

Cassava Production Trends and Analysis

Higher National Diploma in
Information Systems Management 23.2F

Data Warehousing & Data Mining

Assignment Report

Coursework-02

KAHNDISM232F-020

Table of Contents

- 1.Problem Definition & Significance
- 2.Chosen dataset and its source
- 3.Dimensional Model Design
- 4.Data Import to MySQL
- 5.SQL Queries & Results
- 6.Dashboard Design & Insights
- 7.Conclusion

1. Problem Definition & Significance

- **Problem Definition**

What is the trend of cassava production, area harvested, and yield over the years? is covered here in the report and charts.

The primary objective is to analyze the trends in cassava production, area harvested, and yield over the years from 1991 to 2020. Understanding these trends is crucial for making informed decisions in agricultural planning, policy-making, and improving farming practices to ensure sustainable cassava production.

- **Significance**

Cassava is a staple food crop in many parts of the world, particularly in Africa, Asia, and Latin America. Analyzing its production trends helps in:

- Enhancing food security
- Identifying areas for yield improvement.
- Allocating resources efficiently.
- Formulating policies to support farmers.
- Understanding the impact of climatic and economic changes on agricultural productivity.

2. Chosen Dataset and Its Source

Dataset

- **Cassava Yield Data**

The dataset used for this analysis provides an extensive view of cassava production trends in Ondo State, Nigeria, from 1991 to 2020. It includes annual production figures measured in tonnes per hectare, offering valuable insights into changes over three decades.

- **Data Source:**

The data was obtained from the Ministry of Agriculture, Ondo State, which supplied detailed annual yield data for cassava. Additionally, climate data such as temperature and rainfall was sourced from the Nigeria Meteorological Agency (NIMET), allowing us to examine the impact of climate variables on cassava productivity.

- **Source**

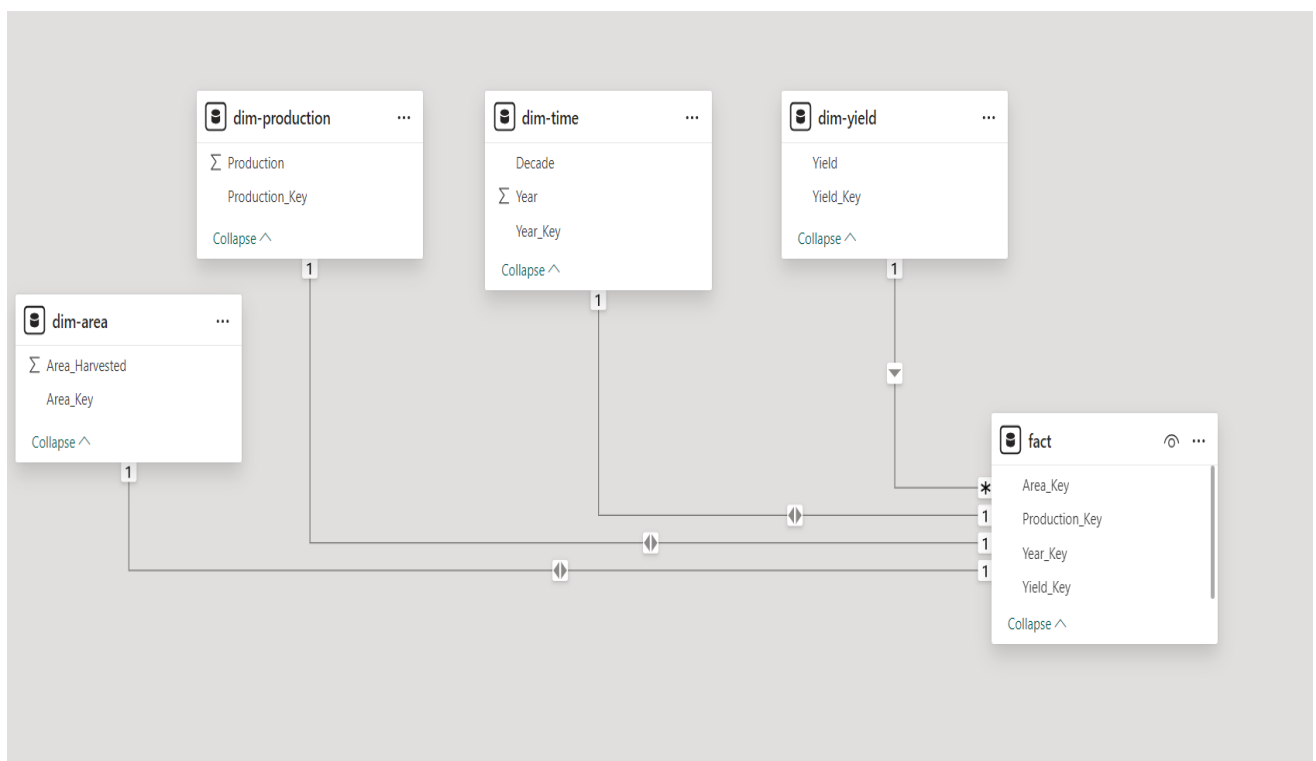
<https://www.kaggle.com/datasets/adekunlejoseph/ondo-state-nigeria-cocoa-cassava-and-weather-data>

3. Dimensional Model Design

The dimensional model for cassava production is structured in a star schema format, consisting of one fact table and four dimension tables. This model is designed to facilitate efficient data analysis by enabling quick access to relevant data through well-defined relationships.

The fact table Fact_Cassava_Production serves as the central repository for key performance metrics, including data related to production, area harvested, and yield. It is connected to four dimension tables, each providing additional context and attributes related to the metrics stored in the fact table.

- Dim_Area- links the harvested area details.
- Dim_Production- stores production figures.
- Dim_Time- provides temporal data, allowing analysis over different time periods, such as by year or decade.
- Dim_Yield- contains yield information.



4. Data Import to MySQL

To import the data into MySQL, follow these steps

Step 01

- After creating the database , Importe the chosen dataset's CSV file into a SQL database.

Step 02

- Using sql queries three dimensional tables and fact table were created.

Step 03

- Tables that created were exported as excel.csv files.

Step 04

- All the tables that exported as csv files were load into the power BI.

Step 05

- Charts are created according to data sheets that we imported.

5.SQL Queries and Results

- Creating table

```
1 • use cassava;
2 • CREATE TABLE Dim_Time AS
3   SELECT
4     Year AS Year_Key,
5     Year,
6     CASE
7       WHEN Year BETWEEN 1990 AND 1999 THEN '1990s'
8       WHEN Year BETWEEN 2000 AND 2009 THEN '2000s'
9       WHEN Year BETWEEN 2010 AND 2019 THEN '2010s'
10      ELSE 'Other'
11    END AS Decade
12  FROM (SELECT DISTINCT Year FROM cassava_crop) AS Years;
```

- Creating dimensional table (Dim_Production)

```
14 • CREATE TABLE Dim_Production AS
15   SELECT
16     @rownum := @rownum + 1 AS Production_Key,
17     Production
18  FROM (SELECT DISTINCT PRODUCTION AS Production FROM cassava_crop) AS Productions, (SELECT @rownum := 0) AS R;
19
```

- Creating dimensional table (Dim_Yield)

```
26 • CREATE TABLE Dim_Yield AS
27   SELECT
28     @rownum := @rownum + 1 AS Yield_Key,
29     YIELD AS Yield
30  FROM (SELECT DISTINCT YIELD FROM cassava_crop) AS Yields, (SELECT @rownum := 0) AS R;
31
```

- Creating dimensional table (Dim_Area)

```

20 • CREATE TABLE Dim_Area AS
21 SELECT
22     @rownum := @rownum + 1 AS Area_Key,
23     AREA AS Area_Harvested
24 FROM (SELECT DISTINCT AREA FROM cassava_crop) AS Areas, (SELECT @rownum := 0) AS R;
25

```

- Creating Fact table(Fact_Cassava_Production)

```

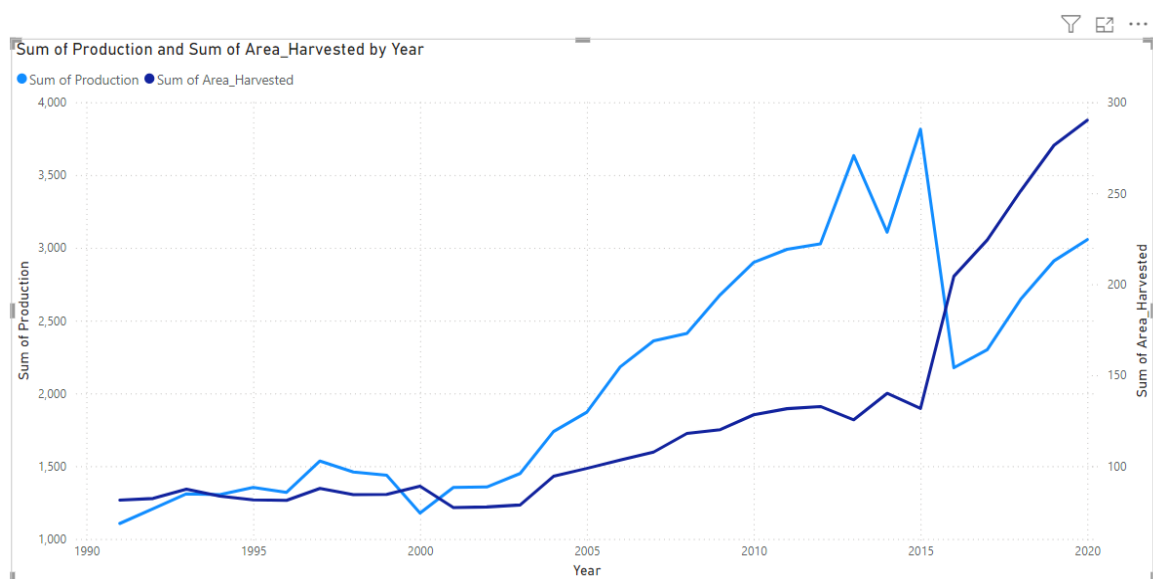
32 • CREATE TABLE Fact_Cassava_Production AS
33 SELECT
34     t.Year_Key,
35     p.Production_Key,
36     a.Area_Key,
37     y.Yield_Key
38 FROM cassava_crop cd
39 JOIN Dim_Time t ON cd.Year = t.Year
40 JOIN Dim_Production p ON cd.PRODUCTION = p.Production
41 JOIN Dim_Area a ON cd.AREA = a.Area_Harvested
42 JOIN Dim_Yield y ON cd.YIELD = y.Yield;

```

6. Dashboard Design & Insights

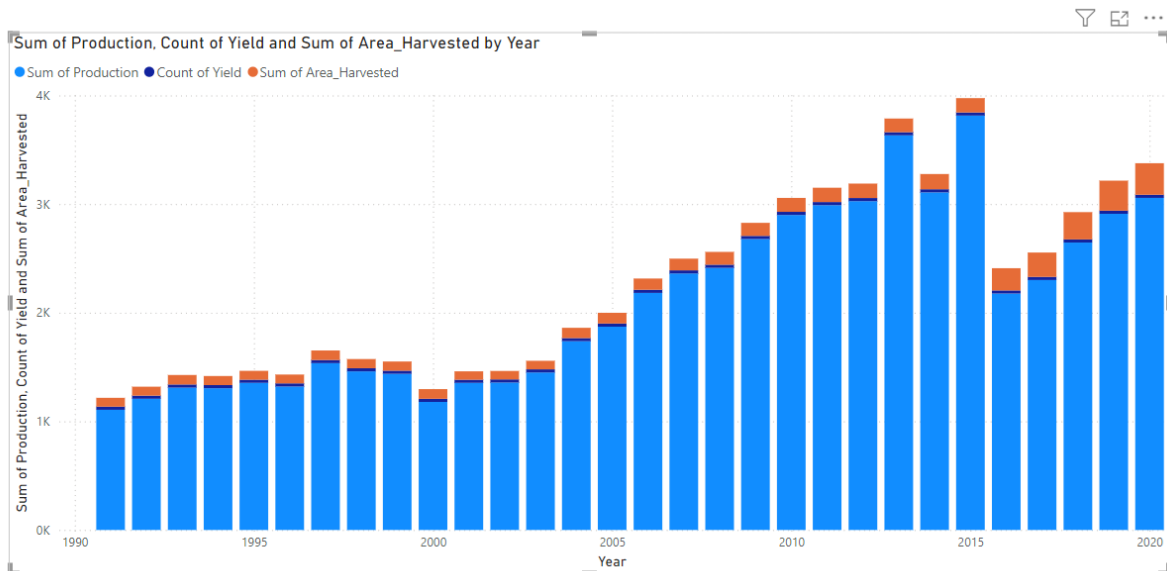
A dashboard is created using Power BI to visualize the data. The following components can be included:

- Line Chart



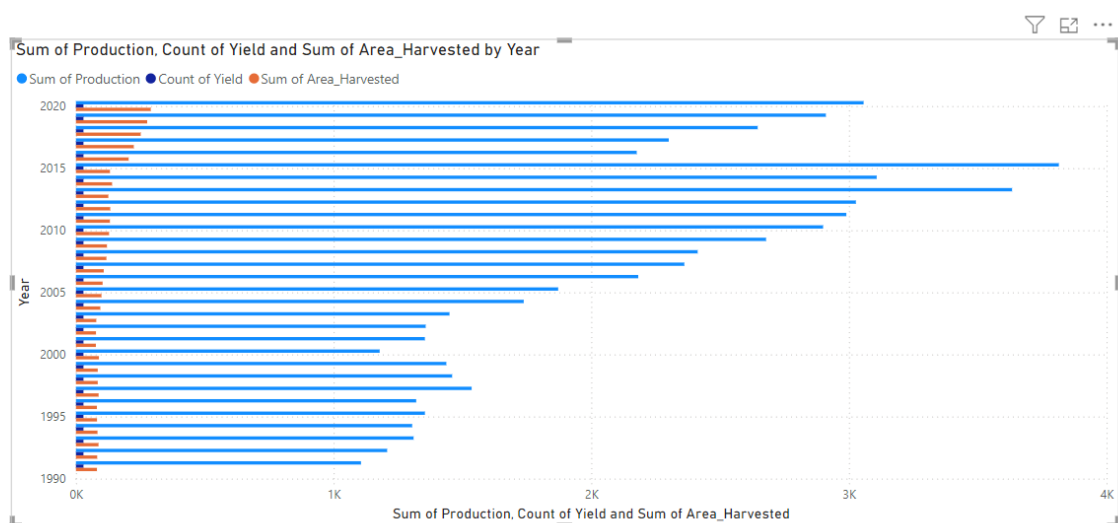
Explanation: This line chart shows the sum of production and the sum of area harvested by year. It helps visualize the trend in production and area harvested over time. The chart indicates a general increase in both production and area harvested from 1990 to 2020, with noticeable spikes in production around the early 2010s and mid-2010s.

- **Bar Chart**



Explanation: The bar chart represents the sum of production, count of yield, and sum of area harvested by year. It shows the comparison between the total production, number of yield occurrences, and the total harvested area for each year. This chart is useful to see the relative contributions of each metric year over year.

- **Stacked Column Chart**



Explanation: The stacked column chart displays the sum of production, count of yield, and sum of area harvested by year, stacked on top of each other. This chart is useful for understanding the cumulative impact of all three metrics over the years, as well as the individual contributions to the total for each year. The steady growth in production is evident, with some fluctuations in yield counts and area harvested

7.Conclusion

This comprehensive report outlines the analysis of cassava production trends, including problem definition, dataset details, dimensional model design, data import process, SQL queries, and dashboard insights. This structured approach ensures a thorough understanding and actionable insights for future planning.