Yash Soni

Toronto, ON

📞 (647) 853-8424 ✉️ yashsonig@gmail.com in LinkedIn ⌂ GitHub

EDUCATION

Wilfrid Laurier University   **Expected Graduation: April 2027**
B.Sc in Data Science | Big Data Concentration | Economics Minor   *Waterloo, ON*

TECHNICAL SKILLS

Languages: Python, SQL, R, Java, C/C++, VBA, JavaScript
Developer Tools: DataBricks, Microsoft Azure, GitHub, VS Code, PowerBI, AWS, Excel, Jupyter Notebook, Eclipse
Data Tools & Libraries: PySpark, Pandas, NumPy, scikit-learn, Matplotlib
Relevant Coursework: Artificial Intelligence, Machine Learning, Data Visualization, Data Structures I/II

EXPERIENCE

Manulife Financial   **September 2025 – December 2025**
Data Analyst Co-Op   *Toronto, ON*

- Leveraged **PySpark and SQL in Databricks** to process financial data for the Global Wealth and Asset Management team, providing insights that helped the Risk Analyst team achieve a **9% return**.
- Automated key data processes, **reducing processing time by 30%** and enhancing scalability, while collaborating with various other departments for seamless integration.
- Designed and implemented reporting processes **using PySpark to monitor and evaluate the efficiency** of core operations and workflows, resulting in a 25% improvement in identifying and addressing process bottlenecks.
- Converting legacy R scripts into Python to **enhance code maintainability** and **align with current tech stack**.
- **Collaborating with cross-functional teams** to gather requirements and deliver actionable insights that informed decision-making for portfolio management.

Wilfrid Laurier University   **September 2024 – April 2025**
Teaching Assistant   *Waterloo, ON*

- Supported 200+ students across DATA100 and MA103 by **delivering tailored walkthroughs & debugging code**.
- Provided individualized assistance in lab sessions, helping students **build functions, troubleshoot statistical models, and interpret regression outputs** in real time.
- **Assessed and graded over 200 assignments** spanning code scripts, math problem sets, and written analysis, ensuring fairness and personalized feedback under tight timelines.

RBC   **May 2024 – August 2024**
Operations Officer Co-Op   *Toronto, ON*

- Processed daily mortgage disbursements **exceeding $2.5M with 97% accuracy** under regulatory timelines.
- Built Excel-based tracking tools to **reduce escalation time** for corrupted records by 20%.
- **Supported reconciliation teams** by flagging data inconsistencies and ensuring audit-ready records.

PROJECTS

Optimal Workout Case Study | *PySpark, Pandas*

- **Engineered a data-driven analytical framework** to **determine optimal workout splits** for achieving a lean, muscular physique, leveraging variables such as height, weight, body fat percentage, and gender.
- Extracted and integrated raw fitness datasets from online repositories using PySpark and pandas, **performing extensive ETL processes** for large-scale data cleaning, transformation, and aggregation.
- Performed EDA utilizing NumPy and Matplotlib to **identify patterns and correlations** between workout regimens and body composition outcomes.
- Developed a modular analytics pipeline to facilitate reproducible analyses and **ensure scalability for future applications**, such as personalized workout recommendation systems.

Credit Card Fraud Detection System | *Python, scikit-learn, Pandas, NumPy*

- **Developed a machine learning model** to **identify potentially fraudulent credit card transactions** by analyzing features such as credit score, transaction history, and spending behavior.
- Applied weighted probability distributions to **calculate risk scores** to identify fraudulent transactions.
- Performed data preprocessing, feature engineering, and model evaluation to **improve prediction accuracy.**
- Gained **hands-on experience in fraud analytics**, supervised learning, and real-world financial data handling.