

Parag Khankari | Data Analyst

Location: Boston, MA | Ph: (513)-428-1085 | Mail: paragkhankari19@gmail.com | [LinkedIn](#)

PROFESSIONAL SUMMARY

Data Analyst with **4+** years of experience in SQL, Python, and large-scale data processing (100M+ records). Proficient in ETL pipeline automation using Airflow, Spark, and cloud-native tools (AWS, Docker, Kubernetes). Skilled in building real-time dashboards with Tableau, Power BI, and Google Data Studio. Strong background in statistical modeling, predictive analytics, and regulatory compliance (HIPAA, GDPR, PCI-DSS, RBI). Known for translating business needs into actionable insights and scalable data solutions. Proven ability to drive data strategy, accelerate decision-making, and deliver measurable business impact.

TECHNICAL SKILLS

Programming Languages: Python (Pandas, NumPy), SQL, JavaScript (Node.js, Express.js)	API Development & Integration: RESTful APIs, OAuth 2.0, JWT, AWS API Gateway
Databases & Data Warehousing: PostgreSQL, MySQL, MongoDB, BigQuery, Snowflake, Redshift	Big Data & Streaming: Kafka, BigQuery, Window Functions, CTEs, Joins, Apache Spark, Batch Processing, Apache Flink
Data Processing & ETL: Apache Spark, Apache Airflow, SQL, Python (Pandas, NumPy)	Cloud & DevOps: AWS, Docker, Kubernetes, GitHub Actions, Azure
Data Visualization: Tableau, Power BI, Google Data Studio, Excel	Security & Compliance: HIPAA, GDPR, PCI-DSS, RBI Regulations, Role-Based Access Control (RBAC), Data Encryption
Machine Learning & Analytics: Scikit-Learn, Statistical Analysis, A/B Testing, Hypothesis Testing	Agile & Collaboration: JIRA, Git, GitHub, Business Requirement Analysis

PROFESSIONAL EXPERIENCE

Data Analyst | Cardinal Health – OH

Sep 2023 – Present

- Redesigned enterprise-scale **healthcare** data workflows across 50M+ records, cutting query execution time by 40% through advanced **SQL** tuning, **stored procedures**, **materialized views**, **indexing**, and **partitioning** in **PostgreSQL**, **BigQuery**, and **Azure SQL**.
- Constructed predictive models using **logistic regression**, **random forest**, and **ROC/AUC** scoring in **Scikit-Learn**, improving accuracy of readmission risk forecasts and reducing model runtime by 50% on **Azure Databricks**.
- Engineered **ETL pipelines** for 100M+ structured records using **Python (Pandas, NumPy)**, **Apache Airflow**, and **Azure Data Factory**, ingesting data from **EHR/EMR**, **SAP**, **flat files**, and **REST APIs** with retry and validation logic.
- Established compliance workflows by embedding **HIPAA**, **GDPR**, and **CMS** logic in pipelines implemented encryption at rest/in transit, **RBAC**, **audit trails**, and **data masking** policies for sensitive **PHI/PII**.
- Streamlined streaming and batch processing with **Apache Spark** on **Azure Databricks**, handling 200M+ weekly records from **Cerner**, **Epic**, and claims systems ensured conformance with **FHIR**, **HL7**, and **OMOP CDM** standards.
- Integrated real-time clinical and operational data from 10+ sources via **AWS API Gateway**, **Azure Event Grid**, and **BigQuery**, enabling automated updates to dashboards and reducing latency by 15% across live reports.
- Developed 20+ interactive **Tableau** dashboards using **calculated fields**, **filters**, and embedded **A/B** testing logic enhanced visibility into **KPIs**, **eQIMs**, and patient flow, reducing manual reporting by 35%.
- Designed 30+ executive-level reports using **MS Excel (pivot tables, VLOOKUP, Power Query)** and **PowerPoint** visualizations, simplifying complex data into clear insights that improved **KPI** tracking and supported strategic decisions across departments.
- Documented **ETL** pipelines, **data** mapping, **validation** specs, and **UAT test** cases in **Confluence**; worked closely with stakeholders in **Agile** sprints using **Jira**, **Kanban**, and daily standups to prioritize tasks and ensure timely delivery.

Business Data Analyst | Accenture – India

Jun 2019 – Aug 2022

- Conducted advanced **CRM data analysis** on 1M+ customer records using **SQL** and **Excel** to identify high-value segments, optimize lifecycle engagement, and support loyalty campaigns, boosting retention rates by 14%.
- Created and maintained **A/B** testing dashboards in **Power BI** and **Google Data Studio** to measure **CRM email** and push campaign performance, increasing customer reactivation by 12% across multiple financial products.
- Built predictive models using **Scikit-Learn** to forecast customer churn and identify cross-sell opportunities, enabling more personalized outreach and targeted marketing campaigns that improved **ROI** by 18% and lowered acquisition costs.
- Standardized and cleaned **fragmented CRM**, transactional, and product datasets from platforms like **Salesforce**, **SAP**, and internal **legacy systems**, streamlining customer journey mapping and enhancing campaign personalization and timing.
- Automated daily refresh of **CRM campaign performance metrics** and engagement data using **Apache Airflow** and **BigQuery**, reducing manual workload by 60% and enabling faster, data-driven marketing decisions.
- Partnered closely with **product**, **marketing**, and **loyalty teams** to define meaningful **KPIs**, assess channel effectiveness, and deliver actionable insights that shaped the direction of quarterly CRM and customer growth strategies.
- Implemented robust authentication and access control protocols including **OAuth 2.0**, **JWT**, and **RBAC** across analytics pipelines to ensure compliance with **PCI-DSS**, **GDPR**, and **RBI** guidelines while protecting sensitive customer data.

EDUCATION

Master of Science in Information Systems - Northeastern University, Boston, MA, USA.

Sep 2022 – May 2024

Bachelor of Computer Science - Savitribai Phule University, Maharashtra, India.

May 2016 – Jun 2020