

# Jake Barnabe *Data Scientist*

✉ jake.barnabe@queensu.ca

☎ (613) 898-7113

🌐 LinkedIn

🐙 GitHub

🔗 Website

## PROFILE

Creative, analytically-minded Data Scientist passionate about uncovering and communicating the story hidden in data. Experienced building data pipelines and communicating insights derived from data. Skilled using Python and R for statistical analysis and modelling. Enthusiastic team member and self-starter. Excellent written and oral communicator.

## EDUCATION

### Master of Data Science

*University of British Columbia*

09/2023 – 06/2024

Vancouver, Canada

- Relevant courses: Statistics, Data Wrangling, Data Visualization, Predictive Modelling, Machine Learning, Databases, Algorithms

### Bachelor of Arts (Honours) - Applied Economics

*Queen's University*

09/2016 – 04/2021

Kingston, Canada

- Relevant courses: Behavioural Economics, Psychology, Statistics, Econometrics

## PROJECTS

### Predicting Player Performance

05/2024 – 06/2024

*UBC MDS Capstone Project with Vancouver Whitecaps FC*

- Constructed an end-to-end pipeline using Python, R, and SQL featuring a series of linear mixed-effects models to predict first-team player performance from match data resulting in an increase in prediction accuracy of 19.5%.
- Identified key factors contributing to improved player performance to inform coaches' decision making.
- Voted "Best Capstone Presentation" by the class.

### Dreamhouse - Real Estate Dashboard

04/2024 – 04/2024

*UBC MDS Group Project*

- Built a dashboard using Dash to present an interactive visualization of housing prices across the top 50 most populous cities in the United States of America.
- The dashboard enables users to filter and explore the housing market across various dimensions including state, city, house type, square footage, number of bedrooms and bathrooms, and more.

### Predicting high-potential FIFA players using individual performance data

11/2023 – 12/2023

*UBC MDS Group Project*

- Predicted FIFA22 players' potential ratings with above 80% accuracy using an RBF SVM model.

### Quanteda - Python package

01/2024 – 02/2024

*UBC MDS Group Project*

- Created a Python package of functions that generate visualizations and metrics to simplify exploratory data analysis of quantitative financial data.

### "Defense wins championships": Coach's fact or fallacy?

04/2024 – 04/2024

*Blog Post*

- Authored a blog post analyzing NCAA men's basketball data using logistic regression to assess the accuracy of a widely-held coaching proverb.


## PROFESSIONAL EXPERIENCE

### Analyst

01/2022 – 08/2023

*Statistics Canada - Canadian Center for Education Statistics*

Ottawa, Canada

- Lead data collection for the 2022 Elementary-Secondary Education Survey (ESES) by corresponding with provincial ministries of education.
- Conducted exploratory data analysis and identified significant impacts of the Covid-19 pandemic on elementary and secondary enrolments.
- Authored the ESES survey release article  to communicate analysis results clearly to the public.
- Maintained the Public Service Employee Survey dashboard in Power BI to visualize data for upper management.

### Project Coordinator

08/2020 – 12/2021

*Royal Canadian Mounted Police - Federal Policing*

Ottawa, Canada

- Lead data validation of records of all staffed and vacant positions within Federal Policing.

- Optimized the data validation process and revised standard operating procedures to reflect new procedure before conducting training sessions with team members.
- Visualized the validated records of Federal Policing's positional data for upper management to inform strategic decision-making.

## **SKILLS**

---

### **Statistics / Research**

Survey & Experiment design, A/B testing, Forecasting, Statistical modelling, Data visualization, Regression, EDA, Bootstrapping

### **Coding Languages**

Python, R, SQL

### **Software & Tools**

GitHub, Spark, TensorFlow, AWS, MongoDB, Snowflake, Docker

### **Machine Learning**

- Supervised learning (Regression, Classification, Neural networks, Gradient descent models)
- Unsupervised learning (Clustering, Dimensionality reduction, Natural Language Processing, etc)