

Yijia (EE-jyah) Wang

Seattle, WA • yijia0725@gmail.com • 276-870-4745 • U.S citizen - No sponsorship required

[Personal Website](#) • [LinkedIn - Yijia Wang](#) • [GitHub - yijia0725](#) • [Project - Housing Analytics](#)

Data Scientist with an M.S. in Applied Statistics (UMich) and expertise in Causal Inference, Experimental Design, and Machine Learning. Automated complex data pipelines to drive 50% efficiency gains in operational workflows. Passionate about leveraging statistical rigor to uncover actionable insights in Product or Marketing Analytics.

EDUCATION

University of Michigan *MS in Applied Statistics* GPA: 3.65/4.0 - [Transcript](#) Aug 2022 – May 2024

Virginia Tech *BS in Statistics, Minor in Mathematics* GPA: 3.96/4.0 - [Transcript](#) Aug 2020 – May 2022

Featured Courses: Causal Inference, Induction, Theoretical Statistics, Experimental Design, Regression Analysis, Bayesian Modeling, Time Series Analysis, Database App Design

EXPERIENCE

Data Scientist *Open House Atlanta Realty & Investment* Nov 2024 – Present

- Implemented an automated pipeline that executes batched Propertyware API calls, generates aggregated maintenance request tables, and maintains data integrity via robust backup logs and cell-wise updates, reducing operation latency for 15 managers from >1 day to 2 hrs
- Architected and deployed a performance-based bonus system for 50+ Make-Ready employees by aligning 7 department heads on statistical scoring criteria, synthesizing multi-source data into transparent, audit-ready evaluation sheets to drive employee incentives
- Designed interactive Looker Studio dashboards integrating BigQuery to track critical KPIs (Occupancy, Rent Collection, etc.) across 10+ visualizations, enabling self-service data access for leadership and reducing time-to-insight by 40%
- Engineered a scalable Streamlit analytics web app for house repair procurement with automated Buildertrend web scraping and secure authentication, replacing fragmented manual logs to enable rapid retrieval of multi-year records
- Developed an NLP pipeline using TF-IDF and LDA to analyze 500+ natural language renovation purchase orders from Renowalk; Identified high-volume item bottlenecks to project a 15% reduction in inventory storage overhead

Student Researcher - Liver Disease Analytics *UMich* mentored by Dr. Jessica Mellinger Aug 2023 – Dec 2023

- Processed ~200 patient demographic & medical records; Conducted hypothesis testing and logistic regression for disease outcome prediction in R
- Authored a reproducible report to optimize future clinical data collection strategies

Research Assistant - High Throughput Phenotyping *Virginia Tech* mentored by Dr. Song Li Nov 2021 – Mar 2022

- Executed data acquisition for computer vision-based crop phenotyping to automate edamame quality assessment
- Annotated precise bounding boxes using VGG Image Annotator (VIA) on hundreds of samples to establish ground-truth labels for model training

PROJECTS

Predictive Modeling for Traffic Accident *UMich* Jan 2024 – Apr 2024

- Developed an end-to-end pipeline to predict accident severity using XGBoost and Random Forest
- Engineered geospatial and temporal features and addressed class imbalance using SMOTE, improving recall on high-severity accidents by 15%

Housing Market Bias Quantification *UMich* Aug 2022 – Dec 2022

- Constructed a bias-detection framework on 15M+ census records using SQL
- Quantified statistical discrepancies in housing evaluations across demographics to simulate regulatory compliance audits

Teaching

Graduate Student Instructor (Statistics) *UMich* Aug 2022 – May 2024

- Taught 300+ students across 5 semesters on statistical theory using R and Python
- Led weekly lab discussions, developed instructional slides, designed homework problems, held office hours, and graded assignments

Teaching Assistant *The Coding School* Jul 2023 – Aug 2023

- Answered questions and facilitated small group discussions for high school students in a college-level Intro to R course.

Math Tutor *Mountain Empire Community College* Mar 2019 – Apr 2020

- Provided 1-on-1 tutoring for entry-level math classes (Algebra, Calculus)

Community Service

Dining Services Associate *Virginia Tech* May 2021 – Dec 2021

- Managed high-volume food services (10 hrs/week) to offset personal living expenses

Volunteer *Service Trip to North Carolina* Mar 2020

- Volunteered to paint and construct affordable housing with Habitat for Humanity during Spring Break

SKILLS

- Statistical Methods:** A/B Testing, ANOVA/Chi-Square, Causal Inference, Bayesian Statistics, ARIMA
- Programming Languages:** Python, SQL, R, MATLAB, JavaScript
- Machine Learning:** Linear & Logistic Regression, K-Means, LDA, XGBoost, Random Forest, SMOTE
- Data Science:** pandas, numpy, statsmodels, scikit-learn, tidyverse, Web Scraping
- Visualization & Tools:** Microsoft Excel, Tableau, Google Looker Studio, BigQuery, Streamlit, Git

Other Interests

Creative Writing, Social Psychology, Board Games, Photography, Travel, Cats, Dogs