

# Larry Miguel R. Cueva

[larrymiguelcueva@gmail.com](mailto: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 Warehousing | Data Modelling | ETL | Data Orchestration | Cloud Infrastructure | Data Analysis | RESTful API  
**Languages & Tools:** Python | SQL | Spark | Airflow | Azure Cloud Services | Git

## EXPERIENCE

<b>Virtuals Protocol</b> <i>Data Engineer, Intern</i>	<b>Dec 2024 – Jan 2025</b>
<ul style="list-style-type: none"><li>Cleaned and processed more than 500k rows of data for various retrieval augmented generated (RAG) AI agents.</li><li>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.</li></ul>	
<b>Creative Dynamix Solutions, Inc.</b> <i>X++ Developer, Intern</i>	<b>Sep 2022 – Oct 2022</b>
<ul style="list-style-type: none"><li>Utilized AnyDesk in tunneling through remote virtual machine for reporting tasks</li><li>Developed and queried data to enhance sales reporting using PowerBI and X++</li></ul>	

## PROJECTS

<b>eda-denoiser-stress-detector</b>   <i>React.js, D3.js, Flask, Scikit-Learn, Tensorflow, Docker</i>	
<ul style="list-style-type: none"><li>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: <a href="https://aristodemus8-eda-denoiser-stress-detector.hf.space/">https://aristodemus8-eda-denoiser-stress-detector.hf.space/</a></li><li>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.</li><li>Validated model performance of 90% AUC &amp; 78% accuracy in biosignal denoising, providing a robust foundation and methodology for future bio-signal research and potential diagnostic tools.</li></ul>	
<b>signal-gender-predictor</b>   <i>SQL, DuckDB, Librosa, Azure Data Lake, Azure Data Factory, Airflow, Terraform</i>	
<ul style="list-style-type: none"><li>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: <a href="https://aristodemus8-signal-gender-predictor.hf.space/">https://aristodemus8-signal-gender-predictor.hf.space/</a></li><li>Automated the ingestion, transformation, and feature engineering of large-scale audio datasets, generating high-quality features for model training and serving.</li><li>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.</li></ul>	
<b>chronic-disease-analyses</b>   <i>SQL, PowerBI, Apache Spark, Airflow, Selenium, S3, DuckDB, Docker, Terraform</i>	
<ul style="list-style-type: none"><li>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: <a href="https://chronic-disease-analyses.vercel.app/">https://chronic-disease-analyses.vercel.app/</a></li><li>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</li></ul>	

## EDUCATION

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