**LILY SU**

**718-775-7391 | lilyxsu@gmail.com |** [**linkedin.com/in/lilyxsu**](https://www.linkedin.com/in/lilyxsu) **| lilyxsu.com | Open to Hybrid in CA, Austin, NYC**

Data Scientist experienced in driving to the core of business logic through analytics-led business strategy.

**CORE COMPETENCIES**

Machine Learning | Deep Business Knowledge | AI Integration in Core Business | Client-facing Communication

**PROFESSIONAL ABILITIES**

* Fine-tuned Large Language Model and performed prompt engineering for document-retrieval applications.
* Architected LangChain application using LLM and vector database APIs for RAG document generation.
* Experienced with large datasets involving hundreds of billions of spending and hundreds of millions of rows.
* Developed data streaming pipelines, data visualization automation scripts, implementing state-of-the-art time-series models implementations, and customer documentation, generating continued business.

**WORK EXPERIENCE**

**Data Science Analyst | Discover Financial Services, Customer Acquisitions Dept. - Remote Jun 2022 – Feb 2023**

*Hired to perform investigative analysis of disparate data to determine the business impact on the underwriting process of consumer credit card applications.*

* Replicated and analyzed underwriting to detect patterns on regional macroeconomic trends.
* Performed consistent half-day turnarounds on identifying the customer journey through analyzing hundreds of millions of transactional customer acquisitions data in wide datasets.
* Provided strong data-backed validation through data reporting on the business impact of mismatches between input and output data of the mainframe to microservices API migration.
* Analyzed large, high-dimensional, and sparse datasets on structured and unstructured data.

**Functional Analyst | T-Mobile Contracted via People 2.0, Procurement Dept. - Remote Jan 2021 – Jun 2022**

*Hired to implement data parity between Oracle and SAP ERP platforms for the Procurement Analytics Cube.*

* Performed model explanatory analysis from unstructured data infrastructure for spend classification AI.
* Led the build-out of a centralized self-service infrastructure for T-Mobile’s procurement organization involving high-dimensional, and sparse datasets involving hundreds of columns, hundreds of millions of rows in order to identify spend logic of hundreds of billions yearly in company spending in SQL and PySpark.
* Headed technical requirements gathering to set technical delivery strategy on Procurement Contract and Supply Chain data as a product for over a hundred high-spend procurement and supply-chain executives, delivering consistently ahead of schedule by months for previously unsolvable work.
* Worked in between financial reporting owners, project managers and across multiple engineering teams to orchestrate the implementation of business logic for scalable OLAP cube that daily recalculated spend metrics.

**Machine Learning Engineer /Data Scientist | Data Analytics NYC, New York, NY Apr 2019 – Jan 2021**

*Supported clients on extracting insights, creating workflows, Al strategy and machine learning applications.*

* Built data pipelines that visualized and aggregated streams of data through an LSTM to predict patterns in daily and weekly foot traffic at the scale of millions of observations per hour, hundreds of million per day.
* Implemented Regression, Random Forest, XGboost, Neural Networks, Deep Learning, NLP, classification, ARIMA time series forecasting models using frameworks including PyTorch, Tensorflow, Scikit-Learn, Pandas, Numpy.
* Advised, assessed scope of work and converted clients to win continued contracts with business clients on bringing value to data through analysis and AI solutions.

**Analytics Consultant | Lily's 3D Products & Services, New York, NY Feb 2012 – Apr 201**9

*Hired by clients to perform analytics, automation and scripting to implementation of products.*

* Performed real-time product development with Fortune 500 clients using 3D CAD software to anatomically adjust and analyze human body dimensions of dress forms for mass apparel production.
* Client-facing in selling data science services to stakeholders to win new collaborations and renewal of contracts.
* Performed Python Developer role and implemented VBA and other Excel scripts to simplify computer-aided design processes through industry research, speeding up workflow by hours.

**RELATED PROJECTS**

**Document Creation Automation | RAG-Based Application Sep – Oct 2023**

*Built generative AI tooling for word document creation using vector databases to add fine-tuning information.*

* Curated a process for extracting, processing and fine-tuning data to custom-fit desired context.
* Engineered prompts for better natural language understanding in replying within context.

**Document Retrieval Application | RAG-Based Application Jul – Oct 2023**

*Built an application that leveraged Large Language Models (LLMs) to summarize, then generate additional documents.*

* Developed prompting and retrieval strategies using a vector database to store fine-tuning data.
* Deployed LLM-based Python application using LangChain to parse through documents.

**Data Analysis and Predictive Strategy on Mobile Location Data | Client-based Work Jun – Aug 2020**

*Given anonymized mobile location data to perform raw data exploration and generate value and insight.*

* Implement an LSTM model to predict daily foot traffic of a radius around a location, prototyping from open-source code to demonstrable results.
* Ran scalable data pipelines extracting data from AWS S3 to process automated jobs on large datasets.

**Covid19CasesNYC.com | Automated Analytics Dashboard Mar 2020 – Jul 2020**

*Deployed a Covid-19 reporting dashboard with up-to-the-minute data on NYC, New York State, and World data.*

* Orchestrated data migration between two data sources, seamlessly integrating new sources to daily batch jobs.
* Performed parallel data processing and set up data pipeline infrastructure and microservices for cloud migration.

**TallyAI Trend Analysis for Small Business | Review Aggregator Web Application Dec 2019 – Feb 2020**

*Built an online review dashboard that uses NLP to give insights for business owners analyzing reviews from Yelp.*

* Researched and implemented NLP systems to analyze trends in location-based review text data.
* Implemented Django application and queried Postgresql database. Created a REST API.

**EDUCATION & CERTIFICATIONS**

**Data Science Fellow Candidate** Correlation One| Oct 2023 - Dec 2023

**Bachelors of Science, Information Computer Science** Park University | Online | Oct 2021 - July 2023

**Terraform Associate Certification** Hashicorp | Online | Nov 2020

**Data Science 9-month Coding Bootcamp** Lambda School | Online | May 2020

**Computer Science & Chemistry Coursework, Post-Bacc Non-degree** Columbia University |New York, NY |May 2017

**Bachelors of Fine Art** Rhode Island School of Design| Providence, RI | Jun 2011

**SPEAKING ENGAGEMENTS**

Implementing and Scaling LLM’s Talk | Austin Python Meetup at Cloudflare Offices October 2023

Plotly Workshop | Machine Learning Night at Fat Cat Fabrication Lab May 2020

Data Science Lightning Talks Organizer and Host | City.Al at Nomadworks Coworking June 2019

**SKILLS**

**Experienced in** Large Language Model fine tuning, Python, PySpark, SQL, PyTorch, Machine Learning, Statistics, Docker