# Lily Lu

Greater Scranton Area, PA | (973)710-2282 | liyinglo914@gmail.com | LinkedIn | GitHub

## EXPERIENCE

**Benco Dental Supply Co.** **Pittston, Pennsylvania**

*Senior Data Analyst* February 2021-Present

- Propel strategic marketing planning using time series forecasting in Python with less than 85 basis points error
- Enhance efficiency of a standard procedure by 60% through automation using a SQL stored procedure and functions
- Strategize human resource planning in high-demand areas by building time series forecasting models in Python
- Empower decision-making and cross-team communication via Qlik Sense dashboards, Excel reports, and SQL Server
- Compute the return on investment of customer acquisition channels using SQL to optimize resource allocation

**Model B** **Washington, D.C.**

*Data Analyst Intern* August 2020-December 2020

- Developed Datorama dashboards for monitoring campaign performance and showcasing client-facing reports
- Charted campaign strategies using web scraping and sentiment analysis to analyze social media trends in Python
- Optimized campaign performance by analyzing KPIs, demographic data, audience targeting, and keywords

## EDUCATION

**George Mason University** **Fairfax, Virginia**

*B.S. in Computational and Data Science* (GPA 3.88/4.0) December 2020

## SKILLS

**Skills**: Statistical Modeling, Time Series Analysis, Machine Learning, Web Scraping, Database Management
**Technology**: Python, R, SQL Server, MATLAB, Java, Qlik Sense, Tableau, Datorama, Microsoft Excel, Git, Jira, Agile
**Languages**: English (fluent), Cantonese (native), and Mandarin (native)

## PROJECTS

**COVID-19 Forecasting using SEIR(CD) Model**

*Summary: Predicted the daily COVID-19 cases and deaths using numerical integration and the SEIR(CD) model (MATLAB).*

- Extracted and formatted the confirmed cases and deaths from COVID-19 in the United States
- Reviewed research to estimate the reproduction number, incubation period, and duration of contagiousness
- Modeled the cases and deaths using the trapezoid method and parameter optimization based on weighted MSE
- Improved model to account for the high and low-risk population based on the age and pre-existing medical conditions

**Marketing Campaign Performance Prediction**

*Summary: Analyzed the most useful KPIs for campaign optimization and prediction through statistical modeling (Python).*

- Processed campaign and demographic data through collection, cleaning, exploratory visualization, and normalization
- Checked assumptions for Ordinary Least Square regression and Multivariate Adaptive Regression Spline models
- Tackled multicollinearity via feature selection using forward selection and Random Forest feature importance
- Evaluated the models through K-Fold cross-validation, model selection, and comparison of feature selection methods

**COVID-19 Social-distancing Monitoring Program**

*Summary: Evaluated the social-distancing feature of farmer's markets to enhance the prevention of COVID-19 (Java).*

- Programmed an application to evaluate the effectiveness of the market space in preventing the spread of the virus
- Simulated market space and human dynamics using multi-dimensional data structures, and techniques of inheritance
- Designed a propagation feature to check the effectiveness of containing the virus both recursively and iteratively

**Serial Killer Data Analysis**

*Summary: Investigated the features of serial killers through web scrapping, text mining, and statistical testing (Python).*

- Scraped, reorganized, and merged unstructured data using pandas, requests, and Beautiful Soup
- Engineered features through natural language processing (NLP) and sentiment analysis using nltk
- Visualized the relationship between victim counts and other variables using seaborn, folium, and matplotlib
- Performed ANOVA tests and assumption checking to test the significance of the difference between groups using SciPy