Aaryen DSouza

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

Technical Skills

Data Analysis & Programming: Python, Java, SQL, KPI Reporting, Trend Analysis, Data Validation, Dashboarding, Node.js
Databases: Oracle SQL, PostgreSQL, MySQL, MongoDB, SSRS, PL/SQL
Data Manipulation & Visualization: Power BI Pandas, NumPy, Matplotlib, Seaborn, Tableau, PySpark, dbt, ArcGIS
Dev Tools: Git, GitHub, Jupyter Notebook, Google Collab, SharePoint, Postman, Jira, VS Code, DataGrip, PyCharm, IDEA
Other Skills: Microsoft Excel, Google Sheets, Data Wrangling, Reporting Automation, Bash Scripting, Windows, Linux

Professional Experience

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

- Designed and implemented **data models** and **Power BI dashboards** using **SQL and DAX**, enabling 5+ university departments to access self-service insights on enrolment, retention, and performance metrics.
- Performed **data extraction, transformation, and quality assurance** on large university datasets (100K+ records) using **SQL and Python**, ensuring accuracy and consistency in reporting.
- Designed reusable **SQL data models** and **DAX** measures to support recurring dashboards and reduce ad-hoc reporting needs.
- Automated parts of the reporting pipeline (Excel + SQL), improving delivery time and reducing manual effort across the team.
- Authored **comprehensive reporting documentation and query guidelines**, cutting onboarding time for new analysts by **25%** and ensuring reproducibility across the OIA team.
- Delivered insights on historical trends, term-over-term shifts, and KPI variations to support strategic planning and data-driven decision-making.

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

- Ensured data integrity by cleaning, preprocessing and validating data in **Oracle SQL** and **MS Excel**, fixing **95%** of data quality issues, and providing detailed analysis reports for price plans, reimbursements and chargebacks.
- Conducted **trend analysis**, **variance analysis**, and **anomaly detection** on high-volume transactional data to identify root causes of discrepancies and reduce recurring issues by **40%.**
- Automated ETL workflows by implementing **Bash/Shell** scripts and **Excel** automation, optimizing data workflows, and reducing manual interventions by 20%, which led to streamlined data onboarding efficiency.
- Developed interactive **Power BI dashboards** to visualize key metrics such as revenue trends, chargeback patterns, and reimbursement cycles, enhancing stakeholder decision-making.
- Built SQL-driven **analytical summaries**, **quality checks**, and **verification reports**, improving visibility into key data metrics.
- Collaborated with business and support teams to interpret data issues, provide actionable insights, and strengthen overall data reliability.

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.
- Optimized loan performance tracking by improving **data processing** and designing **3 interactive dashboards**, enabling better risk assessment and lending pattern identification.

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

- Created a hybrid AI-driven caching system combining **ARIMA** for time-series forecasting and LLM for contextual analysis, improving cache efficiency and reducing latency in high-traffic scenarios.
- Processed **1M+** records with automated preprocessing and built ARIMA/Auto-ARIMA models achieving **RMSE 4.41–4.67**, generating **30-day** traffic forecasts.

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

- Developed a stress prediction model for a mobile app by analyzing **20,500+** user behavior metrics, optimizing and training ML models on digital wellbeing data including screen time, game time, app usage, and social media activity.
- Trained and enhanced model performance by optimizing **Random Forest** and **XGBoost** hyperparameters, increasing classification accuracy to **77%**.

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*