Indian Public Distribution System: An Analysis

A BUSINESS ANALYTICS PROJECT REPORT

Submitted to
DR. PIYUSH CHAUHAN
Associate Professor
Department of Computer Science & Engineering
Symbiosis Institute of Technology, Nagpur Campus

Submitted by
SUJAL JUNGHARE
VI SEM
PRN: 22070521089
Department of Computer Science & Engineering
Symbiosis Institute of Technology, Nagpur Campus

Under the Guidance of
AMIT MAKODE SIR

Course Name: *Business Analytics*
Course Code: *T2228*

॥वसुधैव कुटुम्बकम्॥
SYMBIOSIS
INSTITUTE OF TECHNOLOGY, NAGPUR

Table of Contents

I. Project Definition and Data Understanding

A. Overview

The Public Distribution System (PDS) is a vital food security initiative in India, aimed at ensuring the availability of essential food grains like wheat and rice to economically disadvantaged sections of society at subsidized rates. Given the scale and logistical complexity of the system, it becomes crucial to monitor and assess its efficiency in distributing allocated resources to the intended beneficiaries. This project focuses on analysing district-wise data on the allocation and distribution of wheat and rice through the PDS. By applying data analysis, regression modelling, and clustering techniques, the study aims to uncover distribution patterns, identify operational inefficiencies, and provide insights that can support evidence-based policy decisions for strengthening the public distribution framework.

B. Objective

To analyse and evaluate the **efficiency of food grain distribution** (Rice and Wheat) under the Public Distribution System at a **district level** over time.

C. Key Questions

- Which states or districts have the highest/lowest distribution efficiency?
- Are allocations being fully utilized (i.e., is distribution close to allocation)?
- Are there patterns across months or regions in distribution performance?
- Any scope to identify anomalies or underperforming regions?

D. Data Dictionary

Column	Description
Month	Month and Year
State	Name of the state
State Code	Unique code for state
District	District name
District Code	Unique district code
Total Rice Allocated	Amount of rice allocated (in metric tonnes)
Total Wheat Allocated	Amount of wheat allocated
Total Rice Distributed	Actual rice distributed
Total Wheat Distributed	Actual wheat distributed

II. Data Collection and Integration

A. Data Source

The dataset was collected from the official India Data Portal, specifically from the Public Distribution System (PDS) allocation section provided by the **Ministry of Consumer Affairs, Food and Public Distribution**.

B. Data Format

The data was downloaded in **CSV format**, containing monthly district-wise allocation and distribution information for **rice and wheat** under the **National Food Security Act (NFSA)**.

C. Data Provenance

- The data is **publicly available**, curated and hosted by the **Indian School of Business (ISB)** under their Open Data initiative.
- Data values are sourced directly from **official PDS records**, ensuring authenticity.
- Timestamped entries and identifiers like state_code and district_code preserve referential integrity.

D. Data Integration

- As the data came in a structured format, no complex data merging or API integration was required.
- Column headers were standardized and cleaned to support analysis.
- Date formats were parsed to support time-series operations.

```
# Standardize column names
df.columns = [
    "month", "state", "state_code", "district", "district_code",
    "rice_allocated", "wheat_allocated", "rice_distributed", "wheat_distributed"
]
```

```
# Convert 'month' to datetime format
df["month"] = pd.to_datetime(df["month"], format="%Y-%m-%d")
```

E. Validation

- The dataset was checked for **missing values, inconsistencies, and zero allocation cases**.
- Structural integrity was confirmed: each row represents one district's data for a specific month.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 9 columns):
 #   Column                                             Non-Null Count  Dtype
---  ------                                             --------------  -----
 0   Month (month)                                      500 non-null    object
 1   State (state_name)                                 500 non-null    object
 2   State Code (state_code)                            500 non-null    int64
 3   District (district_name)                           500 non-null    object
 4   District Code (district_code)                      500 non-null    int64
 5   Total Rice Allocated (total_rice_allocated)        500 non-null    float64
 6   Total Wheat Allocated (total_wheat_allocated)      500 non-null    float64
 7   Total Rice Distributed (total_rice_distributed)    500 non-null    float64
 8   Total Wheat Distritibuted (total_wheat_distributed) 500 non-null   float64
dtypes: float64(4), int64(2), object(3)
```

III. Data Cleaning and Preparation (Python)

A. Handling Missing Values

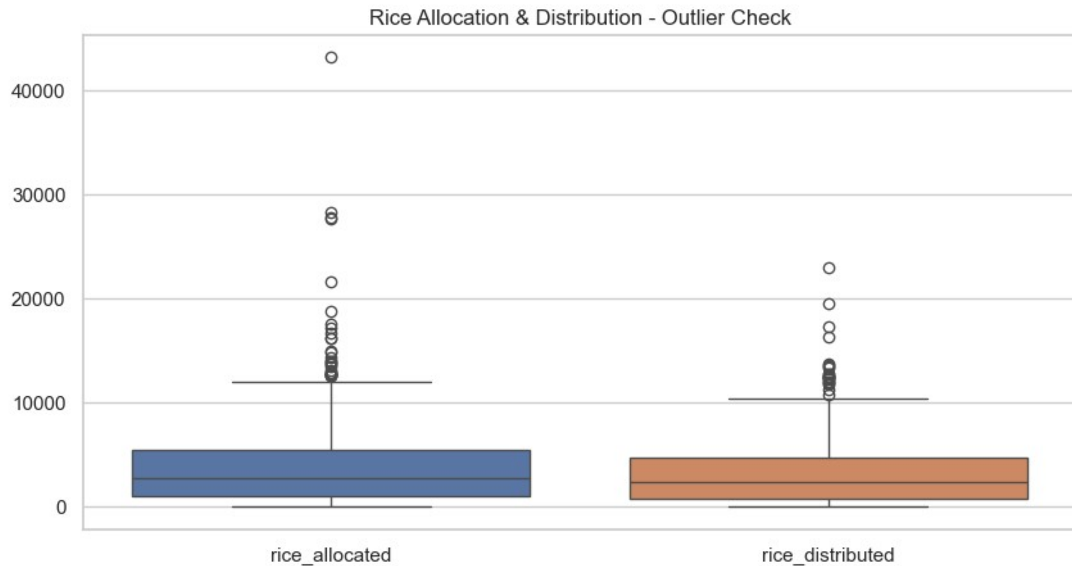
The dataset was checked for missing values across all columns. No critical missing data was found, so no imputation or removal was needed.

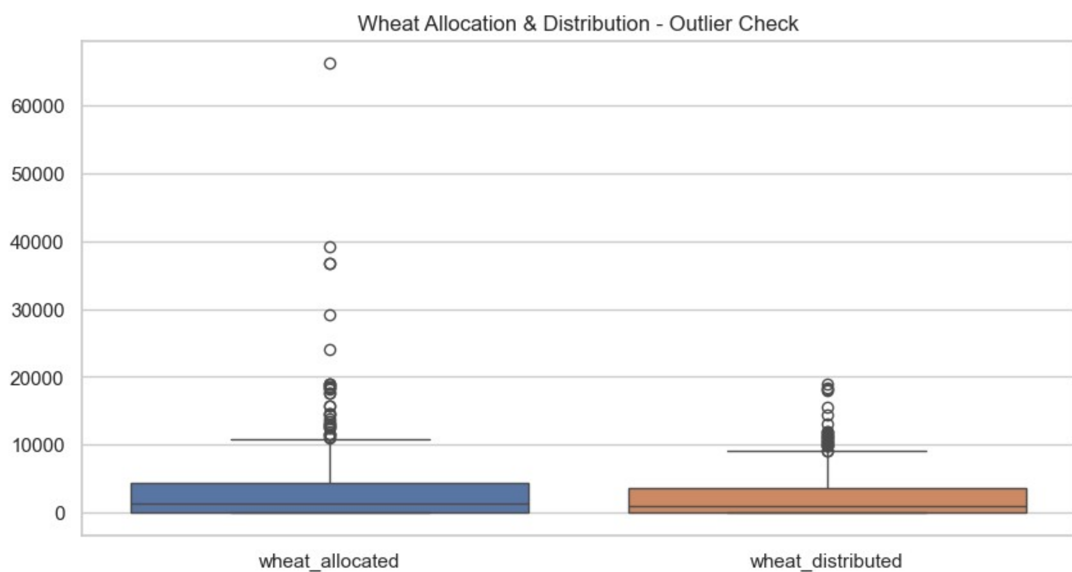
```
# Check for missing values
df.isnull().sum()

Month (month)                                        0
State (state_name)                                   0
State Code (state_code)                              0
District (district_name)                             0
District Code (district_code)                        0
Total Rice Allocated (total_rice_allocated)          0
Total Wheat Allocated (total_wheat_allocated)        0
Total Rice Distributed (total_rice_distributed)      0
Total Wheat Distritibuted (total_wheat_distributed)  0
dtype: int64
```

B. Address Outliers

Boxplots were used to visually inspect potential outliers in the allocation and distribution of rice and wheat. Since values reflect district-level variations (e.g., population), outliers were retained as meaningful.





C. Correct Data Types and Formats

The month column was originally in string format and was converted to datetime format for accurate time-based analysis.

```
# Convert 'month' to datetime format
df["month"] = pd.to_datetime(df["month"], format="%Y-%m-%d")
```

D. Normalize / Standardize Numerical Variables

Standardization was not required at this stage as the project focused on ratio-based efficiency metrics. However, normalization can be considered in later modelling stages.

```
from sklearn.preprocessing import StandardScaler

# Normalize numerical variables (optional)
scaler = StandardScaler()
scaled_columns = ["rice_allocated", "wheat_allocated", "rice_distributed", "wheat_distributed"]

df_scaled = df.copy()
df_scaled[scaled_columns] = scaler.fit_transform(df_scaled[scaled_columns])
df_scaled.head()
```

E. Encode Categorical Variables

Encoding wasn't necessary for the current analysis. However, for predictive modelling or clustering, label encoding for state and district may be used.

```
# Example: Label Encoding for 'state' and 'district'
from sklearn.preprocessing import LabelEncoder

le_state = LabelEncoder()
df["state_encoded"] = le_state.fit_transform(df["state"])

le_district = LabelEncoder()
df["district_encoded"] = le_district.fit_transform(df["district"])
```

F. Create Derived Features and Aggregations

Two new features were created to evaluate efficiency:

- ***rice_efficiency*** = rice_distributed / rice_allocated
- ***wheat_efficiency*** = wheat_distributed / wheat_allocated
  (Handled safely to avoid division by zero)

G. Documentation for Reproducibility

- All cleaning steps were implemented sequentially in a Jupyter Notebook.
- Code cells are annotated for readability and reproducibility.
- Transformations are modular and easy to trace or revert if necessary.

IV. Exploratory Data Analysis (Python)

A. Generate Descriptive Statistics

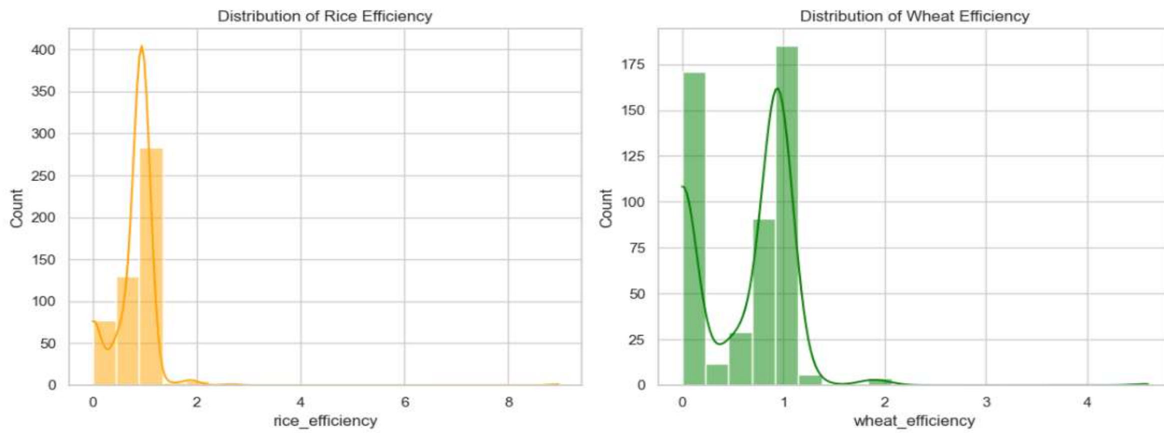
Basic statistics such as mean, median, and standard deviation were computed to understand the central tendencies and dispersion of numerical fields.

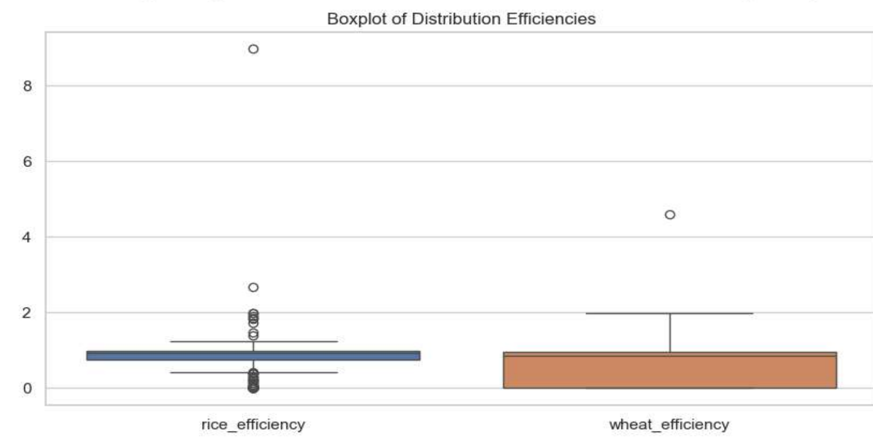
	month	state_code	district_code	rice_allocated	wheat_allocated	rice_distributed	wheat_distributed	rice_efficiency	wheat_efficiency
count	500	500.000000	500.000000	500.00000	500.000000	500.000000	500.000000	500.000000	500.000000
mean	2019-07-14 08:00:57.599999744	18.716000	363.922000	4021.82116	3331.833100	3276.582920	2348.740600	0.804816	0.602625
min	2017-08-01 00:00:00	1.000000	2.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	2018-08-01 00:00:00	9.000000	199.250000	1007.50750	0.000000	793.887500	0.000000	0.752635	0.000000
50%	2019-07-01 00:00:00	20.500000	370.000000	2785.54500	1344.375000	2365.770000	910.535000	0.929911	0.859052
75%	2020-05-08 18:00:00	24.250000	521.250000	5439.58750	4344.272500	4698.620000	3657.310000	0.983491	0.959105
max	2021-11-01 00:00:00	36.000000	723.000000	43221.43000	66225.130000	23004.580000	18864.800000	8.964878	4.600000
std	NaN	9.823459	201.244644	4531.76887	5721.292039	3373.371144	3263.282777	0.515176	0.489325

While most districts have efficiency rates close to or slightly below 1 (ideal case), some extreme values indicate over-reporting or data inconsistencies. The presence of values >1 might be due to errors or exceptional aid inflow.

B. Visualize Distributions

Histograms and box plots were created to examine the distribution and detect any skewness or anomalies in allocation, distribution, and efficiency metrics.

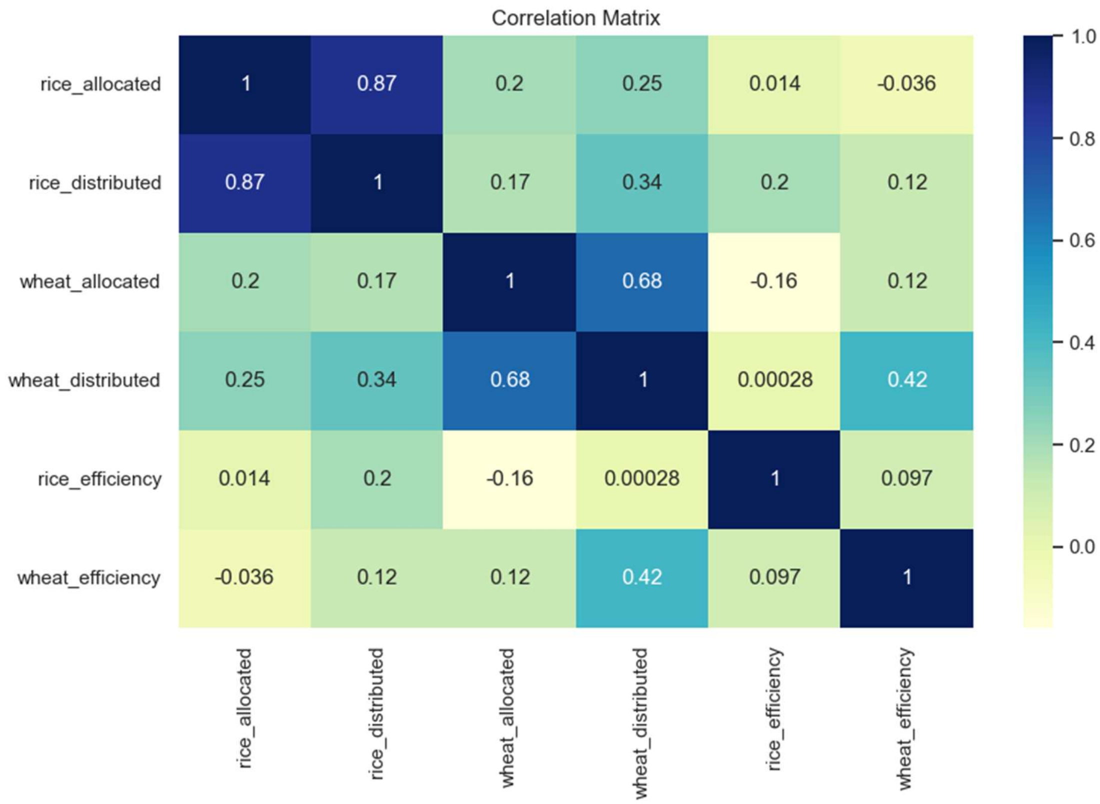




- Distribution of rice efficiency is slightly more right-skewed than wheat.
- Multiple outliers are present in both, suggesting exceptional or erroneous data.

C. Identify Correlations and Relationships

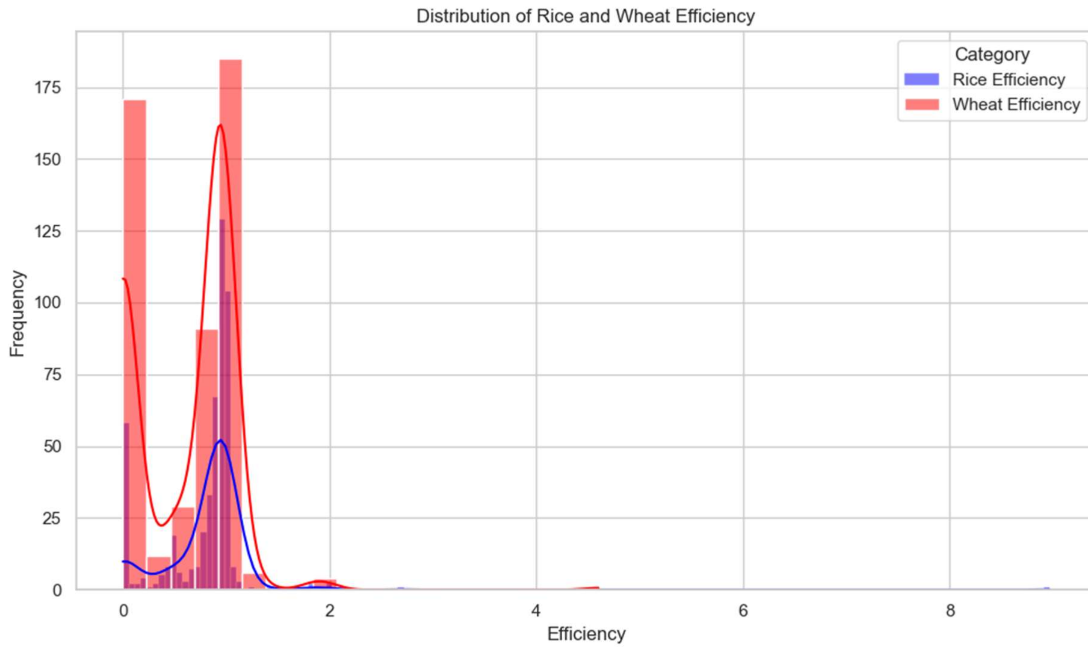
A correlation heatmap was used to identify relationships between allocation and distribution quantities. This helps reveal how rice and wheat allocations influence actual distributions.



- Strong **positive correlation** between allocated and distributed quantities for both rice and wheat.
- Moderate cross-correlation between wheat and rice variables, hinting at a possibly coordinated distribution strategy across commodities.

D. Statistical Properties: Skewness & Kurtosis

Normality was tested using distribution plots and skewness.



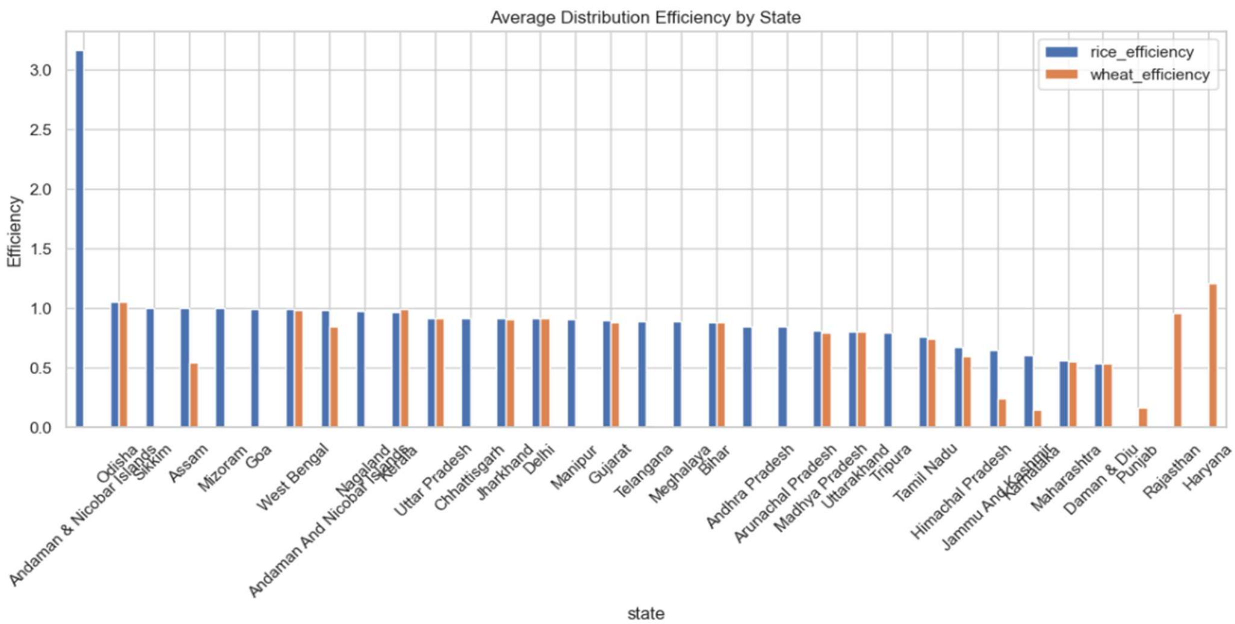
- **Rice Efficiency**:
  - **Very high skewness (7.71)** implies a **heavily right-skewed** distribution, where most values are low but some districts show extremely high efficiency (possibly erroneous).
  - **Extremely high kurtosis (124.73)** indicates a **very peaked distribution** with extreme outliers, likely due to inflated efficiency values above 1.
- **Wheat Efficiency**:
  - **Moderate skewness (0.89)** suggests a slight right skew, indicating more realistic variation in wheat efficiency.
  - **High kurtosis (7.38)** still suggests some peakedness and outliers, but **far less extreme** than rice.

Conclusion:

- Rice efficiency data is **highly non-normal** and requires **data cleaning or transformation** before using it for predictive modelling.
- Wheat efficiency is **closer to a normal distribution** but still not ideal.

E. Segment Data for Comparative Analysis

Data was segmented by state and month to analyze efficiency trends. This helps in comparative analysis across regions and over time.



This segmentation can uncover regional disparities in food distribution effectiveness and helps identify states needing better distribution mechanisms.

F. Documented Findings

- **Most states are close to ideal efficiency (~1)**, with some anomalies.
- **Outliers** need investigation—possible data entry issues or emergency distributions.
- **Strong linear relationship** between allocated and distributed quantities.
- Distribution behaviours exhibit **positive skew** and **moderate kurtosis**, justifying additional normalization before modelling.
- Segmentation reveals **performance variations across states**.

V. Statistical Analysis (Python)

A. Hypothesis Testing – Rice vs Wheat Efficiency

Objective:
Determine whether the average distribution efficiency of **rice** differs significantly from that of **wheat** across Indian districts under the Public Distribution System (PDS).

Method:
Two-sample independent **t-test** (assuming unequal variances).

```
from scipy.stats import ttest_ind

# Hypothesis:
# H0: Mean rice_efficiency == Mean wheat_efficiency
# H1: Mean rice_efficiency != Mean wheat_efficiency

t_stat, p_val = ttest_ind(df["rice_efficiency"], df["wheat_efficiency"], equal_var=False)
print(f"T-statistic: {t_stat:.4f}, P-value: {p_val:.4f}")
```

```
T-statistic: 6.3631, P-value: 0.0000
```

Interpretation:
Since the **p-value is well below 0.05**, we reject the null hypothesis.

Conclusion:
There is a **statistically significant difference** between the mean distribution efficiencies of rice and wheat.
This supports earlier observations that **rice efficiency is generally higher and more consistent**, while **wheat distribution is more variable**—potentially due to regional dietary preferences or logistical differences.

B. Regression Analysis – Allocated vs Distributed Quantity

1. Rice

We performed a simple linear regression to examine how well the **allocated rice quantity** predicts the **distributed quantity** under the PDS system.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:       rice_distributed   R-squared:                       0.761
Model:                            OLS   Adj. R-squared:                  0.761
Method:                 Least Squares   F-statistic:                     1588.
Date:                Fri, 18 Apr 2025   Prob (F-statistic):          4.97e-157
Time:                        23:38:49   Log-Likelihood:                -4412.7
No. Observations:                 500   AIC:                             8829.
Df Residuals:                     498   BIC:                             8838.
Df Model:                           1
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const            664.4364     98.689      6.733      0.000     470.539     858.334
rice_allocated     0.6495      0.016     39.854      0.000       0.617       0.682
==============================================================================
Omnibus:                      122.847   Durbin-Watson:                   1.980
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             2347.866
Skew:                          -0.513   Prob(JB):                         0.00
Kurtosis:                      13.566   Cond. No.                     8.10e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 8.1e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Regression Summary:

- **R-squared**: 0.761 — Indicates that 76.1% of the variance in rice distributed is explained by the rice allocated.
- **Regression Coefficient (slope)**: 0.6495 — On average, for every additional ton of rice allocated, ~0.65 tons are actually distributed.
- **Intercept**: 664.44 — Suggests a base level of distribution even when no rice is allocated, potentially due to pipeline deliveries or prior stock.
- **P-value** (for slope): 0.000 — Statistically significant relationship.
- **F-statistic**: 1588 — Strong overall model fit.
- **Durbin-Watson**: 1.98 — No major autocorrelation in residuals.
- **Condition Number**: 8.1e+03 — Slight concern of multicollinearity or numerical issues, though manageable.

Interpretation:
Rice distribution has a **strong linear relationship** with the allocation, with high efficiency and minimal unexplained variation. This reflects the **better-managed rice supply chain** under PDS.

2. Wheat

A similar regression was done for wheat allocation and distribution.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:      wheat_distributed   R-squared:                       0.465
Model:                            OLS   Adj. R-squared:                  0.463
Method:                 Least Squares   F-statistic:                     432.0
Date:                Fri, 18 Apr 2025   Prob (F-statistic):           1.50e-69
Time:                        23:39:15   Log-Likelihood:                -4598.1
No. Observations:                 500   AIC:                             9200.
Df Residuals:                     498   BIC:                             9209.
Df Model:                           1
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
const            1053.5189    123.738      8.514      0.000     810.407    1296.631
wheat_allocated     0.3887      0.019     20.785      0.000       0.352       0.425
==============================================================================
Omnibus:                      179.942   Durbin-Watson:                   1.932
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             5075.949
Skew:                          -0.932   Prob(JB):                         0.00
Kurtosis:                      18.498   Cond. No.                     7.66e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 7.66e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Regression Summary:

- **R-squared**: 0.465 — Only 46.5% of the variance is explained, indicating a weaker model.
- **Regression Coefficient (slope)**: 0.3887 — For each ton allocated, ~0.39 tons get distributed.
- **Intercept**: 1053.52 — High base level even at zero allocation suggests potential reporting or operational inefficiencies.
- **P-value**: 0.000 — Statistically significant.
- **F-statistic**: 432 — Model is statistically valid but weaker than rice.
- **Durbin-Watson**: 1.93 — No autocorrelation.

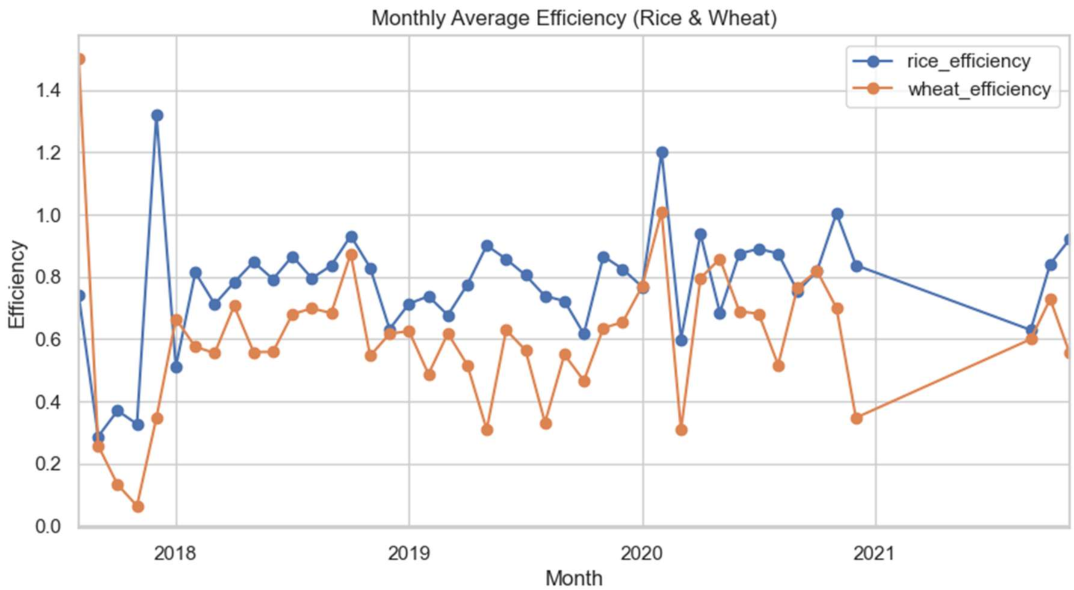
- **Condition Number**: 7.66e+03 — Same caution on potential numerical instability.

Interpretation:
Wheat distribution shows a **significant but weaker relationship** with allocation. The **lower R-squared** and **lower slope** hint at possible inefficiencies, leakages, or variable state-level preferences for wheat.

C.Monthly Efficiency Trend (2017–2021)

To complement the regression analysis, we calculated and plotted monthly average efficiency for both rice and wheat across all districts:

Efficiency = Distributed Quantity / Allocated Quantity



Observations:

- **Rice Efficiency** consistently stays between **0.6 and 1.0**, indicating a stable and well-managed supply chain. Occasional spikes over 1.0 likely stem from prior-month stock utilization or backlogged distributions.
- **Wheat Efficiency**, on the other hand, is **more erratic and consistently lower**, often ranging between **0.4 and 0.7**, with frequent drops below 0.5. This reaffirms the earlier regression finding of **lower predictability and weaker alignment with allocation**.
- Notable dips in both commodities (e.g., early 2020) may correspond to disruptions like **COVID-19 lockdowns**, highlighting systemic stress responses.

Insight:

Rice distribution shows **higher and more consistent efficiency** over time, while wheat reveals **operational inefficiencies or lower priority**, possibly due to regional dietary preferences or supply constraints.

D. Statistical Tests: ANOVA – Difference in rice efficiency across top states

Objective: To test whether there is a statistically significant difference in rice distribution efficiency among the top-performing states.

- **Null Hypothesis (H_0)**: The mean rice efficiency is the same across all selected states.
- **Alternative Hypothesis (H_1)**: At least one state has a significantly different mean rice efficiency.

```
import scipy.stats as stats

top_states = df["state"].value_counts().head(3).index.tolist()
subset = df[df["state"].isin(top_states)]
anova_result = stats.f_oneway(*[group["rice_efficiency"] for name, group in subset.groupby("state")])
print(f"F-statistic: {anova_result.statistic:.4f}, P-value: {anova_result.pvalue:.4f}")
```

```
F-statistic: 6.5804, P-value: 0.0019
```

Test Results

- **F-statistic**: 6.5804
- **p-value**: 0.0019

Conclusion

- Since the **p-value (0.0019) < 0.05**, we **reject the null hypothesis**.
- This means there **is a statistically significant difference** in rice efficiency across the selected top states.
- The variation in distribution efficiency is not just due to chance — some states are performing **better or worse than others**.

VI. Advanced Analytics (Python)

A. Predictive Modelling

To understand how well rice allocation predicts actual rice distribution, a **linear regression model** was developed using scikit-learn. The dataset was split into training and testing subsets.

```
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error

# Feature and target
X = df[["rice_allocated"]]
y = df["rice_distributed"]

# Split into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Initialize and train model
model = LinearRegression()
model.fit(X_train, y_train)

# Predict
y_pred = model.predict(X_test)

# Evaluate performance
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)

print(f"Mean Squared Error: {mse:.2f}")
print(f"R² Score: {r2:.2f}")
print(f"Mean Absolute Error: {mae:.2f}")
```

```
Mean Squared Error: 1809426.98
R² Score: 0.85
Mean Absolute Error: 963.44
```

Performance Metrics:

- **R^2 Score: 0.85**
  → The model explains **85% of the variance** in rice distribution, indicating a strong linear relationship between the allocated and distributed quantities.
- **Mean Squared Error (MSE): 1,809,426.98**
  → Represents the average of the squared differences between actual and predicted values. A higher value is expected due to the large unit scale (metric tonnes).
- **Mean Absolute Error (MAE): 963.44**
  → On average, the model's predictions deviate by approximately **963 metric tonnes** from actual values.

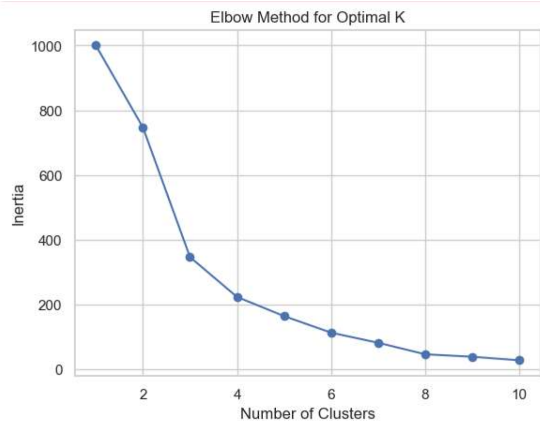
Key Insights:

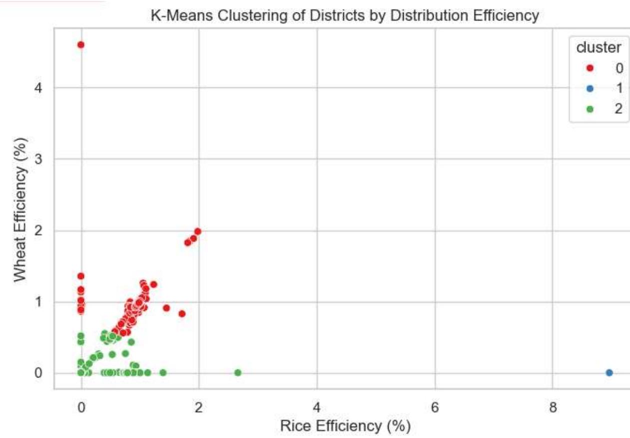
- The model demonstrates a **strong predictive ability** for rice distribution based on allocation, suitable for operational forecasting and supply planning.
- The absolute errors are reasonable given the scale of operations, though operational adjustments might be needed in highly volatile regions or months.

- Further refinement (e.g., adding regional or temporal variables, or transforming skewed data) could potentially enhance prediction accuracy.

B. Segmentation & Clustering

To identify patterns in district-level efficiency, **K-Means clustering** was applied on rice and wheat efficiency values.






```
Cluster Centers (original scale):
[[0.85806014 0.95588575]
 [8.96487805 0.        ]
 [0.68441646 0.0775126 ]]
```

Clustering Summary:

❖ **Elbow Method Plot**

- **Inertia** (within-cluster sum of squares) sharply decreases until around **k = 3**, after which the rate of decrease flattens.
- **Optimal number of clusters (k)** is **3** — a classic elbow shape, suggesting 3 distinct efficiency groupings among districts.

❖ **K-Means Cluster Plot**

- You plotted **Rice Efficiency (%)** vs. **Wheat Efficiency (%)** for each district.
- 3 clusters:
  - **Cluster 0 (Red)**: Moderate rice and wheat efficiency (~0.8–1.5% range)
  - **Cluster 1 (Blue)**: A clear outlier district with very high rice efficiency (~9%) but low wheat efficiency

- **Cluster 2 (Green)**: Lower performing districts with low efficiencies in both rice and wheat

❖ **Cluster Centers (Original Scale)**:

Cluster 0: Rice ~ 0.86%, Wheat ~ 0.96%

Cluster 1: Rice ~ 8.96%, Wheat ~ 0%

Cluster 2: Rice ~ 0.68%, Wheat ~ 0.07%

Key Insights:

- The **majority of districts fall into Cluster 0 and Cluster 2**, indicating a division between moderately efficient and poorly efficient districts.
- **One district (Cluster 1)** is an extreme outlier for rice efficiency. Recommend investigating this district — could signal either an exemplary operation or a data anomaly.
- The **overall distribution of efficiencies is skewed toward the lower end** for both rice and wheat, suggesting opportunities for PDS policy review and operational improvements.

VII. Decision Analysis

Based on the distribution efficiency analysis of rice and wheat under India's Public Distribution System, the following actionable recommendations are proposed:

- **Prioritize operational interventions** in low-efficiency districts to improve logistics, inventory management, and last-mile delivery.
- **Investigate outlier districts** with exceptionally high rice efficiency for potential best practices or reporting anomalies.
- **Implement cluster-based distribution policies**, providing intensive oversight to low-efficiency districts and maintaining existing practices in high-efficiency ones.
- **Adopt predictive allocation planning** using regression models to forecast distribution volumes and adjust allocations proactively.
- **Integrate socio-economic and seasonal data** into future analyses for more accurate demand forecasting and efficiency monitoring.
- **Strengthen governance and data integrity protocols**, adopting real-time stock tracking and periodic operational audits.

These strategies aim to enhance food grain distribution effectiveness, reduce systemic inefficiencies, and support evidence-based decision-making in PDS operations.

VIII. Research Work

A. *Evaluation of the Role of PDS in Nutritional Security in India*

- **Authors:** N. George and A. Pandey
- **Published In:** Journal of Social Policy (2020)
- **Summary:** This study evaluates the Public Distribution System's (PDS) role in promoting food and nutritional security in India. It highlights operational inefficiencies that limit the system's potential benefits and suggests that reforms targeting productivity, transparency, and sustainability are necessary for improving the system's effectiveness in reducing hunger and undernourishment.

B. *Utilization Behavior in Rajasthan's Public Distribution System*

- **Authors:** R. Khera
- **Published In:** Economic & Political Weekly (2019)
- **Summary:** This paper investigates household-level utilization patterns within Rajasthan's PDS framework, emphasizing that underutilization is driven more by supply-side constraints than by lack of demand. It reveals that PDS availability

influences cereal consumption behavior and recommends policy adjustments to address procurement and distribution inefficiencies.

C. *Public Distribution System as a Tool for Food Security and Poverty Reduction*

- **Authors:** R. Ghabru and S. Singh
- **Published In:** Indian Journal of Public Administration (2020)
- **Summary:** The study examines PDS's contribution to food security and poverty alleviation across India. While acknowledging the program's critical role in providing subsidized food grains to vulnerable populations, it identifies persistent challenges such as leakages, inclusion-exclusion errors, and issues in beneficiary identification. The paper calls for better operational transparency and streamlined targeting mechanisms.

D. *Impact of the Public Distribution System on Poverty Levels in India*

- **Authors:** J. Thomas and A. Jacob
- **Published In:** Journal of Development Studies (2021)
- **Summary:** Using National Sample Survey (NSS) consumption data, this study assesses PDS's impact on poverty levels in India. It finds that PDS benefits disproportionately accrue to the poorest deciles, reducing food insecurity and consumption shortfalls. Additionally, it observes secondary benefits for middle-income households in select states and emphasizes extending PDS coverage and efficiency.

E. *Impact of the National Food Security Act (NFSA) on Rice Consumption Patterns*

- **Authors:** A. Publishers
- **Published In:** Food Policy Reports (2022)
- **Summary:** This paper evaluates the impact of NFSA implementation on rice consumption patterns across Indian states. It notes that early adopter states experienced significant increases in rice intake following NFSA's rollout. The study advocates scaling up successful state-level models nationally to improve food security outcomes and recommends continuous assessment of NFSA's long-term effects.

IX. Conclusion

In this project, a comprehensive analysis of the Public Distribution System (PDS) in India was undertaken with a focus on the allocation and distribution of essential food grains — wheat and rice — at the district level. The objective was to examine distribution patterns, measure efficiencies, and uncover insights that could inform policy and operational improvements.

The analysis commenced with **data preprocessing and exploratory data analysis (EDA)** to understand the distribution patterns and detect inconsistencies or irregularities in the dataset. Descriptive statistics and visualizations provided an initial overview of allocation versus distribution trends across districts.

Subsequently, a **linear regression model** was developed to evaluate the relationship between the quantities allocated and actually distributed for rice. The model achieved an impressive **R^2 score of 0.85**, indicating a strong predictive relationship, while performance metrics like Mean Squared Error (MSE) and Mean Absolute Error (MAE) helped gauge model accuracy.

To further categorize districts based on their distribution efficiency, **K-Means clustering** was applied after standardizing efficiency metrics. The **Elbow Method** determined the optimal number of clusters as **three**, revealing distinct groupings of districts:

- Moderately efficient districts
- Less efficient districts
- An outlier district with exceptionally high rice efficiency

This clustering highlighted operational disparities and pointed toward areas requiring policy attention or deeper investigation.

Overall, the project successfully combined statistical modelling, machine learning techniques, and data visualization to offer a meaningful assessment of PDS distribution efficiency at the district level. The insights derived can support decision-makers in identifying operational gaps, improving resource allocation, and enhancing overall system efficiency.

Future studies could build upon this foundation by incorporating additional variables like population demographics, infrastructure indicators, or time-series trends to achieve a more holistic view of PDS performance nationwide.