

Prabhnoor Kaur Karwal

+1 6047231657 | prabhi.karwal4@gmail.com | [LinkedIn](#) | [Portfolio](#)

EDUCATION

University of British Columbia

Bachelor of Science, Major in Data Science
Dean's List

Vancouver, BC

GPA: 3.95/4.33

2024-2025, 2025-2026

TECHNICAL SKILLS

Languages - SQL (joins, aggregations, window functions, CTEs, subqueries), Python (pandas, NumPy, TensorFlow, PyTorch, scikit-learn, Altair, Matplotlib), R (ggplot2, statistical analysis, machine learning), HTML, CSS, Java (Object Oriented Programming)

Tools and Frameworks - Tableau, Excel, PowerBI, MySQL, PostgreSQL, Git, GitHub, CI/CD, Databricks, AWS, Azure, APIs

WORK EXPERIENCE

Computational Research Assistant | UBC Faculty of Pharmaceutical Sciences

May 2026 - Present

- Built a dataset for ML model training by querying the ChEMBL API, applying multi-step filtering and data validation pipelines in R to extract and integrate molecular descriptors and plasma protein binding values, with over 95% data completeness
- Developing a Random Forest classification model to predict the milk-to-plasma ratio for drugs, engineering an end to end R package and a Shiny web application to standardize the full API-to-prediction workflow for use by pharmaceutical researchers

Program Assistant | UBC Masters of Data Science

May 2026 - Present

- Automating the creation of a multi-calendar academic session schedule through Python scripts using the Google Calendar API to programmatically build, populate, and publish calendars, eliminating manual entry and reducing setup time significantly
- Coordinating program administration and event planning logistics while managing collaborative workflows using GitHub Issues and pull requests to track deliverables, review contributions, and ensure timely, organized project execution

Machine Learning Intern | RJH Biosciences

January 2026 - March 2026

- Engineered a structured dataset for FLT3 pathogenicity classification by parsing HGVS variants into genomic features using regex; merged clinical datasets to eliminate duplicate leakage and improved minority class representation from 7% to ~23%
- Built a ML pipeline training 3 models achieving 95.9% recall, deployed it via a [web application](#) enabling real time predictions

Centre Assistant | Kumon-Sullivan

September 2025 - January 2026

- Regularly tutored students on Mathematics and Reading concepts, explaining concepts in easy ways and clarifying doubts
- Tracked performance data, maintained structured records and prepared organized student's progress reports weekly

WorkLearn Project Specialist | UBC Chemical and Biological Engineering

September 2025 - January 2026

- Operated XRF analyzer and processed 10 - 20 samples monthly adhering to project protocols and generating analytical data
- Prepared concise analytical summaries and maintained accurate documentation to support project planning and data integrity

Salaried Notetaker | Centre for Accessibility, UBC

October 2024 - June 2025

- Provided comprehensive notetaking for MATH 100, DSCI 100, and CHEM 121, ensuring accessibility for all students
- Demonstrated exceptional written communication skills by delivering clear and concise notes within 48 hours of each lecture and Collaborated with the Centre for Accessibility to support students with diverse learning needs

Data Analyst Intern | Radar International

February 2023 - April 2024

- Performed data preprocessing, cleaning and validation on structured datasets (10,000+), ensuring data quality standards
- Built Excel and Python dashboards translating cleaned datasets into weekly KPI summaries, enabling data driven decisions

PROJECTS (selected)

Telecom Customer Churn Analysis Dashboard | Python, Tableau | [Link](#)

- Performed churn analysis on 7,043 customer records; engineered CLV and other features to enable risk profiling.
- Trained XGBoost churn prediction model achieving 83.2% AUC; applied SHAP to identify top churn drivers, Identified 52.9% churn rate among 0-6 month and quantified \$139K+ in monthly revenue at risk via interactive Tableau Public dashboard

Credit Risk Analysis & Default Prediction | PostgreSQL, Python | [Link](#)

- Designed a normalized relational schema in PostgreSQL and wrote 15+ analytical SQL queries using CTEs, window functions, and conditional aggregations to profile 150K+ borrower records across risk tiers in a financial services context
- Engineered features from SQL outputs, trained an XGBoost classifier achieving 99.5% AUC to predict loan default probability

NLP & Survival Analysis for HR Retention - BOLT Dathathon (Semi Finalist) | Pandas, Matplotlib | [Link](#)

- Built a full pipeline across 5 datasets, engineered features, ran VADER sentiment NLP on Google reviews, computed Kaplan-Meier retention curves, and applied logistic regression and chi-square tests to model high-performer exit probability
- Delivered three quantified recommendations projecting 15-20% reduction in avoidable exits and \$2M+ in cost savings

Vancouver City FC - Revenue Growth Analysis | Python, VS Code | [Link](#)

- Performed ETL, cleaning, and preprocessing on 3 key datasets, ensuring accurate data assets for stakeholder facing analysis
- Developed data driven visualizations using Matplotlib and did trend analysis to support revenue reporting across key areas

Breast cancer classification model | Pandas, Scikit-learn, Matplotlib, JupyterNotebooks | [Link](#)

- Built K-NN classifier for breast cancer diagnosis achieving 95% test accuracy, utilizing scikit-learn pipelines with feature standardization and grid search hyperparameter optimization across 10-fold cross-validation

- Implementing stratified train-test split to handle class imbalance, resulting in only 3 false negatives (5.7% miss rate)

Newsletter subscription prediction | *R, JupyterLab, Tidyverse, Ggplot* | [Link](#)

- Cleaned and standardized game datasets, performed mean imputation, and visualized usage behaviors
- Built and tuned a KNN model with standardized features, 5-fold cross-validation, and upsampling ($K = 1-40$), achieving 87% precision and identifying 15 - 25 year-old high playtime users as the most likely to subscribe

Financial risk calculator | *Java, VS Code, GitHub, JSON* | [Link](#)

- Built a Java application with interacting classes, implemented JSON-based data persistence with read/write methods
- Developed an interactive Swing UI with buttons, dropdowns, and input validation, enabling users to add, filter, delete, save, and load financial reports, supported by unit tests (JUnit) and version control through GitHub

Infection Mortality Patterns Dashboards | *Altair, GitHub, JupyterNotebooks* | [Link](#)

- Worked in a team to analyze, clean and transform IHME and FAO datasets to explore how mortalities for different diseases are affected through different factors such as research and development, funding and pesticide usage
- Designed interactive Altair dashboards with filters and linked views to support exploratory data analysis and explain all mortality trends

VOLUNTEER LEADERSHIP ROLES

Data Science Student Representative | *UBC Data Science Club*

March 2026- Present

- Monitored and analyzed student engagement in events, maintaining records to inform club planning and improve participation
- Coordinated event logistics, facilitating communication between students, faculty and external partners for smooth execution

Corporate Relations Coordinator | *UBC Computer Science Student Society*

September 2025 - Present

- Fostering partnerships with tech companies and sponsors, managing outreach and communications for events and funding
- Assisted with the outreach for UBC's Technical Career Fair 2026, a premium tech recruitment event featuring 30+ employers



science.coop@ubc.ca | 604-822-9677