**Olivia Ray McCauley**

olive.ray.mccauley@berkeley.edu • (612) 913-2477 • <https://www.linkedin.com/in/oliviarmccauley/>

**EDUCATION**

**University of California, Berkeley**

B.A. in Data Science, B.A in Rhetoric (Honors). Class of 2023, Cum Laude (GPA: 3.88)

M.S. in Data Science. Class of 2024

2023 Best Honors Thesis, UC Berkeley Department of Rhetoric

**PROFESSIONAL EXPERIENCE AND PROJECTS**

**Data Engineering Intern, *Branch Metrics***  **Summer 2023**

* Designed a cost reporting dashboard for the Data Platform team, which displays resource usage over time, cost incurred by top clients, and cost breakdown by product type. Used Apache Airflow to automate a DBT SQL query and automatically ingest new cost data. Wrote LKML files to aggregate data for interactive dashboard visualizations.
* Wrote a 30+ unit tests for a Java class validating client inputs for a product request prior to generation of an API call. Unit tests deliver clear messages about errors and test failures to ensure codebase health.

**Data Science Intern, *Dell Technologies* Summer 2022**

* Supported pilot launch of Dell’s proprietary diagnostic software in the Services Engineering Department, monitoring data engineering practices prior to the global launch as Dell transitioned to Azure DataBricks.
* Designed autoupdating SQL dataset connecting telemetry, tech support, and field service data, created analytic dashboards on PowerBI and DataBricks. Automatic updates allow dataset maturation and reduce team workload.
* Created an automated statistical testing pipeline for product KPIs with Azure DataBricks. Significance testing indicates reliability of performance metrics comparisons between diagnostic softwares. Implemented ANOVA, Kruskal-Wallis, Whitney-Mann U, and Chi-Squared tests using scipy for hypothesis testing.
* Presented preliminary insights to VP of Services Engineering and handed off statistical testing pipeline to team for further development. Testing results import automatically into readable dashboards for non-technical audiences.

**Machine Learning Intern, *CU Data Partners*****Summer 2021**

* Designed a machine learning model to assess delinquent loan payment risk, which has become the basis for Red Wing Credit Union’s automated risk assessment system in the loan approval process.
* Implemented and integrated a machine learning pipeline with automatic ETL to query updating SQL databases, import and clean incoming data with Pandas in Azure Data Studio Jupyter Notebooks.
* Optimized scikit-learn model performance with feature engineering techniques (OneHotEncoding, Lasso Regularization, and PCA) to reduce dimensionality. Developed a logistic regression classifier to predicting risk of loan payment delinquency using loan payment data and credit union executive’s topical expertise.

**Pacman with Reinforcement Learning, *Intro to Artificial Intelligence***  **Fall 2022**

* Implemented autonomous an offline planning agent using Markov Decision Processes, and found the optimal set of actions using value and policy iteration with epsilon-greedy action selection
* Programmed a Q-Learning and approximate Q-Learning agent to identify the value of visiting a particular game state, and utilize the learned policy to win against ghost agents with a 90% success rate.

**Undergraduate Student Instructor**, ***UC Berkeley Department of Data Science*** **Summer 2021**

* Mentored ~40 students to achieve mastery in structured data manipulation, distribution center and spread, experiment design and random sampling for statistical testing, the Central Limit Theorem, bootstrapping sampling and confidence intervals, Bayesian probability, residual analysis, KNN, OLS Regression.

**Debate Coach, *Campolindo High School* 2020 - 2022**

* Designed and taught introductory debate curriculum emphasizing clear communication and formal argumentation structure to maximize rhetorical skill and competitive outcomes.
* Coached a group of 25+ students to become consistent tournament finalists; students won 4 tournaments in the 2020-2021 season, made semis and finals numerous times in the 2022 season.

**SKILLS**

**Programming Languages**: Java, Python, SQL. **Packages**: scikit-learn, Pandas, Matplotlib, Seaborn, NumPy, SciPy, JUnit, IO.File, PySpark, SciPy. **Platforms:** IntelliJ, JupyterHub, Azure DataBricks, SQLServer, GitHub, PowerBI, Tableau, Docker, Kubernetes, Bazel, Flink, AWS, DBT. **Statistical Methods/Skills:** A/B Testing, Bootstrapping, Linear/Logistic Regression, Decision Trees, RandomForest, Clustering, Statistical Testing and Automation. Ordinary Least Squares, Bias-Variance Tradeoff, Convex Optimization, Stochastic Gradient descent, K-Fold Regularization & Cross Validation, Bayes Nets, Hidden Markov Models. Markov chains, Markov chain Monte Carlo, Probability Densities. **NLP:** Statistical Language Models, Authorship Attribution, N-gram analysis.