647-884-9895
pauljared48@gmail.com

Jared Paul

Data Scientist

Portfolio
github.com/jrpaul08
linkedin.com/in/paulj08

Education

McMaster University | Bachelor of Engineering, Software Engineering Graduated: 06/2025

Skills

Tech Stack: Python, SQL, JS, Pandas, Scikit-learn, PyTorch, Power BI, PySpark, React, Flask, MongoDB, MySQL, Matplotlib, Seaborn
Soft Skills: Analytical Thinking, Problem Solving, Communication, Collaboration, Adaptability, Research, Time Management

Technical Experience

Preteckt | Data Scientist 09/2023 – 02/2025

- Performed exploratory data analysis (EDA) on high-frequency telematics data to identify sensor correlations, data anomalies, and fault behavior patterns, producing well-documented reports that combined clear narratives and visualizations to communicate insights and support data-driven decisions.
- Collaborated with vehicle technicians to validate sensor anomalies and diagnostic insights, improving an alerting system generating 300+ daily maintenance alerts with over 80% precision.
- Developed a data preprocessing pipeline to handle missing and extreme sensor values through imputation, outlier filtering, and feature scaling, then applied PCA to enhance data quality and cluster separation for GMM-based anomaly detection.
- Conducted in-depth analysis of GMM internals, examining component means, covariances, and mixture weights to interpret sensor relationships, identify unstable clusters, and refine anomaly thresholds, strengthening model robustness in production.

McMaster University | Research Assistant 05/2023 – 09/2024

- Conducted research on the Nurse Scheduling Problem (NSP), analyzing existing algorithms and methodologies.
- Developed mathematical models and algorithms addressing nurse availability, shift preferences, and regulatory constraints.

Clinimedia | Freelance Full Stack Developer 05/2025 – 09/2025

- Built a full-stack booking platform enabling dental clinics to schedule and manage media days for marketing content.
- Implemented dynamic booking requests, admin approvals, and automated notifications.
- Designed a RESTful API and responsive React frontend, with MongoDB for secure, scalable data management.

Personal Projects

MULTI-LABEL CHEST X-RAY CLASSIFICATION | *Python, PyTorch, NumPy, Matplotlib, React, Express.js* 09/2024 – 07/2025

- Built a multi-label classification pipeline to detect 13 thoracic diseases from the NIH ChestX-ray14 dataset using data preprocessing, augmentation, and class rebalancing to handle abnormal data.
- Trained and optimized a MobileNetV2 model with hyperparameter tuning and dynamic thresholding, achieving 0.80 AUC, 50% accuracy, and 43% recall.
- Performed EDA and model evaluation using ROC-AUC, precision, and recall to validate diagnostic performance.
- Developed a diagnostic report generator with bounding box visualizations that communicates valuable insights.

RETAIL CUSTOMER SEGMENTATION — RFM ANALYSIS | *Python, TS, pandas, scikit-learn, matplotlib, React, Recharts* 08/2024

- Performed end-to-end RFM analysis and exploratory data analysis (EDA) to derive Recency, Frequency, and Monetary features and uncover actionable customer segments.
- Detected and separated outliers using the IQR method to preserve high-value customers while maintaining clustering quality.
- Implemented K-Means clustering with feature scaling, using the Elbow method (inertia) and Silhouette score to determine the optimal number of clusters, resulting in four primary segments plus three specialized outlier clusters.
- Derived key business insights, including ∼25% of customers generating most revenue and a 35% “Nurture” segment showing growth potential, with an interactive dashboard enhancing interpretability for targeted marketing and retention.

NBA Oracle | *Python, JS, Pandas, Scikit-learn, Playwright, BeautifulSoup4, Flask* 06/2025 - 09/2025

- Developed an end-to-end machine learning pipeline to predict NBA game outcomes from over 3,400 games across three seasons.
- Built a web scraping system using BeautifulSoup and Playwright to asynchronously collect and structure team and player statistics into a time-series dataset.
- Engineered features, including rolling averages of team performances and temporal aggregates capturing team form, offensive/defensive efficiency, and recent trends.
- Trained a Ridge Classifier with sequential feature selection and time-series cross-validation, achieving 71% prediction accuracy.