**AgriData Explorer: Understanding Indian Agriculture with EDA**

**Slide 1: Project Overview**

* **Project Title:** AgriData Explorer: Understanding Indian agriculture with EDA
* **Skills Taken Away:** Python scripting, data cleaning, Exploratory Data Analysis (EDA), SQL, Power BI.
* **Domain:** Agricultural, mainly focusing on crop production statistics, yields, and farming area in India.

**Slide 2: Problem Statement**

* India's agricultural sector faces challenges due to complex, fragmented, and inaccessible data.
* Current difficulties for farmers, policymakers, and researchers in accessing, analyzing, and making informed decisions.
* **Project Goal:** Create a data visualization platform to integrate agricultural data from different states and districts.
* **Expected Outcome:** Provide insights into crop production, yields, and cultivation areas, identifying trends, gaps, and regional disparities.

**Slide 3: Business Use Cases**

* **Farmers:**
  + Explore historical crop production and yield data for informed cultivation decisions.
  + Analyze regional productivity to identify areas for improvement (e.g., soil health, irrigation).
* **Policymakers:**
  + Assess regions with low productivity to allocate resources or subsidies effectively.
  + Identify crops with fluctuating yields for better crop insurance and risk management policies.
* **Researchers:**
  + Analyze the impact of climate, soil, and irrigation on crop yields over time.
  + Identify potential areas for agricultural innovation (e.g., high-yielding varieties, pest management).

**Slide 4: Project Stages Overview**

The project follows a structured approach, encompassing:

1. **Database Setup (MySQL)**
2. **Data Collection and Cleaning (Python Script)**
3. **Database Population (Python Script)**
4. **Exploratory Data Analysis (EDA) & Analytical Queries (Python & SQL)**
5. **Power BI Integration and Visualization**

**Slide 5: Stage 1: Database Setup (MySQL)**

* **Objective:** Set up a relational database to store agricultural data.
* **Steps:**
  + **1.1 Database Creation:** Create the Project2\_Agri\_India database.
  + **1.2 Table Definition:**
    - state\_master: Stores state codes and names.
    - district\_master: Stores district details, linked to states.
    - crops: Stores unique crop names.
    - years: Stores unique years.
    - agri\_production: Main fact table with detailed crop area, production, and yield data for various crops (Rice, Wheat, Oilseeds, Sugarcane, Cotton, Fruits, Vegetables, Potatoes, Onion, Fodder, etc.) by district and year.
  + **1.3 Sample Data Insertion:** Insert initial sample data into all tables for testing and setup.

**Slide 6: Stage 2: Data Collection and Cleaning (Python Script)**

* **Objective:** Load raw CSV data and prepare it for database insertion.
* **Steps:**
  + **2.1 Import Libraries:** Use pandas, mysql.connector, matplotlib, seaborn, plotly.express.
  + **2.2 Define MySQL Connection Function:** Establish database connection details.
  + **2.3 Load CSV Data:** Read agricultural data from ICRISAT-District Level Data - ICRISAT-District Level Data.csv into a Pandas DataFrame.
  + **2.4 Rename Columns:** Standardize CSV column names (e.g., 'RICE AREA (1000 ha)' to rice\_area) to match MySQL schema.
  + **2.5 Validate and Reorder Columns:** Ensure DataFrame columns match database schema for correct insertion.

**Slide 7: Stage 3: Database Population (Python Script)**

* **Objective:** Insert cleaned data into the MySQL database.
* **Steps:**
  + **3.1 Connect to MySQL:** Establish an active connection.
  + **3.2 Insert Master Data:** Extract and insert unique states, districts, crops, and years into their respective master tables.
  + **3.3 Insert Agricultural Production Data:** Iterate through the DataFrame, inserting each row's values into the agri\_production table, handling missing values (fillna(0)) and preventing duplicates (INSERT IGNORE).
  + **3.4 Close Database Connection:** Properly close the connection.

**Slide 8: Stage 4: Exploratory Data Analysis (EDA) & Analytical Queries**

* **Objective:** Derive insights from the data using Python for visualizations and SQL for specific questions.
* **Key Python Visualizations (EDA):**
  + Top 7 Rice Production States (Bar Plot)
  + Top 5 Wheat Producing States (Bar & Pie Chart)
  + Oil Seed Production by Top 5 States (Bar Plot)
  + Top 7 Sunflower Production States (Horizontal Bar Plot)
  + India's Sugarcane Production From Last 50 Years (Line Plot)
  + Rice Production Vs Wheat Production (Last 50y) (Line Plot)
  + Rice Production By West Bengal Districts (Bar Plot)
  + Top 10 Wheat Production Years From UP (Bar Plot)
  + *Additional plots for Finger Millet, Sorghum, Groundnut, Soybean, Area vs Production, and Rice Yield by Districts.*
* **Key SQL Analytical Questions:**
  + Year-wise Trend of Rice Production Across States (Top 3)
  + Top 5 Districts by Wheat Yield Increase Over the Last 5 Years
  + States with the Highest Growth in Oilseed Production (5-Year Growth Rate)
  + District-wise Correlation Between Area and Production for Major Crops (Rice, Wheat, and Maize)
  + Yearly Production Growth of Cotton in Top 5 Cotton Producing States
  + Districts with the Highest Groundnut Production in 2020
  + Annual Average Maize Yield Across All States
  + Total Area Cultivated for Oilseeds in Each State
  + Districts with the Highest Rice Yield
  + Compare the Production of Wheat and Rice for the Top 5 States Over 10 Years

**Slide 9: Stage 5: Power BI Integration and Visualization**

* **Objective:** Connect processed data to Power BI for interactive dashboard creation.
* **Steps:**
  + **5.1 Connect to MySQL:** Use "Get Data" in Power BI Desktop to connect to Project2\_Agri\_India (requires MySQL Connector/Net).
  + **5.2 Load Data:** Select and load all relevant tables (agri\_production, crops, district\_master, state\_master, years).
  + **5.3 Create Data Model:** Establish relationships between tables in Power BI's Model View.
  + **5.4 Design Dashboard and Visualizations:** Create interactive reports using various Power BI visuals (bar charts, pie charts, line charts, scatter plots, KPIs).
  + **5.5 Implement Interactive Filters and Slicers:** Add slicers for Crop Type, Region, and Year for dynamic data exploration.
  + **5.6 Enhance Visualizations:** Format visuals, configure interactions for a cohesive user experience.

**Slide 10: Project Deliverables & Results**

* **Project Deliverables:**
  + Python scripts (data extraction, cleaning, visualization).
  + SQL queries (database creation, population, analysis).
  + Structured and normalized MySQL database.
  + Interactive Power BI reports and dashboards.
  + Detailed documentation.
* **Results:**
  + Fully interactive dashboard displaying key agricultural metrics.
  + Identified patterns in crop production, high/low yield areas, and seasonality effects.
  + Data-driven recommendations for farmers on crop selection and productivity improvement.