

Zheyuan Fan

Toronto, Ontario | (647)-505-8370 | bradyfannn@gmail.com
<https://www.linkedin.com/in/zheyuan-fan> | <https://www.github.com/fanzhey1>

EDUCATION

University of Toronto

Toronto, Ontario

Master of Engineering, Emphasis in Analytics. (GPA: 3.77/4.00)

Sep 2021 - Nov 2022

Relevant Courses: Cloud-based Data Analytics; Process Data Analytics; Foundations of Machine Learning; Data Mining in Engineering; Financial Engineering; Statistical Models in Empirical Research; Healthcare Engineering.

Honours Bachelor of Science, Specialist in Statistics & Minor in Economics.

Sep 2017 - Jun 2021

Relevant Courses: Methods in Applied Statistics; Advanced Data Analysis; Time-series Analysis; Statistical Methods in Machine Learning; Spatial Data Analysis; Experimental Design; Applied Multivariate Statistics; Financial Economics.

SKILLS & CERTIFICATIONS

Programming Languages : **SQL** (subqueries, joins), **Python** (pandas, scikit-learn, TensorFlow, Matplotlib), **R**(tidyverse), **SAS**.

Database Management : **Azure** (Azure Data Factory, Azure Machine Learning), **Spark**, **Snowflake**, **Databricks**, **Hadoop**.

Visualization & MS Tools : **Tableau**, **Excel** (vlookup, PivotTable), **PowerPoint**.

Certifications : Microsoft Azure Fundamentals Certification, SAS Certified Advanced Programmer for SAS 9.

WORK EXPERIENCE

Data Analyst Intern - JD.com, Inc. (largest B2B e-commerce company in China)

May 2020 - Aug 2020

- Assisted lead User Experience Researcher to conduct data analysis through high-customer-unsatisfying-rate commodities' data using **SQL** and **Excel**, to lower bad-review rate of various products sold on the e-commerce website.
- Investigated customers' text reviews for reasons that caused high product return rate; compiled reports and assisted brands' representatives to identify key product features to display on the e-commerce website; explained technical concepts to non-technical staff members in plain language.
- Spearheaded the display re-design with **12** brands' representatives of products' descriptions on the website, including the front and main pictures, descriptions, and Q&A sections; monitored recent reviews closely and constructed feedbacks to brands' representatives, leading to **25%** bad-review rate reduction on their products on average.

PROJECTS

Customer Analysis Using Tableau [GitHub](#)

Nov 2022 - Nov 2022

- Analyzed a sales dataset on GitHub using **Tableau**, generated different types of charts including a butterfly chart, a donut chart, a scatter to present revenue by state, age, month, region, gender, and customer's sensitivity to discount.
- Developed an integrated dashboard to show the overall revenue insights, which includes an optional category filter.

Creating Automated ETL Pipeline in Azure Data Factory & Analyzing Copied Dataset

Jul 2022 - Aug 2022

- Created a pipeline in **Azure Data Factory** that copies a csv file from Azure Blob storage account to Azure SQL database; established a scheduled trigger that runs the pipeline periodically.
- Analyzed a job-gender dataset in Azure SQL Database using **T-SQL** to get information about earnings by occupation & gender.

Canadian Election NLP Sentiment Analysis [GitHub](#)

Apr 2022 - May 2022

- Conducted sentiment analysis in **Python** on 2021 Canadian election Twitter data to assess public opinions about the Canadian political landscape, and predicted reasons for negative tweets.
- Produced visualizations charts and raw data preparation using **TF-IDF** tokenization in feature processing.
- Trained ML algorithms (**Logistic Regression**, **Decision Tree**, **Random Forest**, **KNN**, **SVM**, **Naive Bayes**, and **XGBoost**) for sentiment classification, best model achieved **94.4%** accuracy. Categorized reasons for negative election tweets.

Retail Transaction Data Analysis [GitHub](#)

Mar 2022 - Apr 2022

- Predicted clients' response to a promotion campaign; analyzed consumers' shopping behaviors using a retail transaction dataset.
- Conducted feature engineering and implemented ML models (**Logistic Regression**, **Decision Tree**, and **Random Forest**) in **Python** and **PySpark**, performed hyperparameter tuning.
- Implemented and visualized model agnostic methods (PDP, completeness vs feature importance graph) to get insights of the effectiveness of promotion campaign against each group, as well as the predictive power (importance) of each feature.

Visualizing Worldwide COVID-19 Data [GitHub](#)

Jan 2022 - Feb 2022

- Developed a **Python** dashboard to present daily visualizations of COVID-19 data collected from European Centre for Disease Prevention and Control, which automatically generates plots to get insights of how the virus is spreading spatially and temporally.
- Generated **7** static and **1** interactive plots regarding global COVID-19 inflections and mortalities, including the comparison of number of cases/deaths between countries/continents.