

Megan Mehta

✉ mehta.megan24@gmail.com ☎ (925) 918-7428 🔗 @megan-mehta 📁 Research 🌐 Github

🎓 Education

Data Science (B.A.) + Society & Environment (B.S.), University of California, Berkeley 08/2021 – 05/2025
Data Science Scholar ✉ | Computer Science Scholar ✉ | Fung Fellow Nominee ✉ | Bronze-Tier Student Innovation Fellow ✉ | Delta Delta Delta
Courses: Inference & Decision Science, Probability, Machine Learning, Databases, Optimization, Data Engineering, Linear Algebra, Data Structures

💼 Work Experience

Machine Learning Scientist, Founding Team, VAIL (Verifiable AI Layer) ✉ 05/2026 | San Francisco, California
• Pioneering Model Informatics by building ML tools measuring agentic deviancy, stability in model behavior, and tool calling consistency

Data Science and Machine Learning Intern, Zoox 05/2024 – 08/2024 | Foster City, California
• Used **Bernoulli, Poisson**, and **Inverse Poisson** frequentist modeling to build **95% confidence intervals** to validate total collision avoidance metrics from four autonomous driving software versions. Built three **Bayesian Beta Binomial** models with an uninformative (Jefferys) and two different informative priors to estimate credible intervals. Analyzed prior-sensitivity and variance reduction
• Evaluated frequentist vs Bayesian model results. Presented statistical findings to the Collision Avoidance team and Director of Autonomy Metrics, enabling **20+ stakeholders** to interpret validity of current self-driving metrics, and isolate performance target metrics
• Used **Python** to consolidate data into **Apache Spark** to build a supervised learning binary classification ML model for agents predicted to interfere with projected driving trajectory in simulations.
• Used **Apache Spark** in **Databricks** to build empirical distributions of speed, acceleration, energy usage, and drive cycles in 5 cities.
• Built **neural networks, gradient boosted trees, ridge and lasso regression** predictive modeling for geographically dependent energy usage
• Used **Looker Studio** with **PostgreSQL** and **dbt** to create manufacturing safety, incident tracking, and badging access dashboards
• Documented SOPs, user guides, and project deliverables on **Confluence** and **GitHub** for stakeholders for all interdepartmental projects

Data Analyst Intern | Business Intelligence & Solutions, Comcast 06/2023 – 08/2023 | Livermore, California
• Conducted **investigative data analysis** and trained linear regression model with **Scikit-learn** to evaluate Q1/Q2 Comcast Business Mobile (CBM) quota performance for 15+ sales teams. Uncovered critical XM360 Order Processing Platform complications, leading to **invisible negative correlations** between CBM sales and average revenue per sale, affecting quotas. Used **Ridge Regression** and **Lasso Regression** to predict sales.
• Redesigned internal commission structures with Finance & Business Development teams to aggressively prioritize and incentivize CBM sales. Used **Excel, Tableau**, and **Python** to calculate a projected **\$4,500,000+ in revenue for Q3/Q4** stemming from our sales strategy
• Partnered with Accounting, Software, and Sales teams to resolve order platform complications, saving sales teams **1,000+ hours (~\$1,500,000+ in labor & revenue)** per quarter. Used **Seaborn** to visualize sales performance, distribution, and revenue metrics used in commission package revamp. Presented proposal recommendations to **12+ Regional Senior VPs, 25+ directors, & CEO's direct reports**

Machine Learning Engineering Intern, Climformatics 01/2023 – 05/2023 | Berkeley, California
• Created machine learning algorithm with **Python, Jupyter Notebooks, Bash, Sabalcore High Performance Cloud Computing**, and 5 datasets with **5000+ rows** to predict how arctic sea ice albedo restoration affects European heatwaves, wildfire risk, and climate patterns
• Conducted exploratory data analysis, data cleaning & processing, and visualizations with **Pandas, Seaborn, NumPy**, and **Matplotlib**
• **Abstract submitted for publication** in the *American Geophysical Union*, presented findings at SP23 Data Discovery Research Symposium

Technical Product Manager, UC Berkeley College of Computing & Data Science ✉ 01/2022 – 05/2025 | Berkeley, California
• **Managed 35+** teammates, including 5 graduate students, and oversaw a program that served **2400+ students** and **1,000+ client projects**
• Hired 10+ colleagues in <1 year. Manage cross-functional teams with **Kanban** boards & **Agile** methods via **Asana**. Led strategic growth initiatives
• Identified KPIs to pioneer internal performance analytics, and helped secure funding to scale programs by **65%**
• Drove internal sprint processes by leading weekly staff-wide standup and data, consulting, and analytics sub-team meetings
• Helped organize **7 program symposiums**, showcasing a total of **1,500+** project deliverables and facilitating networking for **1,600+** students
• Institutionalized operational SOPs, internal, student and client-facing guidebooks. Standardized project deliverable expectations and quality metrics and documented inter-departmental legacy knowledge dependencies to enable continuity

Program Manager, UC Berkeley Center of Entrepreneurship & Technology ✉ 04/2023 – 05/2025 | Berkeley, California
• Helped organize the 2023, 2024, & 2025 annual summit event exclusively serving **top 2%** ranked Student Innovation Fellows. Connected student founders with industry-relevant mentors, startup accelerator resources, and pre-seed fundraising avenues.
• Created the premier Berkeley-centered innovation sticker contest to facilitate active engagement and marketing efforts with students
• Organized 9 alt-meat, renewable energy, and biotech entrepreneurship challenges serving **350+** Student Innovation Fellows
• Cross-functionally collaborating with production, outreach, and finance teams for marketing campaigns to scale the program by **50%**

📄 Research Publications

Artificial Enhancement of Arctic Sea Ice Albedo for Fire-Weather Risk 07/2023
American Geophysical Union, Climformatics, UC Berkeley Data Discovery
• Created machine learning algorithm with **Python, JupyterNotebooks, Bash, Sabalcore High Performance Cloud Computing**, and 5 datasets with 5000+ rows to predict how arctic sea ice albedo restoration affects European heatwaves, fire risk, and climate patterns
• Conducted exploratory data analysis, data cleaning & processing, and visualizations with **Pandas, Seaborn, NumPy**, and **Matplotlib**
• **Abstract submitted for publication** in the *American Geophysical Union*, presented findings at SP23 Data Discovery Research Symposium

🧠 Skills

Languages, Python, Java, C++, C, SQL, Markdown, LaTeX, Regex / English, Gujarati, Hindi, Italian, Spanish

Tools, Git, Bash, Databricks, Apache Spark, Jupyter Notebook, VSCode, Looker, IntelliJ, High Performance Cloud Computing, Excel, Canva, Tableau, Deepnote, Replit, Asana, Notion, Airtable, Jira, Figma, Cursor, Claude Code, Codex

Libraries/Frameworks, Scikit-learn, Pandas, NumPy, Seaborn, PyTorch, TensorFlow, Sci-Py, Regex, Matplotlib, GitHub, JUnit, NLTK