

Rebecca Han

(289)-992-8168 • RebeccaRH228@gmail.com • rebeccarh228.wixsite.com/rh-data-science

SKILLS

Languages/Libraries: C, CSS, Java, SQL, Python (pandas, Scikit-learn, spaCy, Pytorch, Transformers, TensorFlow)

Technologies: AWS, Hadoop, Microsoft Office 365, MongoDB, MSSQL, MySQL, PostgreSQL, R, Power BI, Tableau

EDUCATION AND CERTIFICATIONS

University of Toronto

Expected Winter 2026

Honours Bachelor of Science (Co-op) in **Machine Learning** and **Data Mining Statistics**

- University of Toronto Scholar Award (\$7,500)

Google Data Analytics Certificate

Obtained 2022 ([Coursera](https://www.coursera.org/certificates/google-data-analytics))

Google LLC

- Successfully completed **6 months** at **10 hours/week** of practice-based assessments designed for introductory-level roles in Data Analytics.
- Demonstrated strong competency in tools used throughout the data life cycle, including managing **spreadsheets**, querying in **SQL**, coding in **R**, and creating presentations through **Tableau**.

EXPERIENCE

Teaching Assistant for Introduction to Computer Science

April 2022 - Present (Contractual)

University of Toronto (UofT)

- Led a cumulative **24 weeks** of tutorials across contracted terms, teaching **Python** and **C** programming languages to university students in a clear and informative manner.
- Evaluated **60%** of student midterms and exams, demonstrating a thorough understanding of how to write efficient and correct code that adheres to common programming practices.

Analytics Intern

January 2023 - December 2023

Canada Revenue Agency (CRA)

- Developed and tested SQL code for the **processing** and **storage** of big datasets with **~28 million** entries.
- Optimized legacy SQL code to reduce runtimes for slow queries by up to **90%**, which increased the team's productivity and number of completed reports per workday by **50%**.
- Created macros in Microsoft Excel VBA to generate organized and informative reports using the fetched SQL data, which was deployed to over **1700** auditors at the CRA to make their workflow more efficient.

PROJECTS

Python KNN model using sklearn Iris dataset

December 2024. <https://github.com/rebebbie/KNN-Model>

- Built a machine learning model in Python that classifies data with **100%** accuracy on a simple testing suite.
- Also learned the following algorithms in a **third-year machine learning course** at **UofT**: linear regression, random forest, class conditionals, naive bayes, logistic regression, principal component analysis, and clustering.

AI Text Summarizer using Python's spaCy library

December 2024. <https://github.com/rebebbie/Notes-Summarizer>

- Solo developed a Python program that utilizes the **spaCy** library to generate two summaries of an input text file, a shortened version that selects the most important sentences, and an abstracted version using **Pegasus**.
- Conducted research into the best Python **pipelines** to use on large text files by experimenting with different **APIs** and models from **Transformers**.

Modelling Car Price Data in Serbia 2024 using RStudio

November 2024 - December 2024. https://github.com/rebebbie/C67_Final_Project

- Led a team of 4 data scientists in the creation of a model to predict a car's price in Serbia given **17 features** and **8414 rows** of raw data, as well as the subsequent Microsoft PowerPoint presentation to explain our findings.
- Final model has Adjusted R-Squared score of **86%** after performing data analysis on relevant variables, removing detected outliers, and performing a **Box-Cox transformation** to adhere to LINE assumptions.
- Conducted market research to understand the potential patterns behind the data, then implemented knowledge to clean the data of all **multicollinearity** and improve the **VIF** score.