

RAJAT KAPGATE

Open to Relocation | +1 (812) 553-2822 | rkapgate@gmail.com | linkedin.com/in/rajat-kapgate | github.com/rajatk9962

PROFESSIONAL EXPERIENCE

Analyst - Strategic Patient Insights Piedmont Healthcare, Atlanta, GA	August 2025 – Present
<ul style="list-style-type: none">Leveraged SQL to analyze provider metrics and national birth rate data, explaining maternity campaign underperformance and enabling a confident \$150K+ reinvestment decision.Developed a GPT-4o powered NLP pipeline using prompt-completion to classify, summarize, and quantify 10,000+ free-text patient survey responses, enabling faster patient experience reporting across 5 departments.Built 2 interactive Tableau dashboards analyzing 1M+ office visit records spanning 3 years to identify trends and seasonal patterns, enabling marketing teams to optimize campaign timing and improve engagement by 15%.	
Research Data Scientist Indiana University School of Optometry, Bloomington, IN	May 2024 – Present
<ul style="list-style-type: none">Analyzed 500+ GB of infant head and eye movement data (NIH Grant EY032897) to identify early markers of eye disorders; performed ANOVA across five age groups, revealing statistically significant motor control variations.Preprocessed time series data using techniques like IQR filtering, rolling sum, low pass filtering and non-max suppression to isolate meaningful head motion segments, enhancing data quality by 40%.Engineered advanced data visualizations, including head movement reconstruction with Unity and Open3D, KDE, polar plots, and correlation maps to analyze infant head dynamics.	
Data Scientist Boehringer Ingelheim Pharmaceuticals, Ridgefield, CT	May 2024 – Nov 2024
<ul style="list-style-type: none">Implemented a FinOps cost analytics dashboard with Streamlit, integrating cloud cost data for better visibility. Monitored Jenkins CI/CD pipelines and performed root cause analysis, uncovering inefficiencies and cutting compute costs by \$200K+.Accomplished a 70% reduction in reporting turnaround for the drug Jardiance by leveraging Large Language Models, LangChain CSV agents, Azure Chat API, and Python-pptx, leading to an increase in decision making efficiency.Orchestrated an ETL pipeline to process 2M+ drug price data points from the Nuro API, optimizing SQL workflows with window functions on AWS RedShift, automating with cron jobs, and storing results in AWS S3.Devised a Retrieval-Augmented Generation (RAG) system on proprietary organizational data using Azure GPT-4o and FAISS for vector-based similarity search, cutting research effort by 40%.	
Data Analyst TCS Research, Mumbai, MH	Jun 2021 – Aug 2023
<ul style="list-style-type: none">Utilized Google BigQuery for crafting intricate database queries and harnessing Power BI and DAX queries to craft impactful dashboards. Managed a high-performing team of IT professionals, resulting in a 40% increase in project efficiency.Improved employee retention by 5% by leveraging SAP data to integrate KPIs (attrition rate, turnover, tenure) with multivariate forecasting, identifying and remediating 3 critical attrition drivers and informing targeted retention strategies.Developed Excel-based financial models incorporating macros and advanced functions (VLOOKUP, INDEX-MATCH) to automate team-level expense reporting, increasing accuracy and reducing manual effort by 40%.Served as a key resource for data science, mentoring six associates and simplifying complex analytics for managers, leading to a 20% improvement in decision-making and team efficiency in visualization and reporting.	

EDUCATION

Indiana University Bloomington, USA	Aug 2023 – May 2025
Master of Science in Data Science	GPA: 3.94/4.00
Coursework: Data Mining, Machine Learning, Advanced Database Concepts, Statistics, Data Visualization, Algorithms	
University of Mumbai, India	Aug 2017 – Jun 2021
Bachelor of Engineering in Computer Engineering	GPA: 3.50/4.00
Coursework: Big Data Analytics, Advanced DB, Elements of Artificial Intelligence, Data Structures, Exploratory Data Analysis	

TECHNICAL SKILLS

- Programming Languages:** Python, R, SQL
- Business Intelligence:** Tableau, Power BI, Excel, Looker Studio, GeoPandas, Seaborn, NumPy, Pandas
- Frameworks and Tools:** Apache Spark, Snowflake, Flask, Jenkins, Hadoop, Docker
- Databases:** MySQL, BigQuery, PostgreSQL, MongoDB, SSIS, SSRS
- Cloud Tools:** Azure (Databricks, Data Factory), AWS (S3, Redshift, Glue, Lambda, Sagemaker), GCP (BigQuery), Kubernetes
- Project Management:** Confluence, Jira, Agile, Waterfall, SDLC, Kanban, CRM, ERP

ACADEMIC PROJECTS

A/B Testing & Marketing Campaign Optimization

- Conducted A/B testing and regression analysis on 365-day Facebook and Google AdWords data to assess conversions and cost efficiency. Used hypothesis testing and cointegration analysis to optimize ad spend allocation and improve ROI.

Vio-Later: Predictive Prevention of Traffic Violations and Road Accidents

- Developed a traffic violation prevention system using New York traffic data, leveraging **GeoPandas** for city map visualization using **ensemble of XGBoost, ANN, Decision Trees** to predict high-risk violations and reduce accidents.

Patient Outcomes (Tableau)

- Unearthed **5 key trends** in MIMIC-III dataset through **data visualization**, providing **actionable insights** into patient outcomes.