ANNIE NGUYEN

nghi.npv@gmail.com | linkedin.com/in/nghi-annie-nguyen | github.com/annie-npv | annienguyen.netlify.app

EDUCATION

Miami University | Oxford, OH
B.Sc. Data Science and Statistics | GPA: 3.87   *Expected May 2025*

SKILLS

- Programming languages: Python, R, SQL, Java, SAS
- Tools: Excel, Tableau/PowerBI, Databricks, Snowflake, AWS, Azure, Spark, Hadoop, Google Analytics

EXPERIENCE

MIAMI UNIVERSITY | Oxford, OH   *October 2024 - Present*
Data Analyst
- Execute SQL queries to collect and analyze data, supporting 100+ data requests for stakeholders
- Conduct statistical analysis using R and SAS, delivering insights for over 10 university-wide projects
- Create interactive Tableau visualizations, boosting data accessibility by 25%

MEDICAL MUTUAL OF OHIO | Brooklyn, OH   *May - Aug 2024*
Business Intelligence Intern
- Migrated real-time and historical EDI Health data display from RShiny to PowerBI, integrating continuous updates from an Oracle database using SQL queries to ensure accurate and timely data retrieval
- Achieved a 30% reduction in process time, significantly enhancing data visualization and user interaction
- Developed comprehensive documentation for the migration of a sales dashboard from Tableau to PowerBI, ensuring a seamless transition and facilitating future maintenance

OXFORD UNIVERSITY CLINICAL RESEARCH UNIT | Remote, UK   *May - Aug 2023*
Statistical Research Assistant
- Integrated and cleaned 727K+ clinical records, fusing quality control and population stratification analysis
- Employed Bayesian mixed-effects linear regression model to deduce the terminal half-life of tafenoquine

MIAMI UNIVERSITY | Oxford, OH   *March 2022 - May 2024*
Statistical Research Analyst
- Conducted studies on the relationship between 14M genetic variations in humans and schizophrenia mice
- Extracted and constructed models using R from 9M+ GWAS data in autism
- Built interactive websites using RShiny to filter and display 9M+ selected chromosomes within the data set

PROJECTS

Analysis of Bank Marketing Campaigns | Github
- Conducted ML analysis on 41,000 observations to predict term deposit subscriptions in banking
- Identified Gradient Boosting Machine (GBM) as the top-performing model, gaining an accuracy of 74.61%
- Addressed class imbalance by downsampling and provided insightful variable analysis to refine strategies

Health Vitality App | Github
- Developed a stroke liability prediction model based on 5,000 observations from a stroke prediction data set
- Created a user-friendly body fat calculation model with health data inputs, referencing 250 body fat data
- Visualized worldwide global health trends from 6,500 data using advanced interactive plots with RShiny

Improving Legal Services Customer Experience: ABA Raw Data for Predictive Analysis | Github
- Visualized key insights from customer demographics, question categories, and sentiment analysis
- Utilized Random Forests model to optimize resources and enhance service quality on 400,000+ inputs

ASK ME ABOUT

- Solitaire card games, Badminton, Building Lego models, and Baking