

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 DBT which I believe it would be beneficial for the roles in Data Engineering/Analytics, I seek.

EXPERIENCE

Virtuels 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

forums-analyses | SQL, DBT, Snowflake, Airflow, Delta Lake, S3, Terraform, PowerBI

- Developing an end-to-end ELT Data Lakehouse pipeline to ingest and model unstructured data from forums Reddit, providing daily public sentiment as a potential for general intelligence.
- Automating a cost-efficient, scalable storage solution by utilizing S3 for primary storage and Snowflake solely for compute-intensive transformations, to directly derive business insights from emerging market trends via public forums. <https://github.com/08Aristodemus24/forums-analyses>

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

- 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 | SQL, PowerBI, Apache Spark, Airflow, Selenium, S3, DuckDB, Docker, Terraform

- 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 | DBT | Snowflake

EDUCATION

Polytechnic University of the Philippines

Aug 2019 – Mar 2025

Bachelor of Science in Computer Science

- 1.9 GPA