BHAVANA N
DATA ENGINEER

TX, USA | (682)423-0842 | bhavanan.upskill@gmail.com | LinkedIn

SUMMARY

- Data Engineer with 4+ years of experience across healthcare and finance, designing high-performance data pipelines and analytics workflows to improve efficiency, compliance, and decision-making.
- Expert in Python, SQL, Snowflake, dbt, SSIS, Apache Spark, AWS Glue, and Azure Data Factory for scalable ETL development and data integration of structured and semi-structured sources.
- Implemented robust data quality, lineage, and transformation frameworks, enabling secure, HIPAA-compliant analytics and reliable downstream consumption for BI and ML pipelines.
- Conducted EDA, hypothesis testing, and predictive modeling using Pandas, Scikit-learn, Power BI, and Tableau to uncover actionable insights, detect fraud, and enhance business performance.
- Applied AI-driven techniques such as anomaly detection, clustering, and NLP-based analytics using TensorFlow and Scikit-learn to automate data validation, detect patterns, and improve predictive accuracy across enterprise datasets.

TECHNICAL SKILLS

Programming Languages	Python, R, SQL
IDEs & Development Tools	Visual Studio Code, PyCharm, Jupyter Notebook
Libraries & Frameworks	NumPy, Pandas, Matplotlib, Seaborn, SciPy, Scikit-learn, TensorFlow, ggplot2
Data Engineering & ETL Tools	Apache Spark, SSIS, dbt, Alteryx, Microsoft Fabric, AWS Glue, Azure Data Factory, SSRS, Collibra
Databases & Warehousing	MySQL, SQL Server, PostgreSQL, Oracle, Snowflake, MongoDB
Cloud Platforms	Amazon Web Services (AWS), Microsoft Azure
Data Analytics & Visualization	Power BI, Tableau, Looker, QlikView, Advanced Excel (Pivot Tables, VLOOKUP, Power Query)
Statistical & ML Techniques	EDA, Hypothesis Testing, Regression, Forecasting, A/B Testing, Clustering, Classification, Feature Engineering, Model Evaluation
Data Handling & Processing	Data Warehousing, Data Transformation, Data Integration, Data Cleaning, Data Wrangling, Data Pipeline Design, Data Governance, Data Storytelling
Other Tools	SAS, MS Visio, Git, GitHub

WORK EXPERIENCE

Data Engineer | **Blue Cross Blue Shield of Texas| TX**   **Jan 2025-Present**

- Spearheaded the design and optimization of large-scale healthcare data pipelines using **Python, SQL,** and **Apache Spark**, integrating EHR, claims, and provider datasets to enhance model accuracy and operational analytics efficiency.
- Architected and automated **ETL workflows** leveraging **AWS Glue** and **dbt**, streamlining ingestion of structured & semi-structured healthcare data & improving scalability, maintainability, & lineage tracking.
- Consolidated and modeled multi-source healthcare data into **Snowflake** and **Hive warehouses**, standardizing schemas and improving query performance for faster analytics, predictive modeling, and enterprise-level BI reporting.
- Developed **AI-driven anomaly detection** and **risk stratification systems** using **Scikit-learn, TensorFlow,** and **PySpark MLlib**, automating pattern detection, compliance alerts, & member health risk predictions with high accuracy.
- Implemented data governance, quality assurance, and metadata frameworks using **Collibra, dbt tests,** and **Great Expectations**, ensuring HIPAA, CMS, and SOC2 compliance while maintaining 99% data consistency and reliability.
- Leveraged **AWS S3, Redshift,** and **Lambda** to build cloud-native, event-driven pipelines, integrating Terraform and Jenkins for CI/CD automation and infrastructure-as-code deployment of secure healthcare data systems.
- Designed interactive **Power BI and Tableau dashboards**, integrating predictive metrics for claims processing, readmission analysis, and patient outcomes to deliver real-time insights and reduce reporting turnaround by 70%.
- Collaborated across **data science, DevOps, and compliance teams**, conducting A/B testing, validating AI outputs, and ensuring robust ML model integration aligned with business and clinical objectives for optimized patient care.

Associate Data Analyst and Programming Intern| **Infinity Management, TX | onsite   Aug 2024-Dec 2024**

- Designed and enhanced interactive **Power BI dashboards** to monitor payroll, P&L, balance sheets, and procurement KPIs, empowering leadership with real-time performance trends and faster data-driven decisions.

- Managed and transformed large-scale structured and semi-structured datasets from **ADP, QuickBooks,** and **ERP systems**, integrating into centralized models for cross-functional financial, HR, & procurement performance analysis.

- Automated recurring data workflows using **Python** (Pandas, NumPy) and scheduling tools, reducing manual effort by 70%, improving efficiency, accuracy, and timeliness of monthly, quarterly, and on-demand reporting.

- Built robust data models and **DAX calculations** within **Power BI,** enabling drill-through reports, real-time slicing, and trend analysis to support financial forecasting, variance analysis, and operational reporting.

- Developed scalable **KPI frameworks** for segmentation by business unit, region, and service line, supporting quarterly business reviews, operational scorecards, and strategic planning sessions across portfolios.

- Established data validation protocols, business rules, and quality controls across analytics pipelines, achieving 98% accuracy, ensuring audit readiness, compliance, and reliable reporting for financial operations.

- Facilitated sessions with stakeholders including **operations leads, managers, and cross-functional teams** to capture reporting requirements, align business goals, improve experience, and increase dashboard adoption by 60%.

- Standardized report formats, metadata definitions, governance practices, and naming conventions, ensuring consistent branding, professional presentation, and reliability across internal and client-facing analytics deliverables.

Data Analyst | Capgemini| India   Oct 2020-Jul 2023

- Developed predictive credit risk models using **Python** and **TensorFlow**, applying regression and machine learning to identify high-risk clients, optimize mitigation strategies, and minimize organizational financial losses.

- Designed and deployed interactive **Tableau** and **Advanced Excel** dashboards, visualizing key credit risk indicators, fraud detection patterns, and financial KPIs to support leadership with timely insights and proactive decision-making.

- Built and maintained robust ETL pipelines using **SSIS** and **SSRS**, integrating datasets from **PostgreSQL** and **Microsoft Azure**, ensuring consistent, accurate, and reliable consolidated reporting across multiple business functions.

- Conducted **statistical analysis** on financial transactions to detect anomalies and fraud, strengthening compliance frameworks and enhancing overall organizational financial risk management practices.

- Implemented scalable **data warehousing solutions** for structured & semi-structured financial datasets, improving data accessibility, storage efficiency, & enabling faster, consistent, & reliable reporting across diverse finance & operations teams.

- Collaborated with **cross-functional teams** including finance, operations, and IT to gather requirements, validate credit models, enforce regulatory compliance, and ensure analytics solutions aligned with governance and risk policies.

- Automated data transformation, cleaning, and preprocessing workflows using **Python** (Pandas, NumPy), reducing manual reporting efforts by 65% while improving the accuracy, timeliness, and efficiency of recurring financial reports.

- Performed **exploratory data analysis** & hypothesis testing to validate fraud detection algorithms, enhancing predictive performance, ensuring regulatory adherence, & enabling proactive strategies for mitigating financial risks effectively.

EDUCATION

Master of science in Advanced Data Analytics
 - University of North Texas, Denton Texas   **August 2023 – May 2025**

Bachelor of Technology in Computer Science
 - SRM University Andhra Pradesh, India   **June 2019 –May 2023**

CERTIFICATIONS

- Azure Data Fundamentals

- GCP- Associate Cloud Engineer