Viraj Shetty

Credit Risk Analyst | Data Scientist
virajvshetty47@gmail.com · +1 4693887873 · Dallas, TX
LinkedIn · GitHub · Portfolio

Experience

Apex Informatics *Dallas, TX* September 2026 - Present
Operational Analyst

- Developed a centralized **Power BI dashboard** tracking **20+ operational KPIs** across project delivery, resource utilization, client performance, and financial operations, reducing reporting effort by **35%**.
- Streamlined **SQL and ETL workflows** to consolidate project, timesheet, client, and financial data from **5+ internal sources**, standardizing operational reporting and improving data accessibility.
- Implemented an **AI-powered text-to-SQL solution** enabling stakeholders to retrieve operational insights using natural language, with automated regression testing through **Hermes** to validate query accuracy.
- Analyzed **resource utilization, capacity, SLA performance, and budget variance** across **15+ client engagements** using Power BI, DAX, and Power Query to support operational planning.
- Automated recurring reporting and **data quality validation** using Python, SQL, Power Query, and Excel, reducing reporting turnaround by **4 hours**.
- Automated **GTM and revenue operations** using **Clay, HubSpot CRM, and n8n**, integrating lead enrichment, prospect segmentation, CRM synchronization, and multistep outreach workflows to streamline pipeline development.

Elevate Credit, *Dallas, TX* August 2023 - July 2026
Data Scientist II

- Standardized the model development lifecycle by reviewing the model approach, target, and segmentation, and replaced email-based reviews with structured leadership alignment meetings, **reducing turnaround time** by **75%**.
- Presented internal & external stakeholders model performance reviews, flagging underperforming models and communicated strategy changes using **Copilot** meeting notes, retired **2** underperforming models in a single review.
- Owned model inventory lifecycle management **10-12** active models, including retirement of deprecated models, delivering real-time status updates to internal stakeholders and bank partners within **24 hours** of any model change.
- Took ownership of coordinating with third-party vendors to validate model replication and compliance with **SR-26-2 & OCC-11-12**, and **facilitated fair lending reviews** to ensure models were free of discriminatory bias.
- Built a fact-and-dimension database architecture from scratch to replace a retiring Tableau/SaaS reporting system, **collaborating cross-functionally** on a scalable schema for current and future credit risk and response models.
- Automated variable stability (PSI) and performance (AUC, KS, lift) tracking for **10** active models via parameterized Jupyter notebooks and an engineered Snowflake-to-Power BI pipeline, reducing execution time by **95%**
- Designed ad-hoc analytical views for **Charge-off benchmarking**, **Roll-rate analysis**, and **Re-parameterization comparisons**, driving proactive monitoring and pre-rollout lift analysis reducing delinquent population by **60%**.
- Enhanced an automated lookup query that retrieves underwriting data by application ID upon **API**-flagged fund/no-fund conflicts, leveraging **Claude** for query refinement, cutting manual investigation time by **85%**.

Loopback Analytics, *Dallas, TX* June 2022 - December 2022
Data Analyst Intern

- Analyzed and tracked client feasibility requirements using **JIRA** and used **Snowflake** to write stored procedures to optimize execution time and space by **25%**, optimized **Power BI** data structure to reduce report load time by **30%**
- Collaborated with a cross functional team, identified **5 key business metrics**, and conducted **Hypothesis testing** on patient data to identify differences in treatment outcomes and improving treatment success rate by **20%**
- Utilized **PySpark** to manage and update NLP pipelines for EHR data and used **Databricks & R** to perform Data Cleaning & **ETL** on socio-economic data from **Census API** reducing the null value in the existing data by **30%**

mUni Campus, *Mumbai, IN* December 2019 - May 2021
Data Analyst

- Collected and integrated data from multiple sources including academic performance and lecture data using **Apache Airflow & AWS S3** and redesigned **API** services for greater scalability, reducing manual effort by **50%**.
- Conducted Root cause analysis for factors affecting student performance using regression analysis and identified factors that impacted test scores, used correlation analysis to detect patterns for developing effective solutions.
- Produced a Python based **YouTube-scraper** which scraped **10K** links and title of a specific keyword.
- Visualized Data and implemented machine learning algorithms **Random Forest** & **KNN** and generated reports. Achieved an average precision of **92%** and recall of **89%** for student performance prediction.
- Performed **ETL** using **pandas** to store the data with optimized **SQL scripts** which increased access speed by **67%**.

Projects

Olist - Brazilian E-Commerce Reporting, *Personal Project*[Code] PowerBI | SQL | DAX

- Developed a **Power BI report** using the **Olist Brazilian E-Commerce dataset** to explore e-commerce activity, organizing information into visual summaries to support analysis and interpretation of marketplace data.
- Presented findings through **data visualizations and summary views**, translating detailed e-commerce records into an accessible report to communicate observed patterns and provide an overview of marketplace activity.

Fraud Transaction Detection, Wells Fargo, *ML Project*[Code] H2o.ai | Python

- Cleaned and prepared **1M** rows of transaction data from Wells-Fargo for analysis by **one-hot encoding** categorical variables and conducted **feature engineering** to identify and select relevant variables
- Implemented **XGBoost** model for fraud detection by using cross-validation techniques such as **grid & random** search to obtain the best hyperparameters, resulting in a robust model with **93%** accuracy

Credit Risk Analytics Reporting, *Personal Project*[Demo] PowerBI | SQL | DAX

- Developed a **six-page Power BI report** using the **UCI Credit Card Default dataset** to analyze portfolio risk, repayment behavior, and delinquency trends, presenting summaries of payment patterns and the portfolio profile.
- Analyzed **delinquency entry rates, roll rates, and repayment status** to track accounts moving into and between delinquency stages across reporting periods, highlighting patterns in payment deterioration and improvements in account status.

Querycraft: Text-2-SQL, *Work Project* [Code] Python / Flask | SQLite | SQLglot

- Built Codex compatible **Querycraft** using **Python, Flask, SQLite** to generate schema-grounded SQL from natural text, developed a JavaScript workbench with schema exploration, editable queries, and CSV export.
- Designed a **SQLGlot validation pipeline** with user-approved, read-only execution, query deadlines, and row limits which was verified functionality through **92 AI automated tests** and browser regression checks.

Education

The University of Texas at Dallas, *Dallas,TX*

MS in Information Technology Management *(Business Analytics Specialization)* August 2021 - May 2023

Relevant coursework: Machine Learning, Advanced Statistics for Data Science, Database Foundations, Big Data

The University of Mumbai, *Mumbai, India*

BE in Computer Engineering June 2016 - November 2020

Relevant coursework: Analysis of Algorithms, Cloud Computing, Natural Language Processing, Data Warehousing

Skills

Programming: R, SQL, PostgreSQL, API, Git, VSCode, Databricks, Linux, Python (Pandas, Matplotlib, Numpy).

Data Science: PyTorch, Keras, Scikit-learn, TensorFlow, XGBoost, LangGraph

Visualization tools: PowerBI, Snowflake, Tableau, Microsoft Excel (VLOOKUP, Macros, VBA, Index-Match)

Cloud: Docker, GCP-BigQuery, Azure (VM, Repository, Storage Management), AWS (S3, Redshift, Athena, Glue),

AI/LLM & GTM Claude, GitHub Copilot, ChatGPT, Perplexity, Databricks Genie AI, Clay, HubSpot