Larry Miguel R. Cueva

larrymiguelcueva@gmail.com | (+63) 970 745 1021 | https://www.linkedin.com/in/michaelcueva | https://github.com/08Aristodemus24

TECHNICAL SKILLS

Core Competencies: Data Cleaning/Preprocessing | Web Scraping | Data Warehouses | Data Modelling | ETL | Data Orchestration | Cloud Infrastructure | Data Analysis
Languages & Tools: Python | SQL | Azure | PowerBI

EXPERIENCE

Virtuals Protocol **Dec 2024 – Jan 2025**
Data Engineer, Intern

- Cleaned and processed more than 500k rows of data for various retrieval augmented generated (RAG) AI agents.
- Developed and wrote scripts automating data ingestion processes of RAG AI agents and pulling raw datasets uploaded by users diverting main workflow to data transformation.

Creative Dynamix Solutions, Inc. **Sep 2022 – Oct 2022**
X++ Developer, Intern

- Utilized AnyDesk in tunneling through remote virtual machine for reporting tasks
- Developed and queried data to enhance sales reporting using PowerBI and X++

PROJECTS

eda-denoiser-stress-detector | *React.js, D3.js, Flask, Scikit-Learn, Tensorflow, Docker*

- Enhanced the accuracy and reliability of bio-signal denoising and stress detection by developing a novel hybrid LSTM-SVM deep learning model, addressing critical challenges in bio-signal data analysis. Link to research: https://aristodemus8-eda-denoiser-stress-detector.hf.space/
- Engineered and deployed a full-stack web application demonstrating the utility and potential of the validated LSTM-SVM model in real world health monitoring applications.
- Validated model performance of 90% AUC & 78% accuracy in biosignal denoising, providing a robust foundation and methodology for future bio-signal research and potential diagnostic tools.

signal-gender-predictor | *SQL, DuckDB, Librosa, Azure Data Lake, Azure Data Factory, Airflow, Terraform, Docker*

- Developed an end-to-end MLOps pipeline for a gender prediction model based on audio signals, reducing cloud operational costs by over 70%, leveraging cloud only for compute during extraction and storage
- Automated the ingestion, transformation, and feature engineering of large-scale audio datasets, generating high-quality features for model training and serving.
- Architected and implemented a scalable multi-stage data pipeline to efficiently process 3.7 billion rows of audio signals, generating high-impact features for voice-based gender prediction, providing a framework for real-time audio analytics and a foundation for voice-based AI applications.

chronic-disease-analyses | *SQL, PowerBI, Apache Spark, Airflow, Selenium, S3, DuckDB, Docker*

- Processed and transformed 20 years of comprehensive US public health data (from 2001-2021) using Spark, consolidating disparate datasets to quantify chronic disease cases and population figures. Link to project: https://chronic-disease-analyses.vercel.app/
- Conducted analysis of chronic disease data to identify most prevalent disease, allowing for potential in more targeted interventions and improving cost efficiency for less prevalent diseases

EDUCATION

Polytechnic University of the Philippines **Aug 2019 – Mar 2025**
Bachelor of Science in Computer Science

- 1.9 GPA