

KRISTINE VILLALUNA

✉ kristinevillaluna6@gmail.com ☎ 403-715-2047 🌐 github.com/kvstats in linkedin.com/in/kristine-villaluna

PROFILE

Statistician with experience working with large real-world survey datasets, performing data cleaning, exploration, analysis, modeling, and reporting using tools such as R, SAS, Python, and PowerBI. Strong statistical background with a specific focus in survey methodology, Bayesian analysis, and machine learning. Passionate about applying statistical methods in social science contexts to gain insights about our society and help inform decision-making by telling stories with data.

SKILLS

R | SAS | Python | SQL | R Markdown | STAN | Microsoft 365 | R Shiny | PowerBI | GitHub | EDA

EDUCATION

2021 – 2022

Toronto, ON

University of Toronto, M.Sc. Statistics

- M.Sc. co-representative of the Statistics Graduate Student Union (SGSU)
- Relevant Courses: Math Stats I & II, Applied Stats I & II, Economic Applications of ML

2015 – 2021

Lethbridge, AB

University of Lethbridge, B.Sc. Applied Statistics ☑

- Concentration: Economics, Minor: Computer Science

WORK EXPERIENCE

05/2022 – present

Ottawa, ON

Mathematical Statistician, Statistics Canada

- Assessed disclosure risk and implemented data reduction methods to achieve privacy standards, ensuring no single entry was identifiable on a Public Use Microdata File (PUMF).
- Prepared detailed documentation for the PUMF user guide, outlining the methodologies used.
- Contributed to the methodological development of survey sampling approaches for the Canadian Health Conditions Monitoring Program by rigorously testing diverse Kish allocation scenarios, resulting in the production of high-quality data for health research.

01/2022 – 04/2022

Toronto, ON

Teaching Assistant, University of Toronto ☑

- Assisted professors in grading and running tutorials for an undergraduate statistic course.

09/2020 – 08/2021

Ottawa, ON

Junior Mathematical Statistician, Statistics Canada

- Investigated bias in estimates resulting from COVID-19, leading to the development of new and improved weighting strategies for Survey of Sexual Misconduct at Work 2019.
- Examined the results of the Public Service Employee Survey 2019 using SQL and PowerBI and presented actionable recommendations to senior management based on the findings.
- Analyzed diverse populations' large datasets using SAS to guide sample and coverage research strategies, resulting in the successful execution of the Census of Agriculture.

05/2020 – 08/2020

Calgary, AB

Data Scientist, Co-op, Livestock Water Recycling

- Migrated PostgreSQL databases from on-prem into the cloud, saving on storage costs.
- Automated sensor data collection via MQTT and IoT for increased efficiency.
- Leveraged Python to build models to predict optimal polymer dose for the LWR system, mitigating the adverse effects of overdosing such as wasted product and monetary costs.

PROJECTS

Toronto Airbnb Prices ☑

- Bayesian hierarchical model in STAN using a room-type top level to predict listing price for Airbnbs in Toronto.

COVID-19 US Vaccination Modeling ☑

- Binomial regression model predicting proportion of the population aged 18+ who are fully vaccinated by US county.

Protester Violence in North America ☑

- Logistic regression model to examine the factors that effect the likelihood of violence at protests within North America.

Latent Dirichlet Allocation (LDA) Topic Modeling ☑

- LDA topic model using Gensim on Department of Justice press release data (2009-2018).