

# TANYA NAHETA

📍 New York, NY | ✉️ tanyanaheta@gmail.com | ☎️ 310-426-4589

🔗 LinkedIn: linkedin.com/in/tanyanaheta | 📁 GitHub: github.com/tanyanaheta | 📁 Portfolio: tanyanaheta.github.io

**Data Scientist with 3+ years' experience applying machine learning, experimentation, and NLP to product, fintech, and Responsible AI use cases.**

## EDUCATION

M.S. Data Science | New York University May 2024

B.A. Economics | University of California, Los Angeles Mar 2019

## EXPERIENCE

Research Assistant, Ethical Foundation Models Project | QLab @ NYU Sep 2024 – Present

- Co-authored a 20+ page research report on foundation models, contributing a core section on governance risks in training data, evaluation methods, and deployment oversight; designed a structured codebook organizing ethical, technical, and social challenges across 30+ AI systems
- Designed and proposed the “AI Balance Sheet,” an original forward-looking evaluation framework that translates qualitative themes into measurable indicators for transparency, inclusion, and public-interest alignment.

Applied Scientist Intern | Zillow Group Sep 2022 – Dec 2022

- Built a multimodal knowledge graph from 80K+ Zillow listing images and 1,500+ textual features using CLIP-initialized embeddings.
- Trained a GraphSAGE model and ran link prediction on unseen nodes, improving recall and attribution accuracy in keyword-label matching.
- Delivered performance diagnostics (macro-F1, cosine similarity) that informed listing relevance and influenced Zillow’s ranking strategy.

Business Systems Analyst | Acclaim Systems Inc. Aug 2019 – Mar 2021

- Built SQL-integrated Power BI dashboards to analyze support ticket trends across five departments, reducing resolution time by 15% and improving issue triage and escalation processes.
- Led CI/CD testing for enterprise-scale data pipelines; authored and optimized SQL queries to validate multi-source datasets before deployment.
- Defined KPIs and developed tailored reporting tools to monitor SLA adherence, ticket volume metrics, and service disruptions, enabling data-driven decision making

## SELECTED PROJECTS

Loan Default Prediction Model/Mortgage Risk Classifier

- Built a logistic regression model to predict mortgage default risk from ~400K borrower and loan records; performed EDA, feature engineering, and binary classification.

Bias Evaluation in Toxic Comment Detection

- Identified fairness disparities in a pretrained toxicity detection model and applied mitigation techniques to improve subgroup parity without sacrificing precision.

Financial Sentiment Analysis

- Extracted strategic signals and sentiment from 500+ company filings using FinBERT to support financial outlook.

## SKILLS

**Languages & Tools:** Python, SQL, R, Git, Power BI, Tableau, Looker, Excel

**Libraries & Frameworks:** scikit-learn, PyTorch, Hugging Face Transformers, Statsmodels, pandas, NumPy, Prophet

**Platforms & Cloud:** BigQuery, Azure, AWS, dbt, Snowflake, Spark, Hadoop