nunique()
        print()
        print()
```

Adaption Strategies

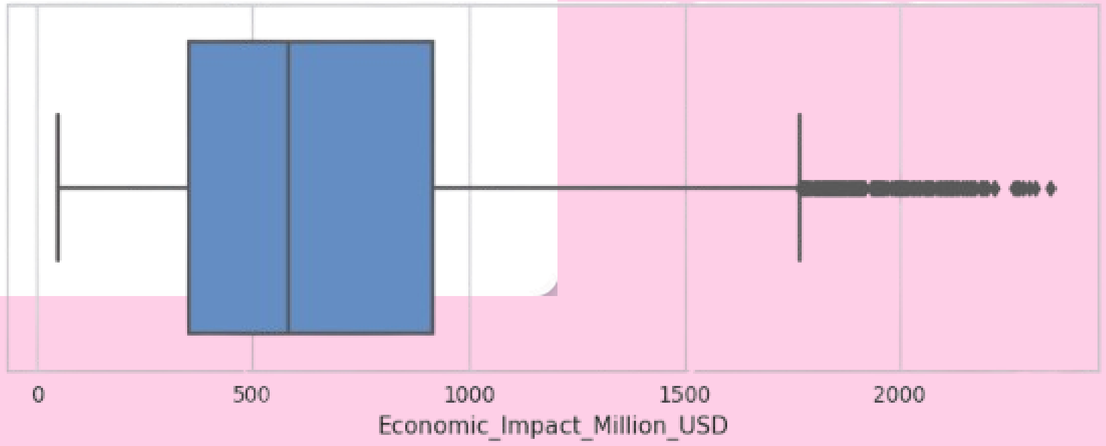
Water Management	2049
No Adaptation	2024
Drought-resistant Crops	1995
Organic Farming	1975
Crop Rotation	1957

Wheat	1047
Cotton	1044
Vegetables	1036
Corn	1022
Rice	1022
Sugarcane	995
Fruits	979
Soybeans	958
Barley	952
Coffee	945

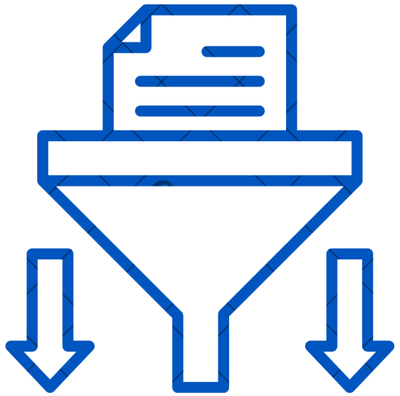
Crop type

Outliers Detection

```
n_rows = 6
n_cols = 2

fig, axes = plt.subplots(n_rows, n_cols, figsize=(15, 20))
fig.suptitle('Box Plot of Numeric Features', fontsize=16)

axes = axes.flatten()

for i, feature in enumerate(numeric_columns):
    sns.boxplot(x=df[feature], ax=axes[i])
    axes[i].set_title(f'Box Plot of {feature}')

for j in range(i + 1, n_rows * n_cols):
    fig.delaxes(axes[j])

plt.tight_layout()
plt.subplots_adjust(top=0.9)
plt.show()
```

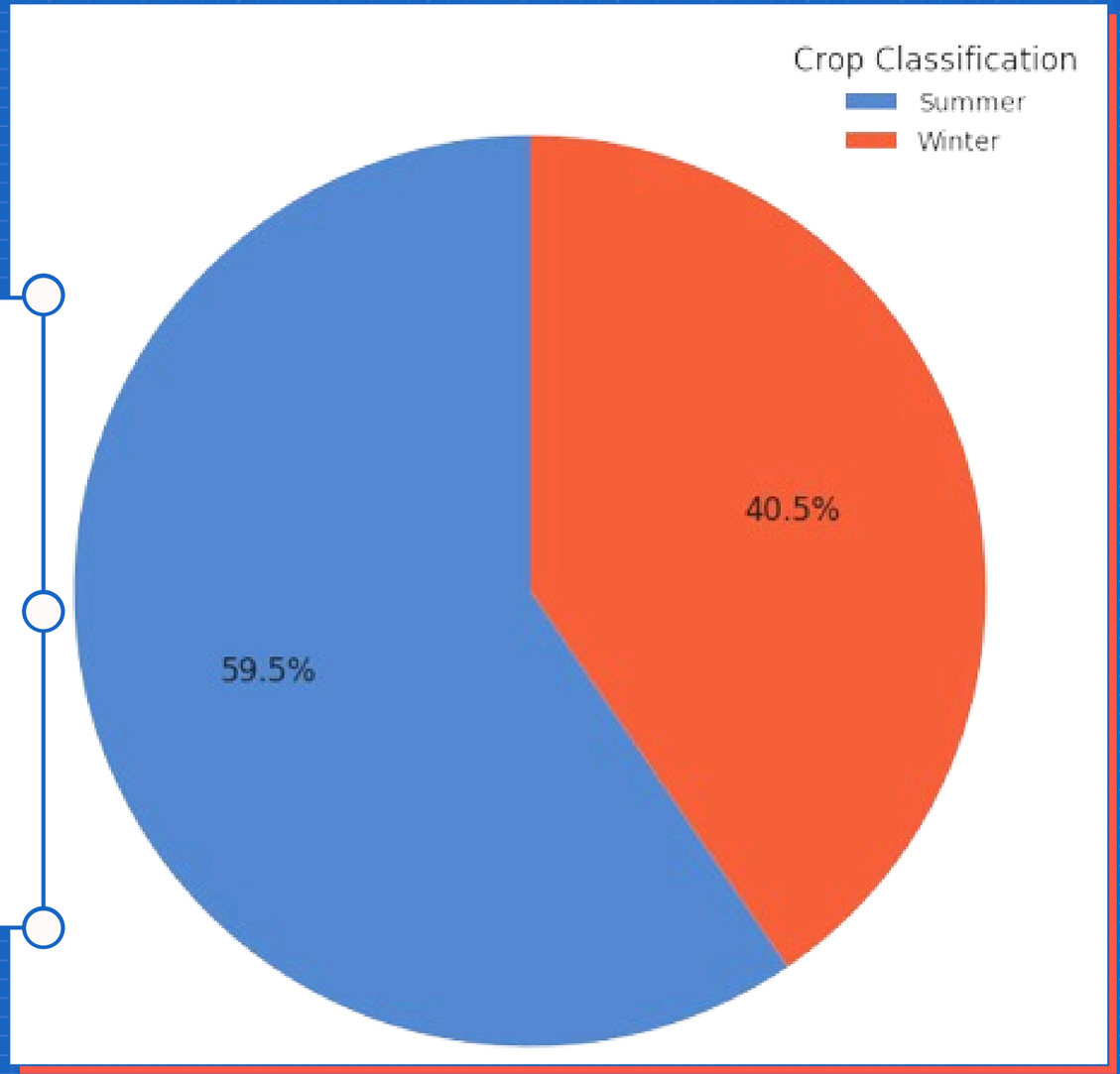


2
Feature Extraxtion

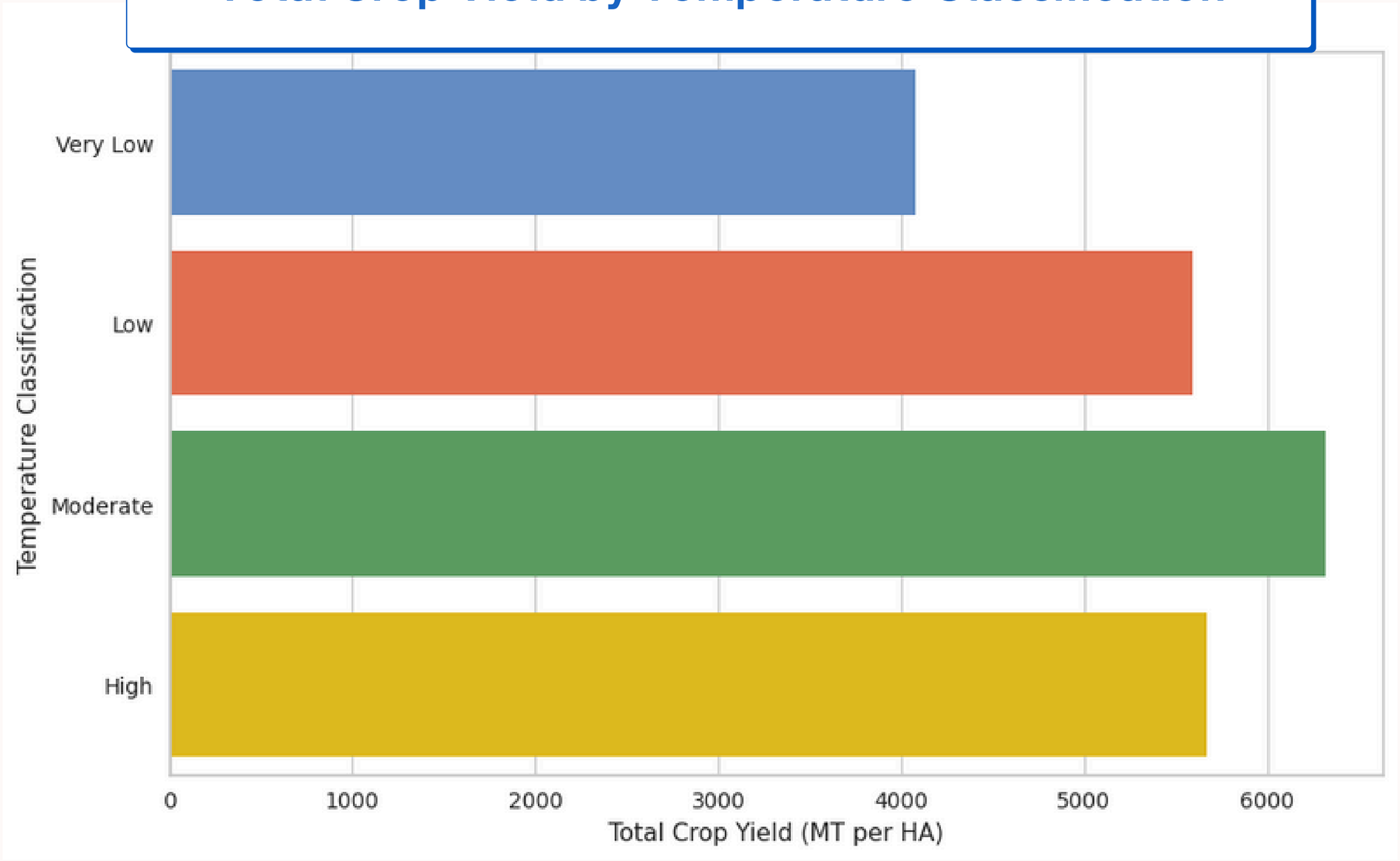


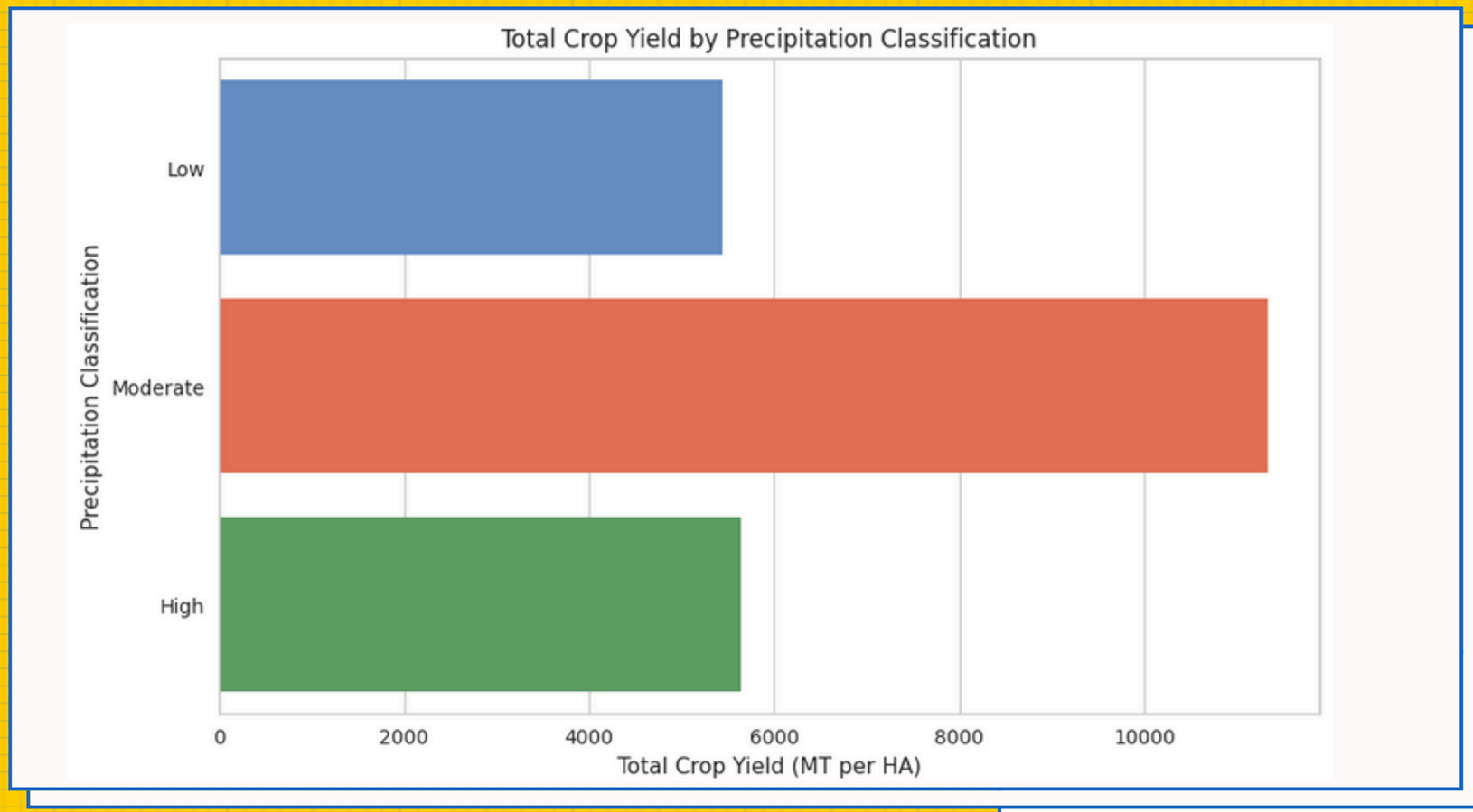
suggest a need to investigate if winter crops can be optimized through improved techniques, irrigation, or crop selection.

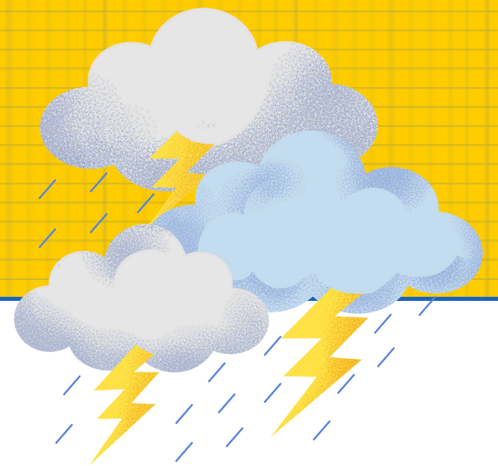
agricultural yield is more favorable during the summer season compared to winter



Total Crop Yield by Temperature Classification

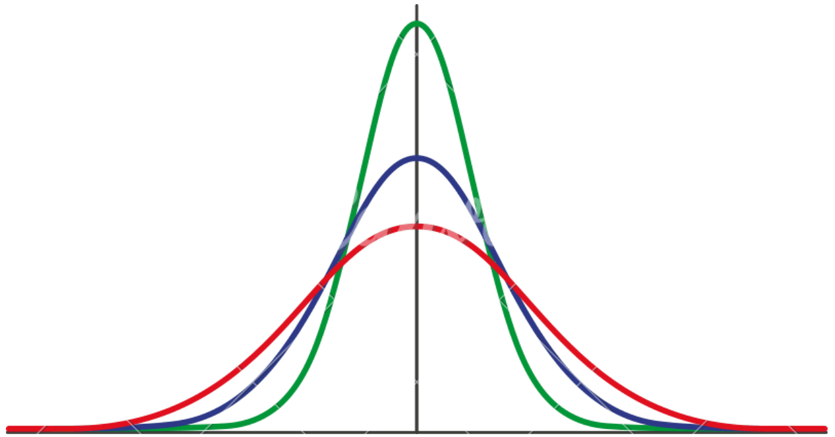




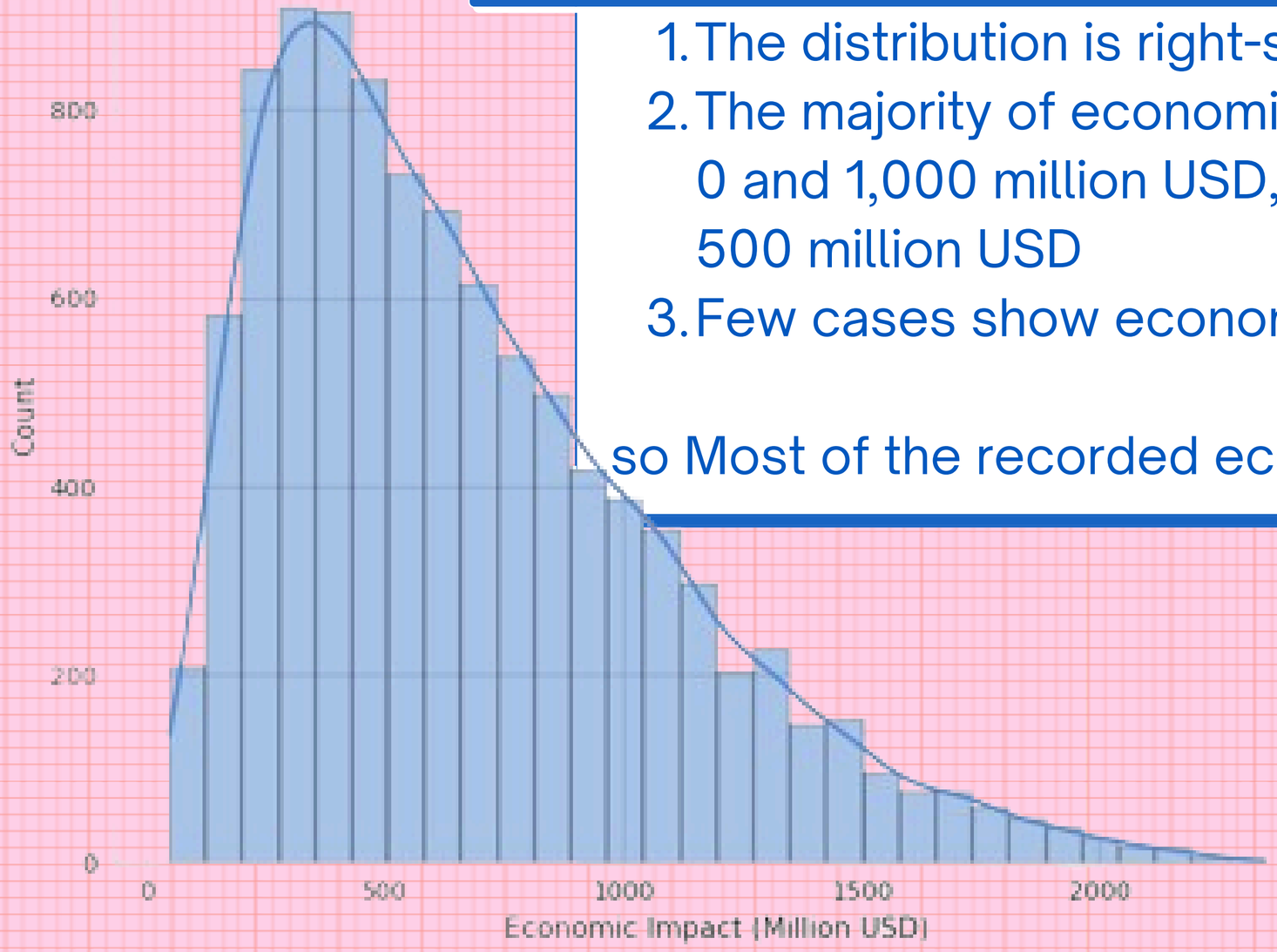


reflect the importance of balanced water availability—neither drought nor excessive water is ideal for most crops

3
Data Distribution



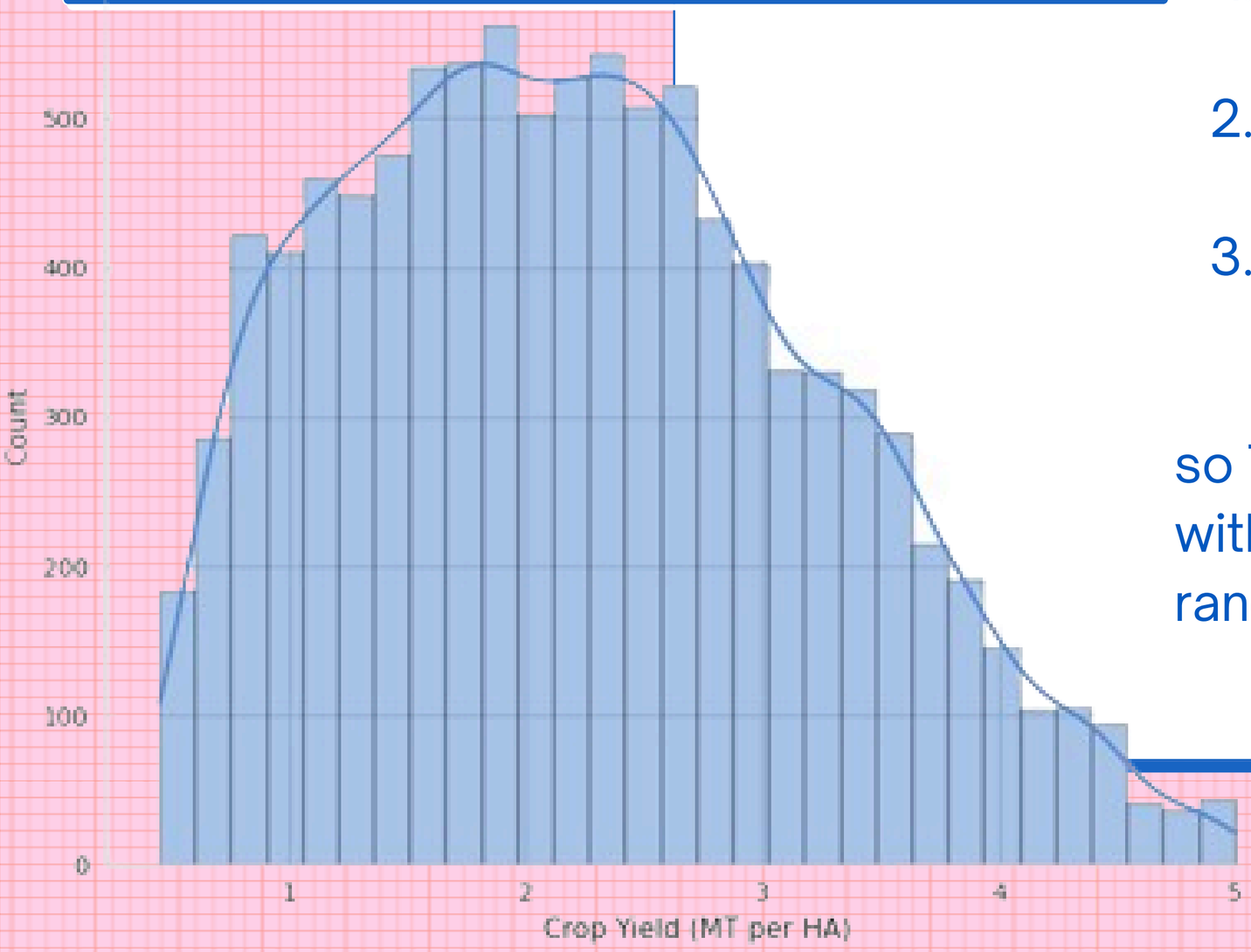
Distribution of Economic Impact (Million USD)



1. The distribution is right-skewed (positive skew)
2. The majority of economic impacts are concentrated between 0 and 1,000 million USD, with the highest frequency around 500 million USD
3. Few cases show economic impacts above 2,000 million USD

so Most of the recorded economic impacts are relatively low

Distribution of Crop Yield (MT per HA)

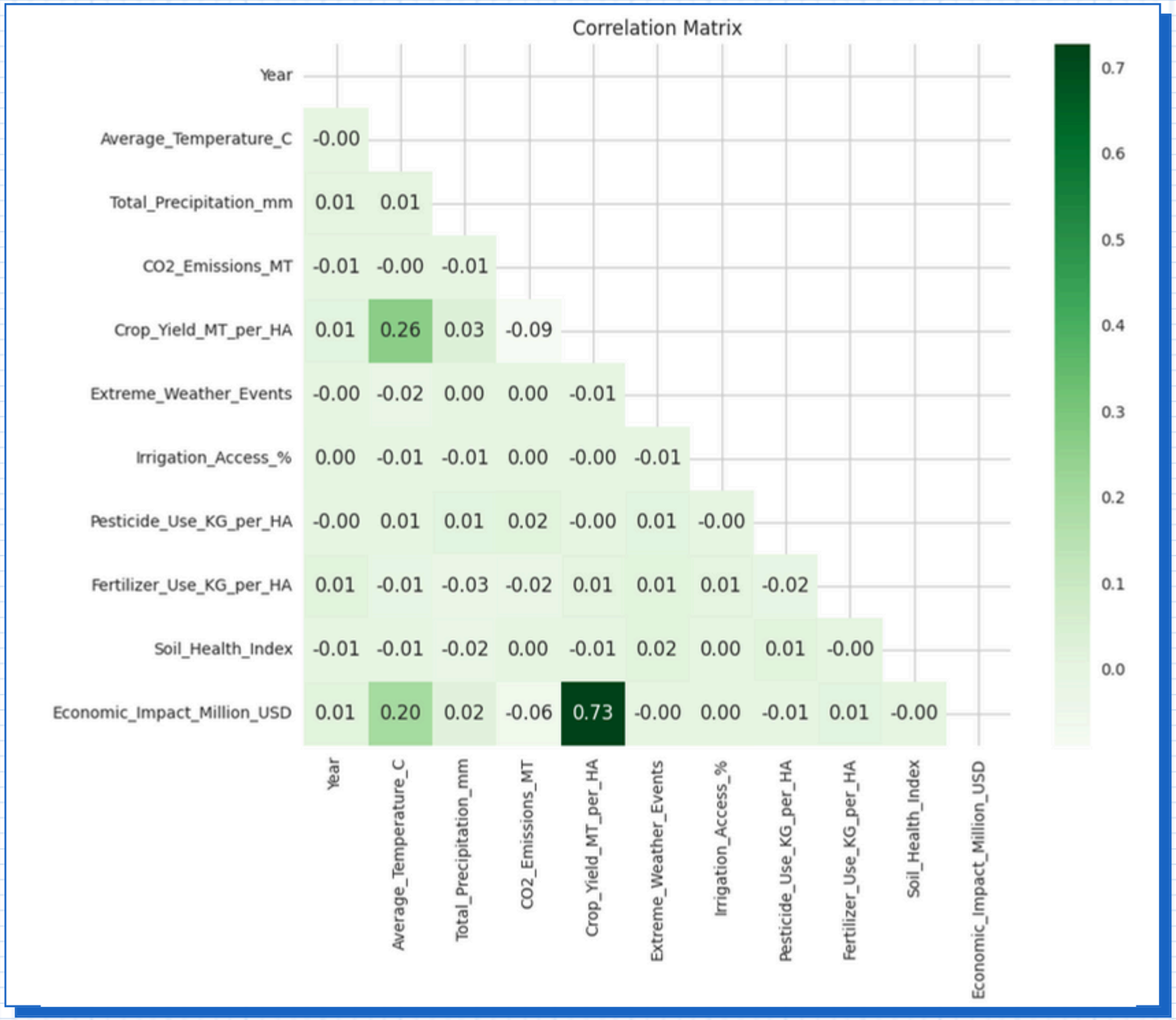


1. The distribution is approximately normal, with a slight left skew.
2. Crop yield values are concentrated between 2 and 3 MT per HA
3. Few observations exist below 1 MT per HA or above 4 MT per HA.

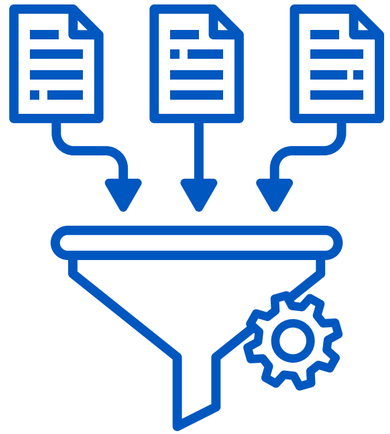
so The crop yield data is fairly consistent, with most yields falling within a narrow range

Correlation matrix

'Crop_Yield_MT_per_HA' with 'Economic_Impact_Million_USD' has the highest correlation value

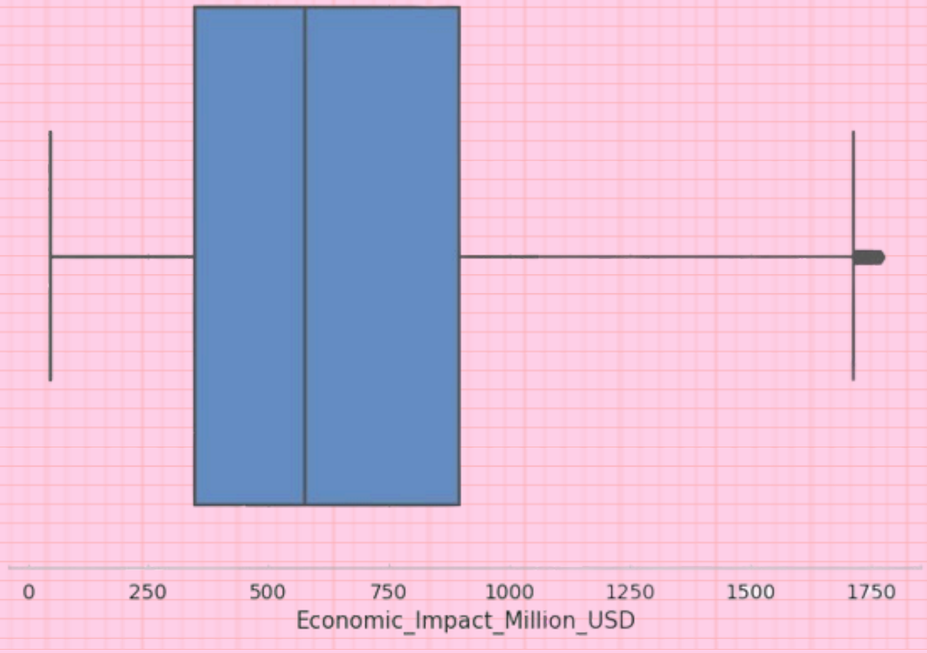


3
Data Preprocessing



Removing Outliers

```
Q1 = df['Economic_Impact_Million_USD'].quantile(0.25)
Q3 = df['Economic_Impact_Million_USD'].quantile(0.75)
IQR = Q3 - Q1
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR
outliers = df[(df['Economic_Impact_Million_USD'] < lower_bound) | (df['Economic_Impact_Million_USD'] > upper_bound)]

df = df.drop(outliers.index)
```



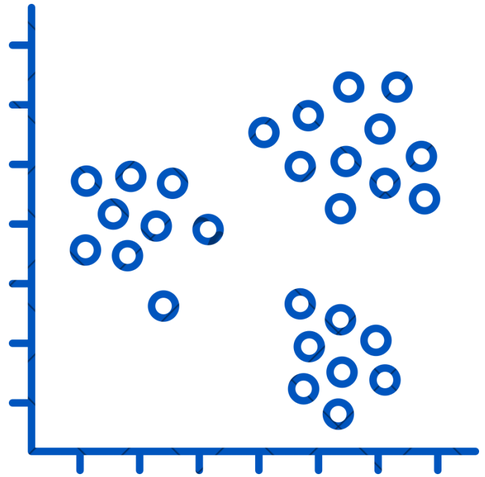
categorical columns encoding

```
label_encoder = LabelEncoder()

for col in categorical_columns:
    clus_df[col] = label_encoder.fit_transform(clus_df[col])

clus_df.head()
```

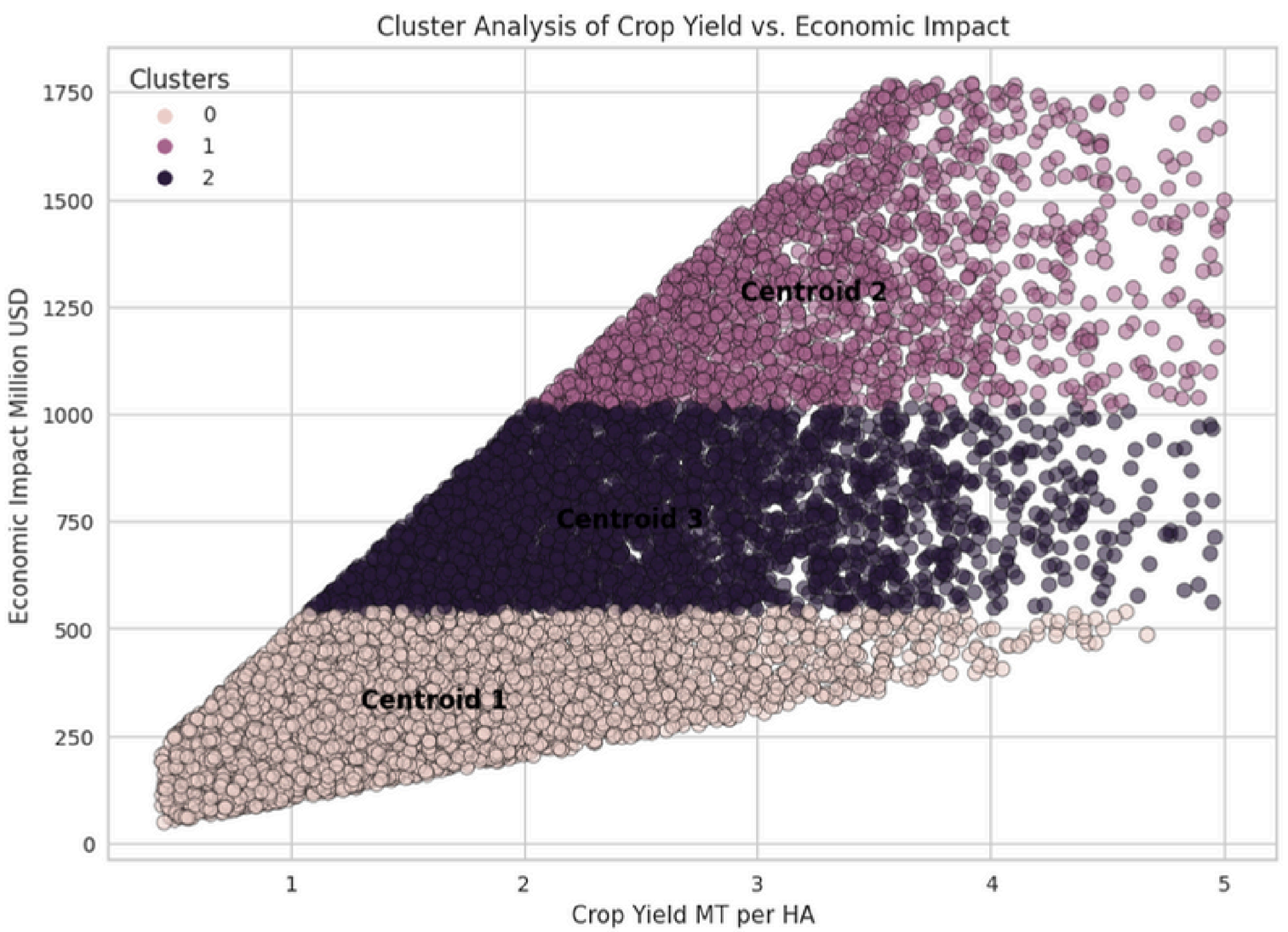
Data Standardization

```
scaler = StandardScaler()

numeric_features = clus_df.select_dtypes(include=['number']).columns
clus_df[numeric_features] = scaler.fit_transform(clus_df[numeric_features])

clus_df.head()
```

4
predictive modeling (K-means)



clusters characteristics




```
Centroid 1: Average Crop Yield MT per HA = 1.61, Average Economic Impact Million USD = 330.23
Centroid 2: Average Crop Yield MT per HA = 3.24, Average Economic Impact Million USD = 1281.80
Centroid 3: Average Crop Yield MT per HA = 2.45, Average Economic Impact Million USD = 751.39
```

Insights on clusters

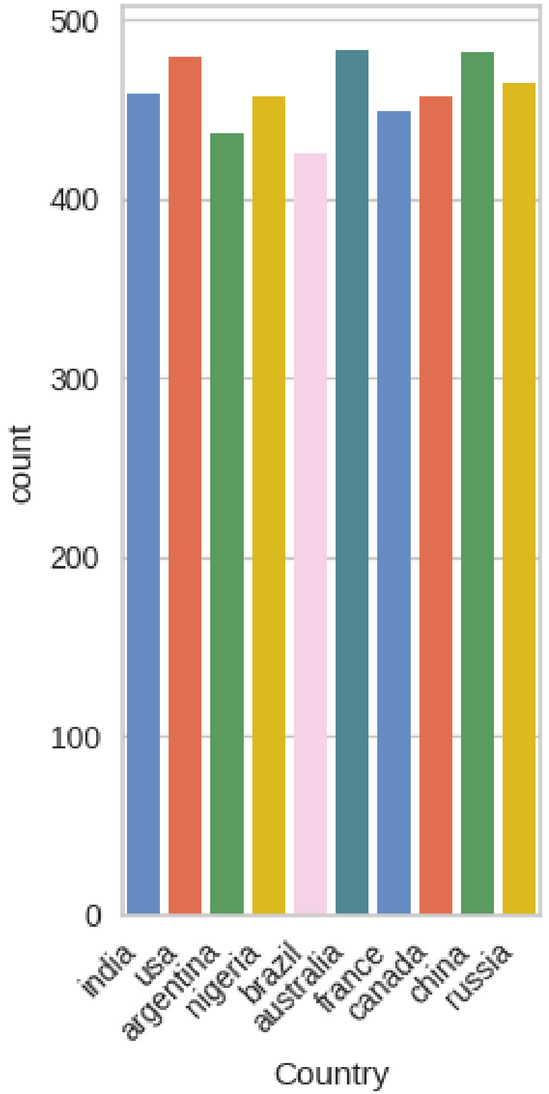
cluster	0	1	2
Year	2007.069357	2007.008380	2007.075116
Average_Temperature_C	13.281117	18.884553	15.856569
Total_Precipitation_mm	1601.844257	1650.634162	1600.470383
CO2_Emissions_MT	15.674622	14.633397	15.072201
Crop_Yield_MT_per_HA	1.613348	3.240931	2.450548
Extreme_Weather_Events	5.028571	4.970391	4.924304
Irrigation_Access_%	55.190410	55.003659	55.314806
Pesticide_Use_KG_per_HA	25.144550	24.811045	24.723999
Fertilizer_Use_KG_per_HA	49.825505	50.634587	49.770667
Soil_Health_Index	65.141117	65.119291	64.409620
Economic_Impact_Million_USD	329.908829	1281.944034	751.042488

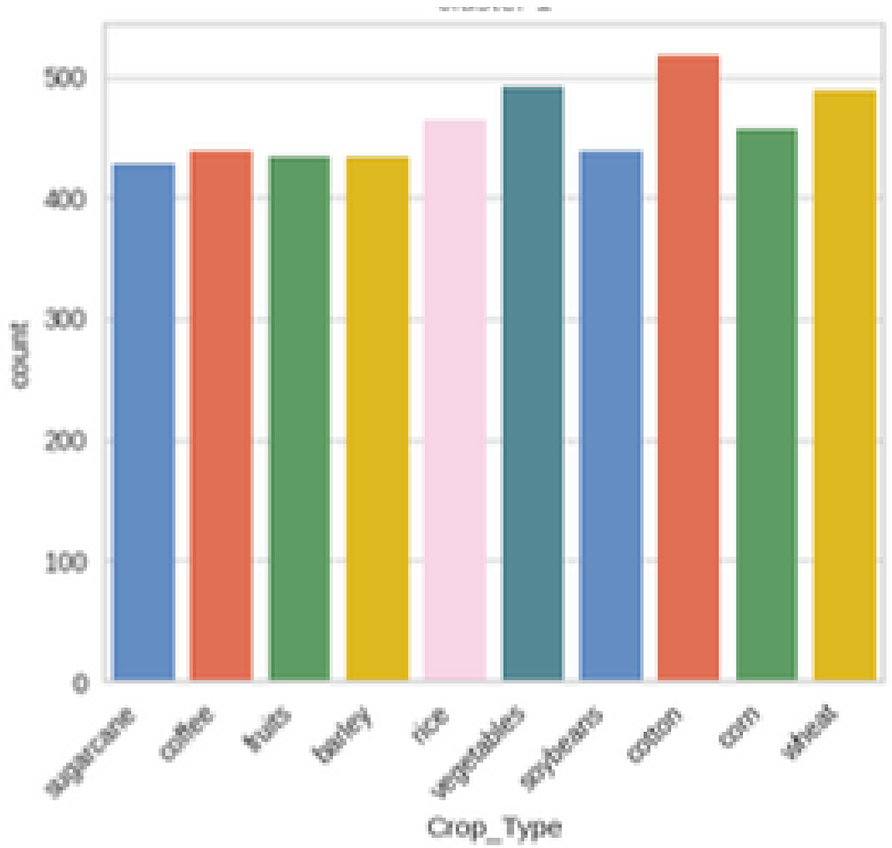
- Cluster 1: Has the highest CO2 emissions, extreme weather events, and pesticide use intensity. However, the lowest average temperature and crop yield lead to the lowest economic impact of the other two clusters.

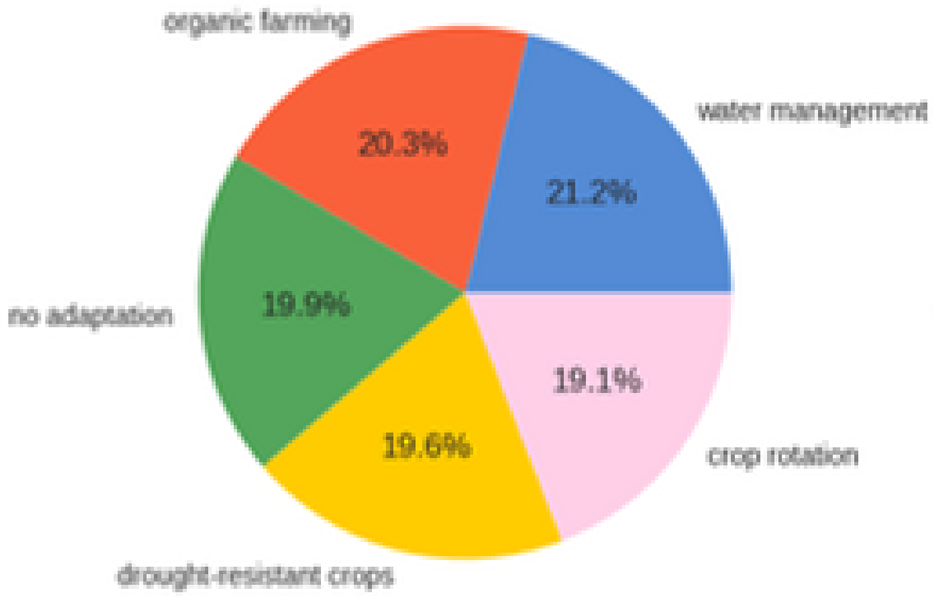
- Cluster 2: Despite the highest average temperature, precipitation, and fertilizer use and the lowest CO2 emissions, the cluster has the greatest economic impact. This may be due to several factors, including lower extreme weather events and the highest crop yields.

- Cluster 3 has a higher average temperature than Cluster 1, the lowest extreme weather events, the highest irrigation access, and much greater crop yield than Cluster 1.

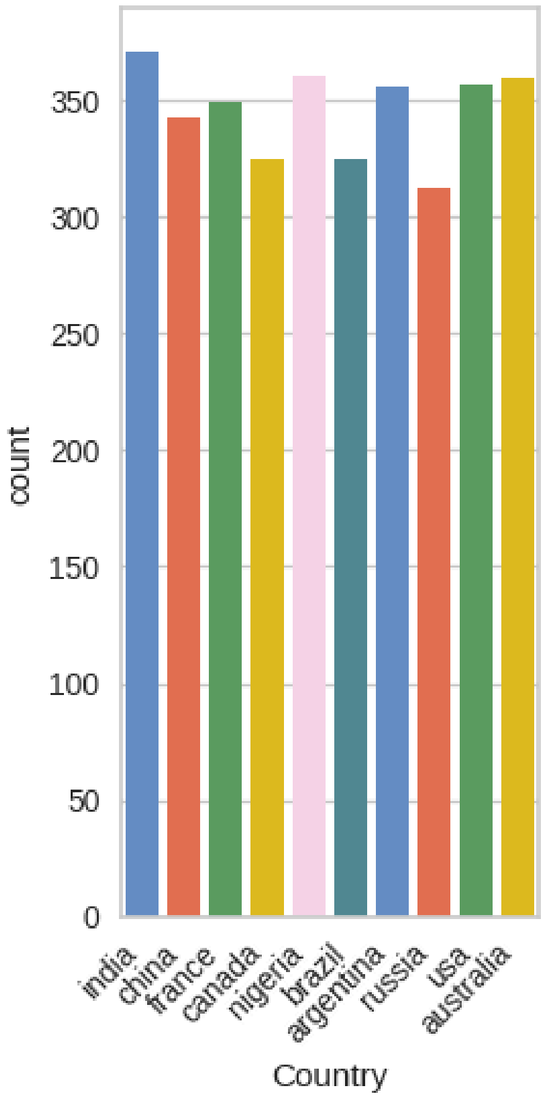
Cluster 1

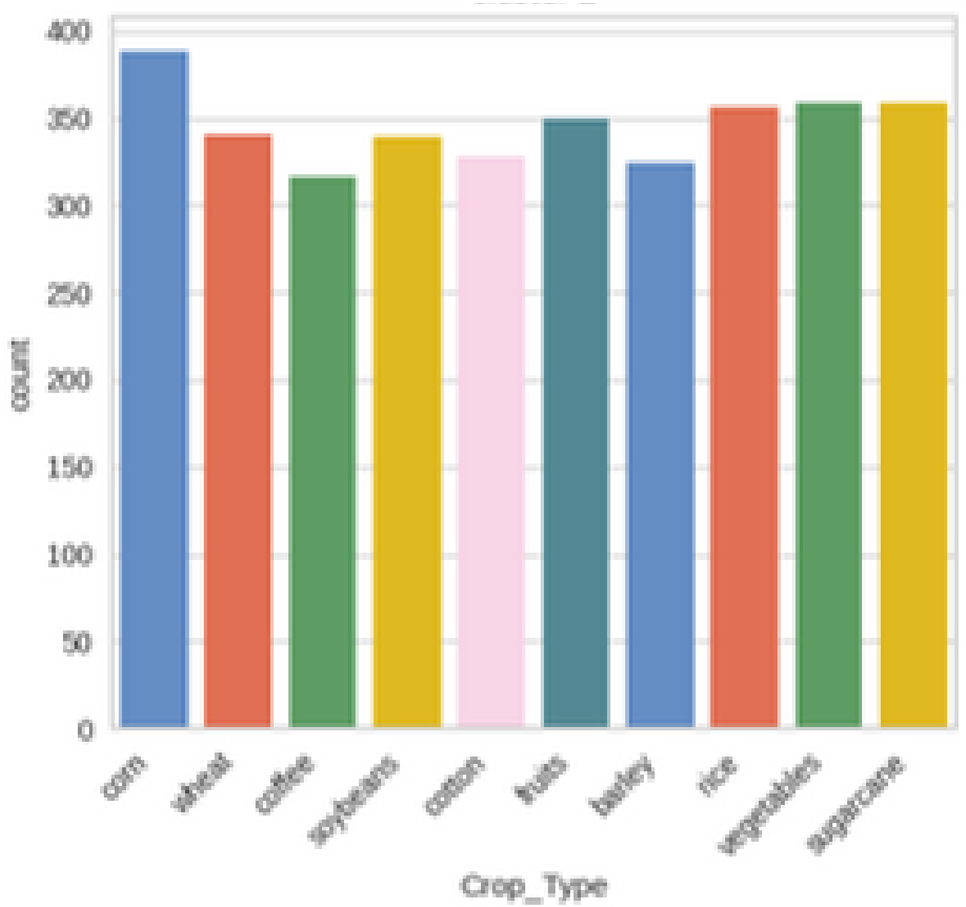

500
400
300
200
100
0
count
india
usa
argentina
nigeria
brazil
australia
france
canada
china
russia
Country

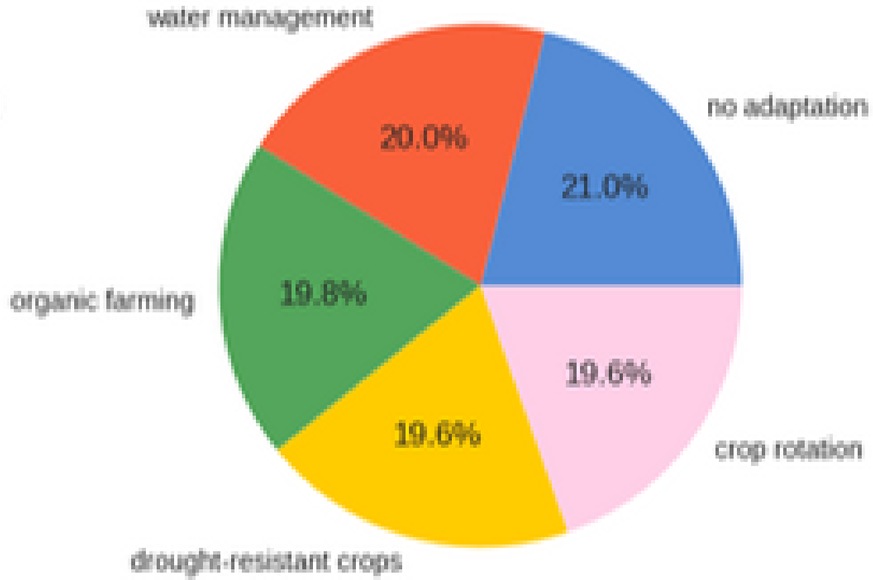

500
400
300
200
100
0
count
sugarcane
coffee
fruits
barley
rice
vegetables
soybeans
cotton
corn
wheat
Crop_Type


organic farming
water management
no adaptation
crop rotation
drought-resistant crops
20.3%
21.2%
19.9%
19.1%
19.6%

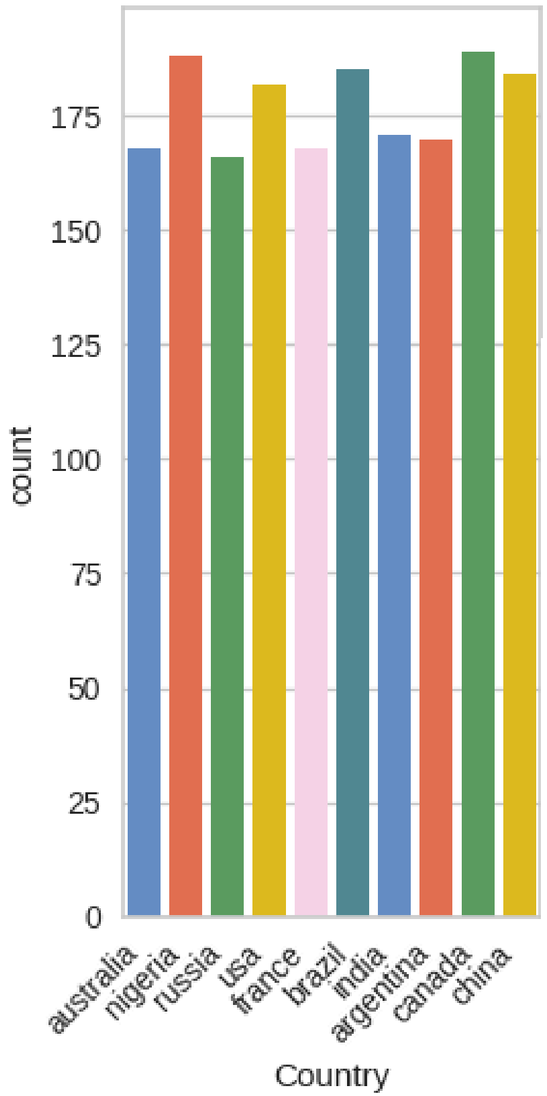
Cluster 2

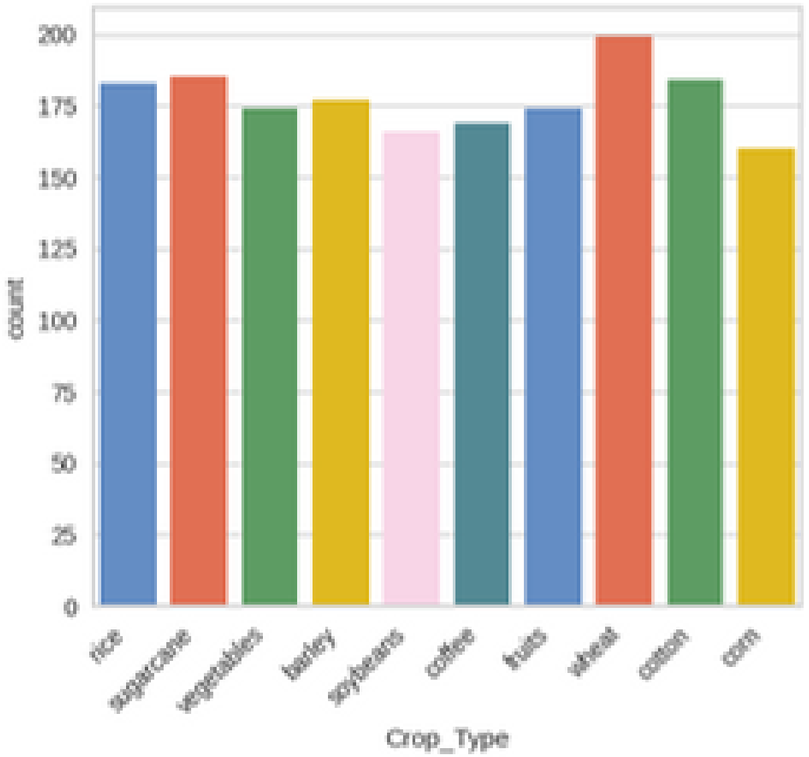

count
350
300
250
200
150
100
50
0
india
china
france
canada
nigeria
brazil
argentina
russia
usa
australia
Country

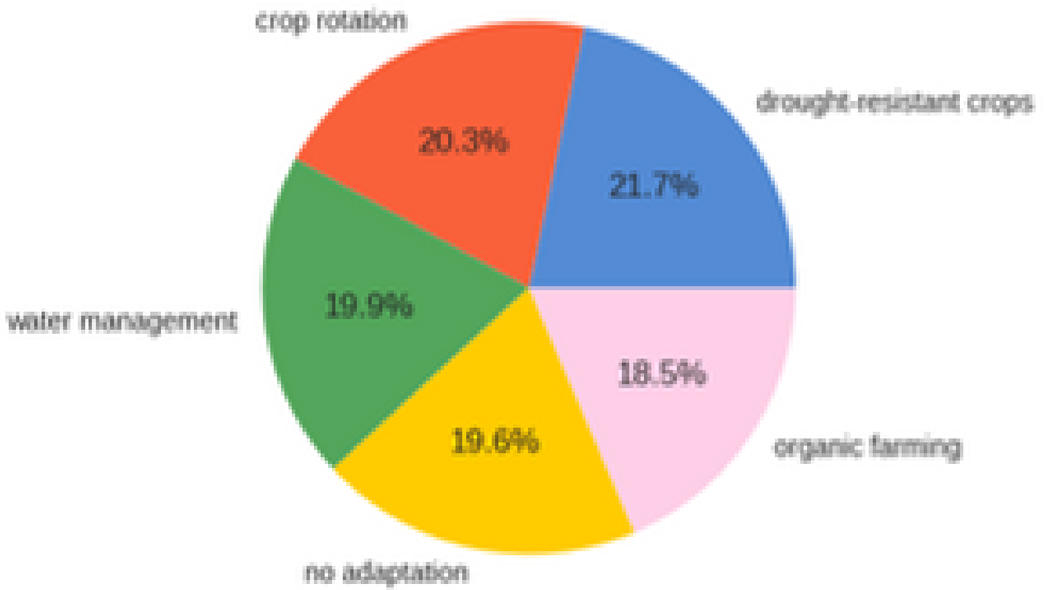

count
400
350
300
250
200
150
100
50
0
corn
wheat
coffee
soybeans
cotton
fruits
barley
rice
vegetables
sugarcane
Crop_Type


water management
no adaptation
organic farming
crop rotation
drought-resistant crops
20.0%
21.0%
19.8%
19.6%
19.6%

Cluster 3


count
175
150
125
100
75
50
25
0
australia
nigeria
russia
usa
france
brazil
india
argentina
canada
china
Country


count
200
175
150
125
100
75
50
25
0
rice
sugarcane
vegetables
barley
soybeans
coffee
fruits
wheat
cotton
corn
Crop_Type


crop rotation
drought-resistant crops
water management
organic farming
no adaptation
20.3%
21.7%
19.9%
18.5%
19.6%

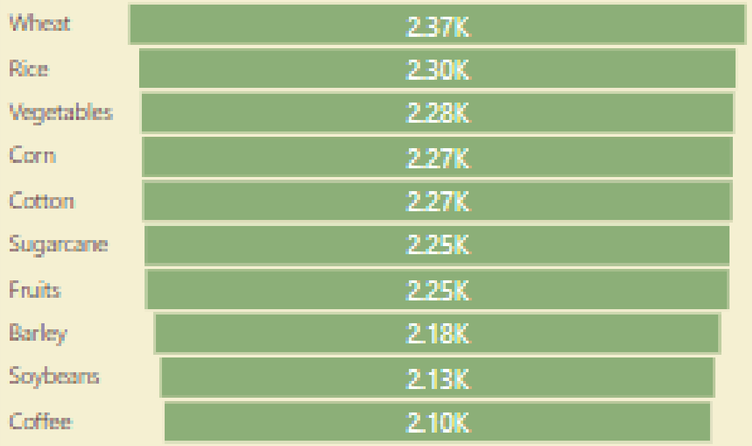
2

Power BI

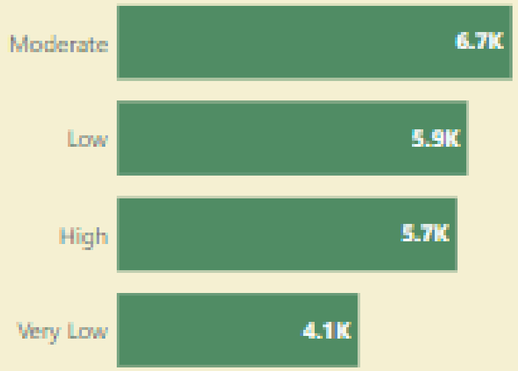
Overall Insights

Home | Overall | Crop_Types | Insights...

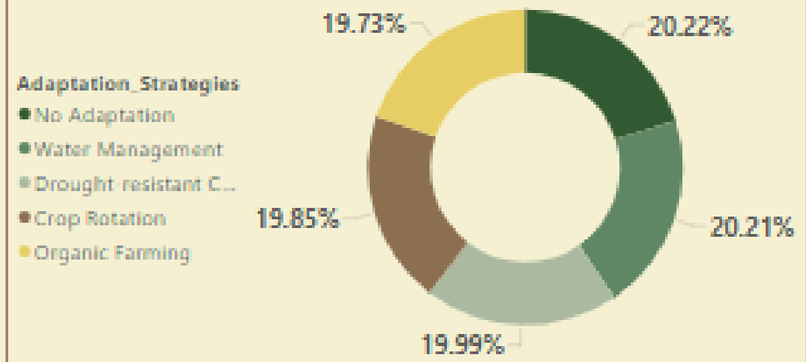
Crop Yield per HA by Crop_Type



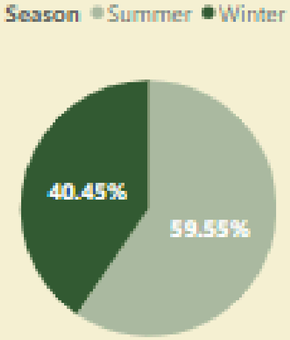
Crop Yield per HA by Temp_classification



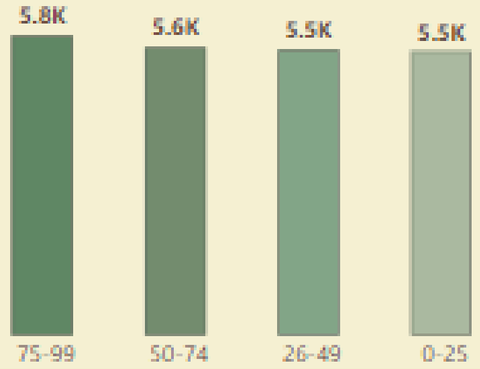
Crop Yield per HA by Adaptation_Strategies



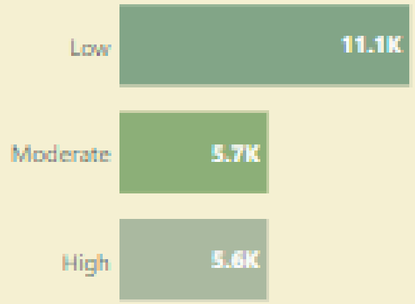
Crop Yield per HA by Season



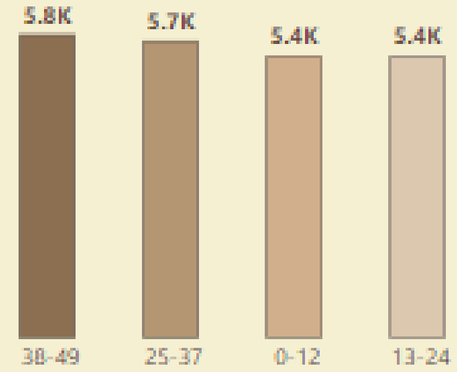
Crop Yield per HA by Fertilizer



Crop Yield per HA by Precp_classification



Crop_Yield by pesticide-classification



Crop Types

Home | Overall | Crop_Types | Insights...

Crops


Barley

Coffee

Corn

Cotton

Fruits


Rice

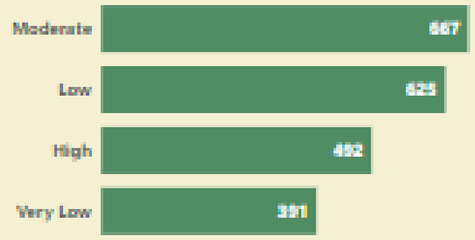
Soybeans

Sugarcane

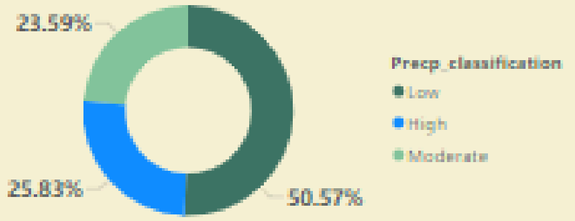
Vegetables

Wheat

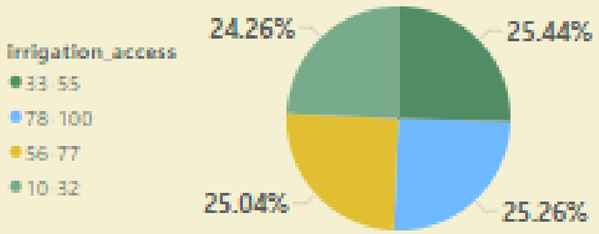
Crop Yield by Temp_classification



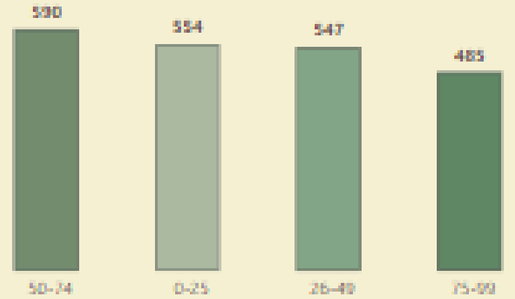
Crop Yield by Precp_classification



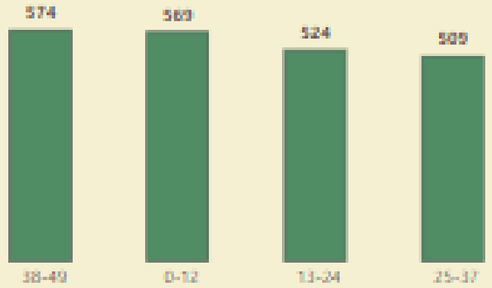
Crop Yield per by Irrigation_access_classification



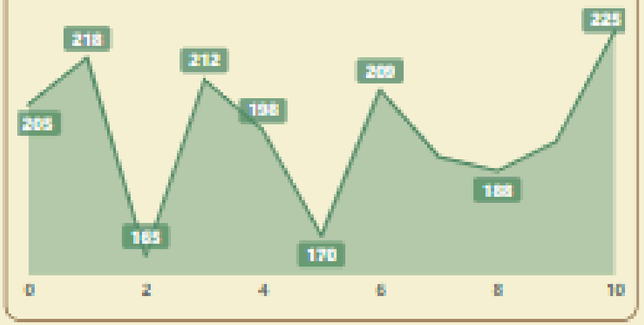
Crop Yield by Fert-classification



Crop Yield by pesticide-classification



Crop Yield by Extreme_Weather_Events



Recommendations

- use greenhouses that maintain ideal temperature by isolating plants from external weather.
- use sensors that monitor the plant's needs for fertilizer, pesticides, and water level.

THANK YOU!