

Eric Villanueva

San Francisco, CA | (415) 596-8558 | ericjr143@gmail.com | [LinkedIn](#) | [Portfolio](#)

EDUCATION

Chapman University | Orange, CA

Aug 2021 – May 2025

B.S. Computer Science | Minor in Business Administration | GPA: 3.783 | Dean's List

TECHNICAL SKILLS

Languages & Querying: Python (Pandas, NumPy, Scikit-learn), SQL (MySQL), R, C++, Java

Analytics & Visualization: Tableau, Power BI, Excel (advanced), Matplotlib, Seaborn

Data & ML: Machine Learning (classification, regression, EDA), Feature Engineering, Statistical Analysis, Data Cleaning

Tools & Environments: Git, Linux, DataGrip, VSCode, Jupyter Notebooks

CERTIFICATIONS

Data Analyst Associate Certification | DataCamp

Data Scientist Associate Certification | DataCamp (Expected July 2025)

WORK EXPERIENCE

Research & Data Operations

Jan 2025 – Present

Stealth Robotics Startup

San Francisco, CA

- Collected and structured operational data from robotic systems performing household task automation to support ML model development and training pipelines.
- Worked with engineering teams to apply standardized data collection protocols, keeping datasets consistent and reliable across evaluation cycles.
- Contributed to iterative model evaluation by surfacing behavioral patterns and performance gaps through systematic data review.

Front-End Web Development Intern

Jun 2020 – Sep 2020

Code Tenderloin

San Francisco, CA

- Built and deployed a website using HTML, CSS, and JavaScript to highlight housing insecurity and social injustices facing San Francisco's unhoused population.
- Documented and analyzed tech employee work culture to surface insights on industry accessibility and barriers for underrepresented communities.

PROJECTS

COVID-19 Global Data Analysis | SQL, Excel, Tableau

- Modeled and joined two tables across 200+ countries to analyze infection rates, death counts, and vaccination rollout over time.
- Wrote queries using window functions, CTEs, and temp tables to track rolling vaccination totals and infection rates by country.
- Built SQL views to store key metrics by continent and country for downstream visualization.

Spotify Streams Analysis | Python, Scikit-learn, Pandas

- Analyzed 4,600+ Spotify songs and found Playlist Count ($R^2 = 0.74$) was a significantly stronger predictor of streams than Playlist Reach ($R^2 = 0.32$).
- Compared polynomial regression models across degrees 2, 3, and 4 on an 80/20 split; degree-2 was optimal as higher-degree models severely overfit the training data.
- Applied GMM clustering with AIC selection and compared DBSCAN vs. K-Means on 2,600+ songs, selecting K-Means ($k=2$, silhouette = 0.49) for cleaner cluster results.