

Dhruvi Kothari

San Francisco, CA | Open to Relocation | +1 (215) 264-4197 | msdakot@gmail.com | [LinkedIn](#) | [Github](#)

Full Stack Data Scientist with 5 years of experience, managing end-to-end data projects in an innovative and fast paced startup, delivering strategic insights to diverse stakeholders. Fast-tracked to Senior Data Scientist in 3 years.

Experience

Senior Data Scientist, UrbanFootprint - Berkeley, CA

Jun 2021 - Present

Managed end-to-end **research design, data analysis and model development** in support of UrbanFootprint's climate change and risk assessment efforts:

- Developed a nationwide statistical probabilistic model in Python saving ~\$200k in coastal flood data purchase costs, supporting risk assessment of \$10B+ municipal bonds for a leading financial insurance provider.
- Built and deployed a productionalized regression model using Databricks to estimate property damage losses for 300M properties in the USA, resulting in a 15% increase in user engagement.
- Implemented and improved a Random Forest predictive model to forecast daily energy consumption with 81% accuracy by leveraging temperature change features and fine tuning hyperparameters.
- Prototyped customer segmentation framework based on energy estimates, identifying five distinct consumer profiles to conduct targeted marketing campaigns for client's energy-efficiency programs in Pennsylvania.
- Spearheaded a high-impact hackathon with 3 cross functional teams to establish actionable KPIs for dashboards; resulting in a comprehensive Product-Requirement-Document (PRD) for resource allocation.

Responsible for enhancing and managing **cloud ETL pipelines and self-servicing dashboards**:

- Optimized Airflow ETL pipeline to handle 1B+ records calculating 90+ metrics, reducing processing time from 18+ to ~2 hours; securing \$1M ARR in new contracts by enabling rapid delivery for a leading consulting firm.
- Optimized large-scale data integration in BigQuery using advanced SQL queries to reduce data scanning by 40% and cutting complex query execution time by 30%, facilitating more cost-effective scaling.
- Built Airflow DAGs to ingest external dataset reducing tech-debt by 100+ hours using dual parallelization. Achieved 80% reduction in refresh times by designing tailored and configuration-driven Tableau dashboards.
- Partnered with the data engineering team to refine product strategy, standardizing SQL logic and data quality using YAML configs and Jinja templates, resulting in a robust framework adopted across 6+ products.

Data Analyst, Fehr & Peers - Walnut Creek, CA

Aug 2019 - Jun 2021

Leveraged **Machine Learning models and Data Visualization** to enable **strategic decisions** in transportation planning:

- Built Gradient Boosting Regression to identify 15 critical factors driving feasibility and ridership growth at proposed CalTrain stations; directly supporting the initiative to increase public transit utilization.
- Applied econometric frameworks (Difference-in-Differences and Regression Discontinuity Design) to assess the success of lane reduction interventions on crash frequency to inform safer-cities initiative.
- Enhanced travel demand forecasting accuracy by 39% by integrating the K-means clustering to identify 10 ZIP code clusters based on travel patterns and locations.
- Design and developed time-series forecast model with Prophet to generate hourly and daily traffic delays; used external weather and events data to improve the model performance.
- Redesigned the vehicle-miles-traveled per person reporting by creating comprehensive Tableau dashboards highlighting commuter trends, resulting in a 10-mile reduction per person through targeted incentives.

Technical Skills

Programming:	Python, R, SQL (PostgreSQL, BigQuery, MySQL), PySpark, Hive, TensorFlow, YAML
DS Concepts:	Data Pipelines/ETL, Experimental Design, Statistical Inference, Time Series, Machine Learning, Regression, Time-Series Analysis, Hypothesis Testing, A/B Testing, Geospatial analysis
Data Visualization:	Tableau, Power BI, Looker, Python Flask, Geospatial Visualization
Tools:	Docker, Git, Jira, AWS, GCP, Databricks

Education

Masters in Information Data Science | University of California Berkeley, CA

May 2020 - Sep 2022

Masters in City Planning & Urban Informatics | University of Pennsylvania, PA

Sep 2017 - May 2019

Bachelors in Urban Planning (Academic Excellence Award Recipient) | CEPT University, India

Jun 2012 - Jul 2016