## 

Deconstructing the DALRRD project

March 2023

**─**

Whitepaper

Written and Published by

-Karabo Eugene Hlahla

- Iman Fasasi

- Onidajo Fikayo Wale-Olaitan

- Paris ihaki

- Haji Rufai

- Chukwunonso Azih

EXPLORE AI

35 BRICKFIELD RD

WOODSTOCK 1ST FLOOR, CAPE TOWN

WESTERN CAPE

7925

**Outline:**

* Executive Summary
* Introduction
* Problem Statement
* Aims and Objectives
* Methodology
* Results
* Conclusion
* References
* Appendices

# INTRODUCTION

# The South African Department of Agriculture, Land Reform & Rural Development (DALRRD) has a vision of equitable access to land, integrated rural development, sustainable agriculture, and food security for all. One of the key pieces towards achieving this is the agricultural marketing information system (AMIS), which collects and provides data on daily commodity prices in horticulture, grain, and livestock categories. However, accessing and interpreting this data has been challenging due to its spread across different websites and formats.

# To address this issue, our team was tasked with automating the scraping and transformation of commodity prices into a dashboard, applying solid exploratory data analysis (EDA) principles to ensure relevant, useful figures and statistics were presented. The resulting dashboard has improved access to critical pricing information for buyers and sellers, contributing to the development of a more sustainable and equitable agricultural sector in South Africa.

# This paper aims to highlight the challenges faced by the South African agricultural sector in accessing and interpreting commodity price data and how the automated data scraping and transformation system developed for AMIS has addressed these challenges. It will also provide an overview of the project, including the methodology used, the benefits of the resulting dashboard, and the impact on the agricultural sector in South Africa. The paper will also discuss the potential applications of this project in other regions and sectors, providing insights into the benefits of automating data collection and transformation for decision-making.

**Problem Statement**

​​The South African Department of Agriculture, Land Reform & Rural Development (DALRRD) is facing challenges in accessing and interpreting critical commodity price data due to its dispersion across various websites and formats. This issue is hindering DALRRD's ability to achieve its vision of equitable access to land, sustainable agriculture, and food security. Our team is currently developing an automated data scraping and transformation system that will convert commodity prices into a structured dashboard, using solid exploratory data analysis (EDA) principles to ensure the presentation of relevant and useful statistics. The goal is to improve access to pricing information for buyers and sellers, contributing to the development of a more sustainable and equitable agricultural sector in South Africa. This paper discusses the proposed methodology, benefits, and potential applications of the project.

# Overview

The South African Department of Agriculture, Land Reform & Rural Development (DALRRD) has the overarching vision of equitable access to land, integrated rural development, sustainable agriculture and food security for all.

**Aim**

Our aim is to develop an automated data scraping and transformation system for the agricultural marketing information system (AMIS) to provide a structured dashboard with relevant and useful statistics on commodity prices. This system will improve access to critical pricing information for buyers and sellers, contributing to the development of a more sustainable and equitable agricultural sector in South Africa.

**Objectives**

1. To identify the different sources and formats of commodity price data in horticulture, grain, and livestock categories in South Africa.
2. To develop a web scraping tool that automatically collects data from different websites and formats.
3. To transform the collected data into a dashboard that applies solid exploratory data analysis (EDA) principles to ensure relevant, useful figures and statistics are presented.
4. To test the dashboard and ensure that it is user-friendly and easily interpretable by buyers and sellers.
5. To evaluate the impact of the automated data scraping and transformation system on the agricultural sector in South Africa.
6. To explore the potential applications of the project in other regions and sectors, providing insights into the benefits of automating data collection and transformation for decision-making.

# Data Sources and Collection

In this project, the following data sources have been identified as relevant:

https://rpo.co.za/slaughtering-statistics/

http://webapps.daff.gov.za/amis/Link.amis?method=GrainMarket

http://webapps.daff.gov.za/amis/amis\_price\_search.jsp

For this project, a web scraping tool that utilizes Selenium was used to extract data from the sources. The tool automatically retrieves data from these websites and converts it into a structured format that can be analyzed and displayed on a dashboard. The data encompasses commodity prices across horticulture, grain, and livestock categories. The web scraping tool has been programmed to obtain data regularly to maintain the accuracy and timeliness of the information presented on the dashboard. Furthermore, the collected data is subjected to rigorous exploratory data analysis (EDA) principles to ensure that it yields relevant and valuable statistics and figures.

# Deliverables for the project

This project has several deliverables that have been identified to achieve the project objectives. These include:

1. Automated pipeline for scraping and transforming data into a database

The web scraping tool has been developed to collect data from the identified sources and format it into a structured database. This automated pipeline ensures that data is collected on a regular basis to ensure the information presented in the dashboard is up-to-date and accurate.

1. Dashboard to display relevant insights and trends: The collected data is transformed using solid exploratory data analysis (EDA) principles to ensure that relevant and useful figures and statistics are presented in a user-friendly dashboard. The dashboard displays insights and trends such as forecasts and regional differences, providing critical pricing information for buyers and sellers in the agricultural sector.
2. Documentation of the pipeline and dashboard: The pipeline and dashboard development process has been documented to ensure that the methodology used in the project is transparent and reproducible. This documentation will serve as a reference for future improvements and modifications.
3. User manual for the dashboard: To ensure that users can easily navigate and interpret the dashboard, a user manual has been developed. The manual provides step-by-step instructions on how to use the dashboard and interpret the displayed insights and trends.

# Tasks and Roles for successful project

## Data engineers:

1. Build and maintain the automated pipeline
2. Ensure data quality and consistency
3. Test the pipeline for accuracy and efficiency

## Data scientists:

1. Develop and maintain the dashboard
2. Identify and visualize relevant trends
3. Ensure the dashboard is user-friendly and responsive

## 

## Project manager:

1. Oversee the project and ensure tasks are completed on time and within budget
2. Communicate with stakeholders to ensure their needs are met
3. Ensure documentation and user manual are complete and accurate

# Tasks Completed:

* Selected tool for web scraping
* Data extracted into local system
* Introduced recursion into script
* Selected output and input file format for preprocessing
* Created a white paper skeleton
* Imported extracted data set into preprocessing script
* Preprocessed the data
* Converted data into processing formats
* Extracted data information from data source

# Data Analysis

Data analysis is the process of systematically examining and interpreting data with the goal of drawing insights and making informed decisions. It involves cleaning, transforming, and modeling data to discover meaningful patterns, relationships, and trends.

The collected data from the web scraping tool was transformed using exploratory data analysis (EDA) principles to ensure that relevant and useful figures and statistics were presented. The data collected included commodity prices in horticulture, grain, and livestock categories. The following analysis was conducted:

1. **Data Cleaning**: The data was cleaned to remove any duplicates, inconsistencies, or missing values that could affect the accuracy of the results.
2. **Descriptive Statistics**: Descriptive statistics were used to summarize and describe the data, including measures of central tendency and dispersion, such as mean, median, mode, standard deviation, and range.
3. **Time Series Analysis**: Time series analysis was used to identify trends and patterns in the data over time, including seasonality and cyclical patterns.
4. **Correlation Analysis**:Correlation analysis was used to identify the relationships between different variables and their impact on commodity prices.
5. **Visualization**: Visualization techniques, including graphs, charts, and maps, were used to present the results of the analysis in a clear and concise manner.

Some of the insights that gained from this analysis include;