Yifan Zhao Email: yifan_zhao@brown.edu Mobile: 401-226-5283

Technical Skills

Languages: R, SAS, Python, SQL,
Software: R Markdown, Microsoft SQL Server, Power BI, LaTeX, GitHub, Excel

Work Experience

COVID-19 Vaccine Effectiveness Research, Brown University *Providence, RI, United States*
Part-time Research Assistant *Jan 2022 - Present*

- Cleaned and manipulated large datasets (16 GB * 100) to generate research cohort that meets the criteria.
- Created looping algorithm in SAS that enabled overnight running and helped reduce manual code operation.

Investment Data Analytics, Ontario Teachers' Pension Plan *Toronto, Ontario, Canada*
Data Analytics Intern *May 2022 – Aug 2022*

- Researched and implemented autoencoder and K-Means clustering algorithm to identify stock market regime. Compared and summarized results with benchmark model and presented to senior stakeholders.
- Created Transaction Cost Analysis and Momentum Trading Analysis dashboards using Power BI that helped traders to understand historical transactional costs. Customized visuals according to business requirements.
- Conducted in-depth data validation with existing reports and created solutions to various data mismatch issues.

Member Services Business Insights and Analytics, Ontario Teachers' Pension Plan *Toronto, Ontario, Canada*
Data Analyst *Feb 2020 – Aug 2021*

- Built a logistic model to predict RFT flags for Quality Assurance team using Python. Improved workflow routing strategy and helped QA team save 10% of process volume with a risk level of less than 5%.
- Improved efficiency in database daily refresh for operational data through building a data mart using SAS and SQL Server and scheduled the program into the production data pipeline.
- Built data visualization reporting tools and dynamic data drill-throughs for team performance tracking using Power BI. Estimated cost of operational transactions for capacity planning and presented to senior leadership.

Data Scientist Intern *May – Aug 2019*

- Saved FTE across internal departments through automating multiple data update processes using Python and Jira REST API.
- Designed a DAX algorithm in Power BI that handled influential data points/outliers in business operational data
- Helped operations team manager better understand and analyze relationship between customer interaction and system transactions by creating dashboards and structured datasets in Power BI Dataflow.

Economic Statistics Methods Division, Statistics Canada *Ottawa, Ontario, Canada*
Survey Methodology Intern *Sep – Dec 2018*

- Simulated different sample design scenarios in terms of stratification level, sample allocation, etc. using SAS and Excel metadata.
- Reduced variation in estimates by implementing and verifying a granular-level industry class (NAICS) stratification method. The method is adopted in the subsequent year.
- Analyzed and evaluated simulated samples and estimates, manipulated data using SAS, and conducted mathematical analysis that explained the differences between sample scenarios.

Academic Projects

SEIR Simulation of COVID-19 Spread *Nov 2021*

- Implemented an SEIR (Susceptible, Exposed, Infected and Removed) simulation model in R for COVID-19 case trend in universities given different school policy conditions based on a research paper.
- Created an interactive Shiny App that allows users to input desired parameters and interpret the resulting graphs.

Time Series Modeling and Forecasting *Apr 2019*

- Implemented and compared several time series models including Winters Seasonal Exponential Smoothing, seasonal ARIMA, and ARIMA with explanatory variables using US Candy Production data

Education

Brown University *Sep 2021 – May 2023*
Master of Science in Biostatistics *Providence, RI, United States*
Coursework: Bayesian Statistics, Survival Analysis, GLM, Longitudinal Data Analysis, Statistical Inference

University of British Columbia *Sep 2015 – Dec 2019*
Bachelor of Science in Statistics *Vancouver, BC, Canada*
Coursework: Experimental Design, Statistical Learning, Sample Survey, Algorithms and Data Structures