# NIKHIL M

Boston, MA | nikhilmagsai17@gmail.com | LinkedIn | Portfolio

## PROFESSIONAL SUMMARY

- Data Analyst with 8+ years of experience in data analytics, business intelligence, and clinical/financial data management across healthcare, finance, and marketing domains.
- Expertise in SQL (MySQL, PostgreSQL, Hive, Snowflake), Python (Pandas, NumPy, Scikit-learn), R, and PySpark for advanced data processing and analysis.
- Experienced in building, automating, and optimizing ETL pipelines using AWS Glue, Airflow, and Azure Data Factory to improve data quality and reduce manual effort.
- Skilled in data visualization and reporting using Tableau, Power BI, Looker, and advanced Excel, delivering actionable insights for stakeholders.
- Strong background in statistical modeling, predictive analytics, clustering, regression, time series forecasting, and A/B testing to support business decisions.
- Hands-on experience with clinical and healthcare data standards (CDISC, MedDRA, CDASH) and HIPAA/FDA compliance for regulated datasets.
- Collaborates effectively with cross-functional teams, translating complex datasets into insights that improve campaigns, operations, and revenue.
- Proven ability to develop predictive models, segmentation frameworks, and dashboards that enhance decision-making and drive measurable business outcomes.

## TECHNICAL SKILLS

**Programming & Analytics:** Python (Pandas, NumPy, Scikit-learn, PySpark), R, SAS
**Databases & Querying:** SQL (MySQL, PostgreSQL, Hive, Snowflake, PrestoSQL)
ETL & Data Engineering: Apache Spark, Airflow, AWS Glue, Azure Data Factory
**Cloud Platforms:** AWS (S3, Redshift, SageMaker), Azure (Databricks, Synapse), GCP (BigQuery)
Data Visualization & BI: Tableau, Power BI, Looker
**Clinical & Healthcare Analytics:** CDISC (SDTM, ADaM), MedDRA, CDASH, EDC Tools, HL7/FHIR Standards
**Statistics & Machine Learning:** Regression, Hypothesis Testing, Time Series Forecasting, A/B Testing, Predictive & Prescriptive Modeling, Classification & Clustering
**Productivity & Collaboration:** Advanced Excel (VBA, Pivot Tables), Jira, Confluence, Git

## PROFESSIONAL EXPERIENCE

**Senior Data Analyst** **Nov 2024 – Present**
**Swift Strategic Solutions | Boston, USA**

- Consolidated 10M+ POS and loyalty datasets with Python (Pandas, NumPy) and SQL, uncovering price elasticity patterns that increased promotion ROI by 15% across beverages and snacks.
- Automated multi-source ETL pipelines in AWS Glue and Snowflake, cutting manual processing by 42 hours monthly and improving campaign readiness speed by 35%.
- Created real-time dashboards in Power BI integrating DAX measures and row-level security, enabling 30+ district managers to monitor promotion performance and react within 24 hours.
- Deployed uplift modeling in scikit-learn to evaluate promotional impact, achieving 18% lift in conversion rates by targeting customer cohorts more effectively.
- Improved data governance by enforcing field-level validation rules in SQL and Snowflake, reducing data errors by 30% and strengthening reliability of financial reporting.
- Partnered with marketing and finance teams to design 12 cross-functional analytics initiatives, influencing $5M budget allocation towards high-performing promotions.
- Implemented time series forecasting models in Azure Databricks to predict sales uplift, enhancing inventory planning accuracy by 20% across 2,000+ stores.
- Delivered standardized promotion effectiveness framework in Tableau, reducing inconsistent reporting across markets by 22% and enabling data-driven decisions at the regional level.

**Business Analyst** **Jan 2021 – Aug 2023**
**WIPRO PVT LTD | Karnataka, India**

- By translating requirements from 50+ stakeholders into Jira user stories, ensured clarity across sprints and cut downstream rework by 28%.
- Complex SQL and Hive queries on 20M+ rows were redesigned to run faster, which lowered dashboard refresh times by 32% and gave analysts near real-time access.
- Senior leadership gained 20% quicker insights when I introduced Tableau dashboards with drill-down KPIs, enabling them to compare regions and product lines interactively.
- Using Postman and Python validation scripts, I detected mismatched values across 15 third-party APIs, improving accuracy of monthly finance reports by 22%.
- Clear data lineage diagrams in Visio standardized how pipelines were documented, reducing onboarding time for new analysts by 35% and minimizing handover gaps.
- A Python-based regression testing framework was rolled out to check BI reports before release, which lowered recurring report errors by 18%.
- Partnering with cross-functional teams, I helped deliver a Conversational BI Assistant on Azure NLP, eliminating 38% of ad-hoc report requests and saving analyst hours weekly.
- During 40+ UAT sessions, I captured defects and feedback early, cutting testing cycle length by 25% while raising end-user adoption.

**Data Analyst** **Jan 2021 – May 2022**
**NTT DATA | Gurgaon, India**

- Used Python (Pandas, NumPy) and R to examine clinical trial and claims datasets, improving accuracy of risk prediction models by 12%.
- Standardized SQL pipelines for survey and claims data, which reduced reporting inconsistencies and supported a 5% increase in market penetration.
- Built segmentation models with R clustering algorithms, enabling targeted outreach to high-value healthcare customers and driving a 25% gain in campaign ROI.
- Simplified patient outcome trends by delivering interactive Tableau dashboards and SPSS reports, equipping 10+ stakeholders with faster decision-making insights.
- Conducted competitor and SWOT analysis using claims data and Excel, uncovering product gaps that directly contributed to a 1.7% revenue lift.

- Explored clinical utilization patterns through Python EDA, identifying behavioral clusters that guided the redesign of healthcare offerings to improve member engagement.
- Produced predictive models with logistic regression and survival analysis in R, raising forecast accuracy of insurance claims by 17%.
- Partnered with compliance and medical teams to embed HIPAA and FDA standards into analytics workflows, safeguarding sensitive healthcare data.

**Data Platform Engineer** **Aug 2018 – Dec 2020**
**American Express | India**

- Designed Spark and Python pipelines to unify promotional, transactional, and cardholder data, shrinking ETL runtimes by over 20% and accelerating campaign analytics.
- Established Hive data models on Hadoop that standardized disparate feeds, boosting cross-team reporting accuracy and enabling smoother integration with BI dashboards.
- Introduced Airflow scheduling workflows that kept marketing datasets refreshed automatically, eliminating manual dependencies and cutting processing delays by 70%.
- Re-engineered Redshift queries and partitioning strategies, which shortened dashboard refresh times by 30% and improved real-time decision-making for executives.
- Devised Python-based validation checks that detected anomalies in financial and campaign data streams, reducing recurring quality issues by 18%.
- Transitioned batch pipelines to AWS S3 and Glue, providing scalable storage for more than 1TB of monthly incoming data while minimizing downtime during migration.
- Enabled real-time insights by streaming engagement and payments activity through Kafka connectors, enriching fraud-detection systems with fresher behavioral signals.
- Partnered with compliance teams to incorporate GDPR and PCI safeguards into pipeline design, ensuring secure handling of sensitive cardholder information.

**Clinical Data Analyst** **Aug 2014 – June 2016**
**Avantsy Technologies Pvt Ltd | Hyderabad, India**

- Extracted and standardized clinical trial and patient interaction data using SQL and Excel, generating weekly reports that improved study enrollment accuracy by 10%.
- Developed Tableau dashboards to visualize patient engagement and trial site performance, enabling stakeholders to optimize study protocols and boost compliance rates by 25%.
- Applied R analytics to segment patient populations and identify high-risk cohorts, which guided clinical outreach and increased study retention by 15%.
- Automated adverse event reporting using Python scripts, reducing manual effort by 70% and enhancing data quality for regulatory submissions.
- Consolidated multi-site trial datasets with SQL joins and Excel pivot tables, producing actionable insights that improved protocol adherence across 5 clinical sites.
- Built predictive models in R to forecast patient dropouts, allowing proactive interventions that reduced attrition by 12%.
- Conducted quality checks on patient and trial data with Python validation routines, detecting inconsistencies and cutting reporting errors by 20%.
- Collaborated with clinical operations and regulatory teams to translate data insights into workflow improvements, ensuring study efficiency and adherence to compliance standards.

## PROJECTS

**Clinical Trial Patient Analytics Platform**

- Processed 500K+ patient records using Python (Pandas, NumPy) and SQL, standardizing multi-site trial data to enhance reporting accuracy by 18%.
- Built predictive models in R and Scikit-learn to identify patients at risk of dropout, enabling proactive interventions that improved retention.
- Created interactive Tableau dashboards for monitoring trial compliance and adverse events, providing real-time visibility for 10+ clinical teams.

**Retail Customer Behavior & Segmentation Engine**

- Ingested 20M+ transactional records with PySpark and SQL, applying K-Means clustering to identify high-value customer segments and boost targeting efficiency by 22%.
- Designed Power BI dashboards with dynamic DAX measures to monitor promotion ROI and campaign effectiveness across regions.
- Integrated data pipelines with Azure ML workflows, enabling retraining of predictive models as new customer data became available.

**Financial Services Fraud Detection Pipeline**

- Engineered ETL workflows with Python, Spark, and Hive to consolidate multi-channel transaction data, reducing data latency by 25%.
- Implemented logistic regression and random forest models to detect high-risk transactions, increasing fraud detection accuracy by 17%.
- Developed real-time dashboards in Tableau and Power BI for fraud monitoring, supporting daily decision-making for risk and compliance teams.

## EDUCATION

**Masters in Business Analytics**
Suffolk University, Sawyer Business School | Boston, MA
**MBA in Marketing & Finance**
NALSAR University, DOMS School of Business | Hyderabad, India

## CERTIFICATIONS

- IBM Data Science Professional Certificate – **Coursera**
- Google Data Analytics Professional Certificate – **Coursera**
- Machine Learning with Python – **Coursera**
- Tableau Desktop Specialist – **LinkedIn Learning**
- Data Engineering on Google Cloud Platform – **Coursera**