**FACULTY OF COMPUTING**

# INTERNSHIP REPORT

January 2024

JULY 29, 2024

STUDENT NAME: P.W. MADDUMAGE

STUDENT NUMBER: IT21007538

# TABLE OF CONTENTS

# 1 INTRODUCTION

## 1.1 Company Overview

Ever since the 1880s, Ceylon Tea became the country's principal and most famous export, forming the backbone of the Sri Lankan economy for many generations. As the industry grew in leaps and bounds, the need to regulate the functions of growing, processing and manufacturing of Tea also arose.

Then, in 1976 the Sri Lanka Tea Board was set-up as the apex regulatory and administrative body of the Sri Lankan Tea industry comprising representatives from both private and government bodies involved in the industry, including cultivators and manufacturers, traders, exporters and plantation employees.

It has the control over all matters connected to the tea value chain starting from tea land management to export, along with all regulatory functions to improve and develop the Sri Lankan tea industry, promotion of Sri Lanka Tea, managing the "Ceylon tea" brand and overall upliftment of Tea Industry and all its stakeholders locally and internationally. Furthermore, they are the custodians of all the information and data related to all parts of Sri Lankan tea industry.

The Tea Board is the apex regulatory and administrative body of the Sri Lankan tea industry carrying out a wide variety of functions with respect to the industry.
This includes, Offering advice and assistance to overseas tea buyers and tea brand owners, Offering advice and assistance to tea exporters, traders, manufacturers and cultivators, Defining, protecting and promoting the 'Ceylon Tea' brand, Defining, protecting and certifying the regional origins of Ceylon Tea, Monitoring and controlling the quality and purity of tea exported from Sri Lanka, Promoting the sale and consumption of Ceylon tea worldwide and Compiling and circulating market data and other information about the Sri Lanka tea industry.

The functions of the Tea Board are carried out through six divisions. These include, The Head Office, The Tea Commissioner's Division, The Promotions Division, The Analytical Laboratory, The Export Division, The Tea Tasting Unit, IT & Statistic Division, Administration Division, Finance Division and Internal Audit Division.

The main objective of the IT and Statistical Division of the Sri Lanka Tea Board is to act as the single, convenient point of access to comprehensive, reliable and up-to-date information relevant to anyone involved or interested in the Tea industry in Sri Lanka.

The Division gathers grower, manufacturer and market information through a wide network of sources including the Board's seven regional offices, each representing one of the seven tea-growing 'districts' of Sri Lanka, the Colombo and London Tea Auctions, the Ceylon Chamber of Commerce and, of course, the

other divisions of the Board itself. Routine information from the Regional Offices is updated real time via the Board's internal data network.

The division also exchanges information with the International Tea Committee, of which the Sri Lanka Tea Board is an Associate Member.

Information collected and analyzed by the Information Technology Division is published as a series of monthly, quarterly and annual reports. These are circulated throughout the Sri Lankan Tea industry and are also available online. In addition, the Division maintains the Tea Sector Database containing information on production, exports and sales.

Timely industry and market data are supplemented by a wealth of historical information in the form of old Tea Reports and other documents dating back to more than a century. Much of this information is readily available at the Tea Board Library, housed at the Board's Head Office in Colombo. The library also contains a plethora of books and other publications of technical and. general nature.

The division is also responsible for the maintenance and timely updating of this the Sri Lanka Tea Board websites. (www.pureceylontea.com and www.srilankateaboard.lk)

## 1.2  Project Overview

1. Production Prediction Model

    I worked on a range of projects involving data analysis and visualization throughout my internship.

    The primary objective of the Production Prediction Model project was to develop a robust predictive model capable of forecasting tea production in various districts of Sri Lanka. By leveraging historical data, the model aimed to provide accurate predictions that could help stakeholders in the tea industry make informed decisions regarding production planning, resource allocation, and market strategies. So, I worked on making a prediction model using historical data and explore various regression and time series models to identify the one that provides the best predictive performance.

2. Power BI report creation

    The next task I have assigned to do is the Power BI report creation on various insights as examples for production, exports, sales data etc. to compare information yearly, monthly or annually. The data was collected from database server and the plantations as raw data. I focused on Power BI themes and templates, aiming to create visually appealing dashboards. I conducted exploratory data analysis (EDA) to understand the dataset better. I developed sample dashboards with relevant measures and parameters for database and server interactions. Utilizing SQL Server Management Studio (SSMS), I validated measures against the original database and reconciled dataset discrepancies with company reports. Additionally, I provided suggestions for visualizations and formulated test cases to ensure dashboard accuracy.

    I executed several Power BI demo projects focusing on hotel, sales and Adventure Works datasets, exploring various visualization techniques and storytelling capabilities. And also, I created stories that helpful for dashboard design. I created power BI themes and templates, that was helpful for designing. I have Completed hands-on training using SSIS, SSAS, and SQL DB to gain proficiency in data integration, analysis, and management using Microsoft SQL Server technologies. I did demo projects ETL (Extract, Transform, Load) processes through an SSIS (SQL Server Integration Services), using SSRS deploy interactive reports.

## 1.3 Glossary

Following are the abbreviations and definitions used in this document.

| Abbreviation | Description |
|---|---|
| SQL | Structured Query Language |
| DB | Database |
| SSIS | SQL Server Integration Services |
| SLTB | Sri Lanka Tea Board |
| KPI | Key Performance Indicator |
| ML | Machine Learning |
| EDA | Exploratory Data Analysis |
| SSMS | SQL Server Management Studio |
| SSAS | SQL Server Analysis Services |
| ETL | Extract, Transform, Load |

# 2 INTERNSHIP INSIGHT

## 2.1 Objectives

The objectives of the internship were multifaceted, aiming to provide a comprehensive learning experience that combined technical skills with industry-specific knowledge. The key objectives included are gaining practical experience as developing hands-on experience in data analysis, predictive modelling, and data visualization, Understanding the Tea Industry as acquiring in-depth knowledge of the tea industry, including its production processes, market dynamics, and regulatory environment, developing Technical Skills as enhancing proficiency in Python for data analysis and modelling, as well as Power BI for creating interactive reports and contributing to the Organization as delivering tangible outcomes, such as predictive models and insightful reports, that could benefit the Sri Lanka Tea Board and its stakeholders.

Furthermore, I aimed to improve soft skills such as communication skills, collaboration and team work which aligned with the company's aim of introducing me to a work culture of team work and collaboration. I was given the opportunity to network with like-minded team members share ideas and broaden my horizons, which helped me to achieve the objectives I had for networking with colleagues.

The internship provided ample opportunities for self-reflection and personal growth. I was encouraged to identify my strengths and weaknesses and, through targeted guidance, learn how to leverage my strengths while addressing areas for improvement

## 2.2 Procedures

- Working hours:

  I was required to work from 8:30 a.m. to 4:30 p.m. during my internship. Sometimes, though, flexibility was necessary, and we concentrated on finishing jobs rather than rigorously following deadlines

- Stand-Up:

  The Daily Stand-up Meetings, which are held every morning at 9:00, are mandatory for our data analytics team to attend. Team members discussed daily duties at these meetings, which ensured alignment and open communication throughout the team.

- Meeting for Daily Review:

  Every day, we convened in the evening to go over the work we finished throughout the day and the difficulties we have encountered.

- Services and Equipment granted

  Throughout my internship, the company provided me with a laptop, mouse, monitor, and headphones. They gave me an allowance for internship. The configuration of our machines included a Virtual Private network (VPN), which enables us to access the business's network from any location in the globe. Gmail was used to schedule meetings and send emails. The primary uses of it includes the for the purpose of communication.

- Experience with data analysis

  Data Collection and Preprocessing: Extracting data from various sources, cleaning it, and preparing it for analysis. This involved working closely with the IT & Statistical Division and other departments within the Tea Board.

  Model Development: Building and refining the Production Prediction Model using Python. This involved iterative cycles of model training, evaluation, and optimization.

  Report Creation: Designing and developing Power BI reports, starting with data modeling and followed by the creation of visualizations and interactive features.

  Collaboration: Regularly collaborating with supervisors and other team members to ensure that the projects aligned with the organization's goals. This included attending meetings, presenting progress updates, and incorporating feedback.

## 2.3  Methodology

The methodologies employed during the internship were guided by best practices in data science and business intelligence. Key methodologies included:

**Data Science Methodology**

- **CRISP-DM (Cross-Industry Standard Process for Data Mining):**

  This methodology was followed for the Production Prediction Model project. It involves six phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. This structured approach ensured that the model development process was systematic and aligned with the project objectives.

- **Model Evaluation Metrics:**

  The performance of the models was evaluated using metrics like MAE, MSE, and RMSE. These metrics provided a quantitative measure of the model's accuracy and were used to compare different models.

**Business Intelligence Methodology**

- **ETL (Extract, Transform, Load):**

  This methodology was used for preparing the data for Power BI reports. The ETL process involved extracting data from various sources, transforming it into a suitable format, and loading it into Power BI for analysis and visualization.

- **Data Modeling:**

  A crucial step in the report creation process, data modeling involved defining relationships between different data entities to ensure that the visualizations were accurate and meaningful.

- **Interactive Dashboards:**

  The Power BI reports were designed with interactivity in mind, allowing users to explore the data from different angles and gain deeper insights.

  Furthermore, during daily stand-up meetings, I kept my supervisor and team members updated on the project's progress while I worked on it. I also solicited their comments and used appropriate techniques, making adjustments as needed along the way. The group members also discussed their updates on their end of the project, which assisted me in choosing the working techniques on the undertaking. I had access to documents and presentations on the projects I was assigned, and I also had the chance to attend team member knowledge transfer meetings. I used a continual learning approach to the projects and solicited feedback on my tasks that finished in order to make sure I'm doing it correct. Additionally, I took care to enhance my knowledge in the relevant fields to better comprehend and complete my job more quickly. This enabled me to do my work more quickly and to a higher standard.

# 3 LEARNING OUTCOME

## 3.1 Learning

The internship provided a wealth of learning opportunities, both in terms of technical skills and industry knowledge. Key learnings included:

**Technical Skills**
- Data Analysis and Modeling: Gained proficiency in using Python for data analysis and predictive modeling. This included working with libraries such as Pandas, Scikit-learn, and TensorFlow.

- Time Series Analysis: Developed a strong understanding of time series analysis techniques, particularly in the context of forecasting agricultural production.

- Business Intelligence Tools: Gained hands-on experience with Power BI, including data modeling, report creation, and the development of interactive dashboards.

- Data Visualization: Learned best practices for visualizing data in a way that is both informative and aesthetically pleasing, ensuring that complex data is presented clearly and concisely.

**Industry Knowledge**
- Tea Production: Acquired in-depth knowledge of the tea production process, including the factors that influence yield, such as rainfall and seasonality.

- Tea Market Dynamics: Gained insights into the global tea market, including export patterns, pricing trends, and the importance of the Ceylon Tea brand.

- Regulatory Environment: Learned about the regulatory framework governing the tea industry in Sri Lanka, including the role of the Sri Lanka Tea Board in maintaining quality standards and promoting the industry.

## 3.2 Measurable Outcome

The internship led to several measurable outcomes, demonstrating the effectiveness of the efforts undertaken:

1. Creation of Production Prediction Model

The final production prediction model achieved a high level of accuracy, with a low MAE and RMSE, indicating its reliability in forecasting future tea production. The model was successfully implemented and is now being used by the Sri Lanka Tea Board to assist in production planning and decision-making.

2. Creation of Power BI Reports

Developed a series of interactive Power BI reports that provide valuable insights into production, exports, and sales. These reports are now being used by stakeholders within the Sri Lanka Tea Board to monitor industry performance. Received positive feedback from supervisors and stakeholders on the clarity and usefulness of the reports. The insights provided have helped inform strategic decisions within the organization.

Tools and technologies: Python, SQL, Power BI, ETL tools, R, Jupiter Notebook, Git, SPSS, Machine Learning, Data mining, Classification techniques

Soft Skills: Presentation skills, Communication skills, Time management skills, Technical writing skills, Self-studying skills, Responsibility, Trustworthiness

## 3.3 Effectiveness of the Effort

I completed the project work during the internship by applying the concepts I had learnt at university. But I required more concepts for the industrial setting, and I was able to connect the concepts I was already learning in the university to the new ones, which made it easier for me to learn more effectively.

During the internship, I found that concepts like databases, SQL Server Management Studio, data mining concepts and machine learning, all of which were covered in the modules I initially studied on campus, were crucial. Mainly the DWBI module and the FDM module was helpful for me in my internship period. Considering that, throughout the internship I studied data warehousing and SSIS, SSRS, and Power BI on

my own during that time and the campus's business intelligence module. That was why it meant a lot to me. I must make an online referral studied many hands-on exercises, documentations, and tutorials to gain a good understanding of those elements.

I was able to advance both my technical and non-technical skills during my internship. It also assisted me in recognizing some of my shortcomings.

### Challenges and Solutions

- Data Quality Issues: One of the challenges faced was dealing with incomplete or inconsistent data. This was addressed through careful data cleaning and preprocessing, ensuring that the final datasets were of high quality.

- Model Complexity: Developing a robust predictive model required balancing complexity with interpretability. The use of ensemble methods and deep learning models was carefully managed to ensure that the final model was both accurate and understandable.

- Stakeholder Engagement: Ensuring that the Power BI reports met the needs of various stakeholders required regular communication and iteration. Feedback was incorporated at every stage to refine the reports and make them more useful.

Reflecting on internship experience and identifying areas for improvement is a valuable exercise. These are several points as I think, where I could potentially enhance or change the approach if I was able to redo the internship.

For the prediction model, implementing stricter data validation processes during data collection to ensure higher quality and consistency will be better than previous data collecting method. This could involve setting up automated data cleaning pipelines or integrating real-time data validation checks. Experiment with more advanced machine learning and deep learning models like XGBoost, CatBoost, or ensemble methods to potentially improve predictive accuracy would be another best practice to improve the accuracy of the model. I will be try utilizing advanced visualization techniques like heat maps, geo-spatial analysis, and dynamic filtering in Power BI to provide more detailed and interactive reports. I will be increasing the frequency and depth of feedback sessions with stakeholders to ensure that the projects remain aligned with their needs and expectations too.

# 4 SAMPLE WORK

## 4.1 <Production Prediction Model>

The goal of this project is to develop a robust production prediction model using historical data. This will explore various regression and time series models to identify the one that provides the best predictive performance. The dataset contains monthly production data along with other relevant features such as district, year, month, total rainfall, sunshine hours and fertilizer usage. The data spans from 2016 to 2022 for given districts. (Galle, Kandy, Nuwara Eliya, Ratnapura, Matara) First, load the data, handle missing values, encode categorical variables, and scale numerical features. Then, split the data into training and testing sets and explore various regression models: Linear Regression, Decision Tree Regressor, Random Forest Regressor, Gradient Boosting Regressor, and Support Vector Regressor. After evaluating all the models, found that the model with the lowest Mean Squared Error (MSE) and highest R^2 score is considered the best model for production prediction. Next, explored time series models: ARIMA and Prophet, to account for the temporal nature of the data.
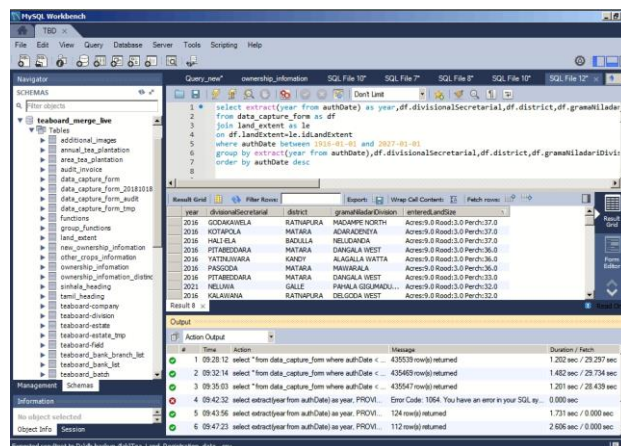


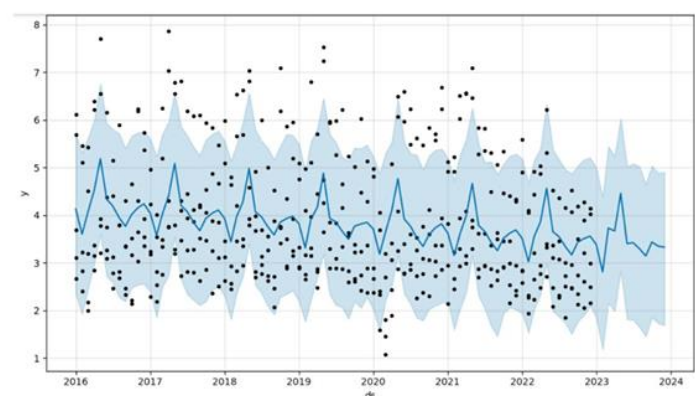*Figure 1 – Data extraction*



*Figure 2 – Production prediction forecast*

## 4.2 <Power BI report creation>

First imported data from various sources into Power BI. Cleaned and transformed the data using the Power Query Editor to ensure accuracy and consistency. Next, model the data by establishing relationships between tables and creating calculated columns or measures using DAX (Data Analysis Expressions). When the data is prepared, designed the dashboard by adding visualizations such as charts, graphs, and maps, ensuring they effectively communicate key insights. Also enhanced the dashboard's interactivity by incorporating slicers, filters, and drill-through options, allowing users to explore the data dynamically. Finally, published the dashboard to the Power BI Service for sharing and collaboration, where it can be accessed by stakeholders online and kept updated with the latest data.



*Figure 3 – Exports dashboard*



*Figure 4 – Production dashboard*