TUAN PHAN

tuanpa141186@gmail.com | 575-621-3955 | Moscow, Idaho, U.S.
[Google Scholar](#) | [LinkedIn](#) | [Data Science Portfolio](#)

Data Science | Applied Mathematics

Ph.D. in Applied Mathematics currently pursuing a career in the **Data Science Industry.** Well-versed in programming languages such as **Python, SQL,** and **Matlab**. Excels at solving complex research and business problems through **data science, machine learning** and **math modeling**. Passionate about continuing to learn and drive further business impact.

KEY SKILLS

- **Programming:** SQL, Python (Base, Pandas, Numpy, Matplotlib, Scikit-Learn, Keras), Matlab

- **Machine Learning:** Linear Regression, Logistic Regression, Decision Trees, Random Forest, KNN, k-means, PCA, Association Rule Learning, Causal Impact Analysis

- **Deep Learning:** Artificial Neural Network, Convolutional Neural Network

- **Others:** Statistics (Hypothesis Testing, AB Testing), Tableau, Github, AWS (S3, Lambda, IAM, EC2, SageMaker, RDS, DynamoDB)

EXPERIENCE

Postdoctoral Fellow   **Mar 2021 - Present**
University of Idaho | Moscow, Idaho

- Directed three research projects in biomedical research, focusing on human heart muscle analysis, bacterio phage interactions and human gut microbiome, leveraging advanced data analytics techniques for comprehensive insights.

- Led a working group (including 3 postdocs), focusing math modeling and data science.

- Served on the Search Committee for the Postdoctoral Fellow position in Mathematical Modeling and Data Science at the Institute for Modeling Collaboration and Innovation, University of Idaho

- Reviewed scientific manuscripts for prestigious journals such as the Journal of Mathematical Biology, Nature Communications, and Frontiers in Applied Mathematics and Statistics, providing expert analysis on data interpretation and statistical methodologies.

PROJECTS

- **Modeling contractile dynamics in murine and porcine heart muscles:** Utilized a math model (consisting of differential equations) to fit mechanical data in murine and porcine left ventricles. The fitted results helped better understand the cooperative mechanisms underlying contractile dynamics in the hearts of small and large mammals. These also led to two grant proposals NIH ($1million) and NSF ($1million).

- **Fruit classification using Convolutional Neural Network (CNN):** Developed a CNN to classify fruit images for a robotic sorting arm prototype, then optimized the model using Dropout, Image Augmentation, and hyperparameter tuning, and explored Transfer Learning with VGG16. The final model achieved a 98% classification accuracy, significantly improving the system's reliability for real-time sorting.

- **"You are what you eat" Customer Segmentation:** Applied k-means clustering on ABC grocery transaction data to split out customers into distinct "shopper types" that could be used to understand customer shopping behaviors over time. The analysis revealed three distinct clusters, including a general group with no specific dietary preference (73.6%), a potentially vegan group (11.8%), and a potentially vegetarian group (14.6%), providing valuable insights for targeted messaging and customer communications.

- **Understanding Alcohol Product Relationships:** A retail client needed to optimize their alcohol section layout and develop targeted promotions. Applied Association Rule Learning to analyze 3,500 alcohol transactions, examining product relationships using metrics such as Support, Confidence, and Lift. The analysis revealed insights for store layout optimization, identified unexpected product associations, and provided a foundation for data-driven decision-making in product placement and promotions.

- **Quantifying Sales Uplift:** A grocery retailer needed to quantify the sales impact of their new "Delivery Club" membership program on customer spending. Applied Causal Impact Analysis using the pycausalimpact library, aggregating transaction data and using non-member customers as a control group to create a counterfactual scenario. The analysis revealed a statistically significant 41.1% uplift in sales for Delivery Club members over a three-month period, compared to what they would have spent without the program.

SELECTED PUBLICATIONS

- **Tuan Anh Phan,** Daniel Fitzsimons, *Modeling the effects of thin filament near-neighbor cooperative interactions in mammalian myocardium*, Journal of General Physiology, Accepted**.**

- Chau Hoang, **Tuan Anh Phan**, Cameron Turtle, Jianjun Paul Tian, *A stochastic framework for evaluating CAR T cell therapy efficacy and variability*, Mathematical Biosciences, Volume 368, February 2024, 109141.

- **Tuan Anh Phan**, Farhana Sarower, Jinqiao Duan, Jianjun Paul Tian, *Stochastic dynamics of human papillomavirus delineates cervical cancer progression*, J. Math. Biol. (2023) 87:85

- Heather Curtsinger, Xianyi Zeng, Zaira Mather, Mary Ballyk, **Tuan Anh Phan**, Ben Niu, Mee Y Bartee, Jianjun Paul Tian, Eric Bartee, *High Levels of Extracellular Potassium Can Delay Myxoma Virus Replication by Preventing Release of Virions from the Endosomes*, J Virol. Jan 2023, Volume 97, Issue 1, e01294-22.

EDUCATION
- **Ph. D. (Applied Mathematics)**: Aug 2015 – Dec 2020, New Mexico State University, USA.

CERTIFICATES / COURSEWORK

- **Data Science Professional Certification (Data Science Infinity):**

  Actionable Learnings: Extracting & manipulating data using SQL. Application of statistical concepts such as hypothesis tests for measuring the effect of AB Tests. Utilizing Github for version control, and collaboration. Using Python for data analysis, manipulation & visualisation. Applying data preparation steps for ML including missing values, categorical variable encoding, outliers, feature scaling, feature selection & model validation. Applying Machine Learning algorithms for regression, classification, clustering, association rule learning, and causal impact analysis for measuring the impact of an event over time. Machine Learning pipelines to streamline the ML pre-processing & modelling phase. Deployment of a ML pipeline onto a live website using Streamlit. Using Tableau to create powerful Data Visualizations. Turning business problems into Data Science solutions.

- **Notable Coursework**: Completed advanced courses in Optimized Financial Mathematics, Foundations of Probability, Statistical Theory and Applications, and Continuous Multivariable Statistical Analysis.