RAUL G. MARTINEZ

@ gio.mtz3@gmail.com 📞 +1 619-734-6231 📍 San Diego, CA in raul-giovanny-martinez-688b1bb4 ⌂ raulgiovannymartinez

WORK EXPERIENCE

Data Engineer III, Thermo Fisher Scientific
📅 June 2021 – Present 📍 Carlsbad, CA

- Managed Data Ops, IT Support, and Data Quality activities for over a dozen ERP systems (Terabytes) on Supply Chain domains across the Enterprise Data Platform, from Raw to KPI Layers, using Databricks as the integration tool and Airflow for orchestration.
- Designed AWS Architecture for real-time monitoring and tracking (PowerBI) of qPCR clinical results for millions of COVID-19 lab samples, to provide actionable insight on pandemic related topics in data science, epidemiology, bioinformatics, and instruments.

Data Engineer II, Illumina
📅 July 2018 – June 2021 📍 San Diego, CA

- Developed data integration and ETL pipelines in the Data Warehouse, to support NGS component tests and sequencing experiments on manufacturing and NPI engineering.
- Advanced our analytics activities through algorithm development for KPI and anomaly detection, data platform reporting (Tableau), and ad-hoc analysis for systems integration.

Purification Development Engineer Intern, Genentech
📅 February 2018 – July 2018 📍 South San Francisco, CA

- Collaborated in virology validation studies related to viral clearance, used in purification process development for the making of therapeutic monoclonal antibodies.
- Streamlined the in-house production for a model retrovirus, with high infectivity, by implementing cell culture, RNA extractions, qPCR, DoE, and statistical 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.
- Tested the different elements for the scientific/biological informatics product including front-end, user experience, software architecture, and relational databases.

Data Analyst, Retrovirox
📅 May 2016 – July 2017 📍 San Diego, CA

- Executed data management, integrity, and documentation for scientific lab experiments and CRO reports in small-molecules research for HIV and Dengue antivirals.
- Implemented data analysis, interpretation, statistical evaluation, and visualizations for high-throughput screening using FACS Flow Cytometry and Microplate Reader outputs.

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 (in Neo4j) to explore drug abuse and overdose patterns, with Cypher queries, across data from Reddit posts and US News articles.

Predicting Helpfulness of Amazon Reviews (2020)

- Analyzed numerical and text data from 200,000 reviews with feature engineering, NLP techniques, and an ablation study.
- Developed a Random Forest Regressor Model to predict the ratio 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

`Big Data` `AI/Machine Learning` `Data Architecture` `Data Operations` `Data Warehouse` `Data Integration` `ETL/ELT` `CI/CD` `Data Quality` `Data Models & Design` `Visualization`

- Programming and Software

`Python` `SQL` `Spark` `AWS` `Databricks` `Airflow` `R` `PyTorch` `Pandas/Numpy` `Scikit-learn` `Object-Oriented` `APIs` `Tableau/Power BI` `GitHub`

RELEV. COURSEWORK

`Data Integration, ETL & Visualization` `Machine Learning` `Hadoop and Spark` `Probability & Statistics with Python` `Data Management Systems`

ACHIEVEMENTS

- AWS Architecture for Real-Time Analytics on qPCR Outputs from TaqPath COVID-19 Diagnostic Tests (2021-2022).
- Illumina NextSeq 1k/2k Product Launch, Involved Since Proof-of-Concept Phase (2018-2020).
- Chancellor's Associate Scholar - UCSD (2014-2017).
- Community College Scholarships - "ASO Campus Involvement" and "Neva Smith" (2014).