PETER CHONG

WLCHONG@UWATERLOO.CA
GITHUB.COM/PETER-CHONG
LINKEDIN.COM/IN/WLCHONG

SKILLS

Languages: Python, C/C++, R, SQL, HTML, CSS, MATLAB & Swift
Technologies: Amazon Web Services, Databricks, Teradata, PySpark, SQLite, Scikit-learn, Git & Bash

EDUCATION

University of Waterloo
Candidate for Bachelor of Mathematics with majors in Statistics and Computational Mathematics

- Relevant coursework: Data Structures & Algorithm, Functional Programming & Databases, Object-Oriented Programming, Mathematical Statistics, Regression, Classification, Data Visualization, Generalized Linear Models

WORK EXPERIENCE

Data Science Associate | XE.com — May – Aug 2021

- Performed entity resolution (ER) and fuzzy string matching to de-duplicate customer records
- Reduced **fuzzy search runtime by 82%** by using NLP methods such as **N-gram** and **TF-IDF**
- Implemented and designed an end-to-end **object-oriented microservice** to validate customers' addresses by leveraging Google Geocode API
- Automated the execution of AWS Sagemaker notebooks using CloudWatch, Lambda and Lifecycle configuration
- Constructed **SQL** queries in AWS Athena to retrieve data from **AWS S3** and projects them in **QuickSight** dashboard

Data Science Analyst | Loblaw Companies Limited — Jan – Apr 2021

- Saved 4 hours of weekly manual work by automating standard report ETL across different teams
- Wrote a **Python script** in Databricks that discovers spam email addresses using Jupyter and regular expressions
- Conducted statistical analysis on store transaction history to better detect anomalies and discover data patterns

iOS Developer | Marlena Books — Sep – Dec 2019

- Developed and integrated Apple Push Notification (APN) using Google's Firebase Cloud Messaging service for backend support with minimal supervision
- Redesigned application's setting page to improve User Experience for individuals with dementia
- Demonstrated quick understanding and strong adaptability by designing black-box tests in a short timeframe for platform upgrades

PROJECTS

Customer Segmentation | *Python, Scikit-learn, RFM* — Aug 2021

- Developed a customer segmentation model by clustering customers into 3 different priority groups using **K-means Clustering** based on **RFM** score

Podcast Recommendation System | *Python, Gensim, NLTK* — Jul 2021

- Built a content-based podcast recommendation engine that recommends 5 most similar podcasts to user
- Web scraped 4,460 podcast data from Apple Podcast using **Selenium** and **BeautifulSoup**
- Applied NLP word embedding models such as **Bag-of-Words**, **TF-IDF** and **Word2Vec** along with **cosine similarity**

Dimension Reduction Visualization | *Python* — Jun 2021

- Explained and compared the strengths of different dimension reduction techniques (**UMAP**, **t-SNE**, **LDA** and **PCA**)

Los Angeles Marathon Finishing Time Estimator | *R, ggplot2, glmnet* — Sep 2020

- Created a marathon finishing time predictor that helps runners obtain their estimated finishing time
- Predicted personal marathon finishing time with an error of less than 6% by implementing 3 different regression techniques