

# Ruby Cheung

(612) 481-6071 • in/rubyccheung • rcheung763.github.io/dsportfolio/

## Relevant Work Experience

---

### Institutional Research Analyst

Wabash College, Crawfordsville, Indiana

November 2022 – Present

- Conducted research on registration tracking, retention, and other institutional priorities; completed special assignments from the Dean of the College.
- Coordinated external reporting and managed ad hoc data requests for accreditation, process improvements, and other campus needs using SQL Servers and RStudio.
- Developed an institutional research (IR) handbook to document processes, improve reporting standards, and reduce the time it takes to complete reporting tasks by 20%.
- Designed and built institution-level dashboards and reports with RStudio to support data-driven decision-making.

### Graduate Research Assistant

Seattle University, Seattle, Washington

June 2024 – September 2024

- Collaborated with the Math and Biology departments to develop and implement sentiment analysis models.
- Developed and deployed Natural Language Processing (NLP) models to analyze user comments on carnivores.com
- Implemented testing and validation procedures to ensure accurate results, currently at 81% accuracy.

### Economic Research Specialist

St. Catherine University, St. Paul, Minnesota

June 2020 – August 2022

- Led economics research teams on projects focused on development economics and labor markets in North Africa and the Middle East.
- Conducted data analysis using hypothesis testing and regression analysis to predict employment rates.
- Developed a GIS-based census map for Sudan and designed Android-based survey instruments using ODK-X with integrated SQLite for data management.

## Technical Competencies

---

- **Programming languages:** R, Python, Stata, SQL, html, CSS
- **Techniques/Methods:** Supervised/Unsupervised Learning, Neural Networks, Regression Analysis, Causal Inference, A/B Testing, Time Series Analysis, Sentiment Analysis, Text Classification, Web Scraping
- **Technology/Frameworks:** RStudio, scikit-learn, TensorFlow, Seaborn, Hadoop, Hive, Spark, MySQL, Tableau, AWS, NoSQL Databases, Selenium, BeautifulSoup, PyTorch, Matplotlib, MongoDB, Keras, Statsmodels, Prophet, SARIMA, ETS

## Education

---

Master of Science in Data Science, Seattle University

May 2025

Bachelor of Science in Economics, St. Catherine University

May 2020

## Relevant Projects

---

### Predicting Recidivism Among Georgia Parolees | Python, scikit-learn, SVM, Gradient Boosting

- Utilized data from the Georgia Department of Community Supervision, built a supervised and an unsupervised model to identifying key variables that can help to predict the likelihood of recidivism, producing actionable insight to mitigate risk
- Developed and evaluated Gradient Boosting and Support Vector Machine (SVM) models, achieving a 73.6% accuracy for binary classification and a 72.6% accuracy with linear SVM, with a focus on reducing false positive rates.
- Applied variable importance techniques that identified correlations between continuous employment and low recidivism rates, highlighting the significance of stable employment for reducing reoffending.

### CarnivoreSpotter.com Sentiment Analysis | Python, TensorFlow, Keras, TextVectorization, Text Classification

- Processed unstructured data by performing feature preprocessing and vectorization to prepare for analysis.
- Developed an NLP pipeline to perform sentiment analysis on user-submitted comments (81% accuracy).