Bilal Bouchemella . Jarvis Consulting

I am a Master's student in Data Science and Business Analytics at HEC Montréal and a Computer Science graduate from McGill University, where I focused on data engineering, data analysis, and machine learning. I currently work as a Data Engineer at Jarvis, where I build data pipelines, develop software components, and work with cloud and analytics technologies. I previously worked as a Data Scientist and an IT Technician, using technologies such as Python and Julia to build analytical workflows and automate data tasks, while gaining experience in machine learning, data processing, and system operations. These experiences strengthened my passion for leveraging data, machine learning, and AI to create practical, reliable solutions that address real-world challenges.

Skills

Proficient: Python, Java, RDBMS/SQL (PostgreSQL), Julia, Pandas/Numpy, Scikit-learn, Linux/Bash, Git, Agile/Scrum

Competent: Cloud platforms (AWS, GCP), Docker, Jira, Power BI, R

Familiar: Spark, DataBricks, Snowflake, C, C++

Jarvis Projects

Project source code: https://github.com/jarviscanada/jarvis_data_eng_BilalBouchemella

Python Customer & Performance Analytics PoC [GitHub]: Analyzed two years of e-commerce transaction data to identify customer behavior, purchasing patterns, seasonality, and customer value. Built an analytics pipeline using Python with Pandas, NumPy, and Matplotlib in Jupyter Notebook, connected to PostgreSQL and Docker. Performed data wrangling and feature engineering including type casting, handling missing values, and creating monthly and behavioral metrics to produce structured analytical datasets. Computed RFM metrics, monthly sales trends, and active user insights to distinguish new versus returning customers, and delivered visual insights to support targeted marketing campaigns, personalized promotions, and customer retention strategies.

Java Grep App [GitHub]: Developed a Java-based command-line application using object-oriented programming principles to replicate the core functionality of the Linux grep command. The application recursively traverses user-specified directories, reads files line by line, and applies user-defined regular expression patterns using Java Streams and Lambda expressions to efficiently filter matching lines before writing results to an output file. Maven was used for project structure and dependency management, with GitHub for version control. The application was manually tested and debugged, and unit tested using JUnit. For deployment, it was packaged as a runnable JAR and distributed via Docker using a custom Dockerfile.

Linux Cluster Monitoring Agent [GitHub]: Built a Linux monitoring agent that collects hardware details and tracks CPU, memory, and disk usage across several machines. Used Bash scripts, Docker, PostgreSQL, and cron to store and centralize system data, helping analyze performance and plan future resource needs.

Highlighted Projects

Machine Learning Framework for Credit Risk Modeling: Designed a machine learning framework to predict credit default risk using financial and behavioral data. Engineered advanced features such as financial ratios, interaction terms, and normalization using Python with Pandas and NumPy to improve model performance. Trained and compared three models including Logistic Regression, Random Forest, and XGBoost using cross validation and evaluated performance with ROC AUC, precision recall, and KS statistic via Scikit-learn. Analyzed model performance and errors to select an optimal approach and support data driven risk decisions, and created visualizations using Plotly to present model results and performance comparisons.

News Coverage Analysis of Gladiator II: Analyzed news coverage of Gladiator II by collecting hundreds of film-related articles using Python and NewsAPI, applying open coding and manual annotation to classify themes, computing TF-IDF scores to extract key terms, and visualizing coverage trends with JupyterLab and Matplotlib to compare the film's visibility with other releases. This project strengthened my skills in text analysis, feature extraction, data collection pipelines, and collaborative workflow management through regular coordination and shared responsibilities.

Hotel Database Application: Developed a full hotel database system as part of a team project by designing the schema, creating and managing SQL tables, and building a Java application with a user-friendly interface for queries, updates, and

operations. Worked closely with teammates where communication and coordination were essential, and took responsibility for organizing tasks, splitting work, and ensuring the database and application components integrated smoothly.

Professional Experiences

Data Engineer, Jarvis Consulting Group (Nov 2025 - present): Developing data engineering solutions by building software components, creating data pipelines, and working with cloud and analytics technologies. Contribute to coding, debugging, and data processing tasks while applying strong problem-solving, communication, and code quality practices across projects.

Data Scientist, R2 (May 2025 - Aug 2025): Built and optimized machine learning models in Python and Julia for time-series predictive analytics on large industrial datasets, and engineered data cleaning and preprocessing pipelines to transform raw sensor data into reliable features. Designed and ran experiments to evaluate models with a balance of accuracy and efficiency, and developed reproducible workflows that supported collaboration across the data science team. Strengthened my experience with Linux-based development, Git for version control, and working with large-scale data in a research-focused environment.

Information Technology Technician, Tennis Canada (May 2024 - Apr 2025): Maintained and upgraded network systems in collaboration with IT teams, provided comprehensive technical support on Mac and Windows to improve staff productivity, and managed the office database to ensure accurate and accessible operational data. Used JIRA to track issues and coordinate tasks, and worked closely with colleagues to configure and maintain over 100 laptops for the 2024 Omnium Banque National tournament, ensuring all devices were fully operational and met event requirements.

Education

HEC Montreal (2025 - present), Master of Science (MSc), Data Science & Business Analytics - Student-Athlete: Varsity Men's Soccer - Relevant Coursework: Machine Learning, Statistical Modelling, and Statistical Learning

McGill University (2021 - 2025), Bachelor of Science (BSc), Computer Science - Student-Athlete: Varsity Men's Soccer Team Captain - McGill Redbird Athlete Award (2021, 2022, 2023): Awarded to varsity student-athletes who excel in their sport - Relevant Coursework: Data Science, Natural Language to Data Science, Applied Machine Learning, Reinforcement Learning, Database Systems, Software Design, Algorithms and Data Structures, Operating Systems, Discrete Structures, Probability, Statistics, and Statistical Computing

Miscellaneous

- Varsity Men's Soccer Team Player of the Year - McGill Athletics (2022)
- University Recruitment Bursary - Fondation Aleo (2021): Awarded to the best student-athletes in the province for excelling in both sport and academics at a Quebec institution
- Semi-professional soccer player with over six years of experience in Quebec. I have been playing since the age of five, developing strong values such as leadership, teamwork, discipline, and resilience through continuous training and competition.