

## EDUCATION

**University of Toronto - Master of Engineering, Mechanical & Industrial Engineering (GPA:4.0)** Sept 2021 - Nov 2022  
Courses: Cloud-Based Data Analytics; AI in Finance; Algorithms & Data Structures; Data Analytics & ML; Deep Learning; MEng Research

**University of Toronto - Honors Bachelor of Science, Dean's List for 2020, 2021** Sept 2017 - June 2021  
Courses: data analysis; Statistical Consultation, Methods for Multivariate Data;; Methods of Applied Statistics, statistical theory;

## SKILLS & QUALIFICATION

**Data Science (5 Years): Feature Engineering, predictive, Statistical & quantitative analysis**, NLP, Image recognition, A/B testing

**Programming:** **Python** (5 Years) [Sklearn, Pandas, PyTorch, TensorFlow, GeoPandas, Numpy, Matplotlib, Seaborn, Beautiful Soup, Shapely], **SQL** (4 Years), **PySpark**(1 Year), **SparkSql**(1 Year), **R**(2 Years) and **SAS**(1 Year)

**Data Engineering:** **Azure** (3 Years) [Data Factory, Synapse, Functions, Machine Learning], **Snowflake** (1.5 Years), **Databricks ML** (1 Year)

**Dashboard:** **Power BI** (2 Years), Tableau (6 Months), Google Analytics(certificate)

**Other Tools:** Visual Studio 2019, Jira, Azure DevOps, GitHub, Confluence, Draw.io, Microsoft Teams

## PROFESSIONAL EXPERIENCE

**Data Analyst Intern | EDPYPF Services Incorporated | Markham** Jan 2023 - Present

- Successfully implemented **Azure platform PoC** to load data from **Datalake** to **Snowflake** DW via **ADF, Databricks, Synapse, Stored Procedure, Eventhub & Azure Functions**. Resulted in company-wide adoption of Azure & Snowflake. Demonstrated ability to integrate various Azure services to build a comprehensive data solution.
- Developed a **DataModeler** class in python for **bank transaction analysis**, including feature engineering, fitting and testing to
- Targeted specific **business categories** for retail banking credit card promotion through clustering and quantitative analysis, clearly **predict customer's behavior** under different categories.
- Delivered a **Customer Segmentation Dashboard** in PowerBI, improving campaign targeting efficiency by **30%**.
- Conducted **SQL Performance Tuning** training for analysts in dealing with modeling, business requirements, and data mart.

**Research Analyst | C-MORE, UoT | Toronto** Jan 2022 - Sep 2022

- Created a **Best Route - Transportation model**, reducing overall risk on the **best route selection** by **20%** in GTA using an extended **SARSA** reinforcement learning algorithm.
- Ensured data accuracy, completeness through **framework design, quality checks & cleaning procedures**.
- Integrated traffic, weather, population data using Python to support **evidence-based model** and insights.
- Provided **precise geospatial data** through **feature engineering** for model **efficiency improvement** by **25%**.
- Proposed a **risk assessment methodology** for **transportation routes** and **visualized road networks** in the GTA.

**Data Analyst, Intern | Chuang Chuang Culture Communication Co., Ltd. | China** May 2019 - Aug 2019

- Used R Studio to analyze customer demographics and behavior, creating profiles that generated over \$800K in revenue since the new consulting product's launch in 2019.
- Collaborated with the marketing team to create potential client lists based on **customer groups** identified using **K-Means** clustering. Achieved **30%** improvement in **email click-through rates** on average.
- Developed a Power BI **customer satisfaction dashboard**, improving customer satisfaction score by **15%**.

## PROJECTS

**Python Data Loading Framework on AWS Cloud | GitHub** May 2023 - Jun 2023

- Developed data pipeline in python to load real time data from AWS S3 buckets for security trading summary report

**Recruitment analytics with Azure Cloud Platform | GitHub** Jan 2022 - Apr 2022

- Built Azure pipelines to transfer files from Blob to Azure SQL DB using ADF; Automated the process with scheduled triggers; Integrated 8 pyspark/sparkSQL Databricks notebooks with ADF.

**Vehicle License Plate Detection | GitHub** Jan 2022 - Apr 2022

- Designed and fine-tuned a VGG16 CNN model with Keras for license plate detection. Achieved 88.1% accuracy in localizing bounding boxes and 76.6% accuracy in identifying letters/digits on license plates. Virtualized results in Python. Improved overall accuracy by 8%.

**Canadian Election NLP Sentiment Classification Analysis | GitHub** Sep 2021 - Dec 2021

- Performed sentiment analysis on 2021 Canadian election Twitter data in Python. Utilized TF-IDF tokenization and ML algorithms (Logistic Regression, Decision Tree, Random Forest, KNN, SVM, Naive Bayes, and XGBoost) achieving 94.4% accuracy.

**Data Visualization – Various Dashboard Development | GitHub** Sep 2021 - Mar 2023

- Various Dashboard Development for case study and proof of concept using Power BI & Tableau