

# MARWIN SOLOMON

marwinsolomon07@gmail.com | +44 7436 683661 | [LinkedIn](#) | [Website](#)

## EXECUTIVE SUMMARY

Data Scientist from LSE with an experience of building explainable machine learning models for Deutsche Bank and Hartree Partners

## EDUCATIONAL QUALIFICATIONS AND ACHIEVEMENTS

- **MSc in Data Science - London School of Economics and Political Science** (Sep '22 - Sep '23)
  - **Attained Distinction along with Graduate Support Scheme Scholarship awarded to meritorious students**
  - Key Modules: Machine Learning & Data Mining, Distributed Computing for Big Data, Data Analysis & Statistical Methods
- **B.Sc. (Hons) Mathematics - St. Stephen's College, University of Delhi** (Jul '18 - Jul '21)
  - Finished with a **CGPA of 9.8/10**, batch rank of **3** out of 42 students and amongst **Top 1%** across the University of Delhi
  - Key Modules: Probability Theory & Statistics, Mathematical Finance, Differential Equations, Real Analysis, Group Theory
- **All India Rank 13<sup>th</sup> (1000+ students)** in MSc in Quality Management Science Entrance 2021, Indian Statistical Institute
- **All India Rank 39<sup>th</sup> (2000+ students)** in Master of Operational Research Entrance 2021, University of Delhi

## WORK EXPERIENCE

- **Hartree Partners – Quantitative Oil Trading Analyst Intern (Location: London, UK)** (Oct '23 – Present)  
*Leading Proprietary Trading Firm in the energy and commodities space with a market capitalization of \$4 billion+*
  - Developed a rolling regression asset agnostic systematic strategy based on macroeconomic and cross-asset indicators
  - Attained an average hit rate of at least 50% over 10 years of historical backtest for 6 different training windows
  - **Skills Gained: Streamlit, Systematic Strategy, Algorithmic Trading, Bloomberg Terminal, Rolling Regression**
- **Scroll Finance - Data Scientist: Contract (Location: London, UK)** (Aug '23 - Oct '23)  
*Leading Series A startup in the second charge mortgage space with a current funding of £4.18 million*
  - **Analysed the London Housing dataset comprising of 3.78M** sold properties from 1995-2022
  - Introduced new features in the dataset to account for comparative property valuation based on similar neighbourhoods
  - Built a model over 1.085M properties with an **average error rate of 10.26% (vs 7% UK Industry Standard)**
  - **Skills Gained: Feature Engineering, Data Cleaning, Big Data, Time Management, Supervised Learning**
- **Deutsche Bank – Graduate Quantitative Researcher: LSE Capstone Project (Location: London, UK)** (Dec '22 - Sep '23)  
*Leading German financial institution with operations spanning across 70 countries*
  - **Elected by a panel of 5 professors to lead a group of students** to construct stochastic models for the company
  - Identified **factors responsible for the movement of FX rates across the G10 markets with EUR & USD as base currencies**
  - Applied Local Interpretable Model-Agnostic Explanations (LIME) algorithm to obtain linear transformation of complex machine learning models to evaluate trading strategy and provide easier explanations to stakeholders
  - Incorporated Permutation Feature Importance (PFI) to **extract the crucial subset of financial datasets**. Helps to ascertain the important features which characterise the data related to any asset class
  - **Skills Gained: Quantitative Research/Analysis, Business Communication, Machine Learning, Python, GitHub**
- **Delhivery - Operations Research Intern (Location: Gurugram, India)** (Jun '22 - Sep '22)  
*India's leading Logistics and Supply Chain company providing services across 2300 cities and towns*
  - Leveraged Python for data cleaning to **construct optimization models** for predicting cost-efficient routes pan-India
  - Analysed the vehicle routing dataset of **2500+ routes** for the company with respect to operational costs and delivery time
  - Achieved **reduction in operational costs by 8%** and **delivery time by 5%** for the delivery services of the company
  - **Skills Gained: Data Analysis, Presentation Skills, Mathematical Modelling, Numerical Optimisation, Technical Reporting**
- **St. Stephen's College - Nagpaul Fellowship Student Researcher (Location: New Delhi, India)** (Aug '19 - Apr '21)  
*Auspicious fellowship in the domain of applied and pure mathematics offered by the Department of Mathematics, St. Stephen's College*
  - **Selected by a panel of 8 professors to lead a group of students** from a competitive **pool of 50+ students**
  - **Modeled the rate of change of deposits based on percentage of quotas** of four major banks in the Greek Market
  - **Developed various epidemic models and graphs based on COVID-19 Modeling** research paper
  - **Skills Gained: Research, Mathematical Modeling, R Programming Language, Data Analysis, Data Visualisation**

## LSE DATA SCIENCE PROJECTS

- **S&P 500 Classification Using PySpark (Domain: Distributed Computing for Big Data)**
  - Aim: To understand the efficiency of machine learning algorithms in terms of time (time complexity) over cloud computing
  - Outcome: **Time complexity of the algorithms to produce result decreases with improvement in computational resources**
  - **Skills Gained: PySpark, Google Cloud Platform, Data Engineering, Matplotlib, Machine Learning**
- **Stock Price Prediction Using Generative Adversarial Networks (Domain: Finance and Investment)**
  - Aim: To analyse the effect of financial news over accurately determining the prices of the stocks
  - Outcome: **Financial news incorporated with GAN models improves the accuracy in predicting the stock prices.**
  - **Skills Gained: Deep Learning, TensorFlow, Time-Series, Seaborn, Sentiment Analysis**

## PERSONAL DATA SCIENCE PROJECT

- **FIFA Project (Domain: Sports Analytics)**
  - Aim: To engineer a model with data visualisation for clustering the players based on similar attacking abilities
  - Outcome: **Identified Erling Haaland as the best player** for transfer on the basis of age, growth potential and overall rating
  - **Skills Gained: K-Means Clustering, Plotly, Regular Expressions, Unsupervised Learning, Data Engineering**

## TECHNOLOGY AND INTERESTS

- Programming Languages: Python (Jupyter Notebook), R, C++, JavaScript
- Libraries: NumPy, Pandas, Matplotlib, Seaborn, Plotly, Scikit-learn, TensorFlow, Keras, BeautifulSoup4, dplyr, ggplot2
- Cloud Computing and Data Engineering: Google Cloud Platform (GCP), Apache PySpark, Hadoop
- Database Systems & Version Control Repository Management: SQL, GitHub
- Web Development, Application Software: HTML, CSS, Streamlit, Microsoft Office, Microsoft PowerPoint, Microsoft Excel
- Football, Music, Skipping, LeetCode Coding, Kaggle Competitions