**YoungKeun Yang** datascientist.young@gmail.com|813-618-0645 |

Land O Lakes, FL | linkedin.com/in/ykyang |www.ykyang.com

**Education**

**University of South Florida** | Tampa, FL (Expected) Aug 2023

M.A. Statistics

**Soongsil University** | Seoul, Korea Aug 2017

B.A. Business Administration

**Technical Skills**

**Programming** | Python (NumPy, pandas, sklearn), R, SQL, Tableau

**Coursework** | Statistics Methods, Design of Experiments, Linear Statistical Models, Multivariate Analysis, Non-parametric Statistics, Mathematical Statistics, Probability Theory, Machine Learning, Data Structure and Algorithms

**Data Science Project**

**Machine Learning Classification: Hotel Cancellation Prediction**

**-** Achieved a top-performing binary classification model with an accuracy of 91%, recall of 88%, and precision of 88% by modeling and comparing popular classification algorithms such as Random Forest, XGboost, and Neural Network

- Transformed a large dataset of 119,390 observations and 32 features into a readily usable format for algorithm implementation by preprocessing

- Unearthed valuable business insights by successfully constructing a model that identified the significance of two key variables: lead time and the number of special requests

**Machine Learning Clustering: Customer Segmentation**

- Implemented feature engineering techniques to create 15 new variables, enriching and unveiling valuable insights for data

- Remodeled the data of 541,909 customers by cohort analysis, RFM analysis, K-means, and Hierarchical clustering

**-** Leveraged data analysis to discern spending patterns and successfully restructured customers into 3-4 distinct groups, providing valuable business insights for strategic decision-making

**Machine Learning Regression: Forecasting Rideshare Price**

- Maximized the performance of models with a dataset of 693,071 observations and 57 features by evaluating several well-known machine learning algorithms, including Linear regression, Random Forest, KNN, Neural Network, SVM, etc.,

- Optimized the model performance resulting in Root Mean Squared Error(RSME) of 2.26, indicating high accuracy through hyperparameter tuning

- Identified the key factor for deciding the rideshare price by surveying the relationship between the variables in the data

**Thesis: Bayesian Estimation of Autocovariance of a Model Error in Time Series (In progress)**

- Investigated the Markov Chain Monte Carlo method for Bayesian estimation for precise Time Series model error prediction

- Developed a new Time Series method with a variant version of GARCH to build a solid model for forecast financial Time Series data such as Stock price or Currency exchange rate

**Work Experience**

**USF Academic Success Center** | Tampa, FL Sep 2022 – May 2023

Statistics Tutor

- Enhanced more than 100 students' understanding of Statistics, with a 100% passing rate by tutoring 16 hours weekly

- Earned director’s referral to the Athletes department and boosted the performance of 8 Athletes’ grades by 20%

**FILA Korea** | Seoul, Korea Mar 2018 – Sep 2019

Sports Marketing Team Staff

- Increased daily revenue by 300% by designing and implementing marketing campaigns based on customer analysis in collaboration with the sales department

- Exceeded KPI of the brand exposure in the media by 20% by analyzing data and planning a new method for publicizing

- navigated unforeseen manufacturing errors and fast-paced situations in sports events by employing simulation using past data

**Brion Company** | Seoul, Korea Jan 2017 – Jul 2017

Sports Marketing Team Internship

- Won bid for Adidas’ marketing operation project, with a budget of $50k, by delivering a presentation based on the analysis of the client’s situation

- Achieved 100% accurate forecasts for non-rainy weather conditions for outdoor marketing events by effectively collecting and analyzing relevant data

- Produced and delivered weekly analytic reports to clients for one year leveraging data visualization for a clear understanding