

Zachary Mays

Data Scientist | 512-653-2088 | zachary.a.mays@gmail.com

PROJECTS

- **ML Claims Processing Pipeline**
 - Lead a team to build an end-to-end SageMaker ML pipeline to process America's medicare and medicaid claims each day.
- **Member Risk and Opportunity Pipeline**
 - Created a Machine Learning model to predict the risk and opportunity of health insurance members for ACO populations with an accuracy of 98% compared to a \$1 million a year vendor.
- **Special Reports on Member Trends**
 - Produced special reports on trends in our member population such as the effect of birth month on vaccination status in children.

SKILLS

Machine Learning: Regression, Cross Validation, Classification, Recommendation, and NLP

Python Libraries: Scipy, Pandas, Seaborn, Sci-kit Learn, Numpy, Pyspark,

Programming Languages: R, Python, HTML, CSS, JavaScript, Bash

Data Management: ETL, Data Governance, Relational Databases

Database Language: SQL

Data Reporting Tool: Tableau, PowerBI

Other: Microsoft Office, Git, Visual Studio Code, Pycharm, SageMaker

EXPERIENCE

GDIT, Remote - *Senior Data Scientist*

September 2022 - PRESENT

- Engineered features from Medicare claims data for predictive modeling in SageMaker using Pyspark and AWS Glue.
- Lead a team to tune models increasing the performance >20% over existing human-lead processes.
- Implemented models in an AWS-hosted ML pipeline that processes over 100 million records daily.
- Advised on the creation of a dashboard to monitor performance drift in QuickSight.

Innovista Health Solutions, Remote - *Market Analyst*

July 2021 - Sept 2022

- Developed a risk assessment and opportunity ML algorithm for Medicare and Commercial members.
- Analyzed population data and found statistically relevant opportunities for improvement.
- Shared data visualizations and whitepapers with C-suite executives and VPs in my department to demonstrate that we could bring AI/ML in-house.
- Advised on the use of Sentiment Analysis of incoming Documents to increase the speed of ETL from sFTP servers into a local SQL Server.

Innovative Genomics Laboratories, San Antonio, TX - *Molecular Technologist*

February 2021 - July 2021

- Lead validation testing, compiled raw data files, and statistically analyzed data for a PGx panel.
- Managed a team of 3 scientists working on 6 different projects for molecular validations, ensured data integrity and consistency across tests.
- Presented findings to the Medical Director to launch the panel.
- Authored detailed test procedures and taught lab personnel to perform the tests.

Texas State University, San Marcos - *Graduate Research Assistant*

January 2020 - December 2021

- Extracted, sequenced, and analyzed over 300 genes and 20 genomes from bacteria associated with an endangered beetle for the USFWS resulting in a publication in an academic journal.
<https://pubmed.ncbi.nlm.nih.gov/34494105/>.
- Leveraged the STAR cluster supercomputer to align and annotate over 100Gb of metagenomic data from insects and their environments using Bash and R scripts such as vegan and phyloseq.
- Performed multiple analyses on genetic, genomic, and metagenomic data using the DADA2 pipeline, nBLAST, and MicrobiomeAnalyst.

EDUCATION

University of Texas, Austin, TX *Post Graduate Program in Artificial Intelligence and Machine Learning*

Google Analytics Professional Certificate Austin, TX - *Google Data Analytics*

Texas State University, San Marcos, TX - Master's in Science - *Bacterial Genetics (2020)*

Texas State University, San Marcos, TX - Bachelor's in Science - *Biology (2018)*