

CARLOS MONSIVAIS

San Diego, CA

619-305-7277

carlos.s.monsivais@gmail.com

Projects Website

<https://carlosmonsivais123.github.io/#>

GitHub

<https://github.com/carlosmonsivais123>

LinkedIn

<https://www.linkedin.com/in/carlosmonsivais/>

- **Programming:** Python (Pandas, Numpy, Matplotlib, Plotly, TensorFlow, TensorFlow Extended, Scikit-Learn, PyTorch, Scrapy, Dash); PySpark; RStudio; Bash; Git
- **Cloud:** GCP (BigQuery, Dataproc, Kubernetes, Cloud Composer, Vertex AI, AutoML, Dataprep, Google Cloud SDK, GCS); AWS (EMR, Lambda, Amazon SageMaker, EC2, RedShift, S3)
- **Databases:** SQL (ML Model Queries, Query Optimization); PostgreSQL; Neo4j Graph Database
- **Software:** Tableau (TabPy, Calculated Field), Power BI, Microsoft (MS) Excel, MS Word and PowerPoint

EDUCATION

Masters, Data Science and Engineering; University of California, San Diego

Bachelor of Science, Statistics; University of California, Davis

Bachelor of Arts, Economics; University of California, Davis

CERTIFICATIONS

Building Batch Data Pipelines on Google Cloud | Google Cloud

ML Pipelines on Google Cloud | Google Cloud

PROFESSIONAL EXPERIENCE

Accenture – San Diego, CA June 2022 - Present

Data Scientist (June 2022 - Present)

- Produced fraud detection models using Machine Learning to detect fraudulent claims for Aetna CVS Healthcare by applying rule based logic and XGBoost classifiers with 98% accuracy, saving approximately \$20,000,000 per year in three markets.
- Achieved industry standard documentation by creating ML Ops processes for Aetna CVS Healthcare for ongoing model analysis and thresholds notifying when a model needs to be retrained.
- Created regression models using a combination of Non-Parametric Statistical methods and Linear Regression to create a recommendation list of the top 10 locations to place 5 new Rolex training facilities and stores within the United States. This automated process saved the team 60 days of delay time by not having to go to 12 different department subject matter experts.
- Saved 8 hours per week by creating efficient data pipelines where API requests from 3 different data sources were automated using AWS Lambda Functions with a rule based system to clean, merge and place data in the correct S3 and Redshift locations.
- Established data validation processes for streaming data being captured through API's with a schema check and data transformations using TensorFlow Data Validation, Python and PySpark manipulations to ensure good data quality.
- Automated client PowerPoint analytics using Python for graph creation in Plotly and text creation that was later embedded on client slide decks to show the results of using the Accenture AI service which saved about 40 hours per month.

CoreLogic – San Diego, CA August 2019 - June 2022

Data Analyst (April 2021 - June 2022)

Business Analyst (August 2019 - April 2021)

- Created outlier detection system using TensorFlow Extended tools, TensorFlow Models and BigQuery Modeling tools through a Vertex AI pipeline for products in the Fraud space, and Home Price Indexes. This created a process for end to end data science work including data extraction, modeling and processing metadata saving 10 hours per week between the engineering and modeling teams.
- Established documentation and processes for ML Ops by saving models in GCS buckets and determining best processes to document outliers for an outlier detection system. As part of the process, I also included how to analyze metadata in TensorFlow Extended by looking at model changes, and data quality changes for each data extract.
- Recognized as a top 10 finalist during "Innovation Challenge" for proposing a solution to include more accurate data impacting zip codes and across more states by implementing clustering techniques to determine high quality and low quality data.
- Saved 16 hours of monthly analyst time by automating manual calculations in the S&P CoreLogic Case-Shiller Home Price Indices, which tracks changes in value of U.S. residential real estate prices.