William Harvey

949-419-5443 • Berkeley, CA • will.harvey@att.net

EDUCATION

University of California, Berkeley | B.S. in Data Science w/ Emphasis in Economics   Expected May 2023
- **GPA**: 3.7/4.0
- **Relevant Coursework**: Principles and Techniques of Data Science, Machine Learning Applications, Data Structures, Linear Algebra, The Structure and Interpretation of Computer Programs, Probability and Mathematical Statistics in Data Science, Data Engineering, Data Mining and Analytics
- **Activities**: *Treasurer*, Phi Gamma Delta; *Orientation Leader,* New Student Services

EXPERIENCE

Data Analyst Intern   May 2022 - Aug. 2022

Data Glacier | Prayagraj, India (Remote)
- Performed exploratory data analysis (EDA) and data wrangling on large datasets and created presentations using matplotlib visualizations to showcase findings.
- Serialized machine learning models using pickle and deployed them into interactive web apps with Flask and Heroku.
- Experience sharing team projects using GIT commands and following agile frameworks such as Scrum and Kanban.

Lair of the Golden Bear Camp Staffer   Jun. 2021 - Aug. 2021

Cal Alumni Association | Berkeley, CA

- Maintained a high standard of cleanliness in all areas of the camp while being an integral part of a 9-person crew responsible for daily cleanup of the camp and preparations for new guests.
- Demonstrated impeccable customer service skills while serving food along with upholding strict food safety and COVID-19 policies to provide a safe dining area.
- Collaborated with peers to create engaging show performances twice a week.

PROJECTS

Version Control System
Implemented key operations of the version-control system "Git" in the Java programming language (e.g., add, commit, reset, checkout).
- Created functionality to add, delete, restore, and move files across directories on a computer.
- Utilized a combination of data structures while accounting for asymptotic runtime complexity in a large program in Java.
- Mastered serialization and persistence in order to retain information in data structures.

Spotify Playlist Multi-Class Logistic Regression
Collected Spotify data and utilized Spotify and Genius API to build a model to predict genre of a song.
- Developed a model using features of a song (e.g., bass and danceability) and lyrics of the song.
- Applied various NLP techniques to the lyrics of each song such as Topic Modeling, Lemmatization, and POS Tagging
- Used this model to classify a song as belonging to the hip-hop, country, or blues genre

COVID-19 Case Prediction Analysis
Extensive report based on selected features to predict number of cases in a state.
- Cleaned a data set of 3,000+ rows on vaccination status, mask usage, and COVID-19 cases per day.
- Created visualizations in the Seaborn library in order to select the most relevant features to train my model on.
- Utilized the Scikit-Learn library in order to train a logistic regression classifier to classify states with high COVID-19 case rates.

SKILLS

Languages and Software: Python, Java, R, Microsoft Office, SQL, Matplotlib, Seaborn, Pandas, Scikit-learn, Jupyter

Skills: Data cleaning, data manipulation, feature engineering, machine learning, data engineering, data visualization, statistical modeling, Natural Language Processing