

Camaron L. Mangham

(678) 654-3434 | camaron.mangham@gmail.com | [linkedin.com/in/clmangham/](https://www.linkedin.com/in/clmangham/) | github.com/clmangham

SKILLS

Programming Languages: Python (Object Oriented, Data Structures/Algorithms), SQL, PostgreSQL, MATLAB, R

Developer Tools: Version control (GitHub, DVC), Docker, Flask, Jupyter Notebook, Unix / Linux Command Line Tools

Frameworks & Packages: Pandas, Numpy, Scikit-learn, Statsmodels, Matplotlib, Seaborn, Altair, Plotly, Pytorch, Pyspark, LangChain, NLTK, Flask, Terraform, Mage, BigQuery, dbt, Tableau

Skills: Machine Learning, Deep Learning, Statistical Methods, Qualitative and Quantitative Research Methods, Experimental Design, Experiment Analysis, Experiment Tracking, A/B Testing, Hypothesis Testing, Data Visualization, Natural Language Processing, Large Language Models, Generative AI, Databases, Cloud Infrastructure (AWS, GCP), ETL, Workflow Orchestration

PROFESSIONAL EXPERIENCE

Duke University School of Medicine – Neurobiology

Durham, North Carolina

Research Data Scientist

Dec 2019 — Dec 2022

- Conducted research using psychophysical and **computational approaches** to investigate the purpose of history dependency of neural representation in the visual areas of the brain, with a focus on understanding their role in guiding behaviors.
- Led **data engineering and analysis efforts** for two concurrent independent research projects.
- Developed and implemented automated pipelines for the conversion and **analysis of 50+ TB of high dimensional neural imaging and behavior data using python and MATLAB scripting**.
- Applied **data visualization, statistical methods, and machine learning** to identify significant relationships between 10+ visual stimuli, neural activity of 100+ neurons, and behavior of many cohorts of participants.
- Generated presentations and reports** that guide decision-making and project management.
- Managed the codebase and collaborated across the organization through **version control via git/GitHub**.

Cornell University – Environment, Health and Safety

Ithaca, New York

Health and Safety Biologist

Nov 2016 – Sep 2019

- Performed **scientific consulting** activities to assist in the design and development of experiments.
- Conducted a **descriptive analysis** to understand the distribution of biological hazards in the institutional research portfolio **using the R programming language**.
- Developed, maintained, and administered biosafety systems and protocols while **collaborating with researchers, engineers, and medical professionals**.
- Trained all levels of students, staff, and researchers** on compliance with health and safety standards, encompassing regulations from entities such as OSHA, EPA, DOH, CDC, DHS, DOT, DEA, USDA, NIH, and State programs.

Georgia State University – Environment, Health and Safety

Atlanta, Georgia

Health and Safety Biologist

Aug 2014 – Nov 2016

EDUCATION

University of Michigan

Master of Applied Data Science

Ann Arbor, Michigan

Dec 2023

Georgia State University

Master of Science in Biology

Atlanta, Georgia

May 2016

University of Georgia

Bachelor of Science in Biology & Psychology

Athens, Georgia

May 2009

PROJECT EXPERIENCE

LLM Cover Letter Generator | [App Link](#)

Developed a Python application using LangChain and OpenAI's Chat GPT 3.5 Turbo to assist users in generating personalized cover letters based on their resume and a job description.

- Utilizes **GenAI and NLP via a Large Language Model** to automatically create personalized cover letters based on user-provided resume and job description.
- Retrieves job description from URL and parses resume for relevant details with a **streamlined input process**.
- Combines **BM25 retrieval and Faiss vector similarity search** to gather pertinent information from both resume and job description.
- Uses a **simple Streamlit web interface** for easy interaction, saving time and effort in the job application process.

Machine Learning Approaches to Dementia Biomarker Identification | [GitHub Link](#)

Used a combination of bioinformatic and machine learning methods to analyze genomic and proteomic data from the Aging, Dementia and Traumatic Brain Injury (TBI) Study.

- Performed **differential gene expression analysis on RNA-seq data** with **Python** to select 32 genes of interest from 50,281 gene samples derived from dementia patients.
- Utilized the curated set of 32 genes to **assess the predictive capabilities of various classification models**, employing multiple metrics for evaluation via **Scikit-learn and Seaborn**.
- Identified the **support vector machine as the optimal model, exhibiting the highest recall and F1 score** for the dementia classification task.
- Conducted a **feature importance analysis**, pinpointing the genes with the greatest predictive power for dementia.

COVID-19 Forecasting with Deep Learning Model | [GitHub Link](#)

Applied deep learning techniques, specifically Long Short-Term Memory (LSTM), to forecast COVID-19 incidence at both the US state and country levels. Notable achievements include:

- Conducted **analysis of correlations** between demographic factors and COVID-19 prevalence using **Python/Pandas/Seaborn**. Identified smoking as the most positively correlated factor, followed by poor health and obesity. Contrarily, median household income exhibited a negative correlation with COVID-19 cases.
- Developed a **geographical visualization using Plotly** illustrating the correlation between the rate of COVID-19 cases and the prevalence of smokers.
- Implemented key components of the **LSTM model using Pytorch** based on established research for COVID-19 forecasting.
- Demonstrated the effectiveness of the model trained on county-level data by **comparing predicted cases and prediction intervals against actual cases with a time series visualization**. Notably challenging periods, such as the holiday season (November, December), were highlighted to showcase the model's performance.

Analyzing Employee Turnover at Acme Aroma | [GitHub Link](#)

Conducted a comprehensive analysis of employee turnover at the fictitious company, Acme Aroma, utilizing data science techniques. Key accomplishments include:

- Engineered an employee turnover model, using **Scikit-learn and Statsmodels**, to assess the likelihood of employees leaving the organization, **leveraging key predictors for turnover risk**.
- Achieved an **F1 score of 0.94 for the negative class (non-turnover) and 0.71 for the positive class (employee turnover)**, demonstrating the model's accuracy. Evaluated the **false positive rate at 9.25%** representing instances falsely predicted as turnover, and the **false negative rate at 15.46%** indicating cases incorrectly predicted as non-turnover.
- Proposed a strategic initiative with a primary focus on work-life balance, **anticipating a 3% reduction in turnover**. The recommendation aimed to address identified pain points and enhance overall employee satisfaction.

Amazon Electronics Reviewer Value | [GitHub Link](#)

Conducted a comprehensive analysis of the "reviewer lifetime value" and expenditure trends within the Amazon electronics domain, utilizing a subset of a dataset comprising 233.1 million reviews spanning all product categories. Key achievements include:

- Implemented distributed processing through **PySpark to optimize processing and conversion this large dataset efficiently**.
- **Visualized a pivotal period of growth** for Amazon (2014-2017), noting a median spending pattern of approximately \$11 by reviewers during this phase.
- Investigated the **predictive significance** of a reviewer's initial Amazon review, revealing that information about their first review provides only marginal value in predicting their Lifetime Value as a Reviewer.