

Parag Khankari | Data Analyst

Location: Boston, MA | Ph: (513)-268-3240 | Mail: paragkhankari1206@gmail.com | [LinkedIn](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

Data Analyst with 4+ years of experience turning complex data into strategic insights across **healthcare**, **finance**, and **enterprise** systems. Proficient in **Python**, **SQL**, **R**, and **cloud ETL** tools (ADF, Databricks, Airflow) for modeling, automation, and data transformation. Skilled in **Power BI**, **Tableau**, and **A/B** testing to evaluate business performance and customer behavior. Experienced in **predictive modeling** and **machine learning** for fraud detection and risk analysis at **EY**. Proven success in stakeholder collaboration, regulatory compliance (**HIPAA**, **GDPR**, **PCI-DSS**), and Agile delivery. Known for translating business needs into scalable data solutions that improve reporting accuracy, drive interactive dashboards, enhance data visualization strategies, and enable impactful decision-making.

TECHNICAL SKILLS

Programming & Analytics: Python (Pandas, NumPy, Scikit-learn, OpenPyXL), SQL, R, Exploratory Data Analysis (EDA), Statistical Analysis, A/B Testing, Hypothesis Testing, Regression Modeling, Forecasting, Data Cleaning, Data Preprocessing.

Databases & Data Engineering: SQL Server, PostgreSQL, MySQL, MongoDB, Snowflake, BigQuery, Redshift.

Data Visualization: Power BI, Tableau, Excel (PivotTables, Power Query, VBA), Google Data Studio, Interactive Dashboards, Data Storytelling, Custom Visuals, KPI Dashboards.

Methodologies & Tools: Agile, Scrum, SDLC, Waterfall, Business, Risk Management, Stakeholder Collaboration.

IDEs & Platforms: Visual Studio Code, PyCharm, Jupyter Notebook.

Machine Learning & Modeling: Scikit-learn, Logistic Regression, Decision Trees, Pattern Recognition, Advanced Analytics, Machine Learning Algorithms, Feature Engineering, Clustering.

Domain Expertise: Healthcare Analytics (ICD/CPT Codes, Claims Lifecycle, Payer-Provider Models), Financial & Credit Risk Analysis, HIPAA/GDPR Compliance.

Big Data & APIs: Apache Spark, Kafka, Restful APIs, OAuth2.0, AWS API Gateway.

Cloud & DevOps: Azure, AWS, Docker, Kubernetes, GitHub Actions.

Agile & Collaboration: JIRA, Git, GitHub, Sprint Planning.

Soft Skills: Effective Communication, Root cause analysis, Time Management.

PROFESSIONAL EXPERIENCE

Senior Data Analyst | Cardinal Health, OH

Sep 2023 – Present

- Engineered **ETL pipelines** using **Azure Data Factory**, **Airflow**, **Databricks** and **Python** to ingest, clean and transform 100M+ records from **Cerner**, **SAP**, and **flat files** on **AWS** (S3, EMR, DMS, RDS), reducing manual data prep by 70%.
- Created modular **SQL stored procedures** and optimized queries using SQL techniques(joins, CTEs, indexing, window functions) across **Databricks**, **Redshift**, **MySQL**, and **T-SQL**, boosting pipeline speed and cutting dashboard latency by 45%.
- Constructed and validated **predictive models** (logistic regression) using Python (Pandas, NumPy, **Scikit-learn**) with **data preprocessing**, feature engineering, and risk variable selection to predict 30-day patient readmission, improving precision by 20%.
- Developed scalable **Power BI** dashboards and **Excel** based reports for OSHA and readmission metrics, using **PivotTables**, **VLOOKUP**, **Power Query**, **VBA** and **Excel Macros**, delivering real-time insights to **EHS**, **compliance**, and clinical teams.
- Architected semantic layer using **star and Snowflake** schema modeling and **DAX** expressions in **Power BI** and **Snowflake**, building 10+ dashboards from **EMR** and **CRM** data, cutting **ad hoc** reports by 40% and speeding clinical decisions by 35%.
- Collaborated in **Agile** sprints with **compliance** and **ops** teams, managing 50+ **Jira** user stories, logging 30+ QA and quality checks in **Confluence**, ensuring 100% **HIPAA/GDPR** compliance across all reporting pipelines.
- Conducted **root cause analysis** on claim rejection trends, combining data from **Cerner** and **SAP** to identify system-level discrepancies, enabling corrective action that improved claim acceptance rate by 18% within 60 days.

Business Data Analyst | EY, India

Jun 2019 – Aug 2022

- Led a **risk analytics** initiative at **EY**, applying **machine learning** techniques (logistic regression, ensemble models) to build **predictive credit risk scoring** algorithms using **Python** (**Pandas**, **Scikit-learn**), reducing false positives in **fraud detection** by 25%.
- Engineered and automated **ETL pipelines** using **Azure Data Factory** and **SQL** to ingest and transform 120M+ records from **SAP ECC** and third-party sources, streamlining data preparation by 45% and accelerating executive reporting.
- Designed and deployed 15+ interactive dashboards in **Power BI** and **Excel** (PivotTables, Power Query, VLOOKUP, Macros) to visualize **KPIs**, **ROI**, and **risk scores**, enabling a \$1.2M budget reallocation across teams.
- Delivered 15+ self-service dashboards in **Power BI**, **Tableau**, and **Excel** by transforming **raw data** from **cloud** sources and **spreadsheets** using **Power Query**, visualizing **KPIs**, **ROI**, and **risk scores**, cutting reporting time by 40% and driving a \$1.2M budget reallocation.
- Partnered with cross-functional teams to define **KPIs**, manage 40+ **Jira** user stories, document compliance workflows in **Confluence**, and led **CRM** data migration from **Salesforce** to **Dynamics 365** with 60% fewer reporting mismatches.
- Created **auditable**, end-to-end documentation of transformation logic, **stored procedure** scripts, and **risk KPIs**, reducing onboarding time for new analysts by 60% and ensuring **SOX** and **RBI** audit readiness.
- Unified ERP and CRM data from **SAP ECC** and **Salesforce** into **BigQuery** to enable cross-domain reporting for 5 global business units, accelerating executive dashboards and reducing manual data joins by 60%.

EDUCATION

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

Sep 2022 – May 2024

Bachelor in Computer Science - Savitribai Phule University, Pune, Maharashtra, India.

May 2016 – Jun 2020