

Agriculture Data Analysis and Dashboard Design Report

Data Warehousing & Data Mining

Assignment Report

Coursework-02

KAHNDISM23.2F 019

Table of Contents

1. Problem Definition & Significance	2
2. Chosen Dataset and Its Source	2
3. Dimensional Model Design	3
4. Data Import to MySQL	3
5. SQL Queries & Results	4
6. Dashboard Design & Insights	6
7. Chart Explanation	8

1.Problem Definition & Significance

Problem Definition:

What is the total standard output (in EUR) for each country?

This analysis's goal is to figure out and assess each European nation's overall standard production in euros during a certain time frame. An essential economic metric for evaluating the productivity and performance of the agricultural sector across nations is the standard output, which expresses the value of agricultural products.

Significance:

Understanding the total standard output is crucial for several reasons:

1. Economic Performance:

In many European nations, the agricultural industry plays a significant role in the national economy. A measure of agriculture's economic contribution that enables cross-national and cross-temporal comparisons is the total standard output. For the purpose of assessing the performance and state of the agricultural economy, this data is crucial.

2. Policy Development:

Accurate data on agricultural output is necessary for policymakers to create and carry out successful programs. Understanding the total standard output facilitates the development of agricultural growth, productivity, and food security-promoting strategies. Additionally, it facilitates the distribution of grants, subsidies, and other types of financial assistance to the agriculture industry.

3. Resource Allocation:

A better way to allocate resources is to understand the whole standard output. Different forms of support may be needed for countries with higher outputs than for those with lower outputs. By using this data, infrastructure development, technical support, and funding can be focused where it is most required.

4. Market Analysis:

For market analysis, the total standard output is a useful statistic. It aids in recognizing patterns, estimating future output levels, and comprehending the supply side of agricultural markets. Businesses, investors, and dealers participating in the agriculture supply chain can benefit from this information.

2. Chosen dataset and its source

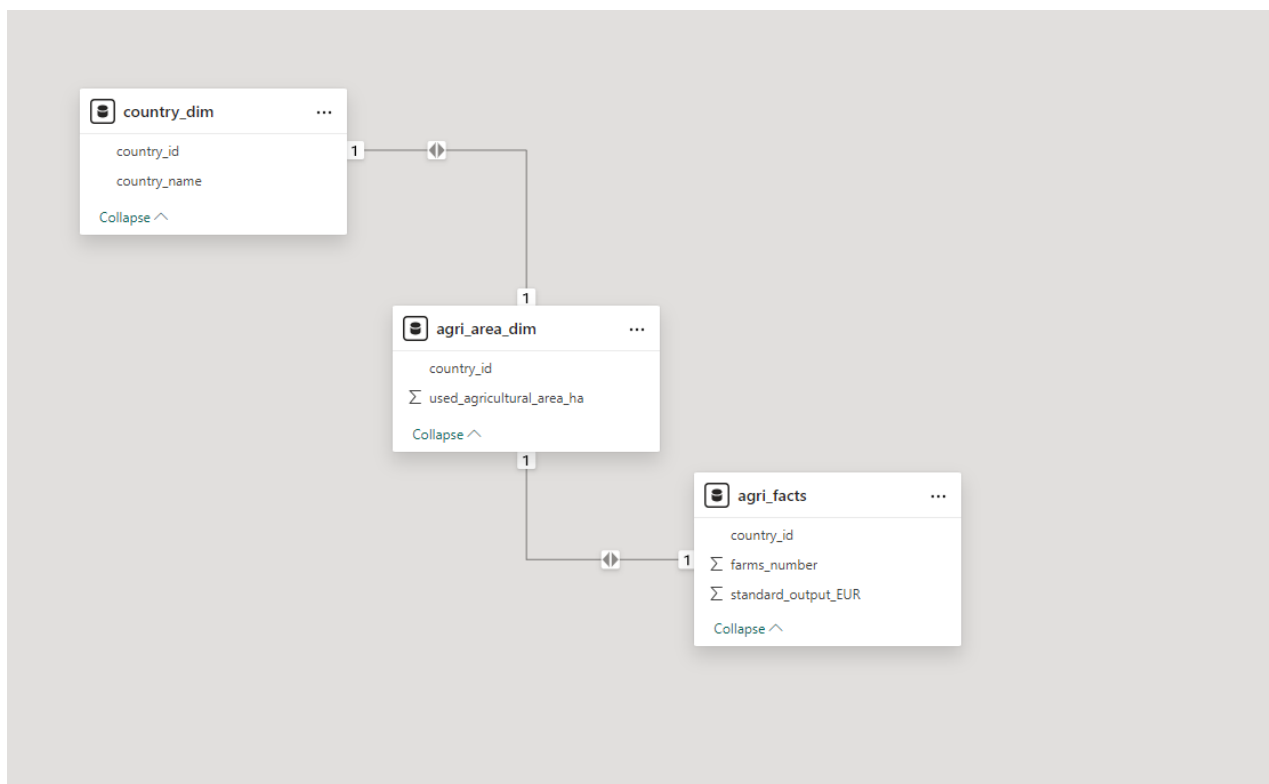
Data source link : <https://www.kaggle.com/datasets/georgeam/european-agriculture-indicatorseurostat-2016>

3. Dimensional Model Design

There are four dimensional tables and one fact table according to the dataset:

1. **dim_country:**
 - 1.1. **Purpose:** This table categorizes data by country, enabling country-wise analysis.
 - 1.2. **Attributes:** country_id (Primary Key), country_name, region, subregion.
2. **dim_year:**
 - 2.1. **Purpose:** This table categorizes data by time, enabling time-series analysis.
 - 2.2. **Attributes:** year_id (Primary Key), year.
3. **dim_agriculture_type:**

- 3.1. **Purpose:** This table provides information about different types of agricultural outputs and structures.
- 3.2. **Attributes:** agri_type_id (Primary Key), agriculture_type.
4. **dim_currency:**
- 4.1. **Purpose:** This table contains information about the currency and conversion rates if applicable.
- 4.2. **Attributes:** currency_id (Primary Key), currency_name, conversion_rate_to_euro.
5. **fact_agriculture_output:**
- 5.1. **Purpose:** This central table stores quantitative data about agricultural output.
- 5.2. **Attributes:** fact_id (Primary Key), country_id (Foreign Key), year_id (Foreign Key), agri_type_id (Foreign Key), currency_id (Foreign Key), standard_output_in_eur, quantity, value_in_local_currency.



4.Data Import to MySQL

Step 01: Importing the Dataset

- After creating the database in MySQL, the chosen dataset's CSV file was imported into the SQL database. This step involved ensuring that the data types and fields in the CSV file were correctly mapped to the appropriate columns in the database table.

Step 02: Creating Dimensional and Fact Tables

- The selected dataset's CSV file was imported into the SQL database following the creation of the database in MySQL. In this phase, it was necessary to confirm that the fields and data types in the CSV file were appropriately mapped to the corresponding columns in the database table.

Step 03: Exporting Tables as CSV Files

- Excel CSV files were exported from the MySQL tables that were produced. In order to prepare the data for loading into Power BI, this step was required. By exporting the tables as CSV files, compatibility problems with the data import into Power BI were avoided.

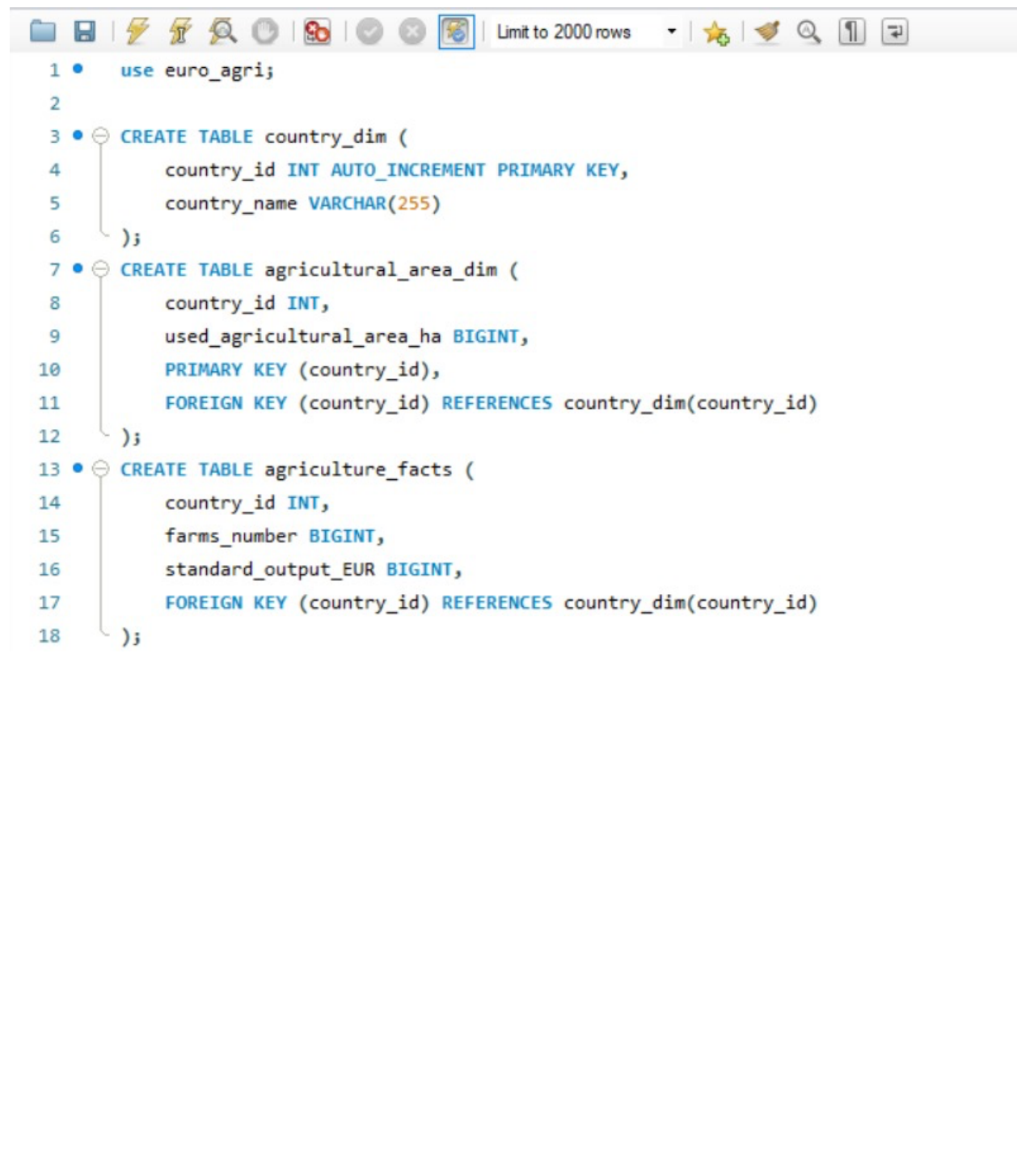
Step 04: Loading Tables into Power BI

- Power BI loaded all of the tables that had been exported as CSV files. The 'Get Data' function in Power BI was utilised to import the CSV files. To make sure the data was prepared for analysis, it was converted and cleaned as needed when it was imported.

Step 05: Creating Charts and Dashboards

- Using the imported data, several Power BI dashboards and charts were made. Bar charts, line charts, map visualizations, and other chart types were among these visualizations. These charts served as a useful tool for analyzing and presenting the notable shifts in farm size and structure that occurred over time in European nations, offering insightful information on the development of the agricultural industry.

5. SQL Queries & Results



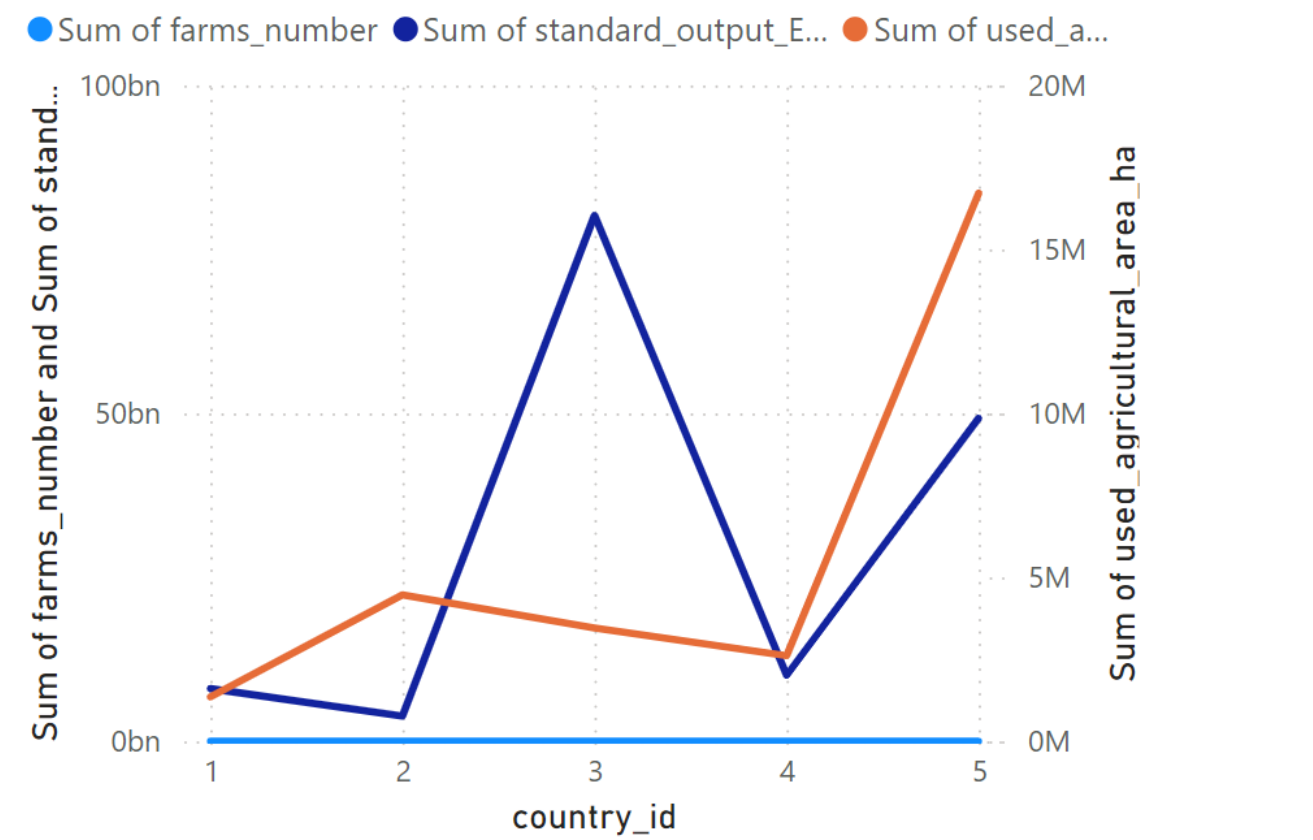
The screenshot shows a SQL IDE interface with a toolbar at the top containing icons for file operations, execution, search, and other database functions. The main area displays three SQL queries for creating tables in a database named 'euro_agri'.

```
1 • use euro_agri;
2
3 • CREATE TABLE country_dim (
4     country_id INT AUTO_INCREMENT PRIMARY KEY,
5     country_name VARCHAR(255)
6 );
7 • CREATE TABLE agricultural_area_dim (
8     country_id INT,
9     used_agricultural_area_ha BIGINT,
10    PRIMARY KEY (country_id),
11    FOREIGN KEY (country_id) REFERENCES country_dim(country_id)
12 );
13 • CREATE TABLE agriculture_facts (
14     country_id INT,
15     farms_number BIGINT,
16     standard_output_EUR BIGINT,
17     FOREIGN KEY (country_id) REFERENCES country_dim(country_id)
18 );
```

```
Limit to 2000 rows
19 • INSERT INTO country_dim (country_name)
20 VALUES
21 ('Belgium'),
22 ('Bulgaria'),
23 ('Czechia'),
24 ('Denmark'),
25 ('Germany');
26
27 • INSERT INTO agricultural_area_dim (country_id, used_agricultural_area_ha)
28 VALUES
29 ((SELECT country_id FROM country_dim WHERE country_name = 'Belgium'), 1354250),
30 ((SELECT country_id FROM country_dim WHERE country_name = 'Bulgaria'), 4468500),
31 ((SELECT country_id FROM country_dim WHERE country_name = 'Czechia'), 3455410),
32 ((SELECT country_id FROM country_dim WHERE country_name = 'Denmark'), 2614600),
33 ((SELECT country_id FROM country_dim WHERE country_name = 'Germany'), 16715320);
34
35 • INSERT INTO agriculture_facts (country_id, farms_number, standard_output_EUR)
36 VALUES
37 ((SELECT country_id FROM country_dim WHERE country_name = 'Belgium'), 36890, 8037986420),
38 ((SELECT country_id FROM country_dim WHERE country_name = 'Bulgaria'), 202720, 3842891030),
39 ((SELECT country_id FROM country_dim WHERE country_name = 'Czechia'), 26530, 80154706065),
40 ((SELECT country_id FROM country_dim WHERE country_name = 'Denmark'), 35050, 10062442040),
41
42 • CREATE TABLE fact_rice_production (
43     Fact_ID INT PRIMARY KEY AUTO_INCREMENT,
44     Year_ID INT,
45     Economic_ID INT,
46     Land_ID INT,
47     Rice_Production DECIMAL(10, 2),
48     FOREIGN KEY (Year_ID) REFERENCES dim_year(Year_ID),
49     FOREIGN KEY (Economic_ID) REFERENCES dim_economic(Economic_ID),
50     FOREIGN KEY (Land_ID) REFERENCES dim_land(Land_ID)
51 );
52
53 • SELECT c.country_name, SUM(a.standard_output_EUR) AS total_standard_output_EUR
54 FROM agriculture_facts a
55 JOIN country_dim c ON a.country_id = c.country_id
56 GROUP BY c.country_name;
57
```

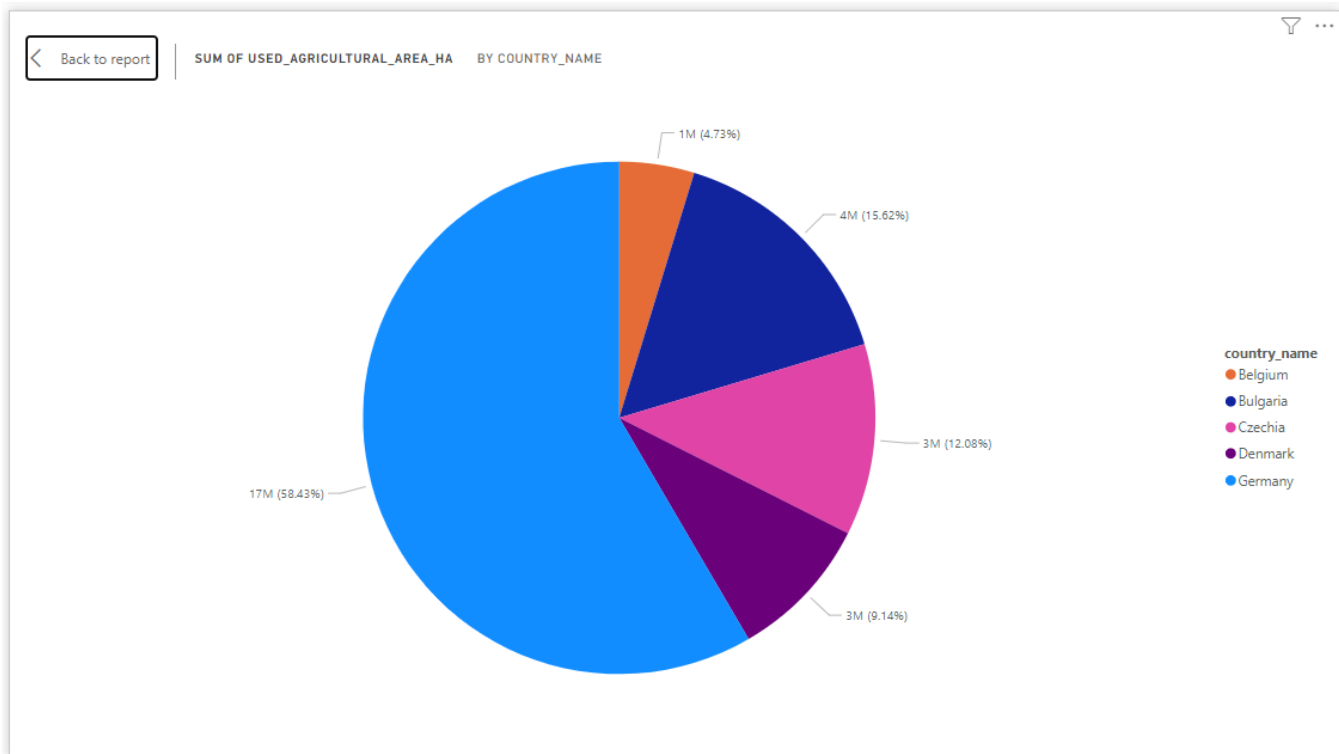

6. Dashboard Design & Insights

Sum of farms number & Sum of used agricultural are ha by country_id

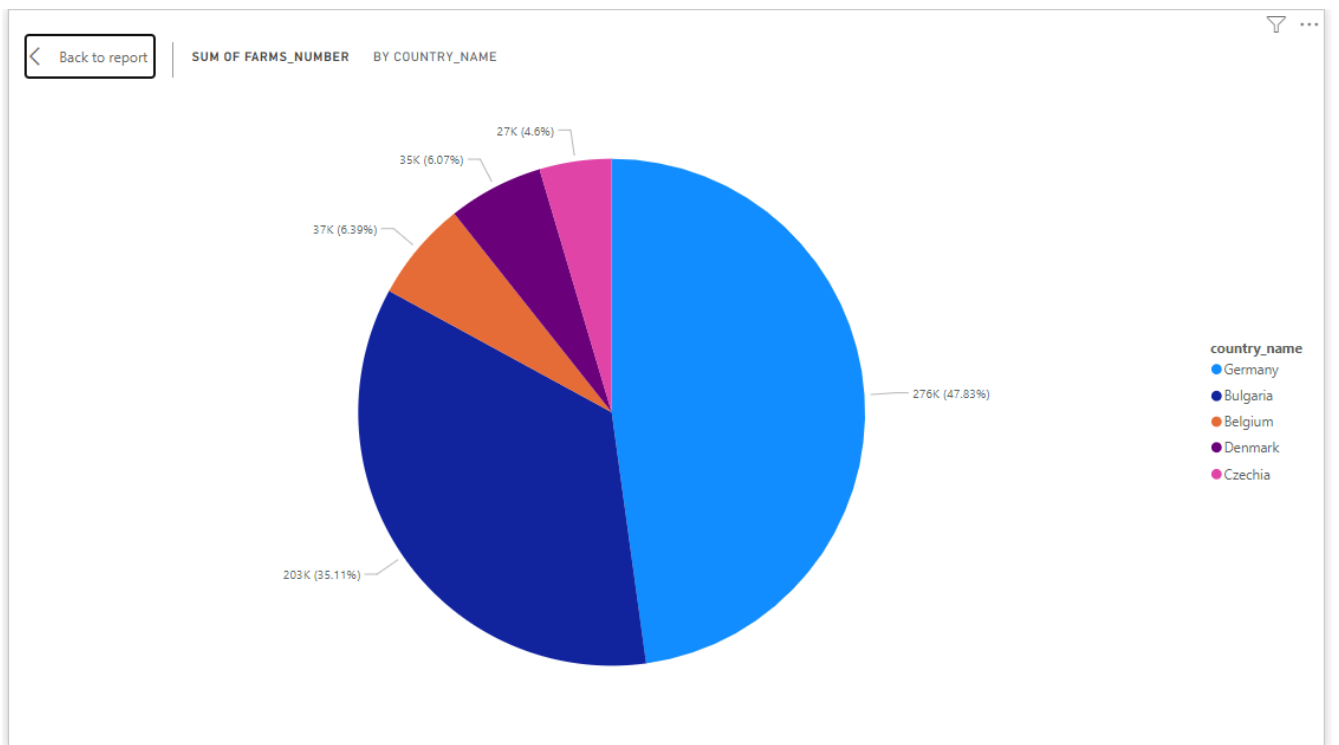


Sum of used_agricultural_area_ha by country_name

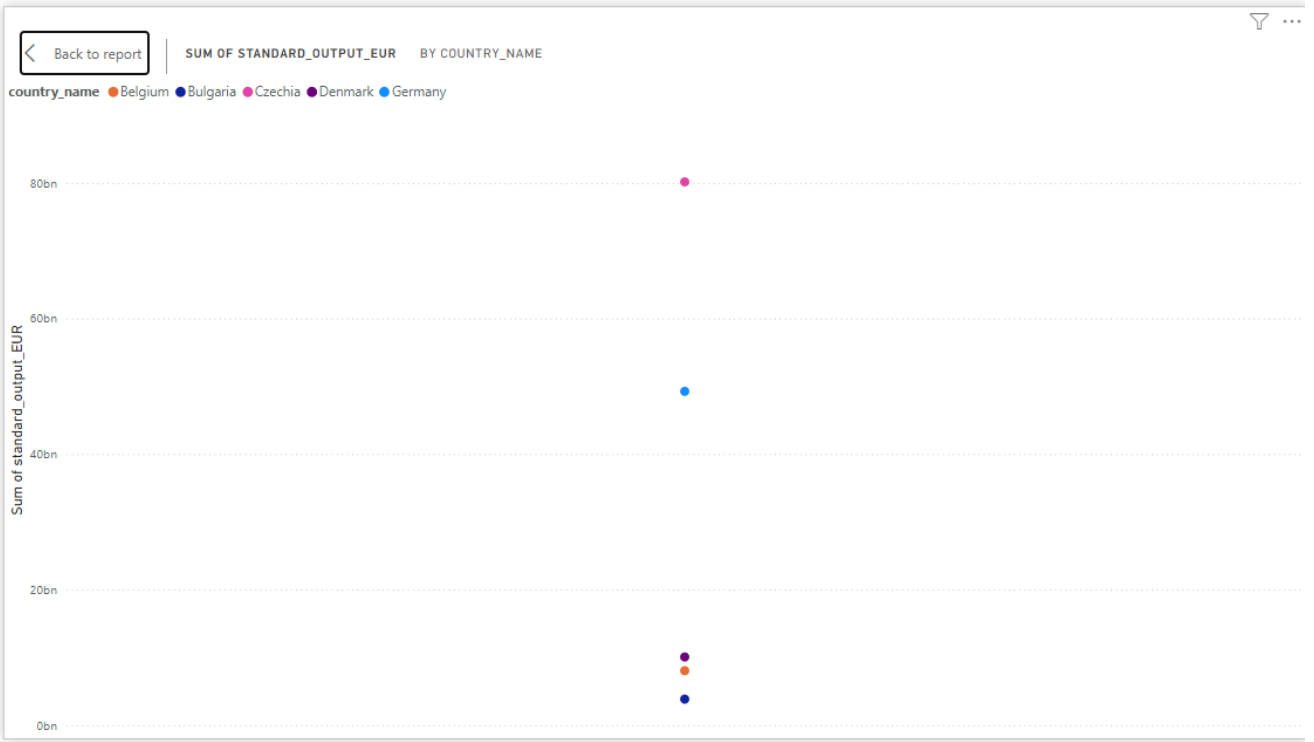
Combined Line Chart



Pie Chart 1



Pie Chart 2



Line Chart

7.Chart Explanation

❖ Pie Chart 1

The "Sum of Farms Number" is broken down in this chart into many categories, which most likely correspond to various nations or areas. Making particular judgements is hard without detailed facts. Still, a bigger piece of the pie usually means there are more farms in that specific category.

❖ Pie Chart 2

This graph shows the data distribution according to "Country Name." Each slice's size indicates what percentage of the total that nation represents. Providing precise insights is difficult in the absence of context or explicit values. A bigger slice, however, would suggest a greater representation of that nation in the dataset.

❖ Line Chart

The association between "Sum of farms number" and "Sum of standard output EUR" across various nations is shown in this chart. Trends across these factors are shown by the lines. It's challenging to pinpoint the precise association between the two in the absence of particular data points or patterns. Generally speaking, though, a positive correlation would suggest that rising values for one variable are related to rising values for the other.

❖ Combined Line Chart

Two line charts are combined in this graphic to show the "Sum of farms number" and the "Sum of standard output EUR" broken down by "country_id." It is difficult to analyze trends or correlations without precise data. On the other hand, the chart attempts to illustrate how these two factors differ amongst nations.