Kevin Bui

📞 (818)-823-1376 ✉ kevinbui806@gmail.com in https://www.linkedin.com/in/kevin-bui 🏠 kevinbui2.github.io

Education

University of California San Diego **Exp. March 2024**
Bachelors of Science in Data Science; GPA: 3.81 *La Jolla, CA*

- Coursework: Principles of Data Science, Theoretical Foundations of Data Science, Practice and Application of Data Science, Exploratory Data Analysis and Inference, Introduction to Data Management, Beyond Relational Data Management, Systems for Scalable Analytics, Probabilistic Modeling and Machine Learning, Representation Learning, Data Structures and Algorithms, Recommender Systems and Web Mining, Business Analytics, Econometrics

Experience

SeeByte **Aug 2022 – Dec 2022**
Data Engineer Intern *San Diego, CA*

- Worked alongside the SeeByte ML Team to facilitate data preparation using Computer Vision Annotation Tool.
- Assisted the team on developing automated tools in data processing for future machine learning projects.
- Trained the model to improve the Automatic Target Recognition of mine-like objects for ROV Autonomy.
- Assisted development of user dashboard that allow for easier use of controlling Mobile Gateway Buoy Vehicle.

Acoustical Engineering Services **Jun 2021 – Aug 2022**
Data Analyst / Acoustical Engineer *Los Angeles, CA*

- Worked at Acoustical Engineering Services to collect, clean, and analyze data for clients, verifying acoustics were in compliance with regulation for each client's proposal.
- Created visualizations for clients' data, clarifying the necessary changes on designs of buildings for compliance.
- Assisted in modeling buildings from clients' design proposals using AutoCAD, testing acoustics.

Projects

Extreme 2018 Electoral Swings Analysis | *Redis, MongoDB, PostgreSQL, LlamaIndex* **May 2023 - June 2023**

- Utilized Psycopg2 to query news articles from SDSC database to clean and store into Redis and MongoDB.
- Trained LlamaIndex model by ingesting queried articles to create an LLM app to answer possible reasons for extreme electoral swings by use of input prompts.

LA County Covid-19 Dashboard | *Tableau* **July 2022**

- Extracted datasets from *data.lacity.org* to create an interactive Covid-19 Dashboard using LA County data.
- Presented different visualizations for users to easily understand information, such as number of positive cases and deaths, and vaccination rates by neighborhood and demographic.

Song Lyrics Generator Using Language Models | *NLTK, GPT-2, Scikit-learn, Genius API* **June 2022**

- Created a Ngram Language Model and utilized GPT-2 to generate new song lyrics given a corpus of an artist's lyrics.
- Utilized the Genius API to scrape all of an artist's lyrics.
- Presented a song's given highest TF-IDF and visualized the lyrics with a word cloud visualization.

Spotify User's Music Statistics | *JavaScript, Spotify Web API* **June 2022**

- Created a website called "A Glimpse of Your Spotify Statistics" which allowed users to view their top tracks, artists, and genres in 3 different time periods.
- Utilized the Spotify Web API to gather users' data and created data visualizations for users' statistics, such as a pie chart to view different genres.

Certifications

- AWS Certified Cloud Practitioner
- Google Data Analytics Certificate

Technical Skills

Languages: Python, SQL, R, JavaScript, TypeScript, Java, XPath, HTML, CSS
Technologies: PostgreSQL, MySQL, Redis, Apache Cassandra, MongoDB, Neo4J, Apache Solr, Docker, Tableau, Power BI, AWS, Linux, JMP, Excel
Other: NumPy, Pandas, Dask, PySpark, Scikit-learn, Matplotlib, Seaborn, D3.js, Git, Bash