Abdullah Naeem

abdullahnaeem@hotmail.ca | Richmond, BC | (604)-355-3806 | LinkedIn | GitHub

TECHNICAL SKILLS

Languages: Python, R, SQL, C, C++, HTML, CSS, MATLAB

Libraries & Frameworks: Pandas, NumPy, Scipy, Scikit-Learn, Dask, PySpark, Hadoop, tidyverse

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

Database & Cloud Tools: MySQL, SQLite, Firebase, Google Analytics 4

Developer Tools: Git, GitHub, VSCode, RStudio, PyCharm, Jupyter, Slack, Jira

EXPERIENCE

Statistics Canada — Montreal, Canada
Statistics Analyst Intern — *January 2023 - April 2023*

- Performed a compression analysis on a variety of file types and reduced file sizes by 40%, leading to cost savings in cloud services costs and speeding up workflows.
- Resolved issues in the internal StatsCan record matching platform (CanSIS) related to repeated record identifiers, improving matching accuracy by 15% and leading to a recommended solution that was implemented by my division.
- Improved an address parser by utilizing regular expressions to successfully identify 365,169 incorrect address formats, boosting the tool's overall accuracy and reliability.

Data Science Student Society — Burnaby, BC
Director of Events — *April 2022 – April 2025*

- Organized and led 5+ events for the Data Science program at SFU, including discussion panels with industry professionals and technical workshops, helping to strengthen community engagement and enhance learning opportunities for Data Science undergraduates.
- Developed targeted outreach strategies that doubled student engagement for events, leading to consistently high turnout.
- Supported the onboarding for new team members, helping to build a more connected student community.

PROJECTS

Analysis on Customer Lifetime Value (CLTV) | *Python, ironSource, Google Analytics 4 (GA4)*

- Estimated CLTV for players of a mobile game developed by a start-up through implementing a Python pipeline to clean, transform, and analyze user behavior data collected from ironSource and GA4 over a 6-week period.
- Developed business insights by visualizing CLTV and churn rate trends using time-series plots, helping to optimize acquisition campaigns for new users.
- Proposed user segmentation strategies and stopping rule thresholds to guide ad spend and budgeting for future ad campaigns.

SFU Transit Time Estimator | *Fall Hacks 2ⁿᵈ Place* | *Link*

- Collaborated with a team of 3 developers to design an application that estimates commute time between W Hastings St, Vancouver and SFU Burnaby using bus arrival data from the TransLink API.
- Parsed XML data by sending GET requests to the API using BeautifulSoup, improving the accuracy of travel time estimates.
- Achieved second place out of 20+ competing teams, demonstrating prototype efficiency and effective teamwork.

Customer Churn Prediction | *Python, Scikit-Learn, Scipy, Seaborn* | *Link*

- Followed the data analysis pipeline using the Scikit-Learn and Scipy frameworks to build a churn prediction model for a telecommunications company using Gradient Boosted Trees, achieving a 90% accuracy score on test data.
- Analyzed high importance features such as monthly bill, data usage, call length, and call frequency to optimize model input during preprocessing.
- Evaluated results from K-Nearest Neighbours, Decision Trees, and Random Forests for model comparison.

EDUCATION

Simon Fraser University — Burnaby, BC
BSc, Data Science major, Minor in Computer Science — *September 2021 – June 2025*

- Department of Statistics and Actuarial Science Award recipient: March 2024

University of British Columbia — Vancouver, BC
BA, Computer Science major — *September 2020 – August 2021*