

# FIRAZ ZAKARIYA

Data Scientist | Analytics Forecasting, ML & Risk Analytics

firaz.zakariya@hotmail.com | +46 734 834 852 | [LinkedIn](#) | Karlstad, Sweden

## PROFESSIONAL SUMMARY

---

Data scientist with 4+ years building forecasting, classification, and scenario-simulation models for operations teams and senior leadership, anchored in statistical rigour, model interpretability, and clear stakeholder communication. Develops gradient boosting and Random Forest pipelines with SHAP-based interpretability, time-series forecasts (ARIMA, tree-based regression with cohort segmentation), and scenario simulations that quantify outcomes under alternative resourcing and macro assumptions, work directly transferable to provisioning, lifetime loss estimation, and capacity decisions under uncertainty. Comfortable across the full model lifecycle, partnering with data engineering on Azure and AWS, and translating model outputs into decisions for non-technical audiences including directors, auditors, and senior executives. MSc Distinction from Cranfield with a research foundation in explainable machine learning

## PROFESSIONAL EXPERIENCE

---

### Technical Data Analyst (Data Scientist)

*Independent Office for Police Conduct (IOPC) | 2024 - 2026*

- **Time-Series Forecasting:** Developed hybrid backlog forecasting models combining cohort-based gradient boosting with 12-month ARIMA projections; validated via out-of-time holdout splits, driving senior leadership capacity-planning decisions.
- **Risk Classification & Interpretability:** Built and deployed Random Forest and Gradient Boosting (XGBoost/LightGBM) classifiers to predict case-delay risk; used SHAP to identify the top drivers of completion-rate variability, enabling proactive prioritisation of at-risk work; evaluated via stratified time-based splits.
- **Forward-Looking Scenario Modelling:** Engineered scenario simulations quantifying throughput and lead-time changes under alternative resourcing assumptions, methodology directly analogous to scenario-driven forward-looking projections used in lifetime loss estimation (ECL) and macroeconomic overlays.
- **Statistical Validation & Experimentation:** Designed controlled A/B experiments using power analysis and minimum detectable effect (MDE) to validate completion-time reductions; translated statistical results into rollout recommendations adopted by operations leadership.
- **Production & Model Lifecycle:** Built reproducible Python modelling pipelines on Azure in partnership with data engineering, covering end-to-end lifecycle (training, validation, scoring, monitoring) with scheduled refreshes; cut manual reporting effort by ~8 hours/week.

### Co-Founder (SaaS Product)

*Cawosh GMS | 2025 - Present*

- Co-founded and launched a subscription SaaS for UK independent garages; owned pricing tier design, plan structure, and add-on offerings from concept to launch.
- Defined the subscriber-facing KPI framework, recurring revenue, active subscribers, engagement, blocked value, churn signals, wired into the operator dashboard for real-time retention and growth decisions.
- Designed retention and repeat-engagement mechanics (automated MOT and service reminders, re-booking loops) to lift subscriber lifetime value and reduce churn in a recurring-revenue model.
- Used Cursor and Claude Code throughout development to accelerate analysis and shipped-feature velocity.

### Lead Project Analyst - Operations Analytics

*KTI Middle East, UAE | 2022 - 2022*

- Built optimisation models for large-scale engineering projects, raising schedule reliability and inspection throughput when resources were constrained.
- Applied clustering (K-means, hierarchical, DBSCAN) to segment projects by complexity profile, enabling differentiated planning across multi-site operations.
- Built simulation models testing schedule reliability across workload and resource scenarios; collaborated with logistics and operations to improve deployment planning accuracy.
- Applied statistical process control and trend analysis to identify process drift, contributing to a 12-15% reduction in rework rates.

- Proposed and prototyped an ETL pipeline (PySpark, AWS Lambda) for scheduled processing of large engineering datasets; transitioned for MSc before deployment.

## KEY PROJECTS

---

### Backlog Reduction Strategy Evaluation - IOPC (2025)

- Modelled alternative backlog-reduction strategies under different resourcing assumptions, comparing them on completion velocity, cost, and risk. Output fed into senior leadership's backlog plan.

### Investigator Workload & Capacity Analysis - IOPC (2025)

- Analysed investigator workload distributions and capacity utilisation across teams; identified imbalances and quantified the impact of alternative allocation strategies on lead time and throughput.

### Case Lifecycle & Stage-Duration Analysis - IOPC (2025)

- Decomposed case lifecycle durations across stages and case types; identified bottleneck phases and quantified time spent in each, used by the process redesign workstream.

### Executive and Ops Performance Dashboard - IOPC (2025)

- Built a Power BI dashboard for senior executive leadership and operations managers (Directors of Operations down to Operations Managers), integrating multiple operational data sources into a single performance view.

### A/B Testing: Final Report Writing Procedure - IOPC (2024)

- Designed and analysed a controlled experiment comparing new vs legacy procedures across investigator teams. Defined primary and guardrail metrics, ran sample-size and power analysis to set the minimum detectable effect, validated a statistically significant reduction in mean completion time, and translated the result into a rollout recommendation adopted by operations leadership.

## TECHNICAL SKILLS

---

**Experimentation & Causal Inference:** A/B test design, sample-size and power analysis, hypothesis testing, ANOVA, guardrail metrics, uplift modelling, Statistical Process Control

**Time-Series & Forecasting:** ARIMA, tree-based regression forecasting, cohort and vintage-style analysis, scenario simulation, out-of-time validation, backtesting, feature engineering, model monitoring

**Machine Learning:** scikit-learn pipelines, cross-validation, hyperparameter tuning, calibration, Gradient Boosting (LightGBM, XGBoost, CatBoost), Random Forest, Clustering (K-means, DBSCAN, hierarchical), Neural Networks (TensorFlow — academic), SHAP / LIME for interpretability

**Statistical Analysis:** Regression analysis, correlation analysis, trend analysis, non-parametric methods, survival analysis foundations

**Programming & Data:** Python (Pandas, NumPy, SciPy, scikit-learn, statsmodels), SQL (window functions, CTEs, query optimisation), PySpark, ETL development, workflow automation

**Cloud & MLOps:** AWS (EC2, S3, Lambda), Azure, Docker, Git, experiment tracking and model versioning workflows

**Visualisation & BI:** Power BI (DAX, Power Query), Plotly Dash, dashboard design for executive audiences

## EDUCATION

---

### MSc Computational Software Techniques in Engineering (Distinction)

Cranfield University, UK | 2022 - 2023

**Thesis:** *Explainable Machine Learning for Flight Delay Classification* — achieved 97% accuracy using LIME and deep neural networks.

### BTech Mechanical Engineering (First Class)

APJ Abdul Kalam Technological University, India | 2016 - 2020

**Bachelor's Thesis:** Composite modelling and analysis for aerofoil skins using ANSYS ACP.

## PUBLISHED RESEARCH

---

*Carbon Fiber Composite Development for In-Ground UAVs with NACA0012 Airfoil Wing* — Elsevier Materials Today: Proceedings (DOI: 10.1016/j.matpr.2021.05.142).