

Joseph Denby

josephdenby.com

joseph.denby0@gmail.com

+1 814 404 0817

SKILLS

- **Languages:** Python (pandas, scikit-learn, etc.), R (tidyverse), SQL, Stan/PyMC3, L^AT_EX
- **Tools:** LLMs, Airflow, Spark, PyTorch, AWS/GCP/Azure, Databricks, Github, LangChain, HuggingFace
- **Domain Knowledge:** Generative AI / NLP, Machine Learning, Deep Learning, Data Visualization, Data Analysis

EMPLOYMENT

- **Tilt** Remote
Staff Data Scientist | Python, SQL August 2023 - February 2026
- **Navy Federal Credit Union** Remote
Lead Data Scientist | Python, SQL, LLMs, LangChain, Azure August 2023 - February 2026
 - **LLM Expert System:** Designed, built, and implemented a domain-expert GenAI chatbot using state-of-the-art techniques in fine-tuning and RAG, deployed via Azure Containers. Allows employees to interactively extract insights from banking regulation documents, massively improving AML & Compliance productivity.
 - **Anomaly Detection:** Designed and built an ML outlier detection system to clean data records and monitor application deployment logs. Saved \$100k+ in labor costs through improved employee productivity.
 - **AML Watchlist Pipeline:** Using Python and SQL, implemented custom address disambiguation and name parsing algorithms to reduce false negatives by over 60%, saving \$300k+ in manual review costs. Collaborated with OFAC and Engineering teams to design and construct a data processing pipeline for screening of 350M+ bank customers.
- **Socure** Remote
Senior Pre-Sales Data Scientist / Manager | Python, R, SQL, Airflow, Spark, Looker April 2021 - March 2023
 - **Selected Clients:** Wells Fargo, Capital One, Chime, DraftKings, Public
 - **Pre-Sales Analytics:** Collaborated closely with sales teammates to support \$10M+ in sales deals through proof-of-concept analysis using Python, R, and SQL. Conducted hands-on machine learning analyses in Python using massive financial datasets to uncover fraud and relieve onboarding friction.
 - **Post-Sales Management:** Led a team of junior data scientists to generate \$5M+ in value for clients using cutting-edge anomaly detection techniques. Leveraged statistical techniques and BI tools to uncover and deliver data insights to business stakeholders.
 - **Pipeline Automation:** Assisted product and engineering teams by automating machine learning model evaluation using Python, R, and Airflow. Queried 500GB+ datasets using SQL on hosted AWS instances to conduct 100+ model iterations without manual intervention.
- **Merrill Lynch** Jersey City, NJ
Data Scientist | Python, SQL, NLP, Scikit-Learn, PyTorch September 2020 - April 2021
 - **E-Communication Supervision:** Built and deployed end-to-end machine learning pipeline using neural networks, including state-of-the-art transformer architectures (BERT, XLNet). Upgraded firm-wide communications compliance monitoring, leading to \$5M+ in yearly savings.
- **Decode_M** New York, NY
Data Scientist | Python, NLP, Web scraping, Bayesian Inference July 2019 - July 2020
 - **Selected Clients:** Airbnb, Lyft, Estée Lauder, Brooklyn Brewery, R/GA
 - **Customer Segmentation:** Advised \$500k+ client marketing campaigns using unsupervised cluster modeling techniques trained on CRM, financial, and text datasets.
 - **Conversation Analytics:** Leveraged unsupervised topic modeling & neural network document embedding (Doc2Vec) to offer brands strategic guidance at scale, uncovering nuanced online conversations from social media, news, and blogs.

EDUCATION

- **University of Chicago** Chicago, IL
Master of Arts in Computational Social Science September 2017 - June 2019
- **Cornell University** Ithaca, NY
Bachelor of Arts in Philosophy & Psychology August 2013 - May 2017