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

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

Experience

Summer Institute Fellow / California Policy Lab Jun. 2022 - Aug. 2022
- Collaborated with a team to construct an ETL (extract, transform, load) data pipeline by merging multiple data frames of nationwide homelessness datasets.
- Analyzed and examined the trend of homelessness in California over the past twenty years and identified the factors contributing to the fluctuations in the homeless population.
- Managed data preparation process of a longitudinal dataset from Department of Housing and Urban Development dating back to 2015, with Continuum of Care as a primary variable consisting of 100,000+ rows.
- Developed and presented findings to a multidisciplinary team, providing important policy implications for addressing homelessness in California and other regions facing similar challenges.

Public Data Intern / National Information Society Agency, NIA Jun. 2021 - Jul. 2021
- Worked as a team to support the government's "Open Data Activation Policy" to promote the use of public data.
- Optimized the data maintenance tool by automating the data inspection process with Python and achieved higher data accuracy from 91% to 99%.
- Provided data maintenance support for the public data portal, www.data.go.kr

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.

Project

Consideration of Pipe Service Life Mar. 2023
- Discovered the main factors in which contributing breakages of the pipe and predicted potential damage.
- Cleaned and preprocessed actual pipes and past breakages information and merged datasets based on a native pipe ID as the primary key variables.
- Conducted data cleaning and preprocessing on client data and merged datasets based on a native pipe id as a primary key variable.
- Identified material and length as the factors in pipe failure through visualization and logistic regression.
- Provided insight into factor contributions to pipe breakages and prediction models for identifying pipe failures.

MLB All-Star Game Prediction, UC Davis Jun. 2022
- Forecasted the 2022 MLB All-Star Game player roster using 2021 season player stats and biometric data.
- Acquired data through web scraping utilizing the Python library, BeautifulSoup.
- Visualized player biometric data to identify relevant patterns and trends for All-Star Game selection.
- Identified key factors such as batting average and age and employed logistic regression and generated player list based on the players' statistics of the first half 2022 season.

NASA Asteroid Classification, UC Davis May. 2022
- Examined the large public dataset of asteroid characteristics and historical collisions from NASA's database to assess high-risk asteroids colliding with Earth.
- Oversampled the skewed dataset using the SMOTE algorithm to balance the data and achieve higher accuracy before model selection.
- Predicted high-risk asteroid collisions with greater accuracy using the decision tree algorithm compared to other employed models, enabling the development of more effective preventative strategies,

Certificate

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