RAUL G. MARTINEZ

gio.mtz3@gmail.com | +1 619-734-6231 | San Diego, CA | raul-giovanny-martinez-688b1bb4 | github.com/raulgiovannymartinez

WORK EXPERIENCE

Data Engineer III, Thermo Fisher Scientific

June 2021 – Present | Carlsbad, CA

- Designed AWS Cloud Architectures, Batch Processing Pipelines, Data QC, and Custom ETL Code for orchestrating big data applications in qPCR COVID-19 clinical results.
- Supported data science efforts for tracking and monitoring pandemic related topics through dashboards and analysis for external customer's specific data and internal product investigations.

Data Engineer II, Illumina

July 2018 – June 2021 | San Diego, CA

- Developed Data Integration, ETL, and Data Modeling solutions for instrument's hardware tests and sequencing data across all NGS platforms.
- Enabled ad-hoc analytics, reporting, and insight generation for key metrics and transformed outputs for the many NGS core components such as optics, micro-fluidics, firmware/software, chemistry. Along stages of product development and manufacturing.

Pharma-Technical Development Research, Genentech/Roche

February 2018 – July 2018 | South San Francisco, CA

- Collaborated in process development/purification in viral clearance studies (virus removal, inactivation) for the production of therapeutic monoclonal antibodies.
- Investigated and optimized virus titer infectivity assays by implementing design of experiments (using SAS JMP), inferential statistics, and regression analysis.

Software Test Engineer, Dotmatics

December 2017 – February 2018 | San Diego, CA

- Collaborated with Software Developers and Application Scientists to troubleshoot customer-specific issues and enhancements for the many scientific/biological informatics product suite offerings by testing software front-end features and components, UX/UI, software architecture, and Oracle relational databases.

Data Analyst, Retrovirox

May 2016 – July 2017 | San Diego, CA

- Implemented data analysis and processing workflows for statistical evaluation (linear regression, Z-factor) in high-throughput screening of thousands of novel small-molecules using FACS Flow Cytometry techniques for measuring antiviral activities.
- Executed data management and documentation for R&D experiments in support to data integrity, reporting, visualization, and experimental design.

PROJECTS AND RESEARCH

PyTorch Deep-Learning Time-Series Forecasting Library (2021)

- Contributed in the development of an efficient and user-friendly open-source library, with state-of-the-art research and benchmark reports, for the PyTorch community.
- Integrated and tested research code for Seq2Seq RNN and DCRNN models, for up to spatio-temporal predictions. Models were tested against traffic data, counts and speed, from CalTrans State Highway network of sensors to evaluate COVID-19 traffic trends.

Drug Epidemic Tracking Through Social Media (2020)

- Built knowledge graph with Data Integration, NLP, and Entity-Matching strategies such as Blocking, Similarity Functions, Feature Vectors, and ML-Matchers.
- Consolidated a graph database schema (in Neo4j) to explore drug abuse and overdose patterns, with Cypher declarative queries, across the data ingested from Reddit posts and US News articles.

Predicting Helpfulness of Amazon Reviews (2020)

- Analyzed numerical and text data from 200,000 reviews in order to extract useful information with feature engineering and NLP techniques. Feature selection was done with an ablation study.
- Developed a Random Forest Regressor Model to predict the ratio between the number of helpful votes with total votes, performance metrics used include MAE, MSE, Precision, and Recall.

EDUCATION

Masters Degree in Data Science and Engineering, GPA 4.0/4.0

September 2019 – June 2021

University of California, San Diego

Bachelors Degree in Bioengineering Biosystems, GPA 3.4/4.0

October 2014 – June 2017

University of California, San Diego
Minor: Entrepreneurship and Innovation

PROFESSIONAL SKILLS

- Domain Related

Data Warehouse | Data Lake | Big Data | Serverless | Data Integration | ETL/ELT Pipelines | Data Quality | Data Models & Design | Visualization | System Design | Software Dev | AI/Machine Learning

- Programming and Software

Python | R | C#/C++ | PyTorch | Pandas | Scikit-learn | Object-Oriented | APIs | Databricks | Hadoop/Spark | SQL/NoSQL | Tableau/Power BI | GitHub | SAS JMP | AWS | Lambda | Athena | Redshift | Batch | Glue

RELEV. COURSEWORK

Data Visualization | Data Integration and ETL | Beyond Relational Data Models | Machine Learning | Data Analysis Using Hadoop and Spark | Probability & Statistics Using Python | Data Management Systems | Python for Data Analysis

ACHIEVEMENTS

- Chancellor's Associate Scholar - UCSD (2014-2017).
- Community College Scholarships - "ASO Campus Involvement" and "Neva Smith" (2014).