

Princewill Egbujor

+1 (510)-224-1232 | San Francisco, CA | princewill.egbujor@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

EDUCATION

University of San Francisco
Master of Science (MS) in Data Science

Jul 2023 - Jun 2024
San Francisco, CA

University of Port Harcourt
Master of Science (MS) in Offshore Engineering

Nov 2013 - Apr 2015
Rivers, Nigeria

Federal University of Technology Owerri
Bachelor of Engineering in Chemical Engineering

Mar 2005 - Mar 2010
Owerri, Nigeria

SKILLS

Programming/ Visualization: Python, SQL, NoSQL, HTML, Linux, PowerBI, Tableau, Looker, Qlik, Excel, PowerPoint, Airflow.
Machine Learning: Regression, Classification, Unsupervised, Deep Learning, NLP, Transformers, Large Language Models (LLM)
Big Data/Database: PySpark, SparkSQL, ETL, Spark, MongoDB, Git, Databricks.
Libraries and framework: PyTorch, Scikit-Learn, NumPy, Pandas, Plotly, Matplotlib, Seaborn, Spacy, SciPy, Selenium.
Web Experimentation: Causal Inference, A/B Testing, Hypothesis testing, Statistical Inference, Time Series Forecasting (ARIMA, Prophet)
Soft skills: Leadership, team player, great adaptability, good communicator.

WORK EXPERIENCE

TruckX Inc | Data Scientist | San Francisco, USA

Oct 2023 – Present

Sharp Turn Logic Detection & Prediction – End-to-end Project

- Crafted a rule-based sharp turn detection logic to capture **99%** of all truck behaviors by analyzing **~1000 hours** of GPS data on **Google BigQuery** using **DataProc** in **GCP Notebooks** to extract insights. Engineered predictive features through exploratory data analysis.
- Developed an ML framework by training **Random Forest** and **Decision Tree classifiers** for predicting sharp turn event for moving trucks.
- Implemented real-time advance alert logic for sharp turns maneuvers, improving driver and client safety by over 11%. [GitHub Repo](#)
- Created an AI-powered document Q/A chatbot leveraging **Generative AI**, **NLP**, and machine learning to streamline trucking compliance procedures. This innovation enhanced driver engagement and improves accessibility to documents. [GitHub Repo](#)

Geofence Recommender system - End-to-end Project

- Currently building an intelligent geofence recommender feature aimed at improving efficiency and route optimization for fleets, automating 80% of geofence implementation through one-click sequence, and significantly cutting down on fuel cost and idle time by 20%.
- Improve client-app feature engagement by **over 200%** by utilizing **Clustering algorithms** to define most influential locations. [GitHub Repo](#)

Dangote Oil Refinery | Operational Readiness Manager - Data Analyst | Lagos, Nigeria

Mar 2016 – Jul 2023

Data Center

- Led the data-driven Operational Readiness Program (ORP) by leveraging **SQL**, **Google Cloud**, and **Apache Airflow** to automate and orchestrate multiple data workstreams at the refinery.
- Developed and implemented a **People Analytics platform** using Python and SQL to integrate and analyze data from various HR streams, supporting over **4000 personnel** during refinery commissioning phase.
- Utilized Microsoft PowerBI to visualize recruitment metrics and identify bottlenecks in hiring processes, resulting in streamlined recruitment procedures and significant time and cost savings.
- Spearheaded the development of predictive analytics models to optimize workforce allocation and reduce downtime, leading to an **18%** reduction in operational downtime and approximately \$5 million in cost savings for the \$9 Billion refinery project.
- Predicted employee attrition rates using advanced statistical and machine learning techniques (**logistic regression and decision trees**), resulting in improved retention strategies and reduced **turnover by 8%**.

Gmunu Limited | Senior Project Engineer | Lagos, Nigeria

Oct 2013 – Mar 2016

Data Analytics for Quality Assurance & Contract Compliance

- Implemented statistical analysis to assess dimension data integrity and consistency, achieving a **13%** improvement in data accuracy, leading to reduction in QA/QC errors and enhancing compliance with design specifications and contractual obligations.
- Conducted regression analysis on project delivery timelines and costs, identifying key factors influencing project efficiency and cost overruns, leading to a **9%** improvement in project planning accuracy and cost estimation.
- Utilized data visualization techniques to identify outliers and anomalies and developed interactive dashboards to present key project insights and trends.
- Developed standardized **data quality metrics and KPIs** to monitor and measure the effectiveness of QA/QC processes, resulting in increased project quality and compliance.

OTHER PROJECTS

Refinery Inventory Management System – Conducted ETL processes and implemented data-driven forecasting and inventory management solutions using **Python Django web framework**, optimizing resource utilization for early trading startup.

ML Models from Scratch – Implemented Decision Trees, Random Forests, Bagging, Boosting (CatBoost, XGBoost), Clustering and Recommendation System. [GitHub Repo](#)