Pranav Grandhe

📞 +1 8043885161 — ✉ Email — 💼 LinkedIn — GitHub — Portfolio

Summary — Analyst with experience in data analysis, reporting, and modeling. Skilled in Excel VBA, SQL, and Snowflake, with the ability to make processes more efficient and provide clear insights. Organized, detail-focused, and comfortable working both independently and in fast-paced environments.

Skills

Programming Languages: Java, Python, SQL, R, C, C++, JavaScript, HTML, CSS, Bash/Shell, VBA.

Software Engineering & Tools: Data Structures, Algorithms, Git, GitLab, Jenkins, Docker, Jira, Figma, Bootstrap.

Machine Learning & AI: Pandas, NumPy, Matplotlib, Seaborn, PyTorch, TensorFlow, Scikit-learn, Deep Learning, Machine Learning, Natural Language Processing (NLP), Retrieval-Augmented Generation (RAG), LangGraph, OpenCV.

Frameworks & Web Development: Flask, Django, FastAPI, Streamlit, Node.js, Express.js, ASP.NET, Spring Boot, Next.js, React.js, Angular, Redux, REST APIs.

Cloud & Big Data: Apache Spark, Hadoop, Apache Airflow, Apache Kafka, AWS, Microsoft Azure, Snowflake, Databricks, dbt, Parquet, ETL Pipelines, Delta Lake, Amazon S3, HDFS, Apache Hive, Data Lake Architecture.

Databases: MySQL, PostgreSQL, MongoDB.

Data Visualization Power BI, Tableau, Microsoft Excel, VBA, Statistical Analysis, Time Series, Regression Analytics, Predictive Analysis.

Experience

Target Agility, *Data Engineer,Bengaluru,IND*   Jan 2022 – Dec 2023

Enterprise Data Pipeline & Query Optimization

– Tackled reporting bottlenecks caused by poorly structured schemas and inefficient SQL queries.

– Redesigned the database layer with 3NF/BCNF normalization, applied indexing/partitioning, and optimized stored procedures, cutting query execution time by 45%.

– Partnered with frontend React and backend Spring Boot teams to expose SQL-driven APIs for real-time analytics dashboards. Improved reporting latency by 70% and delivered high-availability, production-ready SQL solutions.

Education

State University of New York at Buffalo, Buffalo, NY   **2024**
Master of Science in Computer Science and Engineering   *GPA: 3.33*

R.M.K. Engineering College, Chennai, TN, India   **2019**
Bachelor of Technology in Computer Science and Engineering   *GPA: 8.1*

Projects

Credit Risk Classification Web App

– Designed a machine learning model to classify credit scores (Good, Standard, Poor) for financial risk assessment.Processed and engineered features from anonymized financial data, including income, debt, and credit utilization.Evaluated six ML models; selected Random Forest for deployment based on highest accuracy (77.5%).

– Deployed a real-time credit scoring web app using Flask and HTML for lender-facing predictions.Enabled data-driven lending decisions by highlighting top risk factors like outstanding debt and inquiries.

Enterprise Data Engineering Workflow on Azure Cloud

– Built an end-to-end Azure data workflow with Data Factory, Data Lake Gen2, Databricks PySpark, Synapse, and Power BI for governed ingestion, transformation, and analytics.

– Automated pipelines reduced manual effort by 90% and processing time by 60%; external Synapse tables powered interactive dashboards for real-time insights.Integrated Azure Key Vault for secret management and leveraged Managed Identity to simplify secure access to storage and analytics services.

NYC Food Inspection Data Pipeline

– Loaded and cleaned 200K+ NYC food inspection records using Pandas, handling missing values, inconsistent column names, and date formats.Designed a star schema with 1 fact table and 5+ dimension tables (restaurants, dates, violations, cuisines, boroughs).

– Built and executed ETL pipelines to transform raw data into structured tables and stored results in a SQLite database.Generated plots with Matplotlib to visualize inspection trends, grade distribution, and average hygiene scores by cuisine and borough.

Twitter Data Pipeline using Apache Airflow

– Orchestrated a scalable ETL with Airflow and Docker on AWS EC2 for continuous ingestion and durable storage in Amazon S3.

– Improved processing efficiency by 35% and deployment speed by 30% with modular DAGs and containerized operators.Achieved 99.9% data availability; reduced data retrieval time by 25% through partitioned layouts and metadata indexing.

Certifications

– Udemy Practical Database Course, & Python for Machine Learning with NumPy, Pandas & Matplotlib.

– IoT Applications using Raspberry Pi, National Small Industries Corporation