

Luan Nguyen

778-668-3009 | luan.nguyen_3@sfu.ca | [Portfolio](#) | [LinkedIn](#) | [GitHub](#) |

TECHNICAL SKILLS

Languages & Tools: Python, SQL, MongoDB, R, Jupyter, Tableau, Snowflake, dbt, Salesforce, MS SQL Server
DS & ML: supervised/unsupervised learning, regression, classification, clustering, A/B testing, time-series analysis
Deep Learning: NLP(tokenization, Word2Vec, sentiment analysis), CV (CNN, image classification)
Automation & Scripting: Tableau API, encryption, task scheduling, web scraping, data report

EXPERIENCE

- Data Scientist Co-op** | *Snowflake, dbt, Tableau, TF-IDF, Word2vec, Cluster Workstream* Sep. 2024 – Dec. 2024
British Columbia, Canada
- Collaborate with cross-functional teams, including marketing, sales, and product engineering, to optimize metrics such as customer acquisition cost, churn, and customer lifetime value through data-driven insights
 - Design, optimize, and maintain data pipelines and ETL processes in Snowflake, dbt, Hevo, and Hex, ensuring data integrity and accessibility for stakeholders across the organization.
 - Apply NLP techniques, including Word2Vec and clustering algorithms, to analyze sales team notes on downsell, uncovering key themes and insights that informed strategic improvements and optimized sales strategies.
 - Conduct rigorous statistical analyses and A/B testing on hiring product, revealing key factors that reduced time-to-hire and improved candidate quality.
 - Create dynamic and interactive data visualizations in Tableau to present insights on customer behavior trends and product performance to senior leadership, enhancing data accessibility and decision-making
- Data Analyst Co-op** | *Databrick, Azure, MS SQL Server, Tableau* Jan. 2024 – May. 2024
Vancouver Coastal Health, Medical Quality Practice Dept. British Columbia, Canada
- Conducted data cleaning and integration using pandas and pySpark, improving data quality and reducing inconsistencies, critical for a new physician performance evaluation system
 - Designed and implemented ETL pipeline using PyODBC to connect MS SQL Server and Denodo, automating data extraction and transformation processes to ensure real-time updates
 - Automated the encryption and distribution of quarterly performance reports for 500+ physicians using Python and Tableau API, reducing manual effort and improving reporting efficiency
 - Collaborated with data scientists and engineers to develop predictive models for patient readmission, length of stay, and mortality rates, enhancing data-driven decision-making and patient care strategies
 - Performed A/B testing and time-series analysis on physician quality improvement projects, identifying trends that influenced strategic initiatives

HIGHLIGHTED PROJECTS

- Data Engine Pipeline** | *Docker, PostgreSQL, dbt, Airflow, Git* | [Source](#) Dec. 2023
- Developed a robust data pipeline for PostgreSQL using Python's subprocess for shell command execution.
 - Optimized SQL with CTEs and dynamic SQL generation using Jinja in dbt, ensuring efficient transformations
 - Integrated Airflow for distributed computing, focusing on scalability, reliability, and resource allocation in workflow
- Probabilistic Models** | *TensorFlow, Probabilistic Model, Git* | [Source](#) Aug. 2023
- Performed feature extraction and hypothesis tests using MANOVA, Chi-square test, identifying relevant attributes
 - Conducted regression and time series analysis, investigate trends and patterns
 - Developed binary classification models (KNN, Naïve Bayes, Logis. Regress., SVM, Neural Net.) using TensorFlow
 - Evaluated models using multiple techniques (train-test split, cross-valid., bootstrapping), achieved 87% accuracy
- Other Projects** | *Scikit-Learn, PyTorch, R, HTML/CSS/JS, C++, Java* | [Source](#) 2018 – 2023
- Web scraping, features extraction/analysis, prediction model, report generation, web/game development

EDUCATION/PROFESSIONAL DEVELOPMENT

- BASc, 4th Year, Computer Science Major, Applied Statistics Minor** Sep. 2025
Simon Fraser University Burnaby, BC
- GPA: 3.5, Dean's Honour Roll 2022-2023
 - Concentration: Machine Learning
 - Relevant coursework: ML, DS, KDD (3rd year); Probabilistic Analysis (3rd year); DSA (3rd year)
 - Data Analytic Certificate: Python/R data wrangling; Tableau data visualization
 - Social Activities: SFU Canadian Cancer Society Chair (2022-2023)
- Data Structure and Algorithm Cert.** Sep. 2023
University of California San Diego remote, Coursera
- Skills: Algorithms (Greedy, Divide-and-Conquer), Dynamic Programming, Debugging, Software Testing