Naman Mehta

mehta.naman.jobs@gmail.com | (773) 808-9229 | LinkedIn/naman101199 | GitHub/naman101199

WORK EXPERIENCE

Jones Lang Lasalle (JLL)   *Chicago, IL*
Senior Business Intelligence Engineer (Client focused)   *Jun '23 - Present*

- Created **PySpark ETL** pipelines consolidating 5+ years of Energy Storage Systems (ESS) activity data into **Azure Data Lake orchestrated using Airflow**, leveraged to design **Tableau** dashboards for monitoring **State of Health** (SoH) of energy storage arrays, driving predictive maintenance strategies with minimum service disruption.
- Led **5** data extraction, warehousing, and analytics initiatives that **reduced operating costs** by **20%** and created customized dashboards.

Data Engineer Intern (Technology team focused)

- Optimized scripts using **PySpark in Azure Databricks** to reduce runtime with no impact on 3400+ customers.
- Built a system to provide **real-time alerts** and weekly summary dashboards in response to **200 GB of data** embedded into **Azure Data Lake** streaming from **IoT devices** in 400+ office locations.
- Enhanced ETL pipelines for **efficient data streaming** of bill processes using **Apache Kafka** across 20+ vendors.

Pangea Tech   *Bangalore, India*
Data Analyst   *Nov '21 - May '22*

- Consolidated $1Bn+ warranty claims data for a Fortune 500 automotive client using **PySpark ETL** pipelines into **Azure SQL Data Warehouse**, powering **Tableau** dashboards to spotlight trends for defect root cause analysis and train **predictive ML models** for claim approval automation, resulting in **savings of over $400k annually**.
- Conducted comprehensive **data tagging and discovery** during an **On-Prem to Azure** project, revealing critical data inconsistencies in the client's legacy systems, leading to targeted resolutions, and ensuring accurate transfer to Azure.

Equity Data Science   *Mumbai, India*
Quant Data Analyst   *Jun '20 - Oct '21*

- Delivered **customized solutions** to track investments, risk, and return **through automated ETL pipelines** and various **Spotfire** dashboards to analyze performance, resulting in a **90% client conversion** and **retention**.
- Architected an **automated email reporting system** for clients to consolidate their portfolio and market data. Generating **additional revenue** from **~70%** of **existing clients**.
- Implemented an **NLP system automating financial document** processing for 20+ investment firms **with $500M+ annual revenue**.

RESEARCH & PROJECTS

Generative AI Researcher (Product – TQL)

- Engineered a **GenAI product** by **fine-tuning Large Language Models (LLMs)** to automate SQL query generation from natural language, **reducing manual data querying time by 80%**.
- **Deployed** the GenAI product using **Streamlit**, enabling users to interact with the model through a user-friendly web interface, further improving accessibility and usability.

Forecasting NYC Yellow Taxi Prices and Demand

- Conducted EDA on ~29 GB of NYC Taxi trip data using **SparkSQL**, to identify price sensitivity and demand trends.
- Trained **Random Forest** and **XGBoost** models on trip activity data to predict taxi demand, achieving **F1 score of 87%**.

Building Algorithmic Trading Infrastructure

- Developed and deployed a robust **real-time trading infrastructure** on **Azure**, generating 15% profit on 2-years of historical data.
- Aggregated data from different sources using **REST APIs** and **web scraping** to ensure comprehensive and accurate datasets for algorithm training and back testing.
- Designed **unit tests** to validate the functionality and performance of various trading algorithms, leading to **20% decrease in system errors** and enhancing overall system stability.

EDUCATION

The University of Chicago   *Chicago, IL*
MS in Applied Data Science
Narsee Monjee Institute of Management Studies   *Mumbai, India*
B.Sc. in Applied Statistics & Analytics (Relevant Courses: Probability, Database Management Systems, Object Oriented Programming)

TECHNICAL SKILLS

- **Programming Languages & Databases:** Python, R, C++, MATLAB, SQL, Neo4j, MongoDB, Snowflake
- **Frameworks**: TensorFlow, Keras, PyTorch, Scikit-learn, Plotly, Seaborn, NumPy, Pandas, NLTK, A/B Testing, Flask, Spark
- **Platform Tools:** Microsoft Azure, Amazon Web Services, Docker, Kubernetes, Databricks, Git, Tableau, PowerBI, Airflow
- **Certifications:** AWS Certified Machine Learning – Specialty, DeepLearning.ai specialization