Seyoung Jung

Los Angeles, CA 90057 | (213) 218-5824 | sjung.data@gmail.com
linkedin.com/in/sjung-stat | github.com/sjung-stat | sjung-stat.github.io/portfolio

PROFESSIONAL EXPERIENCE

Comscore, Inc.   Los Angeles, CA
Associate Data Scientist   January 2022 – Present
- Utilized python to develop random forest regression models using large scale data which predicts ratings of a telecast to reduce reporting time by 50% for local stations
- Maintain automated daily reporting process for anomaly detection in SQL and python, reducing 10 hours of manual working time per month
- Collaborate with cross-functional teams to develop new methodologies and integrate them into production process

Independent Contractor   Los Angeles, CA
Statistical Consultant   December 2021 – Present
- Prepare data and perform longitudinal data analysis to determine effectiveness of planting protocol
- Write data analysis report and help clients improve experimental design
- Clients: The Watershed Nursery (Aug 2022 – present), Santa Clara Valley Water District (Dec 2021 – Feb 2022)

Davis Sensory Institute   Davis, CA
Data Science Intern   June 2021 – September 2021
- Used python to perform sentiment and clustering analyses on unstructured product survey data to help clients understand their new products and improve them
- Hosted NLP and machine learning workshops for non-technical employees to achieve effective collaboration

SKILLS
- **Programming Languages / Tools:** Python, R, SQL, AWS, Git, Linux, Tableau, Excel, PowerPoint
- **Data Science & Miscellaneous Techniques:** Deep Learning, Natural Language Processing, A/B Testing, Data science pipeline (cleaning, wrangling, visualization, modeling, interpretation), Statistics, Time Series, Experimental design, Hypothesis testing

PROJECTS *(github.com/sjung-stat)*

Restaurant Fake Review Detection   June 2021 – August 2021
- Utilized SQL and python to build restaurant fake review detection classifiers to increase reliability of review platforms
- Achieved a 91% of accuracy with deep learning (CNN, LSTM), XGBoost, random forest, and NLP

Credit Card Fraud Detection   June 2020 – September 2020
- Utilized R to develop fraud detection models to reduce the unexpected loss of businesses and customers
- Achieved a 0.053 log loss with CatBoost when a baseline model generated a 0.152 log loss

E-Commerce Revenue Prediction   April 2020 – June 2020
- Used R to build a revenue forecasting model to enable efficient cash flow management
- Obtained a 1.54 RMSE with XGBoost with target variable ranging from 0 to 23.86

EDUCATION
- **University of California – Davis**   Davis, CA
  Master of Science in Statistics (Data Science Specialization)   December 2021

- **University of California – San Diego**   San Diego, CA
  Bachelor of Science in Statistics   March 2017

- **Relevant Coursework**: Machine Learning, Statistical Data Science, Statistical Inferences, Data Structures & Algorithms