Ritesh Somashekar

✉ ritesh.chandra33@gmail.com ☒ 📞 +1 571-245-7770 ☒ **in**LinkedIn ☒ 🧳Portfolio ☒ ⭕Github ☒

Technologies

- **Programming Languages:** Python, SQL, R, JAVA, C, Scala, Pyspark
- **Data Engineering Tools:** PowerCenter, Precisely Data360, Precisely Assure, AWS (S3, Lambda, Athena), Databricks, PySpark, Microsoft SQL Server, MySQL,Git, Jupyter, Snowflake, Microsoft Power Platform (Power Apps, Power Automate), Kafka, Kubernetes, Apache Airflow, Selenium.
- **Data Analysis & Visualization:** Pandas, NumPy, Scikit-Learn, Tableau, Power BI, Excel (Formulas, VBA)
- **Techniques:** Data Pre-processing, ETL, Data Warehousing, Statistical Analysis, Agile methodologies,Data Visualization, Machine Learning, Web Scraping

Experience

Data Engineer — *Bangalore,India*
Carelon Global Solutions (Elevance Health) ☒ — *May 2022 – Aug 2023*

- Engineered **20+ agile ETL pipelines for provider and consumer analytics,** with a specific focus on **claims balancing workflows**, using **Python, Informatica PowerCenter, and a CI/CD-integrated framework with Precisely Assure, Data360, and AWS (S3, Lambda, Athena)**. Orchestrated pipelines via Airflow with embedded data validation and lineage tracking, improving reliability, observability, and driving measurable business value.
- Built a **Python-based automation workflow** for daily ingestion of data from **NPPES repositories**, seamlessly integrated into existing ETL pipelines. Improved data accuracy to 74%, reduced processing overhead, and contributed to significant cost optimization.
- Built custom **web crawlers to parse over 400,000 healthcare documents in XML/JSON format** — enhancing real-time data availability for downstream applications and **slashing manual processing by 70%.**
- Formulated **Data Analysis Expressions (DAX) for data optimization and Advanced SQL queries comprising of stored procedures to extract data from various data sources, improving efficiency by 60% in the Edward Profiling** project and reduced manual interventions.
- Developed **Tableau and Power BI dashboards** to present provider insights and performed root cause analysis for performance anomalies.

Associate ETL Engineer — *Bangalore,India*
Carelon Global Solutions (Elevance Health) ☒ — *October 2020 – Apr 2022*

- Played a key role in **the Seven Plus Locations project** by identifying key metrics and key performance indicators (KPIs) using Python.
- Designed and developed **advanced Power BI dashboards** and reports to transform complex data into actionable insights showcasing different trends based on **Exploratory Data Analysis and Data Profiling.**
- Developed algorithms for **automating medical invoice processing**, improving claims balancing accuracy and **reducing error rates by 70%.**

Additional Experience

Graduate Research Assistant — *Fairfax, VA*
Costello College of Business - GMU ☒ — *August 2024 – Present*

- Designed **Python and R-based algorithms** to **investigate social cohesion and community coping** behaviors **during disaster scenarios** using large-scale real-world datasets.
- Leveraged graph-based analytics to uncover pattern insights and utilized ensemble models to forecast trends, achieving an RMSE of 0.89.
- Visualized findings using **R and Tableau** to support the use-case and Research goals. Rendered the same on OpenCV reports.

Education

George Mason University(GMU) ☒ — *Aug 2023 – May 2025*
MS in Data Analytics and Engineering, GPA:3.93/4.0
Related Courses: Data Analytics, Business Analytics, Data Mining, Viz using Tableau & Power BI, Advance Machine Learning, Neural Networks, Natural Language Processing, Applied Statistics

Projects

Agentic AI-Based Ticket Automation System ☒ | Agentic AI |SecureGPT | AWS | JIRA |NIST Compliance | Agile

- Designed an autonomous AI system that monitors cybersecurity anomalies and auto-generates JIRA service tickets aligned with SLA policies.
- Implemented agent-to-agent handshakes to streamline ticket triage with 93% automation accuracy.

Analyzing Diabetes Dynamics using Machine Learning ☒ | Python | R | EDA | Web Scraping | MongoDB | Databricks

- Conducted cluster-based segmentation and applied time-series forecasting to identify pre-diabetic patterns in U.S. population groups.
- Achieved strong model performance with an RMSE of 0.87 and MAE of 1.45, highlighting predictive accuracy for healthcare intervention planning.