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: Oracle SQL, PostgreSQL, MySQL, MongoDB, PL/SQL
Data Manipulation & Visualization: Pandas, NumPy, Matplotlib, Seaborn, Tableau, Power BI, PySpark, dbt
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: Windows, Linux/Unix (Ubuntu, CentOS), Shell Scripting, Microsoft Excel, Google Sheets

Professional Experience

Student Data Analyst Intern | *Power BI, Oracle SQL, MS Excel, Python, SharePoint* May 2025 – Present
Office of Institutional Analysis, University of Windsor *Windsor, Ontario, Canada*

- Designed and implemented **data models and dashboards in Power BI and SQL**, 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.
- Authored **comprehensive reporting documentation and query guidelines**, cutting onboarding time for new analysts by **25%** and ensuring reproducibility across the OIA team.

Software Engineer | *Power BI, Oracle SQL, dbt, Python, MS Excel, Bash/Shell Scripting, Java, Linux, Jira* 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.
- Reduced issue resolution time by 40% for 100+ customer-reported issues by performing in-depth root cause analysis (RCA) on legacy app server and Oracle database, leading to faster issue identification and customer satisfaction.
- Automated ETL workflows by implementing **Bash/Shell** scripts, optimizing data workflows, and reducing manual interventions by 20%, which led to streamlined data onboarding efficiency.

Projects

Loan Insights Dashboard | *Power BI, SQL, Data Visualization, Dashboard Design, Financial Analytics, Git* Jan 2025
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* Nov 2024
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.
- Automated data preprocessing for 1M+ records, implementing normalization, and missing value imputation leveraging **Python** and **Pandas**, to guarantee reliable data for predictive models, and conducted exploratory data analysis (EDA) on 500K+ web traffic data to identify trends and visualize key insights with **Matplotlib.**
- Built and optimized ARIMA models to forecast future page views, achieving high accuracy with RMSE values of **4.41** for ARIMA and **4.67** for Auto-ARIMA, and generated forecasts for the next 30 days.

Stress Level Prediction Using Machine Learning | *Python, Pandas, Matplotlib, Seaborn, Scikit-learn, Git* Nov 2024
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.
- Refined data quality by implementing data preprocessing, encoding, and normalizing 7+ numerical features utilizing **Python** and **Pandas**, ensuring high-quality inputs for predictive modeling.
- 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*