Dairy Supply Chain Sales Dataset

Readme File

1. Introduction

Sales data collection is a crucial aspect of any manufacturing industry as it provides valuable insights about the performance of products, customer behaviour, and market trends. By gathering and analysing this data, manufacturers can make informed decisions about product development, pricing, and marketing strategies in Internet of Things (IoT) business environments like the dairy supply chain.

One of the most important benefits of the sales data collection process is that it allows manufacturers to identify their most successful products and target their efforts towards those areas. For example, if a manufacturer could notice that a particular product is selling well in a certain region, this information could be utilised to develop new products, optimise the supply chain or improve existing ones to meet the changing needs of customers.

This dataset includes information about 7 of MEVGAL[1]'s products. According to the above information the data published will help researchers to understand the dynamics of the dairy market and its consumption patterns, which is creating the fertile ground for synergies between academia and industry and eventually help the industry in making informed decisions regarding product development, pricing and market strategies in the IoT playground. The use of this dataset could also aim to understand the impact of various external factors on the dairy market such as the economic, environmental, and technological factors. It could help in understanding the current state of the dairy industry and identifying potential opportunities for growth and development.

---

[1] MEVGAL is a Greek dairy production company

2. Citation

Please cite the following papers when using this dataset:

[1] I. Siniosoglou, K. Xouveroudis, V. Argyriou, T. Lagkas, S. K. Goudos, K. E. Psannis and P. Sarigiannidis, "**Evaluating the Effect of Volatile Federated Timeseries on Modern DNNs: Attention over Long/Short Memory**," in the 12th International Conference on Circuits and Systems Technologies (MOCAST 2023), April 2023, Accepted

3. Dataset Modalities

The dataset includes data regarding the daily sales of a series of dairy product codes offered by MEVGAL. In particular, the dataset includes information gathered by the logistics division and agencies within the industrial infrastructures overseeing the production of each product code. The products included in this dataset represent the daily sales and logistics of a variety of yogurt-based stock. Each of the different files include the logistics for that product on a daily basis for three years, from 2020 to 2022.
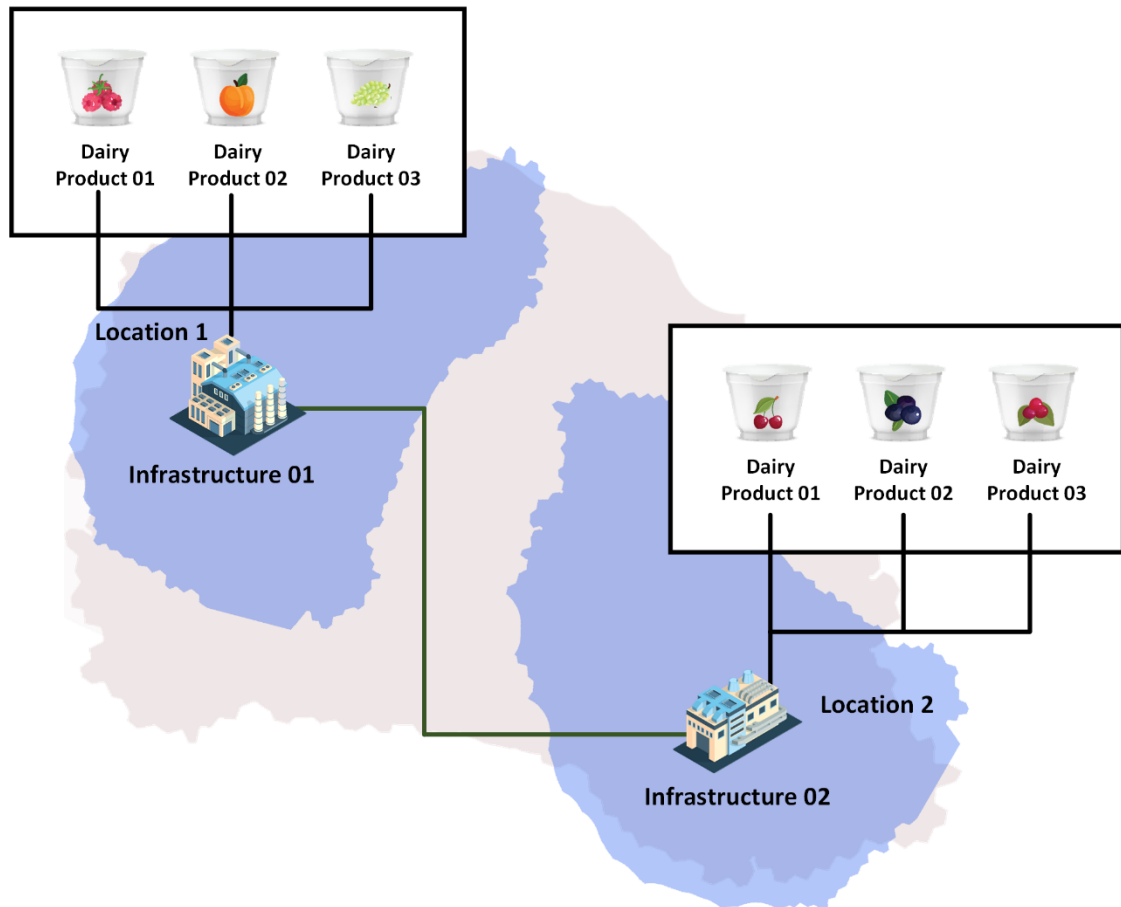
3.1 Data Collection

The process of building this dataset involves several steps to ensure that the data is accurate, comprehensive and relevant.

The first step is to determine the specific data that is needed to support the business objectives of the industry, i.e., in this publication's case the daily sales data.

Once the data requirements have been identified, the next step is to implement an effective sales data collection method. In MEVGAL's case this is conducted through direct communication and reports generated each day by representatives & selling points.

It is also important for MEVGAL to ensure that the data collection process conducted is in an ethical and compliant manner, adhering to data privacy laws and regulation. The industry also has a data management plan in place to ensure that the data is securely stored and protected from unauthorised access.

The published dataset is consisted of 13 features providing information about the date and the number of products that have been sold. Finally, the dataset was anonymised in consideration to the privacy requirement of the data owner (MEVGAL).

File	Period	Number of Samples (days)
product 1 2020.xlsx	01/01/2020–31/12/2020	363
product 1 2021.xlsx	01/01/2021–31/12/2021	364
product 1 2022.xlsx	01/01/2022–31/12/2022	365
product 2 2020.xlsx	01/01/2020–31/12/2020	363
product 2 2021.xlsx	01/01/2021–31/12/2021	364
product 2 2022.xlsx	01/01/2022–31/12/2022	365
product 3 2020.xlsx	01/01/2020–31/12/2020	363
product 3 2021.xlsx	01/01/2021–31/12/2021	364
product 3 2022.xlsx	01/01/2022–31/12/2022	365
product 4 2020.xlsx	01/01/2020–31/12/2020	363
product 4 2021.xlsx	01/01/2021–31/12/2021	364
product 4 2022.xlsx	01/01/2022–31/12/2022	364
product 5 2020.xlsx	01/01/2020–31/12/2020	363
product 5 2021.xlsx	01/01/2021–31/12/2021	364
product 5 2022.xlsx	01/01/2022–31/12/2022	365
product 6 2020.xlsx	01/01/2020–31/12/2020	362
product 6 2021.xlsx	01/01/2021–31/12/2021	364

product 6 2022.xlsx	01/01/2022–31/12/2022	365
product 7 2020.xlsx	01/01/2020–31/12/2020	362
product 7 2021.xlsx	01/01/2021–31/12/2021	364
product 7 2022.xlsx	01/01/2022–31/12/2022	365

3.2   Dataset Overview

The following table enumerates and explains the features included across all of the included files.

Feature	Description	Unit
Day	day of the month	-
Month	Month	-
Year	Year	-
daily_unit_sales	Daily sales - the amount of products, measured in units, that during that specific day were sold	units
previous_year_daily_unit_sales	Previous Year's sales - the amount of products, measured in units, that during that specific day were sold the previous year	units
percentage_difference_daily_unit_sales	The percentage difference between the two above values	%
daily_unit_sales_kg	The amount of products, measured in kilograms, that during that specific day were sold	kg
previous_year_daily_unit_sales_kg	Previous Year's sales - the amount of products, measured in kilograms, that during that specific day were sold, the previous year	kg
percentage_difference_daily_unit_sales_kg	The percentage difference between the two above values	kg
daily_unit_returns_kg	The percentage of the products that were shipped to selling points and were returned	%
previous_year_daily_unit_returns_kg	The percentage of the products that were shipped to selling points and were returned the previous year	%
points_of_distribution	The amount of sales representatives through which the product was sold to the market for this year	

previous_year_points_of_distribution	The amount of sales representatives through which the product was sold to the market for the same day for the previous year	

Table 1 – Dataset Feature Description

4. Structure and Format

4.1 Dataset Structure

The provided dataset has the following structure:

```
Product Sales Dataset
|    Readme.docx
|
+---product 1
|        product 1 2020.xlsx
|        product 1 2021.xlsx
|        product 1 2022.xlsx
|
+---product 2
|        product 2 2020.xlsx
|        product 2 2021.xlsx
|        product 2 2022.xlsx
|
+---product 3
|        product 3 2020.xlsx
|        product 3 2021.xlsx
|        product 3 2022.xlsx
|
+---product 4
|        product 4 2020.xlsx
|        product 4 2021.xlsx
|        product 4 2022.xlsx
|
+---product 5
|        product 5 2020.xlsx
|        product 5 2021.xlsx
|        product 5 2022.xlsx
|
+---product 6
|        product 6 2020.xlsx
|        product 6 2021.xlsx
|        product 6 2022.xlsx
|
+---product 7
         product 7 2020.xlsx
         product 7 2021.xlsx
         product 7 2022.xlsx
```

Where:

Name	Type	Property
Readme.docx	Report	A File that contains the documentation of the Dataset.
product X	Folder	A folder containing the data of a product X.

product X YYYY.xlsx	Data file	An excel file containing the sales data of product X for year YYYY.

Table 2 - Dataset File Description

5. Acknowledgement

References