

# Trevor Santiago

[trevorjsantiago1@gmail.com](mailto:trevorjsantiago1@gmail.com) | (209) 485-0240 | Sacramento, CA | [trevorsantiago.com](http://trevorsantiago.com)

## WORK EXPERIENCE

---

### ALLDATA

Nov. 2021 – Present

*Data Scientist*

*Elk Grove, CA (Hybrid)*

- Developed and implemented a [Python package](#) for a novel string similarity score, replacing a deep learning model to map vehicle information between standards and extract vehicle components from free text. This innovation increased efficiency by approximately 80% while maintaining accuracy.
- Built an automated cloud-based data pipeline in GCP that processes millions of rows of data from all US-based AutoZone stores monthly and prepares it for ingestion into a highly prioritized new product.
- Currently designing and training a deep learning Tensorflow model as part of an NLP pipeline to estimate labor hours given vehicle information and a work description
- Awarded “Extra Miler of the Year” for 2023 for my work on multiple data pipelines that added millions of records into two new products.

### New York Mets

Jan. 2021 – Aug. 2021

*Data Science Intern*

*Remote*

- Minimized runs scored against the Mets by building an outfield defensive alignment model combining results from a hit classification model and a catch probability model using Scikit-Learn and XGBoost.
- Constructed a multi-dimensional distribution for estimating the likelihood of a hit outcome using SciPy, integrated as a component of the defensive alignment model.
- Created visualizations for manual validation of the outfield defensive model’s performance using Matplotlib.
- Developed a novel dataset of MLB venue outfield wall distance measurements using data interpolation and spline fitting for use as a feature in the hit classification model and for future use by the Mets’ analyst team.

## EDUCATION

---

### University of San Francisco

August, 2021

*MS, Data Science*

*San Francisco, CA*

- ReadTheSign Project:
  - Acted as CTO for a 7 person team developing a web application for translating videos of American Sign Language gestures to text using PyTorch.
  - Deployed the web application to AWS Elastic Beanstalk using Docker.
- Coursework:
  - Machine Learning, Deep Learning, Case Studies (Network Analysis, Topic Modeling), SQL, Distributed Computing (Spark), Time Series Analysis, Design of Experiments (A/B testing), Data Structures and Algorithms, Entrepreneurship (Web App Development), Data Ethics.
- 3.81 GPA

### University of California, Santa Barbara

June, 2020

*BS, Mathematical Sciences*

*Santa Barbara, CA*

- Earned a minor in Statistical Science.

## CERTIFICATIONS, SKILLS & INTERESTS

---

- **Certifications:** [Data Engineering Foundations](#); [Google Cloud Platform for Machine Learning Essential Training](#); [Neural Networks and Deep Learning by deeplearning.ai](#)
- **Skills:** Python; SQL; GCP; Docker; Git; Postgres; Scikit-Learn; PyTorch; Tensorflow; PySpark; Pandas; Flask.