

## EDUCATION

---

**Worldquant University BootCamp** LA, USA  
Applied Data Science Lab Mar 2023

Major in Data Science

Relevant Coursework: API Design; Data Science; Data Visualization; Statistics; ML, SQL, Python

**Explore Data Science Academy** Cape Town, SA  
Higher Certificate in Information Technology Jul 2022

Major in Data Science

Relevant Coursework: Data Analysis, Software Engineering; Operating Systems; Algorithms; Artificial Intelligence

## PROFESSIONAL EXPERIENCE

---

**Sabaja Limited** Nairobi, KE  
Data Scientist August 2023 – Present

- A/B Test to check whether a change in the loading page of an ecommerce website has a significant impact on the conversion rate of users visiting the page.
- The application of NLP to aid in the development of chat bots that are based on large pretrained models like GPT and BERT.
- The application of time series analysis to build a forecasting API that utilizes a saved time series analysis model. The api was powered by Fast API.

**Twiga Chemical Industries** Nairobi, KE  
Data Analyst Oct 2022 – April 2024

- Daily Sales Pattern Analysis to derive insights.
- Data selection and model explanation. Analysis of present data to derive insights on sales and product movement based on defined marketing regions.
- Part of the team as a Backend Developer in the development of CRM Software.

**Explore AI** Cape Town, SA  
Data Science Intern May 2022 – July 2022

- A team-led deep learning project to aid in plant disease identification.
- Model Design and development.
- Data selection and model explanation.
- NLP models and implementation
- Data pipeline design and development using Python, Docker and SQL

## SKILLS

---

**Programming Languages:** Python, R, C++, Javascript, HML, CSS

**Big Data & Machine Learning:** Hadoop, Spark, MongoDB, Python(Scikit-learn, Pandas, Numpy..)

**Data Science & Miscellaneous Technologies:** A/B Testing, ETL, Data Science pipeline (cleaning, preprocessing, visualization, modeling, interpretation), Statistics, Time Series, Experimental Design, Hypothesis testing, OOP, APIs, Excel, Git

## **PROJECTS**

---

### **Plant Disease Classification**

- The objective of the project was to classify tomato leaves into healthy and unhealthy.
- Utilized deep learning neural networks to classify image data from Kaggle.
- Leverage transfer learning to improve model accuracy and save time.
- Model was deployed to streamlit for presentation and evaluation.
- Key technologies utilized include Tensorflow, Pandas, Numpy and Streamlit

### **Forecasting Stock Data ( Time Series Analysis)**

- Developed and trained a time series forecasting model using stock data.
- The model was deployed using FastAPI
- Leveraged GARCH model and AlphaVantage API ( source of data).
- Key technologies utilized include Pandas, Scipy, Numpy, OOP, Python
- Data pipeline that involves modeling, orchestration and deployment.