RAJAT KAPGATE

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

SUMMARY

Data Analyst with 3+ years of experience delivering end-to-end analytics from raw data to dashboards and decision-ready insights. Strong in SQL and Python with BI tools (Tableau, Power BI/DAX) and experience integrating multi-source datasets, building KPI frameworks, and running segmented variance/root-cause analysis. Proven impact including a 15% engagement lift, a $150K+ reinvestment decision, and $200K+ cost reduction.

TECHNICAL SKILLS

- **Programming Languages:** Python, SQL, R
- **BI & Visualization:** Tableau, Power BI (DAX), Excel (Power Query/Power Pivot), Looker Studio
- **Data & Analytics:** Pandas, NumPy, GeoPandas, Matplotlib, Seaborn, A/B testing, forecasting, KPI design
- **Data Platforms:** BigQuery, Snowflake, Redshift, PostgreSQL, MySQL, MongoDB
- **Pipelines & Cloud:** Airflow, dbt, Spark, Docker; AWS (S3/Glue/Lambda), Azure (Databricks/Data Factory), GCP
- **Delivery:** Jira, Confluence, Agile/Scrum, SDLC

PROFESSIONAL EXPERIENCE

Analyst - Strategic Patient Insights | Piedmont Healthcare, Atlanta, GA **August 2025 – Present**

- Enabled a confident **$150K+ reinvestment** by diagnosing maternity campaign underperformance by crafting **SQL queries**, blending provider **KPIs, ICD-10 EHR** encounters, and national birth-rate benchmarks with segmented **variance analysis**.
- Spearheaded development of Open Insight, a real-time patient review **datastore**, using **dbt** to build validated marts linking reviews with EHR attributes and the **GPT-4o** API to summarize reviews, cutting reporting cycles by 50% across 5 specialties.
- **Built** 2 interactive Tableau dashboards analyzing **1M+ office visit records** from **Epic EHR** 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**

- 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**

- Implemented a FinOps cost analytics dashboard in **Streamlit** integrating multi-cloud cost data (AWS, Azure, GCP); built and monitored **Airflow pipelines** to ingest up to date spend datapoints, with root-cause analysis to reduce compute costs by **$200K+**.
- Accomplished a **70%** reduction in **reporting** turnaround for the drug Jardiance by leveraging automation using **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**

- 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.
- Led a Digital Pathology cloud migration by building a GCP Dataflow ETL pipeline processing 100M medical records into BigQuery; delivered Power BI/DAX dashboards for operational visibility and improved team delivery efficiency by 40%.
- 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%.**

EDUCATION

Indiana University Bloomington, USA **Aug 2023 – May 2025**
Master of Science in Data Science GPA: 3.94/4.00

University of Mumbai, India **Aug 2017 – Jun 2021**
Bachelor of Engineering in Computer Engineering GPA: 3.50/4.00

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.