

Justin Yee

Morgan Hill, CA | +1 (408) 784-2046 | Justinyee2005@gmail.com | <https://www.linkedin.com/in/jhyee> | <https://github.com/JhcYee>

Summary

Data Science graduate from UC Davis with professional experience building automated ETL pipelines and Python/SQL-based data workflows. Experienced in designing Tableau and Power BI dashboards that cut reporting turnaround time by **43%** and reduced data latency by **88%**. Skilled in statistical analysis, data validation, and translating complex datasets into clear, stakeholder-ready insights. Seeking an entry-level Data Analyst or Business Intelligence Analyst role to apply strong SQL, Python, and dashboard-design skills to data-driven decision-making.

Experience

Tatari | *Data Science BI Intern*

Jun 2026 - Sep 2026

- Cut data latency by **88%** and eliminated manual reconciliation across **4** disparate data sources by engineering an automated ETL pipeline through SQL and Python-based data pipeline design.
- Slashed reporting turnaround time by **43%** by replacing **6** fragmented dashboards with one self-service Tableau dashboard by designing customizable KPI views by breakout level and timeframe for cross-functional stakeholders.
- Delivered real-time metric visibility for internal teams and external clients with zero manual pulls by owning **5** live Tableau dashboards through end-to-end dashboard design and stakeholder reporting.

Davis Data Science Club | *Project Lead*

Jan 2026 - May 2026

- Built an AI-assisted study platform on schedule for a six-member cross-functional team by owning system design and sprint planning through hands-on implementation in Python, SQL, and front-end development.
- Sustained consistent development velocity while reducing integration bottlenecks across all platform components by streamlining team workflows through a scalable architecture design.

The Psychological and Measurements Lab | *Research Assistant*

Sep 2025 - Dec 2025

- Transformed unstructured academic sources into structured, analysis-ready datasets by extracting DOI-linked author contact data with Python (requests, BeautifulSoup) and Pandas, enabling the research team to analyze a large collection of records for subsequent studies.
- Boosted research outreach productivity by **100%** by validating scraped, unstructured data through Python-based workflows with heuristic pattern detection and Power BI reporting.

Learnbotix | *Data Engineer*

May 2025 - Nov 2025

- Converted structured and unstructured research data into scalable, analysis-ready reporting datasets by engineering automated data pipelines through PostgreSQL database design.
- Improved evaluation efficiency by **72%** by generating real-time analytical insights and visual summaries through Python-based data processing, feature extraction, and AI engineering.
- Drove down manual data preparation time by over **50%** and raised dataset reliability by detecting anomalies and missing values through automated data validation and quality-check workflows.

The Human Development and Media Lab | *Research Assistant*

Sep 2024 - Dec 2024

- Uncovered behavioral trends and risk factors in parent-child technology use (ages 1–10) by performing exploratory data analysis and communicating findings through interactive Power BI dashboard reporting.
- Linked parental mediation strategies to a **22%** reduction in screen-time-related risk factors among children ages 1–10 by evaluating intervention outcomes through EDA and interactive Tableau dashboard analysis.

Education

University of California, Davis

Sep 2023 - Jun 2026

Bachelor of Science, Data Science

- **Achievements:** Dean's List Scholar (2x)
- **Coursework:** Applied Machine Learning, Data Processing Pipelines, Algorithms for Data Science, Optimization, Probability Theory, Data Structures, Databases, Applied Time Series, Applied Unsupervised Learning

Projects

Heartifact

Dec 2024 - Mar 2025

- Engineered an end-to-end ECG classification pipeline on 20,000 cardiologist-annotated beats from the MIT-BIH database, benchmarking band-pass against high-pass filtering and extracting QRS-centered windows with per-beat normalization to make noisy physiological signal model-ready.
- Identified the best-performing approach by comparing four supervised classifiers on identical inputs, reaching **99.05%** test accuracy and a **1.3%** false negative rate against a **5%** clinical-risk target.

Cyclistic Case Study

Sep 2023 - Dec 2023

- Surfaced seasonal commuter vs. leisure usage patterns across 12+ months of rider data by analyzing behavioral differences through Python (Pandas, NumPy) statistical analysis.
- Projected a lift in casual-to-member conversion by delivering targeted marketing recommendations through interactive Tableau dashboards visualizing ride frequency, trends, and trip duration.

Skills

- **Languages:** Python, R, SQL
- **Libraries & Tools:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Tableau, PostgreSQL, Git, Power BI, Excel, Databricks, Salesforce
- **Competencies:** Data Cleaning, Feature Engineering, Exploratory Data Analysis (EDA), Dashboard Design, Data Pipeline Engineering, Data Modeling, Data Automation