SANG OK SUH

ASPIRING DATA ANALYST

WORK EXPERIENCE

Volunteer Research Associate

UCLA Semel Institute for Neuroscience and Human Behavior | *Mar 2018 – Oct 2018*

- Extracted peak information and 1/F slope from brain wave EEG data using Python.
- Performed data cleaning and manipulation in R for analysis preparation.
- Used Wilcoxon rank-sum test to differentiate children with autism from those with normal development: significant peak frequency differences in certain regions of the brain.
- Visualized peak differences using ggplot package in R and Tableau.
- Found significant predictors of ASD in children using Random Forest in R.
- Prepared data (extracting, cleaning, wrangling) for grant proposals.

Data Intern

First Capitol Consulting | *Sep 2017 – Dec 2017*

- Increased productivity gains greater than 50% by assisting with accurate data entry using Microsoft Excel (V-Lookup, H-Lookup, Index/Match, Concatenate).
- Utilized Excel functions to find and reduce redundancies among clientele HR data.
- Identified, revised, documented, and produced formal reports to communicate errors in data workbooks to senior analysts.

APPLIED PROJECTS

Red Wine Quality Analysis

STATS 131: Python and other Technologies for Data Analysis | *Mar 2018 – Jun 2018*

- Designed and compared multinomial logistic regression models and Random Forest classification models (scikit-learn) in Python to predict wine quality based on 11 variables.
- Random Forest model resulted in a 75.07% (> MLR: 65.21%) accuracy rate using 4 variables.
- Used Python packages numpy, pandas, matplotlib, and seaborn for exploratory analysis.

Thyroid Eye Disease Analysis

STATS 199: Independent Research | *Mar 2018 – Jun 2018*

- Assisted the professor's statistical consult with UCLA Jules Stein Eye Institute by analyzing longitudinal data and identifying predictors of Ocular Surface Disease Index score for thyroid eye disease patients by applying linear modeling in R.

UCLA DataFest: Indeed User Data Challenge

Finalist | *Apr 2018*

- Managed a 4-member team in a two-day hackathon to define questions and approaches to analyze three sets of large, uncleaned user data.
- Used R and Tableau to clean, analyze, and visualize Indeed's market share of job listings for various job sectors, implementing external data from The Bureau of Labor Statistics.
- Visualized and compared median salaries between Indeed and Glassdoor's job listings for different job sectors.

Expedia User Data Challenge

STATS 141SL: Practice of Statistical Consulting | *Jan 2018 – Mar 2018*

- Performed logistic regression in R to analyze user data in JSON format to identify factors that influence mobile-users' decisions to reserve hotels.
- Found predictors that were highly correlated to successive bookings by mobile users.
- Developed a solution algorithm that allows brand hotels to compete with the high demand of lower costing, non-branded hotels with high ratings during specific travel seasons.

CONTACT

sangoksuh@hotmail.com

213 507 2011

Los Angeles, CA

linkedin.com/in/sangoksuh

SKILLS

R

Python

SQL

Tableau

Advanced Excel

EDUCATION

B.S. Statistics

UCLA, Jun 2018

GPA: 3.72/4.00, *cum laude*

RELEVANT COURSEWORK

Statistical Programming with R
Data Analysis and Regression
Design and Analysis of Experiment
Statistical Models and Data Mining
Computational Statistics with R
Computation and Optimization
Monte Carlo Methods
Python for Data Analysis
Practice of Statistical Consulting
Probability
Mathematical Statistics
Linear Models
Linear Algebra