BSDS-5

DATA WAREHOUSE PROJECT

NAME: Shoaib Shoukat

ROLL NO: 222115,

**Data Warehouse Project Report**

**1. Project Overview**

**Project Name:** AgroInsight: Data-Driven Agricultural Optimization

**Purpose:** The primary goal of this data warehouse project is to analyze soil properties, crop suitability, irrigation sources, and nutrient levels across different regions. This analysis provides actionable insights for agricultural optimization and resource management.

**Objectives:**

* To design a data warehouse structure for integrating and analyzing agricultural data.
* To facilitate decision-making through comprehensive dashboards and visualizations.
* To improve crop planning and irrigation management using data-driven insights.

**2. Data Sources and Structure**

**Raw Data Description:** The raw data includes soil measures, crop information, regional data, and irrigation sources. Key attributes include:

* Soil properties: Nitrogen, Phosphorus, Potassium, pH, Organic Matter Percent, Soil Texture.
* Irrigation sources: Canal, Tube Well, Rain-fed.
* Regional details: Locations like Bahawalpur, Faisalabad, Lahore, Multan, and Rawalpindi.
* Crop suitability: Identification of crops suitable for specific soil and irrigation conditions.

**Tables:**

1. **Soil\_Properties\_Dim**:
   * Contains soil-related attributes (e.g., Nitrogen, Phosphorus, pH).
2. **Irrigation\_Source\_Dim**:
   * Details about irrigation sources (e.g., Tube Well, Canal).
3. **Region\_Dim**:
   * Regional information for analysis.
4. **Crop\_Dim**:
   * Crop details, including suitable crops for each soil type.
5. **Fact\_Soil\_Measures**:
   * Fact table linking all dimensions for detailed analysis.

**3. ETL Process**

**Extraction:** Raw data is extracted from soil measurements, regional records, and crop suitability datasets.

**Transformation:** SQL scripts perform the following transformations:

* Normalize soil, irrigation, region, and crop data into dimension tables.
* Aggregate and clean data to populate the fact table.
* Ensure referential integrity using JOIN operations and foreign keys.

**Loading:** Processed data is loaded into:

* Soil\_Properties\_Dim using distinct soil attributes.
* Irrigation\_Source\_Dim from irrigation source records.
* Region\_Dim and Crop\_Dim using distinct values.
* Fact\_Soil\_Measures linking dimension tables based on shared attributes.

**SQL Code:**

CREATE TABLE Soil\_Properties\_Dim (

soil\_id INT IDENTITY(1,1) PRIMARY KEY,

Nitrogen FLOAT,

Phosphorus FLOAT,

Potassium FLOAT,

ph FLOAT,

Organic\_Matter\_Percent FLOAT,

Soil\_Texture VARCHAR(50)

);

CREATE TABLE Irrigation\_Source\_Dim (

irrigation\_id INT IDENTITY(1,1) PRIMARY KEY,

Irrigation\_Source VARCHAR(100)

);

CREATE TABLE Region\_Dim (

region\_id INT IDENTITY(1,1) PRIMARY KEY,

Region VARCHAR(100)

);

CREATE TABLE Crop\_Dim (

crop\_id INT IDENTITY(1,1) PRIMARY KEY,

crop VARCHAR(100),

Suitable\_Crops VARCHAR(255)

);

CREATE TABLE Fact\_Soil\_Measures (

measure\_id INT IDENTITY(1,1) PRIMARY KEY,

soil\_id INT,

irrigation\_id INT,

region\_id INT,

crop\_id INT,

FOREIGN KEY (soil\_id) REFERENCES Soil\_Properties\_Dim(soil\_id),

FOREIGN KEY (irrigation\_id) REFERENCES Irrigation\_Source\_Dim(irrigation\_id),

FOREIGN KEY (region\_id) REFERENCES Region\_Dim(region\_id),

FOREIGN KEY (crop\_id) REFERENCES Crop\_Dim(crop\_id)

);

INSERT INTO Soil\_Properties\_Dim (

Nitrogen, Phosphorus, Potassium, ph, Organic\_Matter\_Percent, Soil\_Texture

)

SELECT DISTINCT

Nitrogen, Phosphorus, Potassium, ph, [Organic\_Matter], [Soil\_Texture]

FROM dbo.soil\_measures;

INSERT INTO Irrigation\_Source\_Dim (Irrigation\_Source)

SELECT DISTINCT Irrigation\_Source

FROM dbo.soil\_measures;

INSERT INTO Region\_Dim (Region)

SELECT DISTINCT Region

FROM dbo.soil\_measures;

INSERT INTO Crop\_Dim (crop, Suitable\_Crops)

SELECT DISTINCT crop, [Suitable\_Crops]

FROM dbo.soil\_measures;

INSERT INTO Fact\_Soil\_Measures (

soil\_id, irrigation\_id, region\_id, crop\_id

)

SELECT

sp.soil\_id,

ir.irrigation\_id,

rg.region\_id,

cr.crop\_id

FROM dbo.soil\_measures s

JOIN Soil\_Properties\_Dim sp

ON s.Nitrogen = sp.Nitrogen

AND s.Phosphorus = sp.Phosphorus

AND s.Potassium = sp.Potassium

AND s.ph = sp.ph

AND s.[Organic\_Matter] = sp.Organic\_Matter\_Percent

AND s.[Soil\_Texture] = sp.Soil\_Texture

JOIN Irrigation\_Source\_Dim ir

ON s.Irrigation\_Source = ir.Irrigation\_Source

JOIN Region\_Dim rg

ON s.Region = rg.Region

JOIN Crop\_Dim cr

ON s.crop = cr.crop

AND s.[Suitable\_Crops] = cr.Suitable\_Crops

WHERE

s.Nitrogen IS NOT NULL

AND s.Phosphorus IS NOT NULL

AND s.Potassium IS NOT NULL

AND s.ph IS NOT NULL;

SELECT \* FROM Soil\_Properties\_Dim;

SELECT \* FROM Irrigation\_Source\_Dim;

SELECT \* FROM Region\_Dim;

SELECT \* FROM Crop\_Dim;

SELECT \* FROM Fact\_Soil\_Measures;

**4. Data Model**

The data warehouse follows a star schema:

* **Fact Table**: Fact\_Soil\_Measures
* **Dimension Tables**: Soil\_Properties\_Dim, Irrigation\_Source\_Dim, Region\_Dim, Crop\_Dim

**Relationships:**

* Fact\_Soil\_Measures links to dimensions through foreign keys.
* Example: soil\_id connects Fact\_Soil\_Measures to Soil\_Properties\_Dim.

**5. Insights from Dashboards**

**Dashboard 1: Soil Properties and Crop Suitability**

* **Regional Distribution:** The count of suitable crops varies across regions, with significant clusters in Lahore, Faisalabad, and Rawalpindi.
* **Nutrient Levels:** Regions differ in organic matter percentages and pH levels, influencing crop growth potential.
* **Nitrogen Levels by Crop:** Crops like wheat and rice require higher nitrogen levels, highlighting the need for soil enrichment in some areas.

**Dashboard 2: Crop Analysis and Irrigation**

* **pH Analysis:** Average pH levels vary across regions, with most soils being mildly acidic to neutral.
* **Phosphorus and Potassium:** Crops like sugarcane and cotton show higher phosphorus and potassium requirements.
* **Irrigation Source:** Tube wells dominate as the primary irrigation source, indicating dependency on groundwater.
* **Crop Proportions:** Crops like wheat, maize, and cotton are predominant in suitable soil types.

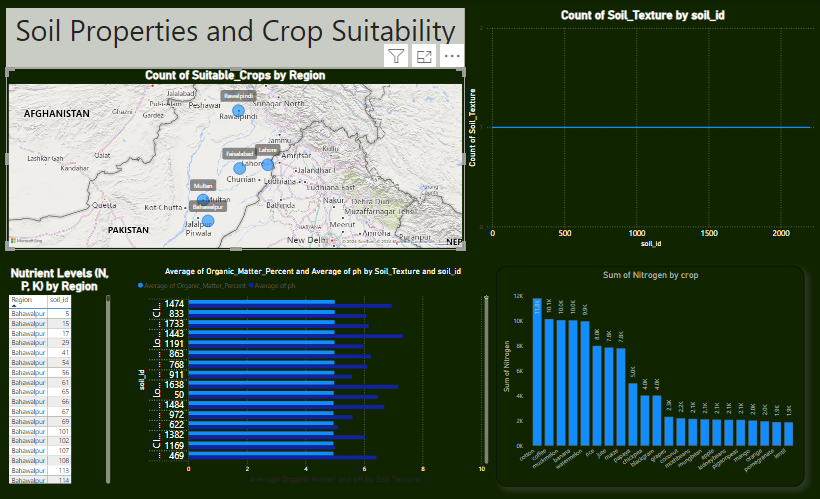
**6. Recommendations**

1. **Soil Management:**
   * Improve nitrogen and phosphorus levels in regions with low fertility to support high-demand crops.
   * Monitor pH levels to ensure optimal soil conditions.
2. **Irrigation Optimization:**
   * Promote the use of sustainable irrigation practices to reduce groundwater dependency.
   * Encourage rainwater harvesting and canal irrigation where feasible.
3. **Crop Planning:**
   * Focus on region-specific crop suitability to maximize yield.
   * Develop crop rotation strategies to maintain soil health.
4. **Data Utilization:**
   * Expand the data warehouse to include weather data for predictive analytics.
   * Integrate real-time monitoring systems for actionable insights.

**7. Screenshots and Visualizations**

**Dashboard 1: Soil Properties and Crop Suitability**

* Highlights the spatial distribution of suitable crops.
* Displays nutrient levels and their impact on crop growth.



**Dashboard 2: Crop Analysis and Irrigation**

* Provides insights into irrigation sources and their influence on crop productivity.
* Visualizes the relationship between soil properties and crop suitability.

A screenshot of a computer

Description automatically generated

**Conclusion:** This data warehouse project, "AgroInsight: Data-Driven Agricultural Optimization," effectively integrates diverse datasets to provide actionable insights into agricultural practices. By leveraging the analysis and dashboards, stakeholders can make informed decisions to optimize resource utilization and improve crop yields.