Aaryen DSouza

437-966-6119 | dsouza96@uwindsor.ca | linkedin.com/in/aaryendsouza | github.com/aaryen-dsouza | Portfolio

Technical Skills

Languages: Python, Java, JavaScript, TypeScript, C, SQL, HTML/CSS
Databases & BI: Oracle SQL, PostgreSQL, MySQL, MongoDB, PL/SQL, Data Modeling, Query Optimization
Data Manipulation & Visualization: Power BI, Tableau, PySpark, dbt, Pandas, NumPy, Matplotlib, Seaborn, ArcGIS
Frameworks: React, Django, Redux, Node.js, Express, WordPress, Material-UI, Bootstrap, Tailwind CSS
Dev Tools: Git, GitHub, Jupyter Notebook, Google Collab, VS Code, DataGrip, PyCharm, IDEA, Postman, Jira
Other Skills: Data Wrangling, Reporting Automation, Shell Scripting, Microsoft Excel, Google Sheets, Windows, Linux

Professional Experience

Data Analyst Intern May 2025 – Present
Office of Institutional Analysis, University of Windsor *Windsor, Ontario, Canada*

- Designed and implemented **SQL-based data models** and **Power BI dashboards**, enabling 5+ departments to access self-service insights on enrolment, retention, performance metrics, and institutional trends.
- Performed **data extraction**, **transformation**, and **quality assurance (QA)** on 100K+ records using **SQL** and **Python**, ensuring high accuracy and consistency across reporting workflows.
- Developed reusable **DAX measures**, trend and variance analysis logic, and forecasting components to support recurring dashboards while reducing dependency on ad-hoc reports.
- Automated reporting workflows with **SQL**, **Excel**, and **Python**, improving refresh reliability, reducing manual steps, and standardizing delivery timelines for key reports.
- Generated actionable insights on enrolment shifts, term-over-term patterns, retention trends, and KPI fluctuations, supporting evidence-based planning and strategic decision-making.

Data Engineer Jul 2022 – Mar 2024
Vermont Information Processing *Mumbai, India*

- Built and maintained **SQL-based ETL pipelines** to ingest, transform, and reconcile large-scale pricing, reimbursement, and chargeback data, delivering reliable **analytics-ready datasets** for downstream reporting and analysis.
- Performed detailed **trend analysis**, **variance analysis**, and **anomaly detection** on transactional datasets to identify recurring discrepancies, uncover root causes, and reduce repeated data-related issues by **40%**.
- Automated **ETL workflows** using **Bash/Shell** scripting, reducing manual intervention by **20%** and improving pipeline reliability.
- Developed interactive **Power BI dashboards** that visualized revenue trends, reimbursement cycles, chargeback patterns, and operational KPIs, strengthening analytical visibility for business and support teams.
- Created SQL-driven **analytical summaries**, **QA checks**, and **validation reports** to improve visibility into key metrics, enhance data reliability, and support internal decision-making workflows.
- Collaborated with engineering, QA, and business teams to troubleshoot data issues, translate requirements into data transformations, and support ongoing improvements to data pipelines.

Projects

Loan Insights Dashboard | *Power BI, SQL, Data Visualization, Dashboard Design, Financial Analytics, Git* GitHub

- Developed an interactive loan data analysis dashboard using **Power BI** and **SQL**, enabling real-time insights into 5+ key financial metrics, borrower behavior, and loan performance, improving data accessibility for decision-makers.
- Enhanced data visualization and reporting by implementing **10+ KPI** tracking, trend analysis, and loan classification, improving financial strategy planning.

Predictive Caching for Web Pages | *Python, Pandas, Matplotlib, Statsmodels, Scikit-learn, Git* GitHub

- Engineered a hybrid caching system that combined **ARIMA**-based forecasting with LLM-driven context analysis, improving cache efficiency and reducing latency in high-traffic environments.
- Processed **1M+** traffic records with automated preprocessing and built ARIMA/Auto-ARIMA models achieving **RMSE 4.41–4.67**, generating **30-day** page-view forecasts.

Stress Level Prediction Using Machine Learning | *Python, Pandas, Matplotlib, Seaborn, Scikit-learn, Git* GitHub

- Analyzed **20,500+** behavioral metrics including screen time, gaming, app usage, and social media activity to develop a digital-wellbeing stress prediction model using ML algorithms.
- Performed preprocessing, feature engineering, and model tuning for **Random Forest** and **XGBoost**, improving predictive accuracy to **77%** and identifying key behavioral drivers.

Education

University of Windsor Windsor, ON
Master of Applied Computing (Grade: 88.44%) *May 2024 – Present*

University of Mumbai Mumbai, India
Bachelor of Engineering - Computer Engineering (CGPA: 8.58/10) *Aug 2018 – May 2022*