Wahid T. Khan

wahid@alumni.emory.edu ● (404) 933-4794 ● linkedin.com/in/wahid-t-khan

IMMIGRATION STATUS	<mark>Permanent Resident/ Green Card holder: I will not need company sponsorship.</mark>

PROFESSIONAL SUMMARY

Data scientist passionate about helping stakeholders make better decisions.

- 4+ years of experience in designing and developing ML projects with Product partners.
- Strong background in predictive modeling techniques and statistical hypothesis testing.
- Experienced writing production code in Python and SQL with test coverage.

PROGRAMMING SKILLS

Languages: SQL, Python, Bash
Libraries/ Frameworks: SQLAlchemy, duckdb, pandas, plotly, statsmodels, scikit-learn
Tools: Snowflake, Hex, Looker, Git, Airflow, Docker

EXPERIENCE

Doximity, Inc, San Francisco CA
Senior Data Scientist   Dec 2024 – Present
Data Scientist   Apr 2022 – Nov 2024

- Boosted user engagement by 40% YoY by optimizing ML scorer to deliver the best social updates. Projects included engineering new model features to incorporate user behavior, exploring XGBoost and other algorithms, and designing A/B/n tests to evaluate models.

- Increased targeted user engagement for email campaigns by 61% YoY by providing actionable insights from analyses and developing reusable data pipelines.

- Built dimensional model to track the flow of high-impact publications data from raw ingestion to users' profiles and newsfeeds. Designed Looker dashboards for Product to stay informed on critical source of data.

California Policy Lab, Los Angeles CA
Data Analyst   Aug 2020 – Mar 2022

- Lead analyst on a study to measure the impacts of California's training programs. Presented results and policy implications to directors at California state agencies. Study was cited in President Biden's 2023 Economic Report to Congress.

- Implemented entropy balancing algorithms to build statistically valid treatment and control groups from complex administrative data.

- Co-taught Git and command line tools as part of professional development for 15 data analysts and data scientists.

PROJECTS

RAGDOll: Retrieval Augmented Generation with DuckDB and Ollama

- Built a fully offline and private AI system to have productive conversations with notes and other personal data using Llama 3.2 and DuckDB. Currently working on improving answer quality for ambiguous prompts.

- The system abstracts away state-of-the-art LLM tooling into declarative configs, making it easy to swap out frameworks and tinker with tuning parameters.

EDUCATION

Emory University, Atlanta GA
Master of Arts in Economics (STEM), GPA: 3.55/4.00   May 2020