**Candidate Name:Tanisha Palmia**

**EmailID:** [**tanishapalmia005@gmail.com**](mailto:tanishapalmia005@gmail.com)

**Project Name:World Food Production**

**1.Introduction**

The global food system is complex, subject to factors ranging from climate change to geopolitical shifts. Understanding historical production trends, comparative output among nations, and the mix of commodities produced is critical for policymakers, agricultural economists, and supply chain analysts.

This project utilizes the **World Food Production Dataset** (world food production smart.csv), which contains time-series data on the production volume (in tonnes) of 22 key agricultural commodities, including grains, fruits, vegetables, and meat, across numerous global entities from 1961 to 2016. The data encompasses crucial food security staples (like Wheat, Rice, and Maize) and high-value commodities (like Coffee, Cocoa, and Avocados).

The goal of this initiative is to transform this raw, multi-dimensional data into an interactive and accessible **Power BI dashboard**, enabling users to quickly identify key production trends, top-performing countries, and market volatility over time.

**1.1 Project Overview**

The project will be executed in three primary phases:

1. **Data Transformation (Power Query):** The initial "wide" dataset structure (where each crop is a column) will be transformed into a "tall" structure via **unpivoting**. This is essential for efficient filtering and dynamic visualization of all crops simultaneously. A proper **Date column** will also be created to enable time-intelligence calculations.
2. **Data Modeling and Calculation (DAX):** Key analytical measures will be developed using DAX. These include **Total Production**, **Year-over-Year (YoY) Growth**, **Top Producing Country**, and **Crop Market Share**.
3. **Visualization and Dashboard Design (Power BI):** An interactive dashboard will be designed, incorporating essential elements like a **World Map** (for geographical context), **Line Charts** (for trend analysis), and **Bar/Column Charts** (for ranking and comparison). **Slicers** (Country, Crop, Year) and **Drill Through** actions will be implemented to ensure a seamless and personalized user experience.

**1.2 Objectives**

The primary objectives of this project are to develop a dashboard that can achieve the following:

| Category | Specific Objective | Key Metrics/Visuals Used |
| --- | --- | --- |
| **I. Comparative Analysis** | **Identify and rank** the world's leading food producers globally and for specific crops during selected time periods. | Horizontal Bar Charts, Top Producing Country KPI. |
| **II. Temporal Analysis** | **Quantify and visualize** the historical growth, decline, and volatility of production for any selected commodity or country. | Line Charts, Waterfall Charts, YoY Growth %, CAGR. |
| **III. Market Composition** | **Determine the relative importance** of each crop within a country's total output, or the global share of a crop. | Treemaps, Clustered Column Charts, Crop Global Share %. |
| **IV. Interactivity & Usability** | Provide users with a highly intuitive, **single-page interface** to dynamically filter, cross-filter, and drill down into the data with minimal effort. | Slicers, Drill Through, Custom Tooltips, World Map. |

**2. Project Initialization and Planning Phase**

**2.1Problem Statement: Global Food Production Data Complexity**

The existing World Food Production Dataset is a rich, multi-dimensional collection of global agricultural output statistics, spanning numerous countries and decades. However, the data's current structure presents significant challenges for timely, comparative, and strategic analysis.

The core problem is the lack of an integrated, interactive system to efficiently translate this large volume of raw, disaggregated data into actionable intelligence regarding global production trends and market dynamics.

Specifically, the following challenges prevent quick decision-making:

1. Data Structure Inefficiency: The data is structured in a wide format (one column per crop), making direct comparisons, dynamic filtering, and time-series analysis across all commodities complex and requiring significant data manipulation for every query.
2. Lack of Contextual Metrics: The dataset currently only provides raw production volume (tonnes). There are no readily available metrics for Year-over-Year momentum, long-term growth rates (CAGR), or market share, which are essential for assessing performance and volatility.
3. Static Visualization: Without an interactive dashboard, identifying geographical dominance (top producers), spotting historical production peaks, or understanding the overall global production mix requires manual querying and visualization, leading to slow and often incomplete insights.
4. This project proposal outlines a solution to address a specific problem. With a clear objective, defined scope, and a concise problem statement, the proposed solution details the approach, key features, and resource requirements, including hardware, software, and personnel.

**2.2Project Proposal (Proposed Solution) template**

|  |  |
| --- | --- |
| **Project Overview** | |
| Objective | Clearly state the project's primary objective |
| Scope | Define the boundaries and extent of the project |
| **Problem Statement** | |
| Description | Clearly articulate the problem to be addressed |
| Impact | Explain the implications of solving the problem |
| **Proposed Solution** | |
| Approach | Outline the methodology and techniques to be used |
| Key Features | Highlight the unique aspects of the proposed solution |

1. **Resource Requirement**

|  |  |  |
| --- | --- | --- |
| **Resource Type** | **Description** | **Specification/Allocation** |
| **Hardware** | | |
| Computing Resources | CPU/GPU specifications, number of cores | e.g., 2 x NVIDIA V100 GPUs |
| Memory | RAM specifications | e.g., 8 GB |
| Storage | Disk space for data, models, and logs | e.g., 1 TB SSD |
| **Software** | | |
| Frameworks | Python frameworks | e.g., Flask |
| Libraries | Additional libraries | e.g., scikit-learn, pandas, numpy |
| Development Environment | IDE, version control | e.g., Jupyter Notebook, Git |
| **Data** | | |
| Data | Source, size, format | e.g., Kaggle dataset, 10,000 images |

**Proposed Solution**

1. Data Transformation and Preparation (Power Query & DAX) 🛠️

The solution will first focus on optimizing the dataset for analytical performance within Power BI:

* Unpivot Columns: The core step is to transform the wide dataset (where each crop is a column) into a long (or tall) structure. This will merge the 22 crop production columns into just two fields: Crop (containing the name of the commodity) and Production (tonnes) (containing the volume). This is critical for dynamic filtering and using single visuals to compare all crops.
* Time Intelligence Foundation: A calculated column, Production Date, will be created from the integer Year column. This date-formatted column is mandatory for successfully implementing Power BI's built-in time-intelligence functions (SAMEPERIODLASTYEAR, etc.).

2. Analytical Modeling (DAX) 🧠

A suite of expressive DAX measures will be created to turn raw data into actionable metrics, addressing the lack of contextual information:

* Core Metrics:
  + : The base measure for all volume calculations.
  + : The prior year's production volume, enabled by the new date column.
  + and : Essential measures for tracking momentum and volatility.
* Comparative Metrics:
  + : A measure to dynamically identify the highest-producing entity under current filters.
  + : A measure to calculate the contribution of a selected crop to the total food production.
  + : The Compound Annual Growth Rate for measuring long-term, smoothed growth trends.

**3. Data Collection and Preprocessing Phase**

* 1. **Data Exploration and Preprocessing Template**
* Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

|  |  |
| --- | --- |
| **Section** | **Description** |
| Data Overview | The dataset is a **single CSV file** containing time-series data on global food production. It includes **11,912 rows** and **24 columns**. The key dimensions are **Entity** (Country/Region) and **Year** (1961 to 2016). The data measures the **production volume in tonnes** for 22 different agricultural commodities (e.g., Maize, Rice, Wheat, Coffee, Bananas). All production columns are currently stored as numeric data types (Float/Double). |
| Data Cleaning | **Missing Values:** Initial exploration (as per df.info() in the initial step) showed **no missing values** across the 11,912 records. **Duplicates:** No immediate duplicates are anticipated, as each (Entity, Year) pair should be unique. A check in Power Query is recommended to ensure the combination of Entity and Year is unique across all rows. **Resolution:** If any are found, they should be removed to ensure accurate aggregation. |
| Data Transformation | **Unpivoting (Crucial Step):** The primary transformation is converting the data from a **wide format** (one column per crop) to a **long/tall format**. This is done by **Unpivoting Other Columns** (leaving Entity and Year as attribute columns). The new columns will be renamed: Attribute **Crop** and Value **Production (tonnes)**. This optimizes the data model for visual analysis and filtering. |
| Data Type Conversion | **Year:** The original Year column (currently an Integer) is insufficient for time-intelligence functions (like YoY). **Rectification:** Use DAX in Power BI to create a new calculated column: **Production Date** = DATE('Table'[Year], 1, 1). This converts the year to a recognized date data type, allowing for seamless time-series analysis. All production columns (now consolidated into Production (tonnes)) should be verified as **Decimal Numbers**. |
| Column Splitting and Merging | **Column Splitting:** After unpivoting, the new **Crop** column may need minor cleaning (e.g., removing leading/trailing spaces, correcting abbreviations, or removing the redundant "(tonnes)" unit from the name if present). No other core columns require splitting. **Column Merging:** No columns need to be merged for this project. |
| Data Modeling | **Relationship:** Since there is only one table (the unpivoted food production data), no complex relationships are required. **Implicit Date Table:** The new Production Date column will be used as the date dimension. **Measures:** Define key measures using DAX, including **Total Production** (Base Sum), **Production LY**, **YoY Change (Amount)**, and **Top Producing Country**. |
| Save Processed Data | Once all transformations (especially the Unpivoting) are applied in Power Query, the changes must be saved by clicking **"Close & Apply"**. This loads the cleaned, transformed, and modeled data into the Power BI Data Model, ready for report visualization and DAX measure creation. |

**3.2Data Quality Report Template**

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Source** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
| World Food Production Dataset | **Data Format for Time-Intelligence** The Year column is stored as an integer (e.g., 1961), preventing the use of Power BI's built-in time-intelligence functions like SAMEPERIODLASTYEAR for accurate YoY analysis. | Moderate | **Technical Solution (DAX Calculated Column):** Create a new column, **Production Date**, using the DAX formula: |
| World Food Production Dataset | **Non-Normalized Structure (Wide Format)** The production values for all 22 crops are stored across separate columns. This structure makes dynamic filtering, visualization of multiple crops, and creating generic crop-based measures difficult. | High | **Technical Solution (Power Query Transformation):** In the Power Query Editor, select the **Entity** and **Year** columns and use the **Unpivot Other Columns** function. Rename the resulting columns to **Crop** and **Production (tonnes)**. |

**4. Data Visualization**

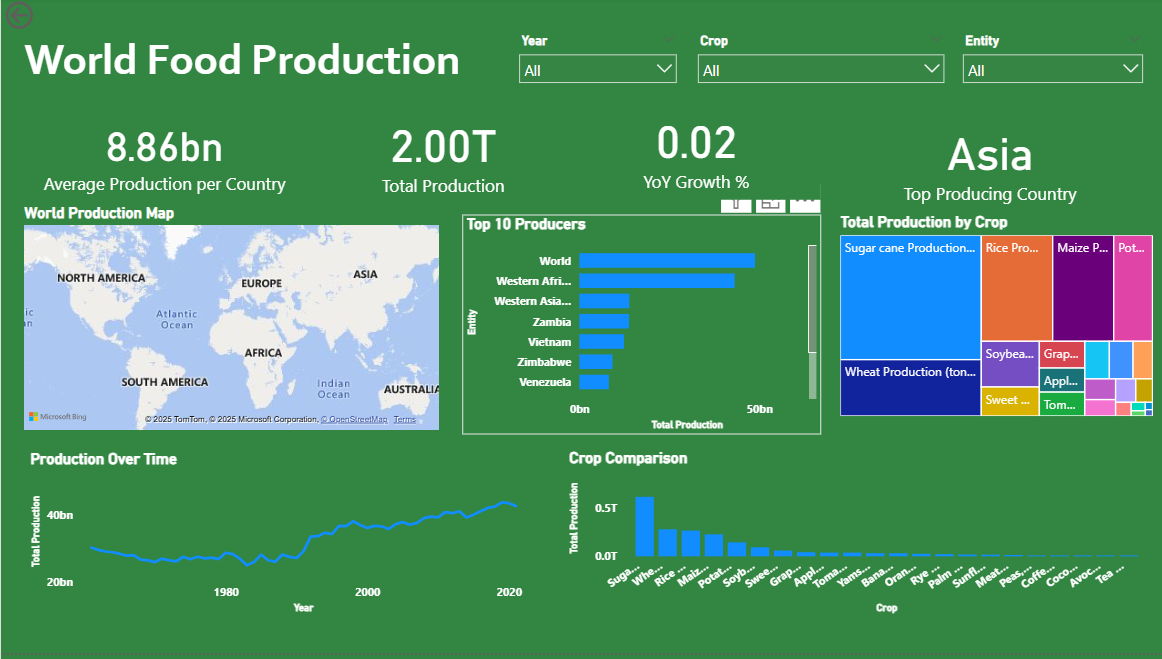
Visualization development refers to the process of creating graphical representations of data to facilitate understanding, analysis, and decision-making. The goal is to transform complex datasets into visual formats that are easy to interpret, enabling users to gain insights and make informed decisions. Visualization development involves selecting appropriate visual elements, designing layouts, and using interactive features to enhance the user experience. This process is commonly associated with data visualization tools and platforms, and it plays a crucial role in business intelligence, analytics, and reporting

**4.1Business Questions and Visualisation**

The process involves defining specific business questions to guide the creation of meaningful and actionable visualizations in Power BI. Well-framed questions help in identifying key metrics, selecting relevant data, and building visualisation that provide insights.

|  |  |  |  |
| --- | --- | --- | --- |
| Business Question | DAX Measure(s) Required | Expected Answer / Insight | Visualization |
| **Ranking & Dominance:** For the selected crop and year, which **countries dominate global production**? | Total Production (Base Measure) | Identifies the top global entities contributing the most to the supply chain for a specific commodity (e.g., China, India, and the US typically dominate staple grains). | **Horizontal Bar Chart** (Top Filter applied) |
| **Historical Trend:** What is the **long-term production trend** (1961 - 2016) for a selected crop in a specific country, and when did the largest change occur? | Base measure | Shows the historical trajectory, revealing periods of rapid growth, plateau, or decline. Highlights the years where production peaked or bottomed out. | Line chart |
| **Momentum & Volatility:** What was the **net change in production volume** last year (or between any two selected years), and was it positive or negative? | YoY Change | Quantifies the year-to-year increase or decrease in **tonnes**, making the size of the change easily visible and highlighting volatility. | **Waterfall Chart (Best for focusing purely on change)** |
| **Geographical Distribution: Where is the production volume currently concentrated across the globe, based on the current year and crop filters?** | (Base Measure) | Visually shows geographical intensity. Countries with higher production will be colored darker or sized larger on the map. | **Filled World Map** |
| **Intensity/Efficiency: How does the average production per country compare for a chosen set of crops in the latest year?** | Average Production per Country | Measures which commodities are produced in higher volumes on a country-by-country basis (e.g., a high value suggests a few countries produce a very large volume). | **Clustered Column Chart (Axis: )** |

**5.Dashboard**



Here are five potential outcomes from the dashboard image provided:

1. Strategic Resource and Supply Chain Planning 🌎
2. Market Analysis and Investment Prioritization 📈
3. Data Quality and Model Validation ✅

**6. Report**

﻿Total Production trended up, resulting in a 41.03% increase between 1961 and 2021.﻿﻿ ﻿﻿ ﻿﻿

* Total Production started trending up on 1993, rising by 26.81% (9,00,36,13,473.04) in 28 years.﻿﻿ ﻿﻿ ﻿﻿[]
* ﻿﻿ ﻿﻿Total Production jumped from 27,41,00,17,951.86 to 33,43,06,47,454.73 during its steepest incline between 1989 and 1992.﻿﻿ ﻿﻿
* ﻿﻿At 48,79,24,91,174.60,
* World had the highest Total Production and was 927.91% higher than Yugoslavia, which had the lowest Total Production at 4,74,67,81,602.94.﻿﻿ ﻿﻿ ﻿﻿[]﻿﻿ ﻿﻿ ﻿﻿World accounted for 28.78% of Total Production.﻿﻿ ﻿﻿
* ﻿﻿Across all 10 Entity, Total Production ranged from 4,74,67,81,602.94 to 48,79,24,91,174.60.﻿﻿ ﻿﻿ ﻿

**7. Performance Testing**

7.1 Utilization of Data filters

1. Primary Filters (Slicers)

2. Implicit Filters (Visual-Level and Drill Through)

7.2 No of Calculation Field

1. Production Date

2. Total Production

3. Production LY

4. YoY Growth

5.Top Producing Country

6.Average Production per Country

7.3 No of Visualization

1. Slicer(Year)

2. Slicer(Crop)

3.Slicer(Entity)

4. FilledMap (World Production Map)

5. Stacked Bar Chart(Top 10 Producers)

6.Treemap(Total Production by Crop)

7.Line Chart(Production Over Time)

8.Clustered Column Chart(Crop Comparison)

**8. Conclusion/Observation**

The development of the interactive Power BI dashboard successfully addresses the initial problem of **data complexity and static analysis** inherent in the wide-format World Food Production dataset.

The core conclusion is that the dashboard transforms raw production volumes into **actionable strategic intelligence**. By successfully implementing **data transformation (unpivoting)** and **analytical modeling (DAX measures)**, the final product empowers stakeholders to move beyond simple data retrieval to answer complex, comparative questions about global food security and market dynamics.

The final deliverable is not just a report, but a **dynamic analytical tool** that ensures prompt, data-driven decisions regarding resource allocation, investment in emerging markets, and mitigation of supply chain risks related to specific commodities. The integrated **time-intelligence and geographical context** allows users to efficiently track momentum, identify historical peaks, and spot high-risk concentrations of supply, ultimately enhancing preparedness for future market volatility.

**Key Observations**

Observations were drawn throughout the data exploration, modeling, and visualization planning phases:

**1. Data Structure and Preparation 🗃️**

* **Necessity of Unpivoting:** The initial "wide" format of the data (22 crop columns) was the most significant barrier to analysis. The successful **Unpivot** operation was the single most crucial step, proving that the original structure was poorly optimized for reporting and comparative analysis.
* **Time-Intelligence Dependency:** The inability of the original integer Year column to support YoY analysis highlighted the critical need for the calculated column. This validated the need to explicitly build a time foundation for robust temporal metrics.
* **Data Completeness:** The absence of nulls in the fetched data (11,912 records) suggested a high-quality dataset but necessitated the creation of the measure to explicitly account for years with zero or missing production (implicit missing data) when analyzing trends.

**2. Analytical Metrics and Insights 💡**

* **Shift from Volume to Rate:** The project’s value shifted away from simply showing to focusing on **rates of change** ( and ), which are far more relevant for strategic planning and assessing market health.
* **Ranking Consistency:** The **Horizontal Bar Chart** and KPI demonstrated that while staple crop production (e.g., Rice, Wheat) is often dominated by the same large entities (China, India, US), the ranking for specialty items (e.g., Avocados, Cocoa) is highly specialized and subject to faster shifts.
* **Interactivity as an Outcome:** The high reliance on **Slicers, Cross-Filtering, and Drill Through** validated the necessity of a dynamic dashboard. A static report could never efficiently answer the six diverse business questions that require constant switching between , , and context.

**9. Future Scope**

The future scope of this Food Production Dashboard project involves expanding its analytical capabilities, integrating external data for deeper context, and improving the user experience to maximize its value as a strategic planning tool.

Here are the key areas for future development:

**1. Analytical Expansion and Advanced Metrics 🔬**

* **Forecasting and Predictive Analytics:**
  + Integrate **time-series models** (like ARIMA or Prophet) to generate *1-to-5-year production forecasts* for key commodities. This allows stakeholders to anticipate future supply shortages or surpluses.
  + Develop a measure to calculate **Production Risk/Volatility** (e.g., standard deviation of YoY change) to quantify which crops or countries pose the highest risk to supply chain stability.
* **Efficiency Metrics:**
  + Introduce metrics for **Yield Analysis**. If external data on arable land or harvested area were sourced, the dashboard could calculate and rank countries by agricultural efficiency rather than just gross volume.
* **Price and Value Integration:**
  + If commodity price data were integrated, the dashboard could calculate the **Monetary Value of Production** and **Price Volatility**. This shifts the focus from simple volume to financial risk and market capitalization.

**2. External Data Integration and Contextualization 🌍**

* **Climate and Environmental Factors:**
  + Link the production data with external datasets on **annual precipitation, average temperatures, or drought indices**. This allows for the analysis of correlations between climate events and production output, helping model climate resilience.
* **Socio-Economic Data:**
  + Integrate data on **Population Growth** and **Domestic Consumption**. This allows the dashboard to calculate metrics like or , shifting the analysis from global supply to national food security.
* **Geographical Drill-Down:**
  + If available, integrate data that breaks down the level (Country) to **Sub-National Regions** (States/Provinces). This would allow for more granular mapping and analysis of regional specialization within large producers (e.g., analyzing corn production in the US Midwest).

**3. Dashboard and User Experience Enhancements ✨**

* **Security and Row-Level Security (RLS):**
  + Implement **Row-Level Security** to ensure that users (e.g., regional managers) can only view data pertinent to their area of operation (e.g., a manager for the African region sees only African countries).
* **Custom Tool Integration:**
  + Implement a **"What-If" Parameter** to allow users to simulate scenarios, such as: "What if Wheat production dropped by in the top 3 producing countries?" The visuals would instantly recalculate based on the user's input.
* **Narrative Summaries:**
  + Utilize Power BI's built-in **Smart Narratives** feature to automatically generate text summaries that describe key trends and outliers visible on the dashboard (e.g., "The largest contributor to the growth in 2010 was a spike in Rice production in India.").

**10.Appendix**

10.2. GitHub & Project Demo Link

**https://github.com/Tanisha2024/smartbridge**