

# Alex Liang

[ala206@sfu.ca](mailto:ala206@sfu.ca) | 778-713-1775 | [alexanalytics.me](http://alexanalytics.me) | [github.com/Alex-Liang01](https://github.com/Alex-Liang01) | [linkedin.com/in/alex-liang-analytics/](https://linkedin.com/in/alex-liang-analytics/)

## Skills

**Programming:** Python, R, SQL

**Databases:** MSSQL, MySQL, PostgreSQL, Snowflake, MongoDB

**Data Visualization:** Tableau, Power BI, Matplotlib, Seaborn, Plotly, ggplot2

**Machine Learning Frameworks:** Scikit-learn, TensorFlow, PyTorch, XGBoost, LightGBM, Caret

**Machine Learning:** Classification, Clustering, Dimension Reduction, Ensembles, Factor Analysis, Generative Models, Natural Language Processing, Neural Nets, Regression, Reinforcement Learning, Time Series

**Data Processing and Data Streaming:** Apache Spark, Apache Airflow, DBT, Kafka

**Statistics:** Experimental Design, Sampling Techniques, Probability

**Dashboard Frameworks:** Shiny, Dash, Streamlit

**Tools:** AWS, Docker, Git, LangChain, LangGraph, Microsoft Office

## Work Experience

### Junior Data Quality Coordinator I Co-op

(May 2023 - Dec 2023)

#### Policy Reporter

- Ensured data quality assurance for a SaaS-based central database of healthcare medical documents from various American insurance payers
- Identified data quality control issues in the database such as duplicates, document mismatches and reversions reducing errors in the database
- Completed Index Mapping project one quarter in advance of schedule, allowing for automation of updating existing policies within the database reducing policy update turnover by 35%
- Provided support on inquiries on SQL queries addressed to the data quality assurance team ensuring efficient resolutions of data-quality related issues
- Successfully managed multiple projects from different supervisors on data quality assurance issues within the database simultaneously, meeting deadlines in a high-pressure environment
- Collaborated with senior data engineers and data quality coordinators to identify optimal ways to efficiently store niche documents within the database ensuring data completeness
- Maintained multiple data quality assurance Slack channels, effectively communicating with both technical and non-technical stakeholders by adapting communication styles as needed

## Personal Projects

### NBA Analytics

(Jan 2025 - Present)

- Extracted, transformed, and loaded player performance data for the 2024-25 season from an NBA API and webscraping into a PostgreSQL database for analysis
- Designed an Entity-Relationship diagram to structure NBA API endpoints into a star schema to optimize Tableau dashboard query performance
- Automated data pipelines using Apache Airflow and Python to ingest NBA data daily, allowing for up-to-date analysis
- Built interactive Tableau dashboards to visualize key metrics in NBA games, allowing for faster insights into team and player performance
- Created a natural language SQL agent using LangChain, LangGraph, and the OpenAI API that translates user input into executable SQL queries and visualizes the results with Matplotlib allowing for faster exploratory data analysis

# Alex Liang

[ala206@sfu.ca](mailto:ala206@sfu.ca) | 778-713-1775 | [alexanalytics.me](http://alexanalytics.me) | [github.com/Alex-Liang01](https://github.com/Alex-Liang01) | [linkedin.com/in/alex-liang-analytics/](https://linkedin.com/in/alex-liang-analytics/)

## Academic Projects

### Maintaining Data Governance through Synthetic Data

(Jan 2025 - April 2025)

Analytics Project (Capstone Course: BUS 439)

- Collaborated in an agile team of five students working with Electronic Arts (EA) to assess whether synthetic data can support data governance in cross-functional teams
- Simulated datasets using Python with varying amounts of missing data, sample size, and association strengths to evaluate synthetic data fidelity and scalability
- Generated synthetic datasets using neural net models and statistical methods using the Synthetic Data Vault library in Python
- Evaluated the fidelity of synthetic data by training LightGBM classifiers and comparing pairwise correlations between synthetic and simulated datasets to assess similarity in feature relationships
- Evaluated the privacy protection of synthetic datasets by applying various privacy evaluations metrics to measure the extent of personally identifiable information present
- Presented findings on how synthetic data trained with neural nets performed better with more complex training data in terms of data fidelity but worst in terms of data privacy

### Beedie Credit Enrollment Analysis

(Jan 2025)

Analytics Project (Capstone Course: BUS 439)

- Analyzed Beedie enrollment data to recommend a strategy to help students graduate on time with a diverse skill set to succeed in the business world
- Ingested data from multiple sources using web scraping into a normalized SQL database
- Wrangled data into a clean form ensuring accuracy and quality for regression machine learning applications using ensembles and regularized regression techniques in Python
- Developed an informative and interactive Tableau dashboard that visualized key findings that students with two or more concentrations graduate faster with a more diverse skill set
- Presented recommendation of enforcing two Business concentrations to graduate to both technical and non-technical stakeholders, winning grand prize out of four teams

### Hotel Booking Cancellations in Python

(Oct 2024 - Dec 2024)

Customer Analytics (BUS 445)

- Analyzed Hotel Booking data to identify important factors that affect hotel booking cancellations
- Gathered requirements through stakeholders engagement before starting analytics to clarify business objectives and ensure that concerns are addressed
- Developed a data pipeline to preprocess and clean data to ensure quality control for data analytics
- Visualized trends in the data using Matplotlib in Python to support data driven decisions
- Built machine learning classification pipelines to uncover hidden hotel cancellation factors
- Presented key findings and recommendations to reduce hotel booking cancellations to managers

## Education

Simon Fraser University

(Sep 2019 - Present)

**Major: Data Science** Bachelor of Science

CGPA (3.39/4.33)

## Certificates

Amazon Web Services: AWS Certified Data Engineer - Associate

(Jun 2025)