

# Sean Sanchez

seansanchez951@gmail.com | [linkedin.com/in/seansanchez951](https://www.linkedin.com/in/seansanchez951) | [github.com/seansanchez951](https://github.com/seansanchez951)

Amsterdam, NL

## Experience

---

### **Data Scientist, DJ Monitor** | Jan 2023 – Present

- Developed ETL data ingestion pipelines, streamlining real-time data collection and aggregation for improved scalability.
- Implemented automated visualization dashboards for weekly field reports on hardware device data.
- Designed similarity matching algorithms for music metadata parent-child relationships, employing vectorized scoring techniques to enhance backend database transparency.
- Optimized PostgreSQL queries and indexes for Data API, resulting in a 92% reduction in query response times (from 8s to 0.6s)

### **Data Analyst Intern, Pathloom** | Aug 2021 – Nov 2021

- Spearheaded collection, cleaning, and analysis of extensive third-party hiking trail data, leveraging Python and geo data libraries to ensure data accuracy and integrity.
- Engineered and imported fully structured databases to AWS, optimizing accessibility for software development teams and facilitating seamless integration into the cloud environment.
- Collaborated with organizational leadership to successfully integrate novel data types for application database schemas.

### **Spare Parts Coordinator, Vanderlande Industries (LAX)** | May 2017 – Oct 2018

- Managed the inventory and purchasing of spare parts for Tom Bradley International Airport's baggage and jet bridge system.
- Created reporting dashboards to service recurring and ad hoc inventory data requests.

## Projects

---

### **Sentiment Analysis on Facebook Messages** | Machine Learning and Natural Language Processing

- Feature engineered raw text data for training deep learning model using EDA, data cleaning, text vectorization.
- Trained Keras model to determine an individuals emotional state given their Facebook messages.
- Technologies used: Python, Keras, Tensorflow, Matplotlib, Google Colaboratory.

### **Reviewing Pitchfork Music Reviews** | Data Visualization

- Analyzed Pitchfork music reviews and built custom data visualizations in D3.js.
- Built website to showcase visualizations along with reporting insights based on project objectives.
- Technologies used: D3.js, HTML/CSS, Python.

### **Counting Cells in Microscopy Images** | Data Science and Machine Learning

- Ranked top 4 in Kaggle competition by creating deep learning model with 84% accuracy.
- Built a U-Net convolutional neural network for biomedical image segmentation.
- Technologies used: Python, Keras, Tensorflow, Scikit-Learn, Google Colaboratory.

### **Binary Classification of Breast Cancer Presence** | Data Science and Machine Learning

- Achieved 98% classification accuracy using stochastic gradient descent method to build machine learning model.
- Trained model to predict the presence of breast cancer in an individual based on a set of health indicators.
- Technologies used: Python, Numpy, Scikit-Learn, Google Colaboratory.

## Education

---

### **B.S. Data Science, University of San Francisco** | 2022

- Computer Science Coursework: Data Structures and Algorithms, Machine Learning, Database Engineering, Object-Oriented Programming with Java, Software Development, Data Visualization, Data Science with R
- Mathematics Coursework: Statistics with Python, Linear Regression, Probability with Applications, Linear Algebra