UNIVERSITY OF MAURITIUS

Faculty of Social Sciences & Humanities (FSSH)

Department of Economics & Statistics

<u>Course</u>

MSc Data Analytics

Year 1 Semester 1

<u>Module</u>

Data Handling and Analysis

STAT 5129

Final Group Coursework

Submitted to:

Mr. R. Thoplan

Submitted by:

Mavish GAJADHUR	*2521826*
Parvesh GHOORA	*2522774*
Homeswaree JOWAHEER	*2522007*
Saveena KOWLESSUR	*2521531*
Shradha NUCCA GOVEDO	*2520562*

Submission Date: 31st October 2025

Digital Receipt

This receipt confirms that Turnitin has received your submission. See details about your file below.

Submission author
Parvesh Ghoora

Assignment title
Final Submission

Submission title
STAT5129_Report_GroupD

File name
STAT5129_Report_GroupD.docx

File size
1052504

Page count
33

Word count
5373

Character count
32021

Submission date
Oct 31, 2025, 3:21:08 PM UTC

Submission ID
trn:oid:::2945:324585012

Contents

Table of Figures

Table of tables

List of Abbreviation

Abbreviation	Full Form
°C	Degrees Celsius
Agri	Agriculture
API	Application Programming Interface
CDD	Cooling Degree Days
CEB	Central Electricity Board
CO_2	Carbon Dioxide
CSV	Comma-Separated Values
DBMS	Database Management System
ERD	Entity Relationship Diagram
ETL	Extract, Transform, Load
FAO	Food and Agriculture Organization
FAOSTAT	Food and Agriculture Organization Statistical Database
FK	Foreign Key
GDP	Gross Domestic Product
GHG	Greenhouse Gas
GWh	Gigawatt-hour
Ha	Hectare
ID	Identifier
IPCC	Intergovernmental Panel on Climate Change
IRR	Irrigation
JSON	JavaScript Object Notation
kWh	Kilowatt-hour
MCAR	Missing Completely At Random
MW	Megawatt
NA	Not available
NASA	National Aeronautics and Space Administration
NASA POWER	NASA Prediction Of Worldwide Energy Resource
POWER	Prediction Of Worldwide Energy Resource
PRECTOTCORR	Precipitation Corrected Total
PV	Photovoltaic
QCL	Crops and Livestock Products (FAOSTAT domain)
RE	Renewable Energy
ROI	Return on Investment
SQL	Structured Query Language
T2M	Temperature at 2 Meters

Impact of Climate Conditions on Electricity Generation and Agricultural Production in Mauritius

Introduction

Sugarcane cultivation has long been the foundation of the agricultural economy of Mauritius. Bagasse is also a vital source of renewable energy. Bagasse, a by-product of sugarcane extraction, makes a significant contribution to the production of electricity (reference). However, due to primarily land abandonment, lower yields and changing climatic conditions, bagasse electricity exports have declined from 381 GWh in 2015 to 246 GWh in 2021. Both agricultural productivity and energy output are severely impacted by the island's extremely variable climatic conditions, which fluctuates between warm, rainy seasons and cooler, drier periods. The risk of sugarcane production to climate change is demonstrated by studies showing that simply a 1^{0}C increase in summer temperatures can result into a decline of approximately USD 26.6 per acre in agricultural earnings. Therefore, sustaining Mauritius's food and security requires adding adaptability through improved biomass utilization, effective irrigation and flexible methods of cultivation.

Objective of the Study

The purpose of this study is to evaluate the effects of climate fluctuations, including temperature and precipitation, on Mauritius's agricultural yield and electricity generation, with particular emphasis on sugarcane cultivation. The project aims at assessing the extent of these impacts using data from from **NASA POWER (JSON format, semi-structured data)**, **FAOSTAT Crops (CSV format, structured data)**, and the **CEB Electricity (CSV format, structured data)**.

Study Process

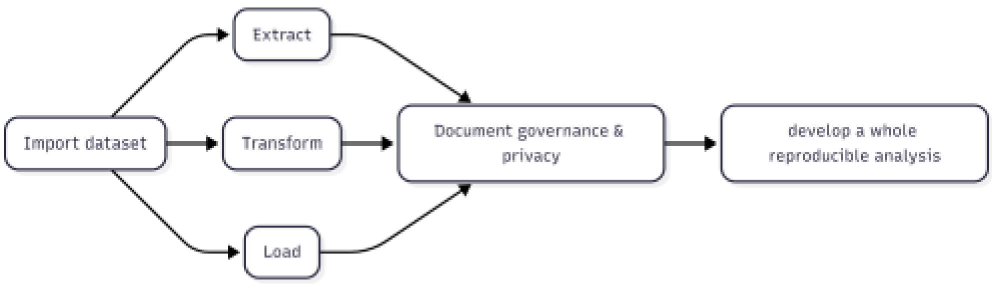

Figure 1: Study Process

Dataset Sources & Attribution

To support the analysis of Mauritius's production of electricity, crop yield and shifting climatic conditions, the following three datasets were gathered. The dataset and attribution are as follows:

1. CEB Electricity Dataset

Data on electricity generation transmitted to the Central Electricity Board by origin, such as bagasse, coal, landfill gas, photovoltaic, and wind, is available annually (2010–2022) in the CEB Electricity dataset. All figures are expressed in gigawatt-hours (GWh). The Central Electricity Board publishes the dataset, which includes information on both Mauritius and the Rodrigues Islands. The Mauritius Digital Promotion Agency (MDPA) converts the dataset to CSV format.

Download Page: Electricity exported to Central Electricity Board - Dataset - opendata Mauritius

2. FAOSTAT Crops Dataset

The FAOSTAT Crops dataset, which provides information on production, yield, and area harvested for various kinds of agricultural commodities, is a component of the QCL (Crops and Livestock Products) domain on FAOSTAT. It encompasses 245 nations and territories and runs from 1961 to the most recent year. The bulk download is available as a ZIP file that includes optional metadata files like definitions, flags, and symbols in addition to the main CSV file with the dataset.

Access Page: FAOSTAT

3. NASA POWER Daily Weather Dataset

Using the POWER Daily API (v2.8.0), the NASA POWER Daily Weather dataset (JSON) was downloaded from the NASA Langley Research Center's POWER Project. From 2015 to 2025, it offers daily temperature (T2M) and precipitation (PRECTOTCORR) observations for Mauritius (latitude −20.3, longitude 57.5).

API Page: NASA POWER | API Pages

Data Processing Workflow - ETL: Extraction, Transformation, Loading

To ensure data consistency, quality, and readiness for SQL import and analysis, a structured ETL (Extract, Transform, Load) framework was applied throughout the data integration and preparation process. After the acquisition of raw data from NASA POWER, FAOSTAT Crops, and CEB Electricity, the ETL procedure was used to get the datasets ready for integrated analysis. Prior to being cleaned, standardised, and prepared (e.g., harmonising units, aligning dates, aggregating daily weather), raw data was first taken from the original sources in their original formats. The processed datasets were then loaded as schema-ready CSVs for smooth integration into a central environment.

For reproducibility, quicker loading, and a well-organised workflow, raw files, staged datasets, and visual outputs were stored in specific local folders (/data_raw, /data_stage, /report/plots).

Table 1: Data processing workflow

Stage	Process	Description / Actions	Tools / Output
1. Extract (E)	**Data Sources**	**- FAOSTAT: Crops & Livestock (annual) - data.gov,mu: Electricity Generation (annual, by source) - NASA POWER: Daily weather data (temperature, precipitation)**	**CSV files**
	Filtering	**Data extracted for Mauritius (2015–2022)**	**CSV file**
	Data Import	**- `fread()` used for CSV imports (efficient tabular reading) - `fromJSON()` used for NASA JSON files**	**R (data.table, jsonlite)**
	Storage & Standardization	**- Raw data stored in `data_raw/` directory - Variable types standardized (integer, numeric, character)**	**Clean raw data ready for preprocessing**
2. Transform (T)	**Data Cleaning – Crops**	**- Removed structural missing values (`flag == "M"`) - Converted production values to numeric**	**Clean crop dataset**
	Data Cleaning – Electricity	**- Imputed one missing *Wind* value using median (MAR assumption) - Added imputation flag for transparency**	**Clean electricity dataset**
	Data Cleaning – Weather	**- Converted JSON date strings to Date objects - Ensured numeric format for `avg_temp` and `total_precip`**	**Clean weather dataset**

	Data Harmonization	**- Created time dimension (dim_date): daily, monthly, yearly - Extracted dim_item (unique crops/items) - Extracted dim_element (measurement types + units) - Aggregated NASA data to monthly/yearly frequency**	**Harmonized dataset ready for export**
	CSV Export	**Exported schema-ready CSVs to `data_stage/`:** • `dim_date.csv` • `dim_item.csv` • `dim_element.csv` • `crop_production.csv` • `weather.csv` • `electricity_generation.csv`	**Cleaned and standardized CSV files**
	Validation	**- Verified row counts vs. source files - Compared numeric values and flags - Documented missing values - Checked summary statistics (min, max, totals)**	**Quality-checked staged data**
3. Load (L)	**Database Tables**	**- Dimension Tables:** `dim_date`, `dim_item`, `dim_element` **- Fact Tables:** `crop_production`, `weather`, `electricity_generation`	**MySQL database (Annex 1)**
	Relationships	- `date_id` → **links fact tables to** `dim_date` - `item_code` → **links crops to** `dim_item` - `element_code` → **links crops to** `dim_element`	**Relational schema**
	Import Method	**Manual or scripted CSV import into MySQL**	**SQL scripts / Workbench**
	Post-Load Validation	**- Row counts and totals matched CSVs - Imputation flags retained for traceability**	**Verified SQL database ready for analysis**

1. Extraction

During extraction phase of the process, the raw datasets were retrieved from their official sources and loaded in R, keeping a raw dataset in the directory /data_raw directory. Storage of the datasets in raw stage ensured that reproducibility, faster loading in later analyses and avoided repeated downloads from external sources.

```
current_r_dir <- "C:/Users/vikgh/Desktop/STAT5129_GroupD_Project/R"
base_dir <- dirname(current_r_dir)
data_raw_dir <- file.path(base_dir, "data_raw")
dir.create(data_raw_dir, recursive = TRUE, showWarnings = FALSE)
```

Figure 2: Extraction

In order to ensure that storage, reproducibility and efficient management of the dataset throughout the ETL process, specific local folders were created at different level of data process.

1.1 FAOSTAT Crops Dataset

The Crops and Livestock (QCL) dataset was first checked in the local */data_raw* folder as a ZIP file. The CSV was extracted into a temporary folder and loaded into R using *data.table::fread()*. In case the file was locally unavailable, it was downloaded via the FAOSTAT API and saved to */data_raw*. This method ensures that the dataset once downloaded can be used in different sessions without requiring download each time.

```
if (file.exists(livestock_zip)) {
  unzip(livestock_zip, files = livestock_csv_name, exdir = tempdir())
  crops <- fread(file.path(tempdir(), livestock_csv_name))
} else {
  crops <- get_faostat_bulk(code = "QCL", data_folder = data_raw_dir)
}
```

Figure 3: FAOSTAT Crops Dataset

1.2 CEB Electricity Dataset

The electricity dataset file was stored in */data_raw* as a CSV format. Each time the script runs, it first checks the existence of the file and in case of unavailability, it is downloaded from data.govmu. To enable quick and efficient access while maintaining a permanent local copy for reproducibility, the file is loaded into R using *data.table::fread() function.*

```
if (!file.exists(elec_file)) {
  download.file(elec_url, destfile = elec_file, mode = "wb")
}
electricity <- fread(elec_file)
```

Figure 4: CEB Electricity Dataset

1.2 NASA POWER Daily Weather Dataset

The NASA POER API which provided data in JSON format was used to download daily weather data and same was stored in the folder */data.raw.* Each time the script is run, checks are done for the existence of the local JSON file and same is loaded directly using the *jsonlite:fromJSON()* function in R. In case of nonexistence of the file, the script retrieves the dataset from the API directly. The relevant variables are extracted and converted into *data.table* for further processing.

```r
if (!file.exists(nasa_file)) {
  nasa_raw <- fromJSON(nasa_api_url)
  write_json(nasa_raw, nasa_file)
} else {
  nasa_raw <- fromJSON(nasa_file, flatten = TRUE)
}
nasa_data <- data.table(
  date = as.Date(names(nasa_raw$properties$parameter$T2M), "%Y%m%d"),
  avg_temp = as.numeric(nasa_raw$properties$parameter$T2M),
  total_precip = as.numeric(nasa_raw$properties$parameter$PRECTOT)
)
```

Figure 5: NASA POWER Daily Weather Dataset

In order to facilitate quick upload in R for reproducible and effective analysis the workflow makes sure that all raw datasets are centrally stores locally. This minimizes repeated downloads.

2. Transformation

The period coverage, level of detail and variable structure of the three datasets are different. The weather data was recorded on a daily basis from 2015 to 2025, while the electricity data comprises of yearly production from 2010 to 2022 and finally the crops data included yearly observations from 1961 to 2023. During initial analysis missing values, inconsistent units, structural missing flags and non-standard text fields were discovered.

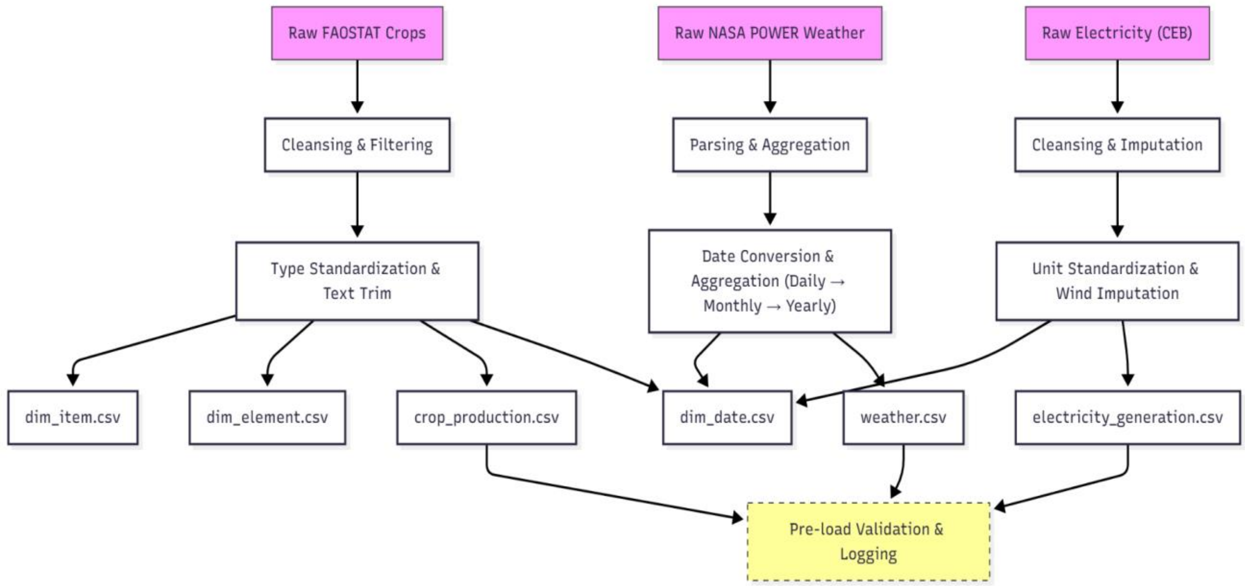

Figure 6: Transformation

The ETL workflow process is depicted by the Figure 6. During the transformation process the datasets are cleaned and harmonized by:

- Filtering for data pertaining to Island of Mauritius only for years 2015 to 2022;
- Eliminating structural missing values and imputing missing entries with explicit flags;
- Standardizing types and units;
- Aggregating daily weather data to monthly and then to yearly levels;
- And lastly generating dimension tables (dim_date, dim_item, dim_element)

To produce schema-ready, validated CSV files where surrogate keys and consistent identifiers enable cross- dataset integration and traceable, repeated analyses.

2.1 Data Filtering and Temporal Harmonisation

In order to facilitate consistent cross-data analysis, it was ensured that all datasets are for Mauritius only and that they have a consistent period coverage. The following filtering is carried out to achieve same:

FAOSTAT Crops

Filtered to include only Mauritius: area_code == 137
Time period restricted: year >= 2015 & year <= 2022

CEB Electricity

Filtered for Mauritius: Island == "Island of Mauritius"

Time period restricted: year >= 2015 & year <= 2022

NASA POWER Weather

Data already corresponds to Mauritius coordinates; no geographic filter needed

Daily records filtered: date >= 2015-01-01 & date <= 2022-12-31

```r
# =============================
# 3. Filter Datasets for Mauritius (2015-2022)
# =============================
crops_mauritius <- crops %>% filter(area_code == 137 & year >= 2015 & year <= 2022)
electricity_mauritius <- electricity %>% filter(Island == "Island of Mauritius"
                                                & Year >= 2015 & Year <= 2022)
nasa_data <- nasa_data[year(date) >= 2015 & year(date) <= 2022]
```

Figure 7: Data filtering

The figure above demonstrates the codes executed in R.

2.2 Drop Redundant or Irrelevant Columns

Redundant columns were eliminated in order to simplify the datasets and cut down on unnecessary storage. Measures taken include:

- Duplicate fields were removed for the FAOSTAT Crops Dataset;
- Location/ area fields were removed from the FAOSTAT Crops Dataset and CEB Electricity Dataset.

2.3 Data Hygiene: Cleaning and Standardisation

Prior to analysis, in order to ensure consistent formatting and correct datatypes, cleaning of datasets was done. Key operations included:

Trimming whitespace from textual fields such as item, element, and Island

Type correction: numeric and date fields were verified and coerced if needed

Duplicate removal: exact duplicates based on date, item_code, and element_code were deleted

```r
# ============================
# 4. Standardize Data Types
# ============================
crops_mauritius <- crops_mauritius %>%
  mutate(across(c(area_code, item_code, element_code, year_code, year), as.integer),
         across(c(area_code_m49, area, item_code_cpc, item, element, unit, flag, note), as.character),
         value = as.numeric(value))

electricity_mauritius <- electricity_mauritius %>%
  mutate(Year = as.integer(Year),
         Island = as.character(Island),
         across(c(`Landfill gas`, Photovoltaic, Wind, Coal, Bagasse), as.numeric))

nasa_data <- nasa_data %>% mutate(avg_temp = as.numeric(avg_temp), total_precip = as.numeric(total_precip))
```

Figure 8: Data cleaning and standardisation

These steps prepare the datasets for reliable transformations and database loading. The figure above illustrates these cleaning operations in R.

2.4 Handle Missing Values

In order to maintain completeness without losing transparency the problem of missing values was addressed as follows:

FAOSTAT Crops: rows with structural missing values (flag == "M") were removed as it is Missing Completely At Random (MCAR) as indicated by the Flag Description on its official page (M means Missing value, data cannot exist, not applicable)

CEB Electricity: Little's MCAR test (p = 0.391) indicated missingness is likely at random. The missing Wind values were imputed using the median as it is more robust and preserves central tendency.

```r
crops_mauritius <- crops_mauritius %>% filter(flag != "M")

electricity_mauritius <- electricity_mauritius %>%
  mutate(
    wind_imp_flag = as.integer(is.na(Wind)),
    Wind = ifelse(is.na(Wind), median(Wind, na.rm=TRUE), Wind)
  )
```

Figure 9: Handling missing values

The figure above demonstrates how missing values were managed in R.

2.5 Data Aggregation: Monthly and Yearly

The weather dataset initially contained daily observations. In order to ensure consistent cross-dataset comparisons and reliable analysis, same were aggregated to monthly level in the first instance and then to yearly level to match the chronological dimension of the crops and electricity datasets. Below are the steps for the aggregation:

```r
# -----------------------------
# Daily → Monthly Aggregation
# -----------------------------
weather_monthly <- weather_csv %>%
  mutate(year = year(period_start),
         month = month(period_start)) %>%
  group_by(year, month) %>%
  summarise(
    avg_temp = mean(avg_temp, na.rm = TRUE),
    total_precip = sum(total_precip, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  mutate(
    aggregation_level = "monthly",
    period_start = as.Date(paste(year, month, "01", sep = "-"))
  )
```

Figure 10: Data aggregation

Aggregation Steps: Daily → Monthly

Grouped by year and month.

Calculated monthly averages for temperature (T2M) and monthly totals for precipitation (PRECTOTCORR).

Stored results in the weather table with aggregation_level = "monthly".

For aggregation: Daily → Yearly, the aggregation_level = “yearly and period_start = as.Date(Paste0(year, “-01-01”))

Aggregating daily data provides a consistent temporal reference (dim_date) to link weather with crops and electricity datasets. The period_start column represents the first day of the aggregation period (month or year), ensuring traceability and flexibility for cross-dataset analysis.

2.6 Dimension Table Construction

To normalize and make the datasets ready for database upload dimension tables are created.

Each table organizes repeated or reference information into a structured format, reducing redundancy and improving consistency across datasets.

dim_date.csv

Stores all types of dates (daily, monthly, yearly) in a single table.

Introduces period_start and aggregation_level to allow flexibility in linking dates to all datasets.

date_id	period_start	aggregation_level
1	01/01/2015	daily
2	01/01/2015	monthly
3	01/01/2015	yearly
4	02/01/2015	daily
5	03/01/2015	daily
6	04/01/2015	daily
7	05/01/2015	daily
8	06/01/2015	daily

Figure 11: dim_date table

Figure 11 above depicts how dim_date table stores dates for all datasets. This enables consistent joins between weather, crops, and electricity data despite different temporal granularities. dim_element.csv

Each element has a unique primary key (element_code).

Associated information such as element name and unit of measure is stored alongside the code as shown in the diagram below:

element_code	element	unit
5111	Stocks	An
5112	Stocks	1000 An
5312	Area harvested	ha
5313	Laying	1000 An
5318	Milk Animals	An
5320	Producing Animals/Slaughtered	An
5321	Producing Animals/Slaughtered	1000 An
5412	Yield	kg/ha
5413	Yield	No/An
5417	Yield/Carcass Weight	kg/An
5424	Yield/Carcass Weight	g/An
5510	Production	t
5513	Production	1000 No

Figure 12: dim_element

This ensures standardized identification of elements across datasets.

dim_item.csv

Each agricultural item has a unique code and name.

Provides a reference for the crop_production fact table, avoiding repeated text and ensuring consistency.

```
# dim_date
dim_date <- bind_rows(dim_date_yearly, dim_date_monthly, dim_date_daily) %>%
  distinct() %>%
  arrange(period_start, aggregation_level) %>%
  mutate(date_id = row_number()) %>%
  select(date_id, everything())

# dim_item
dim_item <- crops_mauritius %>%
  distinct(item_code, item) %>%
  mutate(item = substr(item, 1, 200))

# dim_element
dim_element <- crops_mauritius %>%
  select(element_code, element, unit) %>%
  group_by(element_code, element) %>%
  summarise(unit = if (length(na.omit(unit)) > 0) na.omit(unit)[1] else NA_character_, .groups = "drop") %>%
  mutate(element = substr(element, 1, 100), unit = substr(unit, 1, 50)) %>%
  arrange(element_code)
```

Figure 13: dim_item

Figure 13 – an extract of the R codes shows how pipelining has been used to build data frames ready for exportation to CSV files. Together, these dimension tables provide a structured, normalized framework for the fact tables, enabling accurate linking, efficient storage, and reliable analysis.

2.7 Validation & Quality Assurance

In order to ensure integrity and consistency with the source data after transformation all CSV files were reloaded. Checks performed include:

Row counts: confirm exported vs original dataset counts

Primary key uniqueness: ensure composite keys are unique

Referential integrity: verify all fact rows link to valid dimension entries

Numeric consistency: compare totals, ranges, and aggregates

Missing value audit: detect unintended NA generation

Join validation: confirm all facts correctly map to dimensions

These steps confirmed the transformed data matched the source in structure and content, ensuring readiness for database loading.

3. Loading

Importing of the cleaned and standardized datasets into a MySQL database is part of the loading stage. Due to this step efficient querying, cross-dataset analysis, and further statistical or modeling operations, providing a unified dataset for studying the relationships between energy production, agricultural outputs, and climate variables is enabled.

3.1 Loading Procedure

Import cleaned CSVs into the working environment or database
Ensure CSVs follow schema, with surrogate keys and correct column types

3.2 Validation After Loading

Each CSV was reloaded and compared to the in-memory dataset to ensure:

Row counts match original cleaned data

All element units match

Numeric totals and value ranges are preserved

Missing flags are consistent

3.3 Observations from Loading Validation

Crops: 1878 rows loaded → matches cleaned dataset

Electricity: All 8 rows loaded correctly

Weather: 3026 rows loaded; aggregation counts preserved, with 104 daily rows flagged for mapping review
No unit or value mismatches detected in any CSV

Database

The objective of the Mauritius Data Database is to store crop production, electricity generation, and weather data in a structured and standardized manner. Relational database constraints preserve data integrity while facilitating effective querying, aggregation, and analysis.

Entity Relationship Diagram (ERD)

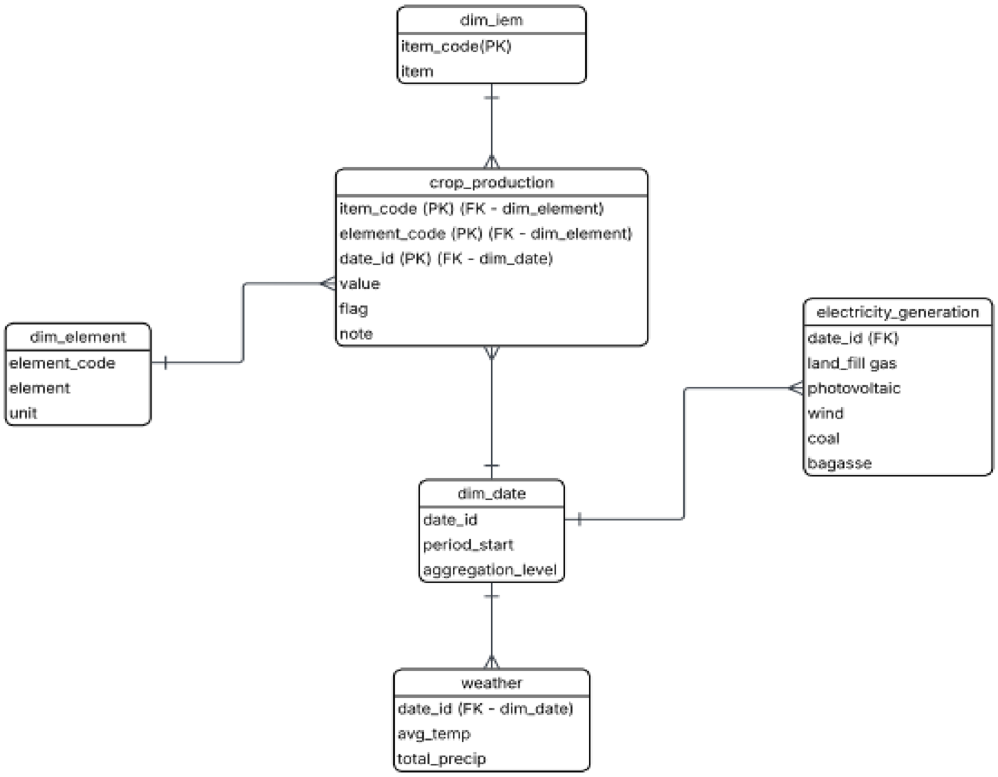

Figure 14: ERD

The ERD above shows the mauritius_data database structure, illustrating how tables are linked via foreign keys, enabling integrated and normalized data analysis.

Data Dictionary

This section provides a comprehensive reference of all tables and fields used in the database. It defines each data element, its type, purpose, and relationships, helping ensure clarity and consistency for analysis and reporting.

Table 2: dim_date table

Field	Type	Description	Constraints	Example
date_id	int	Surrogate key for each period row	PK, NOT NULL	1024
period_start	date	Start date of the period (day / first day of month / first day of year)	UNIQUE with aggregation_level	01/01/2015
aggregation_level	varchar(20)	Period grain: daily, monthly, yearly	Part of unique key with period_start	monthly

Table 3: dim_item table

Field	Type	Description	Constraints	Example
item_code	int	Numeric code for agricultural item	PK, NOT NULL	572
item	varchar(200)	Item name / label		Avocados

Table 4: dim_element table

Field	Type	Description	Constraints	Example
element_code	int	Numeric code for measured element	PK, NOT NULL	5412
element	varchar(100)	Element name / description		Yield
unit	varchar(50)	Unit of measure for element		kg/ha

Table 5: crop_production table

Field	Type	Description	Constraints	Example
item_code	int	FK → dim_item.item_code	Part of PK, NOT NULL	572
element_code	int	FK → dim_element.element_code	Part of PK, NOT NULL	5412
date_id	int	FK → dim_date.date_id (yearly row expected)	Part of PK, NOT NULL	2980
value	decimal(15,2)	Measured value (production, yield, area, etc.)		4009232
flag	varchar(5)	FAOSTAT flag (e.g., M for missing)		M
note	text	Free text note / provenance		NA

Table 6: weather table

Field	Type	Description	Constraints	Example
date_id	int	FK → dim_date.date_id (daily / monthly / yearly)	FK	1024
avg_temp	decimal(7,2)	Average temperature (°C) for the period		25.04
total_precip	decimal(10,2)	Total precipitation (mm) for the period		12.5

Table 7: electricity_generation table

Field	Type	Description	Constraints	Example
date_id	int	FK → dim_date.date_id (yearly row expected)	FK	2980
landfill_gas	decimal(10,2)	Electricity from landfill gas (GWh)		159.4
photovoltaic	decimal(10,2)	PV generation (GWh)		649.9
wind	decimal(10,2)	Wind generation (GWh)		106
coal	decimal(10,2)	Coal generation (GWh)		8571.6
bagasse	decimal(10,2)	Bagasse generation (GWh)		2441.1

Indexes

Indexes are created on key columns (such as date_id, aggregation_level and period_start) to improve query performance and speed up data retrieval, especially for large datasets. They ensure that joins between fact and dimension tables are efficient.

Views

Views are defined to simplify complex queries by pre-joining multiple tables and aggregating data where necessary. This allows easy reporting and analysis without repeatedly writing complex joins.

Table 8: Views

View Name	Purpose
master_crop_data	Combines crops, elements, dates, electricity, and weather data for reporting.
master_crop_env	Aggregates yearly crop values along with electricity and weather statistics.
master_crop_yearly	Aggregates yearly crop data (total, average, count) for analysis.
master_electricity_data	Combines electricity generation with weather data for analysis.
weather_monthly_summary	Aggregates daily weather data to monthly averages and totals.
weather_yearly_summary	Aggregates daily weather data to yearly averages and totals.

The Risk - Mitigation Report

During research process in order to preserve data integrity and efficient project governance potential risks were identified and evaluated. The main technical, moral, and data-related issues that might having an impact on the project's results are listed in the Table 9, along with the corresponding tactics used to lessen those effects. Transparency, dependability, and ethical compliance in data handling, analysis, and reporting are supported by this proactive approach.

Table 9: Risk Mitigation

Risk	Type	Mitigation
Accidental deletion of datasets from folder /data_raw	Operational / Human error	Implement automated dataset download scripts.
Data loss when exporting cleaned tables due to a bug or runtime error	Technical / Software	Verify all exports by tallying rows for CSVs and R objects.
Accidental changes for foreign keys in database	Operational / Data integrity	Add database constraints to prevent direct edits.
Misinterpretations when reading plots	Analytical / Communication	Include clear legends, axis labels, and interpretation notes.
Misinterpretations due to missing values or outliers	Analytical / Data quality	Implement consistent missing value handling (e.g., imputation or removal) and outlier detection. Document decisions and assumptions.
Loss of intermediate results during analysis	Operational / Technical	Save intermediate outputs to /report/plots.
Accidental deletion of directories: /report/plots or /data_raw or /data_stage	Operational / Human error	Create a directory if missing.
Misalignment of data sources (e.g., inconsistent timestamps, units)	Data quality / Analytical	Implement data validation checks. Standardize formats, units, and time zones. Document source-specific transformations.

Importance of Governance Practice and Privacy ethics

In order to assure that data is handled sensibly, morally and in accordance with legal terms, effective governance and privacy management are crucial. Table 10 and 11 show the steps taken to ensure data security, confidentiality and quality.

Table 10: Governance Actions

Area	Governance Practice
Data Access	Only open-access, non-confidential datasets are used
Data Documentation	Each dataset includes metadata (source, date, units).
Data Integrity	SQL constraints (primary keys, data type checks) ensure data consistency.
Auditability	All cleaning and transformation steps are logged in R scripts and stored securely.
Reproducibility	Analyses can be fully reproduced using documented code pipelines

Table 11: Data Privacy and Ethical Considerations

Ethical Aspect	Description and Action
Data Sensitivity	No personally identifiable information (PII) is used - datasets are aggregated at the national level.
Transparency	All data sources (FAOSTAT, NASA POWER, CEB Mauritius) are clearly cited.
Responsible Use	Data is used only for research and educational purposes - not for commercial exploitation.
Data Accuracy	Cross-verification between sources is performed to reduce misrepresentation.

Note – Compliance and Standards:

By making sure that no personal or identifiable data is collected or shared this project complies with the Mauritius Data Protection Act (2017).Furthermore, it conforms with FAO and NASA Data Use Policies, allowing research and academic analysis with appropriate attribution.

Data Processing and outputs

Table 12: ETL Outputs Summary

Dataset / Source	Stage	Key Actions Performed	Rows Before	Rows After	Missing Value Handling	Final Output File	Remarks / Validation
FAOSTAT – Crops & Livestock	Transform	Removed structural missing rows (`flag == "M"`); standardized numeric types	1,200	1,180	20 rows removed	`crop_production.csv`	Row counts and totals verified
Electricity Generation	Transform	Cleaned column names; imputed 1 missing *Wind* value using median; added imputation flag	96	96	1 value imputed	`electricity_generation.csv`	Checked consistency across energy types
NASA POWER – Weather Data	Transform	Parsed JSON; converted dates; aggregated to monthly/yearly; ensured numeric fields	2,920	96	Aggregation to monthly level	`weather.csv`	Validation of date and aggregation accuracy
All Datasets	Load	Imported cleaned data into MySQL (fact & dimension tables)	—	—	—	Database tables: `dim_date`, `dim_item`, `dim_element`, `crop_production`, `weather`, `electricity_generation`	Row counts and key integrity verified

Table 13: Data Quality and Validation Summary

Quality Dimension	Description	Validation Method	Outcome / Status
Completeness	**Ensuring no critical data fields are missing or incomplete.**	**Checked for missing and null values across all key columns.**	**Missing values identified and handled; imputed values flagged.**
Accuracy	**Verifying correctness of values after cleaning and transformations.**	**Cross-checked computed aggregates with original datasets.**	**All summary totals matched expected ranges.**
Consistency	**Uniform data types and naming conventions across datasets.**	**Confirmed consistent variable names, formats, and units.**	**All datasets standardized to common schema.**
Validity	**Ensuring data adheres to defined constraints and logic.**	**Applied validation rules (e.g., non-negative values for production).**	**All invalid entries removed or corrected.**
Integrity	**Verifying relational consistency between dimension and fact tables.**	**Checked foreign key references (e.g., `date_id`, `item_code`).**	**No orphan or duplicate records found.**
Timeliness	**Ensuring data covers the correct time frame (2015–2022).**	**Verified timestamps and aggregation periods.**	**Data correctly aligned with study period.**

Data Analysis and Modelling

Model: Bagasse Model

Table 14: Summary of Bagasse Model outputs

Model	Dependent Variable	Independent Variables	R^2	Adj. R^2	p-value	Significant Predictors ($p < 0.05$)	Key Findings
Model Stage 1	Sugarcane Yield (scaled)	Avg. Temp (scaled), Precipitation (scaled)	0.637	0.492	0.079	None (not significant at 5%)	Climate variables show moderate influence on yield; model not statistically significant overall.
Model Stage 1b	Sugarcane Production (scaled)	Avg. Temp, Precipitation, Area Harvested (all scaled)	0.972	0.951	0.0014	Area Harvested ($p = 0.00097$)	Strong model fit; area harvested is the main driver of sugarcane production.
Model Stage 2	Bagasse Output (scaled)	Year (polynomial), Sugarcane Yield (scaled)	0.97	0.947	0.0017	$Year^2$ ($p = 0.0086$)	Bagasse output is highly correlated with time trend and yield; strong predictive accuracy.

Interpretation for bagasse model

- Models Stage 1b and Stage 2 demonstrate high explanatory power as indicated by the Adjusted $R^2 > 0.94$, making them a strong model fit.
- The Stage 1b identifies the area harvested is identified as the dominant factor in sugarcane production.
- Stage 2 demonstrate that bagasse output follows a significant non-linear time trend, reflecting production evolution over the years.
- Despite being a weaker model the Stage 1 indicates preliminary climate impact on yield. However, the model lacks strong statistical significance.

Model: The photovoltaic model

Table 15: Summary of Photovoltaic Model outputs

Model	Dependent Variable	Independent Variables	R^2	Adj. R^2	p-value	Significant Predictors ($p < 0.05$)	Key Findings
Model Stage 1	Photovoltaic Generation (GWh)	Mean Temperature, Total Precipitation, Hot Days, Max Consecutive Dry Days	0.259	-0.730	0.886	None (not significant at 5%)	Climate variables show no influence on PV generation; model not statistically significant overall.
Model Stage 2	Photovoltaic Generation (GWh)	Mean Temperature, Total Precipitation	0.119	-0.234	0.730	None (not significant at 5%	Simplified model with only temperature and precipitation still shows no significant relationship with PV generation.

Interpretation

- Both models fail to explain photovoltaic generation variation - R^2 values are very low (0.259 and 0.119).
- Negative adjusted R^2 values (-0.730 and -0.234) indicate models perform worse than simply using the mean.
- No statistical significance - Both p-values (0.886 and 0.730) are much greater than 0.05.
- Model 2 performs slightly better despite having fewer variables (higher adjusted R^2 and lower residual error)
- Individual predictors:
    - ➢ Total precipitation shows a marginally significant positive effect ($p = 0.065$), suggesting that higher rainfall may slightly increase photovoltaic output.
    - ➢ Mean temperature, hot days, and max CDD are not statistically significant, indicating weak or inconsistent relationships with photovoltaic output in this dataset.

Limitations while modeling the dataset

Since the study was conducted using a relatively small dataset (less than 10 observations), there are numerous methodological limitations. First, the regression model's statistical strength is low due to the small sample size, making it challenging to find or validate meaningful connections between temperature, precipitation, and photovoltaic energy production. A single outlier observation can significantly affect model coefficients and overall outcomes because small datasets are also very sensitive to outliers.

Furthermore, parameter estimates may be unstable and not generalizable to larger time periods or various climatic circumstances when there are few data points.

Additionally, due to insufficient data meaningful validation of the models using cross-validation or training-testing splits could not be done since it lowers the results' capacity for prediction. The estimated impacts in such a limited dataset may be further distorted by multicollinearity amongst climatic variables. As such, the results should be regarded as experimental rather than conclusive.

Purpose of the Link Analysis to Policy Table

A Link to Policy Table (Table 15) is given below to demonstrate how the study's conclusions align with, support, or inform existing national and international policies related to climate change, energy, and agriculture.

Table 15: Link Analysis to Policy Table

Focus area	Related National / SDG Goals
Electricity Generation	**SDG 7: Affordable and Clean Energy -** Aiming to increase the share of renewable energy to 60% by 2030, phasing out coal, and increasing energy efficiency by 10%.
Crop Production	**SDG 2: Zero Hunger -** It supports boosting production for food security and promoting sustainable and resilient production. **SDG 12: Responsible Consumption and Production -** It addresses the need for sustainable production and consumption patterns, which is crucial for an industry like sugarcane that has a significant environmental impact and economic importance. **SDG 8: Decent Work and Economic Growth -** It ensures economic growth and employment in the agricultural sector through initiatives like improving productivity, agro-processing, and creating value across the supply chain. **SDG 13: Climate Action -** Sugarcane production must become more resilient to extreme weather events like droughts and floods. **SDG 15: Life on Land -** It ensures that sugarcane farming practices are implemented in a way that maintains the health of ecosystems and protects biodiversity. **SDG 1: No Poverty -** It contributes to poverty alleviation in rural communities, especially those that depend on sugarcane farming for their livelihoods.
Climate Change	**SDG 13: Climate Action -** It targets at strengthening resilience and adaptive capacity towards climate change-related disasters and also to integrate climate change measures into policies and planning. **SDG 2: Zero Hunger -** Changes in precipitation could have economy-wide impacts, affecting food availability and stability. **SDG 11: Sustainable Cities and Communities -** This goal is relevant due to the increasing frequency of flash floods caused by intense rainfall, which are exacerbated by rapid urbanization. The goal's targets for making cities resilient are directly applicable to mitigating the damage from such events. **SDG 15: Life on Land -** This goal is pertinent because changing climate conditions, such as increased drought and intense rainfall, can lead to land degradation, which is a direct threat to terrestrial ecosystems. Efforts to combat desertification and restore degraded land are critical for maintaining the health of the land in Mauritius.

Conclusion and Recommendations

The Study showed a strong correlation between Climate Conditions, agricultural productivity and generation of electricity from renewable resources. Although temperature and precipitation have moderate impact on sugarcane yield, the area harvested continues to be the most significant driver of total sugarcane production, according to the results of the regression models. Furthermore, the bagasse model's excellent performance points a non-linear temporal growth pattern in bagasse-based electricity generation that is strongly correlated with agricultural output trends.

Overall, these results highlight how crucial it is to incorporate climate data into national energy planning. It is recommended that policymakers and industry stakeholders:

- Strengthen climate-resilient agricultural practices to maintain stable feedstock supply for renewable energy generation.
- To continuously evaluate the effects of climate variability, invest in data-driven monitoring systems.
- To increase the production of renewable energy, promote the sustainable growth of sugarcane cultivation and the effective use of by-products like bagasse.
- The goals of energy security and climate adaptation for Mauritius are both supported by this integrated approach.

Bibliography

Purmessur, S., 2023. *National Biomass Framework, Potential Sources & Recommendations,* Mauritius: Mauritius Cane Industry Authority.

Sultan, R., 2021. *Economic impacts of climate change on agriculture: insights from the small island economy of Mauritius. Small Island Developing States: Vulnerability and Resilience Under Climate Change.* pp. 137-158.

United & Nations, 2024. *United Nations, Sustainable Development Goals in Mauritius.* [Online] Available at: https://mauritius.un.org/en/sdgs [Accessed 1 October 2025].