

Yaser Selvam

MSc Data Science and Analytics | 2+ YoE | SQL | Audience Segmentation | Data Analyst

London, UK | +44 7769486140 | yaseruk259@gmail.com | linkedin.com/in/yaserselvam | github.com/yaserselvam | yaserselvam.vercel.app

PROFESSIONAL SUMMARY

Level 7 certified Data Analyst, 2+ years in SQL, Python and Power BI across charity fundraising, retail and enterprise data. Genuinely inquisitive: questions what the data and analysis mean before issuing a report, not just what they say. Designs segmentation for around 330 campaign selections a year, and turned a data-handling analysis into org-wide AI policy: the discovery report behind it scored 10/10 and was adopted as the basis for further work.

RELEVANT SKILLS

Functional Competencies: Audience Segmentation | Campaign Selections | Suppression Logic | ETL Pipeline Development | Data Modelling | Query Optimisation | Data Quality Assurance | GDPR Compliance | Dashboarding | Stakeholder Reporting | RFM Analysis | A/B Testing | Hypothesis Testing | AI-Augmented Workflow Design | LLM Evaluation

Technical Tools:

Databases & Query Languages: SQL (T-SQL / SQL Server) | VBA | BigQuery | Snowflake | DuckDB

Programming & Statistical Analysis: Python (pandas, NumPy, SciPy, statsmodels, Matplotlib) | PySpark

BI & Visualisation: Power BI | Tableau | Excel (advanced, MOS certified)

Cloud & Data Engineering: Azure Data Factory | Azure Databricks | dbt | Fivetran

Analytics Platforms: GA4

AI-Assisted Development: Claude | LLM evaluation & rubric design

Version Control: Git | GitHub

WORK EXPERIENCE

Data Analyst, WaterAid

Sep 2025 – Present

- Design the audience segmentation and campaign selections behind around 330 campaign selections a year, across email, mail, phone, SMS and online activity.
- Separated 12 fields that are genuinely personal data from those that only resemble it; the resulting restrictions document unblocked 5 previously refused briefs and is being built into organisation-wide AI tool policy.
- Authored the discovery report on the end-to-end selections workflow: scored 10/10, adopted as the basis for further work, and praised in writing by the department head.
- Corrected two figures in the team's business case, one cutting the claimed saving by a third after separating elapsed time from effort, and sourced every rate to a documented method.
- Built an AI-augmented, human-verified SQL workflow on 8 prompt templates, now in controlled trial across a three-analyst team.

AI Training and Evaluation Contributor (Freelance), Outlier AI

Feb 2025 – Aug 2025

- Evaluated 100+ pairs of LLM responses (~200 model outputs) for correctness, reasoning, relevance and instruction-following, producing 100+ written rationales supporting human-feedback training data.
- Completed ~40-50 prompt-writing, annotation and revision tasks, creating challenging prompts and correcting factual, reasoning and instruction-following errors before submission.
- Applied 5-7 rubric criteria across 100+ evaluations covering factual accuracy, reasoning quality, relevance, completeness and instruction adherence, using a Data Science background on technical tasks.

Programmer Analyst, Cognizant Technology Solutions

Nov 2022 – Jun 2023

- Developed SQL and supported PySpark transformations across ~6 pipelines at a Fortune 500 IT firm, consolidating 4 retail sources into 20-25 curated tables that became the reporting base for 15-20 analysts.
- Refactored a daily sales and inventory pipeline with a senior engineer, replacing full-history scans with incremental processing, cutting runtime from 90 to 60 minutes.
- Helped implement ingestion-stage validation for inconsistent date formats and missing fields, ending manual reruns that had been hitting a few times a month.
- Triaged ~12 failed production runs using Azure Data Factory and Databricks logs, resolving ~8 without senior escalation and restoring loads before the reporting window.

Junior Data Analyst, Northern Lights Solutions (NLTVC Sdn Bhd)

Jan 2022 – May 2022

- Built the reporting base for 150 customer sites, consolidating 5 SQL source tables and ~10,000 records into a single validated dataset in place of scattered files.

- Analysed ~1,300 support cases to identify 7 recurring fault categories and the sites failing repeatedly, ranking where engineering effort would pay back most.
- Supported delivery of a 3-page Power BI dashboard tracking 9 KPIs that replaced manual spreadsheet comparison for 8 stakeholders.
- Helped consolidate 5 data sources and automate the monthly refresh, cutting reporting preparation from 4 hours to 90 minutes, a 60% reduction.

PROJECTS

GA4 Product Analytics: dbt on BigQuery and Snowflake

dbt, BigQuery, Snowflake, Fivetran, Power BI, GA4 | github.com/yaserselvam/ga4-dbt-product-analytics

- Modelled the GA4 e-commerce dataset with dbt on BigQuery (15/15 tests), then ported the build to Snowflake with Fivetran ingestion (16/16 tests).
- Quantified the purchase funnel across 61,252 users, identifying view-to-cart as the dominant leak at 20.5% against 77.4% cart-to-checkout, directing effort to the product page.

AI-Augmented Customer Segmentation

SQL, DuckDB, Claude | github.com/yaserselvam/ai-augmented-segmentation

- Built RFM segmentation over 1,067,371 public e-commerce transactions, scoring 5,832 customers into 9 segments and showing 24.9% of customers drive 70% of net revenue.
- Delivered an auditable win-back selection, waterfalling 5,832 scored customers to a 625-record audience against £1.27M of lapsed value, with audit checks catching 4 real defects.

Product A/B Test: Design, Analysis & Decision

Python, statsmodels, SciPy | github.com/yaserselvam/product-ab-experiment

- Designed a two-arm retention experiment across ~90,000 users, powered at 80% to detect 1pp, with a Sample Ratio Mismatch check clearing the 49.56/50.44 split before any outcome was read.
- Measured a 0.82pp drop in 7-day retention (95% CI [-1.33, -0.31], $p = 0.0016$), returning a do-not-ship decision on a change that would have cost retention.

EDUCATION

University of Leeds

Sep 2023 – Nov 2024

MSc Data Science and Analytics, Merit

Highlight: Dissertation evaluated treatment effects across multiple outcomes using statistical models, focused on hypothesis-testing accuracy, error rates and correlation structure.

Meenakshi Sundararajan Engineering College

Aug 2018 – Jul 2022

BE Computer Science and Engineering, First Class with Distinction

COURSES AND CERTIFICATION

- Microsoft Office Specialist: Excel Associate (Microsoft 365 Apps), Microsoft, May 2024
- Microsoft Fabric Analytics Bootcamp, Microsoft / ANS, April 2026
- Microsoft Certified: Power BI Data Analyst Associate (PL-300), Microsoft, exam booked September 2026
- Microsoft Certified: Fabric Analytics Engineer Associate (DP-600), Microsoft, exam booked November 2026