# Deconstructing the DALRRD Daily Commodity Prices project

Whitepaper

Written and Published by

Team 8 @ExploreAi

-

May 2023

# **Contents:**

- Introduction
- Problem Statement
- Aims and Objectives
- Methodology
- Results
- Conclusion
- References
- Appendices

#### Introduction

The South African Department of Agriculture, Land Reform, and 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.

#### **Problem Statement**

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

The dispersion of critical commodity price data across various websites and formats 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 agricultural sector in South Africa.

### **Aims and Objectives**

The aim of this paper is 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. The specific objectives of this project include:

- Identifying the different sources and formats of commodity price data in horticulture, grain, and livestock categories in South Africa
- Developing a web scraping system that automatically collects data from different websites and formats
- Transforming the collected data into a dashboard that applies solid exploratory data analysis
   (EDA) principles to ensure relevant, useful figures and statistics are presented.
- Testing the dashboard and ensuring that it is user-friendly and easily interpretable by buyers and sellers
- Evaluating the impact of the automated data scraping and transformation system on the agricultural sector in South Africa
- Exploring 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.

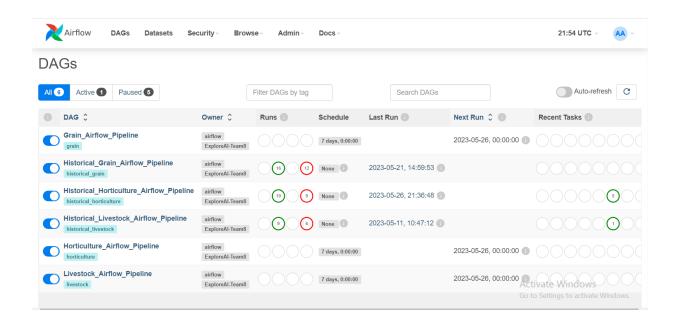
## Methodology

A web scraping system that utilizes Selenium was set up to extract data from the sources. The system automatically retrieves data from these websites and converts it into a structured format that can be analyzed and displayed on a dashboard. The system is monitored using Apache Airflow, a powerful task scheduler that allows for the loggability and monitorability of the data pipelines that interact with the data source. The data sourced encompasses commodity prices across horticulture, grain, and livestock categories. The web scraping system 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 it yields relevant and valuable statistics and figures.

#### **Results**

The project has achieved its objectives by identifying the different sources and formats of commodity price data in horticulture, grain, and livestock categories in South Africa, developing a web scraping system that automatically collects data from different websites and formats, transforming the collected data, and applying solid exploratory data analysis (EDA) principles to achieve insights.

Our web scraping data pipeline interface:



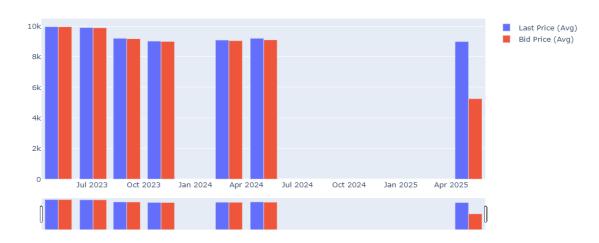
Here we can see the interface which our pipeline automation tool takes. From this UI, we can monitor our data scraping system in the form of data pipelines which then load to our database.

Some of the insights gained include;

#### Grain



#### Average Price (Last/Bid Price) over the Past Year for Unsettled BEAN Contracts



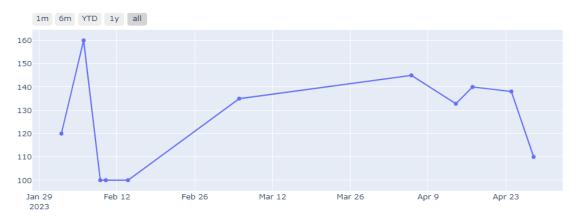
This chart shows the average price (Last/Bid) for each contract, which is yet to reach its maturity date, calculated over a period of 1 year. These average prices are then plotted against their respective maturity dates or contract dates. This chart provides a comprehensive view of the 1-year average price of all existing contracts.

The insight gained from this observation is that it allows for a holistic understanding of the price trends and patterns of the contracts over time. By calculating the average price for each contract over a year-long period, the chart provides a smoothed representation of the price behavior, reducing the impact of short-term fluctuations or noise.

#### **Horticulture**



#### Closing Price



This chart shows the closing prices for horticulture products. This chart allows users to explore price information at a detailed level, including various packaging and presentation options for the products. The goal is to identify specific price trends that can assist in decision-making processes. The chart utilizes a dynamic filtering technique implemented through a six-level filtering system. This approach enables efficient exploration of the dataset while ensuring that any combination of filter values produces relevant output.

#### Livestock:cattle



This chat is designed to compare different class pairs within the cattle livestock dataset. This chart allows users to select one or two classes of cattle and presents their price performance over time in a comprehensive view. The purpose of this chart is to facilitate an easy and effective comparison between different cattle class pairs.

By focusing on price performance over time, users can analyze historical trends, seasonal variations, and specific events that have influenced price movements. The interactive nature of the chart provides flexibility and customization options, catering to user preferences and increasing usability

#### Conclusion

We have successfully automated the scraping and transformation of commodity prices in the horticulture, grain, and livestock categories, and created a comprehensive dashboard for use by the South African Department of Agriculture, Land Reform, & Rural Development (DALRRD)'s AMI system users. Through our collective efforts, we have achieved the following outcomes:

- Enhanced Data Accessibility: By automating the data collection process, we have significantly improved the accessibility of agricultural market information. Buyers and sellers can now easily access up-to-date and accurate data on commodity prices, empowering them to make informed decisions.
- User-Friendly Dashboard: We have developed a user-friendly dashboard that presents the commodity price data in a clear and intuitive manner. The dashboard incorporates solid exploratory data analysis (EDA) principles, allowing stakeholders to easily interpret trends, forecasts, and regional differences.
- Stakeholder Empowerment: Our project will now empower stakeholders, such as retailers
  and farmers, by providing valuable insights and statistics. They can now set and negotiate
  prices more effectively, leading to improved market efficiency and fairness.
- Contribution to DALRRD's Vision: Our project aligns with DALRRD's overarching vision of
  equitable access to land, integrated rural development, sustainable agriculture, and food
  security for all. By improving the accessibility and usability of agricultural market
  information, we have taken a significant step toward realizing this vision.
- Foundation for Future Advancements: Our project serves as a strong foundation for future advancements in agricultural market analysis. DALRRD can build upon our work, incorporating additional features and expanding the scope of data to further enhance decision-making processes and support sustainable agricultural practices.

As a group, we have demonstrated our commitment to leveraging technology and data analysis to drive positive change in the agricultural sector. Our successful completion of the project showcases our collective skills in data scraping, transformation, and dashboard development, as well as our ability to apply EDA principles effectively. Our project has made a meaningful contribution toward improving agricultural market transparency, efficiency, and fairness. We are proud of our achievements and the positive impact our work will have on stakeholders within the agricultural industry.

#### **Recommendations**

The app as currently constituted can draw historical data a maximum of 365 days for some of the charts. This limitation was set to ensure usefulness of information published. However, we look forward to a situation where the range of historical data required can be set by the user for increased flexibility.

We also see a possibility for AI/ML models to be built for this app to make possible the short term prediction of the prices.

# **Appendices**

Appendix 1: List of Websites Used for Data Collection

- 1. <a href="https://rpo.co.za/slaughtering-statistics/">https://rpo.co.za/slaughtering-statistics/</a>
- 2. <a href="http://webapps.daff.gov.za/amis/Link.amis?method=GrainMarket">http://webapps.daff.gov.za/amis/Link.amis?method=GrainMarket</a>
- 3. <a href="http://webapps.daff.gov.za/amis/amis\_price\_search.jsp">http://webapps.daff.gov.za/amis/amis\_price\_search.jsp</a>

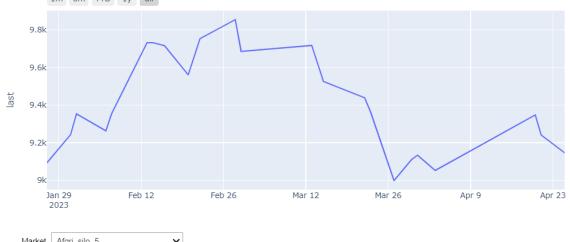
#### Appendix 2: Definitions

- Selenium Selenium is an open-source framework and a popular tool for automating web browsers. It provides a programming interface to interact with web elements, simulate user actions, and perform automated testing of web applications.
- Apache Airflow Apache Airflow is an open-source platform used for orchestrating, scheduling, and monitoring complex workflows or data pipelines. It allows you to programmatically define and execute workflows as directed acyclic graphs (DAGs)

Appendix 3: EDA Screenshots



# Average Last Price of BEAN Contracts Versus Upload Date (Period = January 27, 2023 To April 25, 2023)





#### Last/Bid Price of '01 Mar 23 BEAN / Afgri silo\_5 ' Contract vs Upload Date

