# Data Analytics Report: Agricultural Query Analysis System
# Big Data Analytics Project

## 1. Dataset Description:

### 1.1 Domain Overview
The dataset represents a comprehensive agricultural advisory system from the Telangana region, capturing farmer queries and expert responses across multiple districts. This system serves as a crucial bridge between agricultural experts and farmers, providing timely solutions to farming challenges.

### 1.2 Dataset Specifications
Total Records: 500 agricultural queries
Time Period: January 2025 (Single month snapshot)
Geographical Coverage: 30+ districts across Telangana
Data Collection: Farmer queries through Kisan Call Center with expert responses

### 1.3 Data Structure
The dataset contains 13 columns capturing multiple dimensions:

Geographical Context:
- StateName, DistrictName, BlockName (Administrative hierarchy)

Temporal Context:
- CreatedOn (Timestamp), year, month

Agricultural Context:
- Season, Sector (AGRICULTURE/HORTICULTURE), Category (Cereals/Pulses/Vegetables/etc.)
- Crop (Specific crop names), QueryType (Technical categories)

Content Context:
- QueryText (Farmer's question), KccAns (Expert's response)

### 1.4 Data Quality Assessment
Completeness: High data integrity with minimal missing values
Consistency: Standardized categorical values across records
Accuracy: Expert-verified responses with technical recommendations
Relevance: All records pertain to agricultural domain challenges

# 2. Observed Insights & Hidden Facts

## 2.1 Visualization Explanations

**Visualization 1:** Top Crops by Query Volume
What it shows: A bar chart displaying which crops farmers ask about most frequently
Simple Explanation:

Paddy (rice) is the most discussed crop with 125 queries

Chillies come second with 85 queries

Maize, Black Gram, and other crops follow
Why it matters: This tells us which crops are most important to farmers and might need more support

**Visualization 2:** Query Distribution by Category
What it shows: A pie chart showing different types of farming categories
Simple Explanation:

Cereals like rice and wheat get the most questions (40%)

Vegetables are second with 30%

Pulses, Fruits, and other categories make up the rest
Why it matters: Shows which farming sectors need more attention and resources

**Visualization 3:** Top Districts by Query Volume
What it shows: A horizontal bar chart ranking districts by number of farmer queries
Simple Explanation:

Jogulamba Gadwal district has the most queries (48)

Jagital and Mahabubabad follow with high numbers

Some districts have very few queries
Why it matters: Helps identify which areas have more active farmers or more farming problems

**Visualization 4:** Query Types Distribution
What it shows: A bar chart showing what kinds of problems farmers ask about
Simple Explanation:

Plant Protection (pest control) is the biggest concern (36%)

Nutrient Management (fertilizers) is second (19%)

Weed Management, Sowing Time, and other issues follow
Why it matters: Reveals the most common challenges farmers face

**Visualization 5:** Monthly Query Trend
What it shows: A line graph showing how queries change through the month
Simple Explanation:

Queries are spread throughout January

Some days have more questions than others

The trend is relatively stable with small variations
Why it matters: Helps plan resources throughout the month

**Visualization 6:** Sector-wise Distribution
What it shows: A bar chart comparing Agriculture vs Horticulture queries
Simple Explanation:

Traditional Agriculture gets 76% of all queries

Horticulture (fruits, vegetables) gets 24%
Why it matters: Shows the current focus of farming activities in the region

**Visualization 7:** Heatmap - Crop vs Query Type
What it shows: A color-coded grid showing relationships between crops and problems
Simple Explanation:

Darker colors mean more questions about that combination

Paddy has dark colors for Plant Protection and Nutrient Management

Chillies show strong patterns for pest-related issues
Why it matters: Helps understand which problems affect which crops most

**Visualization 8:** Crop Diversity in Districts
What it shows: A horizontal bar chart showing how many different crops each district grows
Simple Explanation:

Some districts grow many different crops (high diversity)

Others focus on fewer crops

This shows farming patterns across regions
Why it matters: Helps in planning crop-specific support programs

**Visualization 9:** Monthly Trend of Top Crops
What it shows: Line graphs showing how queries for top crops change during the month
Simple Explanation:

Different crops have different query patterns through the month

Some crops get more questions at specific times

Shows seasonal patterns for different crops
Why it matters: Helps anticipate when farmers will need help with specific crops

**Visualization 10:** Block-level Analysis
What it shows: A bar chart showing query distribution within the top district
Simple Explanation:

Even within one district, different blocks have different query numbers

Some blocks are more active than others
Why it matters: Helps target resources at the local level

# 3. Recommendations

## 3.1 Strategic Agricultural Extension Recommendations

### 3.1.1 Targeted Pest Management Programs
Immediate Action: Develop district-specific integrated pest management (IPM) packages for top 3 crops (Paddy, Chillies, Maize) based on query patterns.

Implementation:
- Create mobile-based alert systems for pest outbreaks
- Distribute pest identification charts in local language
- Train local agricultural officers on latest pest control techniques

### 3.1.2 Enhanced Nutrient Management Support
Medium-term Initiative: Establish soil testing mobile vans for frequent district rotations.

Expected Impact:
- Reduce nutrient-related queries by 40%
- Improve crop yields through precise fertilizer recommendations
- Cost savings for farmers through optimized input usage

## 3.2 Technology Integration Recommendations

### 3.2.1 Digital Knowledge Repository
Development: Create a centralized digital platform with:
- Crop-specific problem-solution database
- Regional language content
- Mobile app with offline capability

Features:
- AI-powered query routing
- Expert response templates
- Multimedia content (videos, images)

### 3.2.2 Predictive Analytics System
Advanced Implementation:
- Develop early warning systems for common problems
- Use historical data to predict seasonal challenges
- SMS-based alert system for preventive measures

## 3.3 Capacity Building Recommendations

### 3.3.1 Farmer Training Programs
Structured Approach:
- Season-based training calendars
- Crop-specific workshops
- Demonstration plots in high-query districts

### 3.3.2 Extension Officer Enhancement
Training Modules:
- Advanced diagnostic skills
- Digital tool proficiency
- Communication effectiveness

## 3.4 Monitoring and Evaluation Framework

### 3.4.1 Performance Metrics
Key Indicators:
- Query resolution time
- Farmer satisfaction scores
- Problem recurrence rates
- Adoption rate of recommendations

### 3.4.2 Continuous Improvement
Feedback Mechanism:
- Regular farmer feedback collection
- Expert performance reviews
- System efficiency audits

# 4. Conclusion

The comprehensive analysis of the agricultural query dataset reveals a well-utilized system providing crucial support to Telangana's farming community. The insights generated highlight specific pain points in crop management, particularly in plant protection and nutrient management for major crops like paddy and chillies.

The recommendations provided are actionable, scalable, and designed to address the identified challenges while leveraging the existing infrastructure. Implementation of these suggestions will significantly enhance the effectiveness of agricultural extension services, leading to improved crop productivity, reduced input costs, and increased farmer incomes.

**Report Prepared By**:I.INDRAKIRAN
**Project: Data Analytics Final Year Project**
**Institution**: MALLAREDDY UNIVERSITY

**Dataset Source**: Government Agricultural Query System, Telangana
**Analysis Period**: January 2025
**Tools Used**: PySpark, Pandas, Matplotlib, Seaborn
**Methodology**: Descriptive Analytics, Pattern Recognition, Multi-dimensional
Analysis