

Larry Miguel R. Cueva

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

SUMMARY

A Computer Science major, driven to grow more in the field of Data Engineering. I am an analytical thinker, skilled in orchestrating scalable ETL/ELT pipelines for analytics and AI. The recent projects & internships I had allowed gain experience in large data processes, using distributed computing frameworks like Apache Spark and I believe it would be beneficial for the roles in Data Engineering/Analytics, I seek.

EXPERIENCE

Virtuals Protocol <i>Data Engineer, Intern</i>	Dec 2024 – Jan 2025
<ul style="list-style-type: none">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. <i>X++ Developer, Intern</i>	Sep 2022 – Oct 2022
<ul style="list-style-type: none">Utilized AnyDesk in tunneling through remote virtual machine for reporting tasksDeveloped and queried data to enhance sales reporting using PowerBI and X++	

PROJECTS

eda-denoiser-stress-detector <i>React.js, D3.js, Flask, Scikit-Learn, Tensorflow, Docker</i>
<ul style="list-style-type: none">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/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 <i>SQL, DuckDB, Librosa, Azure Data Lake, Azure Data Factory, Airflow, Terraform</i>
<ul style="list-style-type: none">Developed an end-to-end MLOps pipeline for a gender prediction model API based on audio signals, reducing cloud operational costs by over 70%, leveraging cloud only for compute during extraction and storage. Link to API: https://aristodemus8-signal-gender-predictor.hf.space/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 <i>SQL, PowerBI, Apache Spark, Airflow, Selenium, S3, DuckDB, Docker, Terraform</i>
<ul style="list-style-type: none">Processed and transformed 20 years of comprehensive US public health data (from 2001-2021) using Spark, extracting and 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

TECHNICAL SKILLS

Core Competencies: Data Cleaning Web Scraping Data Warehousing Data Modelling ETL Data Orchestration Cloud Infrastructure Data Analysis APIs
Languages & Tools: Python SQL Azure Cloud Services Spark Linux

EDUCATION

Polytechnic University of the Philippines <i>Bachelor of Science in Computer Science</i>	Aug 2019 – Mar 2025
<ul style="list-style-type: none">1.9 GPA	