# MARIO AGBAN

661.877.1621 • Santa Clarita, CA • mario.agban@berkeley.edu • linkedin.com/in/mario-agban • github.com/mario-agban

## EDUCATION

**University of California Berkeley** | B.A Applied Mathematics and Data Science 2018 - 2022

## SKILLS

- **Programming**: Python, SQL, R, Regex, JS, HTML, CSS
- **Data Pipeline:** dbt, Fivetran
- **Product Analytics**: Google Analytics, Amplitude, Segment
- **Business Intelligence:** Tableau, Looker, Power BI, Jupyter Notebook
- **Spreadsheets**: Excel, Sheets
- **Packages:** Ggplot, Pandas, Seaborn, Matplotlib, NumPy, Sci-Kit Learn, TensorFlow, Pytorch, Ploty, Scipy
- **Statistical Techniques:** Hypothesis Testing, A/B Testing, Confidence Intervals, Bootstrap, Modeling

## EXPERIENCE

**Analytics Engineer Fellow** June 2024 - Now

Modern Data Lab | **Online**

- Leveraged the modern data stack to extract, clean, and transform real-world datasets using SQL and presented insights using Tableau
- Coursework included Excel, PostgreSQL, Snowflake, Tableau, dbt, Looker, and statistics

**Coding and Web Development Instructor** August 2022 - June 2024

CodeCampus | Summit High School District, Bay Area

- Taught high school computer science and robotics courses SME with fluency in Python and JavaScript
- Developed new curricula: effective lesson plans, hands-on projects, and designed assessments that met theNext Generation Science Standards, Common Core State Standards, International Society for Technology Education Standards, and Computer Science Teachers Association Standards

**Data Analyst Intern** April 2022 - July 2022

Lead Researcher Dr. Alessandro Selvitella | Project | Purdue University, Indiana

- Analyzed and cleaned COVID-19 case data and Foursquare location data using Python and Pandas/NumPy
- Visualized infection rates and machine learning-generated clusters via interactive maps created with Matplotlib
- Algorithmically scored locations based on infection rates and ranked by contribution to coronavirus cases
- **Results:** Used exploratory data analysis and ML techniques to examine different types of popular venues present in a Texas county that made an impact on the COVID-19 infection rate and created a ranked list of the most "dangerous" types of venues.

**Data Analyst Intern** Sep 2021 - Dec 2021

LEED Green Building Council | Project | Metro Area, Washington D.C.

- Inspected data to derive insights and identified correlations and patterns to apply statistical methods to analyze data in MySQL Workbench
- Developed reports and dashboards the business stakeholders, APPs, programmers, and analysts to pull from the enterprise data repository
- **Results:** Migrated 38% of the company data from Yellowfin to Power BI

**Data Analyst Intern** June 2021 - Sep. 2021

CarpeMed Travel | Project | San Francisco, CA

- Extracted, scraped, cleaned, and analyzed raw tabular data from hospitals in CSV and JSON formats
- Designed pipelines to automate extraction, transformation, and loading of data into repositories of new medical data for heat map visualizations

## PROJECTS

**Meta-analysis of Hydroxychloroquine Case Study** May 2020 - July 2020

Data Science Research Thesis | Project | UC Berkeley, CA

- Analyzed randomized and observational studies of hydroxy to determine its effectiveness in the treatment of positive COVID-19 patients
- Programmed an algorithm using iPython that investigates the claims of several medical testing labs, accounting for contextual information about each study and summarizing the findings into one recommendation
- Study Conclusion: Hydroxychloroquine is ineffective of its treatment to COVID-19