Aarav Mahajan

5th Year, Statistics and Data Science

+1-672-968-0912 | mahajanaarav2020@gmail.com | linkedin.com/in/amahaj02 | github.com/amahaj02

Education

University of British Columbia — Vancouver, BC

Bachelor of Science, Major in Statistics, Minor in Data Science — Expected May 2027

- **Relevant Courses:** Databases in Data Science, Intermediate Algorithm Design, Object-Oriented Programming, Containerized Workflows for Data Science, Applied Machine Learning, Cloud Computing for Data Science
- **Activities:** UBC Academic Essentials (Math Course Facilitator), Undergraduate Statistics Society (Events Director), AMS Elections (Polling Officer)

Work Experience

Data Platform Engineer Co-op — Sep. 2025 – Aug. 2026

Sanofi AI Centre of Excellence — Toronto, ON

- Built regression testing and production monitoring for an enterprise AI data-search platform using RAG, tracing 19 agent methods and 4 scorers with W&B Weave, GitHub Actions, Amazon Bedrock and Snowflake Cortex.
- Led infra upgrades for a production data platform, migrating 40+ Lambda functions and 25+ Airflow pipelines to Python 3.13 and rebuilding EMR Serverless deployments with ECR, Docker, Terraform and GitHub Actions.
- Automated Snowflake infrastructure deployment with least-privilege access controls for a Streamlit app serving 3,000+ monthly users, cutting data-publishing time from 30+ days to 7 days.
- Developed and deployed an Airflow-based ingestion pipeline (Lambda, dbt) integrating Informatica CDGC into Snowflake, enabling real-time observability across 8,000+ data assets for enterprise reporting.
- Won the annual Snowflake × Sanofi Hackathon, building the data layer and two semantic views powering Cortex Analyst tools in a sales and inventory Cortex Agent now under evaluation for enterprise-wide use.

REX Committee Events Director — Sep. 2024 – Apr. 2025

Undergraduate Research Opportunities, UBC — Vancouver, BC

- Recruited 90+ graduate research mentors and built a Gale-Shapley matching tool pairing them with 400+ undergrads, increasing mentor participation by 40%.

Department Liaison — Sep. 2024 – Apr. 2025

UBC Statistics — Vancouver, BC

- Represented 550+ students across faculty and administration, resolving 90+ academic and registration cases per term while reducing response times by 25% and increasing departmental engagement by 20%.

Research Assistant — Sep. 2023 – Apr. 2024

Undergraduate Research Opportunities, UBC — Vancouver, BC

- Implemented physics-informed neural networks (PyTorch) for inverse differential-equation problems, optimizing physics-based loss functions to reach $\sim 0.1\%$ error vs. analytical solutions across 80+ robustness experiments.

Technical Projects

URL Shortener | *Kubernetes, Prometheus, Grafana, Redis, PostgreSQL, cert-manager* — Mar. 2026 – Apr. 2026

- Deployed a fault-tolerant, high-availability URL shortener on Kubernetes with ingress-nginx load balancing, Flask/Gunicorn multi-pod architecture, and TLS termination via cert-manager.
- Instrumented full observability stack with Prometheus scraping and Grafana dashboards tracking 4 golden signals (latency, traffic, errors, saturation), with live Discord alerting.

AWS Infra Automation Suite | *Terraform, Lambda, EventBridge, IAM, GitHub Actions* — Dec. 2025 – Sep. 2026

- Built a reusable Terraform bootstrap layer for S3 remote state and GitHub Actions OIDC, provisioning least-privilege IAM roles with permissions boundaries through approval-gated workflows.
- Implemented CI/CD for scheduled Lambda workloads—automated testing, Docker builds, and Terraform deployment, with EventBridge triggering and CloudWatch observability

Telehealth Disparities Analysis: Socioeconomic and Demographic Insights — Jan. 2025 – Apr. 2025

- Implemented z-tests in SQL with Python result extraction, finding Black non-Hispanic individuals represented 26.0% of telehealth usage versus 12.4% of the population ($p < 0.0001$).
- Engineered a relational database schema from raw RANDS and Medicare datasets, conducted multi-table joins and z-tests in SQL (MySQL and OracleDB), and extracted result tables via Python for statistical interpretation.

Technical Skills

Languages: Python, SQL, R, JavaScript, TypeScript, Bash, HCL
Libraries: boto3, Snowpark, PySpark, scikit-learn, FastAPI, NumPy, pandas, Matplotlib
Tools: Git, Docker, GitHub Actions, Terraform, AWS (Lambda, S3, EMR, ECR, CloudWatch, DynamoDB, SSM, IAM), OpenShift (ROSA), Snowflake, Airflow, dbt, schemachange, Spark, PostgreSQL, MySQL, MongoDB, Hugging Face, Jira
Certifications: Snowflake SnowPro Associate