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

Consideration of Pipe Service Life

- Distinguished main factors contributing to pipe breakages and predicted potential pipe failures with machine learning algorithms.
- Conducted data cleaning and preprocessing on client data related to pipe characteristics and breakages.
- Identified material and length as important factors in pipe failure through box plots and logistic regression modeling.
- Provided insight into factor contributions to pipe breakages and potential prediction models for identifying pipe failures.

MLB All-Star Game Prediction, UC Davis

- Analyzed dataset to predict which players will be selected for the MLB All-Star Game.
- Scraped multiple sources, including player stats and biometric data, using the BeautifulSoup library.
- Employed data cleaning, feature engineering, EDA, and logistic regression modeling.
- Visualized player biometrics using a violin plot to identify trends relevant to All-Star Game selection.
- Achieved 90% accuracy using logistic regression, identifying key factors like batting average and age.

NASA Asteroid Classification, UC Davis

- Estimated prediction modeling to assess the risk of high-risk asteroids colliding with Earth.
- Managed a large public dataset of asteroid characteristics and historical collisions from NASA's database.
- Utilized SMOTE algorithm to balance an unbalanced dataset and fitted to decision tree model.
- Concluded decision tree model accurately predicts the risk of high-risk asteroid collisions, which can aid in developing preventative strategies.

Certificate

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