Jong Wook. Choe
582 Grove Street, San Francisco, CA, 94102 | (530) 760-6057
1994wook@gmail.com | Linkedin | Portfolio

Experience

Summer Institute Fellow / California Policy Lab Jun. 2022 ~ Aug. 2022

- Cleaned a longitudinal, Continuum of Care level analytic dataset of Department of Housing and Urban Development dataset dating back to 2015 each of 100,00+ rows.
- Merged multiple data frames in R while minimizing data loss.
- Analyzed twenty years data with R to propose and answer the policy-relevant research questions on California Homelessness project.

Public Data Intern / National Information Society Agency, NIA Jun. 2021 ~ Jul. 2021

- Provided data maintenance support for the public data portal, 'www.data.go.kr'
- Constructed data analysis on 25,000+ rows of mask inventory data at drug stores crawled from open-source API and visualized data mapping with Python.
- Examined mask sales volume to classify potential inventory shortages by applying effective machine learning technique.

Marketing Intern / Daily Beer, Inc Jul. 2018 ~ Aug. 2018

- Created statistical report on new opening delivery business to provide insights on commercial analysis and market research.
- Organized and shaped social networking business campaign based on research.

Education

Berkeley Data Analytics Boot Camp Oct. 2023

- Advanced understanding in Excel, Python and R programming, JavaScript charting, SQL database, Tableau, machine learning.

University of California Davis, Davis, USA Jun. 2022

- Bachelor of Science in Statistics; minor in Mathematics

Project

MLB All-Star Game Prediction, UC Davis

- Analyzed dataset which factors contribute to a player's selection for the MLB All-Star Game and predict which players will be selected for upcoming event.
- Scraped from multiple sources, including player statistics, biometric data, and historical All-Star Game rosters using BeautifulSoup library.
- Employed data cleaning and preprocessing, feature engineering, exploratory data analysis, and logistic regression modeling. The player's biometric data were visualized using violin plot to identify any patterns of trends that could be relevant to All-Star Game selection.
- Achieved an accuracy of 90% using logistic regression model while identifying several key factors that are strongly associated with All-Star Game selection, including batting average, on-base percentage, and age.

NASA Asteroid Classification, UC Davis

- Estimated prediction modeling to estimate the risk of high-risk asteroids colliding with Earth based on various predictor variables.
- Managed large public dataset of asteroid characteristics and historical asteroid collisions, which was obtained from NASA's database.
- Utilized the SMOTE algorithm to address the issue of an unbalanced dataset and fitted to a decision tree model to identify the relationship between the response variable (risk of collision) and the predictor variables.
- Concluded decision tree model accurately predicted the risk of high-risk asteroids colliding with Earth based on the selected predictor variables, which could help in developing strategies to prevent or mitigate asteroid collisions.

Certificate

- **Google Analytic** individual Qualification
- Completed Coursera's **Google Data Analytics Specialization**, **Stanford Machine Learning**, and **SQL for Data Science** certificate programs.