

PRANAY SARODE

DATA ENGINEER | PYTHON | SQL | PYSPARK | AZURE DATABRICKS | ETL/ELT

Pune, India | +91-9970990122 | pps.campion@gmail.com | linkedin.com/in/pranay-sarode-36a67750

PROFESSIONAL SUMMARY

Data Engineer with 3+ years of core data engineering experience across financial services and telecommunications, with earlier analytics/BI experience in fintech. Hands-on in Python, SQL, PySpark, Apache Spark, Azure Databricks, Azure Data Lake, Azure Data Factory, AWS S3, SSIS, data quality, ETL/ELT, and CI/CD. Experienced in building batch ingestion and transformation workflows, validating and reconciling enterprise datasets, optimizing SQL/Spark processing, and delivering analytical data products to business stakeholders. Strong focus on reliability, traceability, documentation, and Agile collaboration.

TECHNICAL SKILLS

Programming & Processing	Python, SQL, PySpark, Apache Spark; Scala/R working knowledge
Cloud & Data Platforms	Azure Data Factory, Azure Databricks, Azure Data Lake Storage, Azure Blob Storage, AWS S3; AWS Glue/Lambda/Redshift exposure
Data Engineering	ETL/ELT, ingestion, batch processing, data validation, cleansing, transformation, enrichment, reconciliation, data warehousing, dimensional modeling
Databases & BI	SQL Server, MySQL; Power BI (DAX), Tableau, Excel/VBA
DevOps & Delivery	Git/GitHub, Jenkins, CI/CD, Docker; Agile/Scrum, code review, source-to-target documentation, runbooks
Additional Hands-on Lab	Hadoop HDFS/YARN, Hive Metastore, Iceberg, Kafka (KRaft), Flink, Airflow, Trino, Prometheus/Grafana on RHEL

PROFESSIONAL EXPERIENCE

Infosys Limited, Pune - Data Engineer

Mar 2021 - Nov 2024

Project 1 - Capital Group Portfolio Analytics | Financial Services (Dec 2022 - Nov 2024)

- Built PySpark, Python, and SQL ETL workflows to ingest, cleanse, transform, and reconcile high-volume investment datasets across CSV, JSON, Parquet, Azure Data Lake, and AWS S3.
- Implemented automated data-quality controls covering schema validation, null checks, business-rule validation, and source-to-target reconciliation; supported quality-controlled and compliance-sensitive deliverables.
- Optimized SQL Server queries and analytical logic for NAV, risk, fund performance, and KPI reporting across large investment datasets.
- Supported migration/modernization of legacy data sources into a centralized Azure analytics platform and used Azure Databricks for distributed processing.
- Automated reporting/deployment activities with Python, Git, and Jenkins; participated in sprint planning, stand-ups, code reviews, architecture discussions, and global stakeholder reviews.

Project 2 - Broadband Customer Churn Analytics | Telecommunications (Mar 2021 - Nov 2022)

- Built PySpark ETL workflows for high-volume telecom usage, demographic, CSV, Parquet, JSON, and log data; performed cleansing, transformation, feature engineering, and preparation for analytics/ML.
- Developed advanced SQL transformations using joins, CTEs, and window functions to derive churn features and customer behavior indicators.
- Implemented ETL testing for schema checks, null handling, reconciliation, and data-quality validation across data-lake and warehouse layers.
- Used SSIS for scheduled batch data movement and processing; automated recurring reporting with Python/Excel VBA and supported dev/test/prod delivery through CI/CD practices.
- Developed churn analysis/predictive models and Power BI dashboards to expose churn-risk segments and KPIs; documented model/pipeline behavior and presented findings to stakeholders.

Mphasis Limited, Pune - MIS Executive / Data Analyst

May 2019 - Feb 2021

Project 3 - Klarna US Customer Behavior & Fintech Analytics

- Analyzed historical transaction data using SQL and Python to derive purchase frequency, spend behavior, payment preferences, cohort trends, and customer segments.
- Built PySpark ETL workflows to ingest, cleanse, and transform structured and semi-structured data from AWS S3/Azure Blob Storage into centralized analytical datasets/warehouse layers.
- Applied schema checks, null handling, record reconciliation, and pipeline-quality checks to improve trust in downstream reporting datasets.
- Developed predictive customer-behavior analysis with Python/Scikit-learn and created Power BI dashboards using DAX for AOV, conversion, cohort behavior, and anomaly monitoring.
- Presented data-driven findings to marketing/product stakeholders to support customer engagement, promotion, and retention decisions.

EARLIER EXPERIENCE

Wipro Limited, Pune - Senior Associate, MIS & Reporting | Dec 2016 - Feb 2019 - Automated MIS/Power BI reporting; SQL/Python trend analysis.

Dhruva Enterprises, Pune - Assistant Resident Engineer, Network Support | Aug 2013 - May 2016 - Network monitoring, troubleshooting, RCA, and technical documentation.

EDUCATION & CERTIFICATIONS

Bachelor of Arts (Economics), University of Pune

Certifications: IBM Data Analyst Professional Certificate (Coursera) | Data Science & Analytics Specialist - ITVEDANT | Hardware & Networking Technician - Silicon Valley Institute