

Samuel Edwards

s_edwards@berkeley.edu | [LinkedIn](#) | <https://samuel-elijah-edwards.com>

EDUCATION

University of California, Berkeley
Bachelor of Arts, **Data Science**

May 2024

Relevant Coursework: Data Engineering, Data Structures and Algorithms, Advanced Business Analytics

SKILLS

Programming Languages: Python, SQL, R, Bash, Java

Cloud Technologies: AWS, Linux

Data Analysis & ML: Data Mining and Analysis, Statistical Modeling, Machine Learning, Data Visualization

Tools & Technologies: Git, Pandas, Polars, MongoDB, Docker, Postman,

EXPERIENCE

Computer Scientist

08/2024 – Present

U.S. Census Bureau

Remote

- Engineered multi-step data tabulation software using **PostgreSQL** and **Python**, applying **data engineering principles** for **data modeling** and efficient database access via Python scripting. Implemented **data governance** (noise injection, automated verification) to ensure high-quality data for stakeholders.
- Contributed to the design and implementation of **PostgreSQL-based single source of truth**, optimizing data structures and developing Python scripts to create custom **table views** for enhanced data analysis.
- Significantly improved **data processing efficiency** by developing a **Python** module utilizing multiprocessing to convert SAS datasets to **Polars DataFrames**, reducing processing times **over 70%**. Also created a multi-step **data verification** process to analyze **data quality** and accuracy.

Information Technology Intern

07/2024 – 08/2024

Hilmar Cheese & Ingredients

Turlock, CA

- Designed and implemented **ETL pipelines** to extract, transform, and load data from SAP ECC, Azure Active Directory, and SAP SuccessFactors into an interactive **PowerBI** dashboard, significantly improving IAM system reporting and data accessibility for stakeholders.
- Automated the resolution process for IAM discrepancies by implementing **Power Automate** workflows, resulting in a **50% reduction in resolution times** and proactively maintaining **data security** by keeping the cybersecurity team informed with timely plans-of-action.
- Analyzed employee permissions** by querying SAP ECC, Azure Active Directory, and SAP SuccessFactors databases to identify over **100 IAM discrepancies**, ensuring **data consistency** and adherence to user access policies across disparate systems.

Full-Stack Web Development Intern

10/2023 – 02/2024

Visdomur Private Equity and Dalness Investment

Remote

- Enhanced website performance and user engagement by refactoring with **React** and optimizing **PostgreSQL database interactions** for efficient **data retrieval and storage**.
- Developed and implemented **JavaScript** event listeners to collect **user interaction and web traffic data** for subsequent analysis and user experience improvements.
- Designed and implemented a **REST API** using **Node.js** to facilitate the transfer of user registration data from the front-end to an **SQL database**.

PROJECTS

Customer Sales Segmentation Analysis (R, Excel, Statistical Modeling, Machine Learning)

- Utilized R for **data pre-processing** and **feature engineering** on a 100K-record transaction and demographic data set to enhance **data quality** for modeling.
- Leveraged statistical modeling in R for **feature selection** and conversion rate prediction, identifying distinct customer segments with a potential for over **10% ROI increase** in campaigns.
- Visualized key trends and patterns using **Excel** to derive **actionable insights**, communicated to stakeholders through a comprehensive PowerPoint presentation.

Stock Valuation Prediction Model (Python, Machine Learning, NLP)

- Developed a **Python-based machine learning model** for stock valuation, incorporating data from the Google Finance API and employing **feature engineering** techniques to enhance predictive accuracy.
- Engineered **data pipelines** in **Python** to integrate historical market data with news sentiment analysis, resulting in a **10% improvement in forecast accuracy**.
- Evaluated and iteratively refined the models performance metrics to optimize prediction accuracy.

National Basketball Association (NBA) Research Project (Python, Latex)

- Conducted comprehensive analysis of an NBA game dataset (2004-2022) using **Python** to extract key performance insights.
- Performed causal inference studies using Python to quantify the impact of stadium attendance on free-point rates, showing a **3% impact** on free-point successes.
- Presented finding through clear **data visualizations** and concise insights, effectively communication technical information to diverse audiences.