# Hannah Nguyen

hannahpnguyen07@gmail.com | linkedin.com/in/hannah-p-nguyen/ | hnguyennn.github.io/

## EDUCATION

**University of California, Santa Barbara** — Santa Barbara, California
*Bachelor of Science in Statistics and Data Science and Applied Mathematics* — *Expected Graduation Date: June 2027*
- Relevant Coursework: Probability and Statistics, Design and Analysis of Experiments, Regression Analysis, Theoretical Linear Algebra, Mathematical Methods of Analysis, Partial Differential Equations, Intermediate Python, Intermediate C++, SAS Base Programming
- Statistics and Data Science GPA: 3.78, Applied Math GPA: 3.82, Overall GPA: 3.81

## EXPERIENCE

**Data Scientist Intern** — October 2024 – Present
*IDXExchange LLC* — *Remote*
- Collaborated on a prediction model as part of a team, then independently took over and continued development after team members exited early.
- Self-taught data cleaning, model building, and evaluation, while balancing coursework in statistics and data science
- Analyzed large datasets and used problem-solving skills to improve efficiency and accuracy in project outcomes.
- Created a comprehensive guide to assist future interns, documented key methodologies and challenges

**Note Taker** — January 2025 – June 2025
*UCSB Disabled Students Program* — *Santa Barbara, California*
- Provided clear and detailed notes on time, ensuring concepts and formulas were legible.
- Supported multiple students, up to five, across different subjects, including Partial Differential Equations, Probability and Statistics, and Regression Analysis.
- Maintained consistency and accuracy, improved understanding of advanced mathematical and statistical concepts.

## PROJECTS

**Impact of Paperclip Positions on Final Distance** | *R, ggplot2, RStudio* — May 2025 – June 2025
- Conducted a Factorial Design experiment testing the effects of paperclip positions on final distance.
- Performed sample size calculations and used linear modeling in R to conduct statistical analysis on data.
- Visualized and analyzed data using ggplot2 and kable tables.
- Practiced technical writing and clear communication, applying topic to broader context and referencing multiple peer reviewed articles, as well as interpreting results from analyses.

**Regression Analysis on Diamond Prices** | *R, ggplot2, RStudio* — May 2025 – June 2025
- Built and interpreted simple and multiple linear regression models, including applying linear transformations.
- Explored relationships between variables through descriptive statistics and visualizations.
- Detected multicollinearity between variables using variance inflation factor (VIF) and correlation matrices.
- Constructed and interpreted confidence and prediction intervals.

**Close Price Prediction Model** | *Python, Numpy, Pandas, Git* — October 2024 - Present
- Continuing the development of a property closing price prediction model, with a focus on hyperparameter tuning and model optimization to improve performance.
- Processed and cleaned numerical data to improve data quality and model performance.
- Experimenting with Random Forest model, using SciKit-Learn, and tuning hyperparameters to adjust to various issues, such as overfitting.
- Documenting progress and challenges in a detailed guide to assist future interns and ensure project continuity.

## TECHNICAL SKILLS

**Languages**: Python (Intermediate), C++ (Intermediate), R (Intermediate), SAS Base Programming (Intermediate), LaTex, HTML/CSS (Beginner)
**Libraries**: pandas, NumPy, SciKit-learn, ggplot2, dplyr, Flask
**IDE/Text editor**: Visual Studio Code, RStudio, Overleaf (for LaTex)
**Data Science**: Probability and Statistics, Design of Experiments, Regression Analysis, Object-Oriented Programming