Piyusha Sayal

LinkedIn | piyusha.2510@gmail.com| +1 (240)-869-3189 | Portfolio

Data Engineer with 3–4 years of experience building scalable data pipelines, orchestrating ETL workflows, and improving data quality across cloud platforms. Skilled in PySpark, Airflow, SQL, and AWS with strengths in automation, data modelling, and distributed processing. Recognized for delivering production-ready data systems that enhance analytics accuracy and operational efficiency.

SKILLS & CERTIFICATIONS

Data Engineering & Big Data: PySpark, Apache Spark, Hadoop, Hive, ETL, Airflow, Snowflake, Matillion, SSIS, Informatica, Databricks
Programming & Cloud Platforms: Python, SQL, R, C++, Stata, AWS (S3, EC2, Glue, Lambda), Azure, GCP
Databases: PostgreSQL, SQL Server, MySQL, MongoDB
Data Tools: Pandas, NumPy, AWS CLI, Git, REST APIs, Power BI, Tableau, Alteryx, Excel
Soft Skills: Stakeholder Management, Problem-Solving, Requirements Gathering, Agile
Certifications: AWS Certified Data Engineer, Academic Scholarship (NJIT), Gen AI (BCG), Big Data (IBM), Data Analytics (Accenture, KPMG)

WORK EXPERIENCE

Optimove   **New York City, NY**
Customer Data Engineer   *August 2024-Present*
- Optimized ETL workflows (SQL, Python, and Airflow), improving pipeline efficiency and reducing the data delivery time by 40%.
- Partnered with cross-functional teams to translate business requirements and resolve 200+ data quality issues for 10+ clients.
- Generated data-quality KPIs to resolve 50+ critical issues, presenting findings to stakeholders to drive a 30% reduction in downtime.

New Jersey Institute of Technology   **Newark, NJ**
Business Development Analyst (Autonomous Truck Driving Model)   *August 2023-May 2024*
- Conducted 20+ industry interviews to identify market trends, informing product decisions for autonomous trucking solutions.
- Served as Entrepreneurial Lead for the NSF I-Corps, uncovering inefficiencies through customer discovery and stakeholder analysis.
- Evaluated customer needs and gaps, delivering strategic recommendations to shift toward semi-autonomous truck capabilities.

Data Operations Analyst   *June 2023-August 2023*
- Executed a comprehensive switch of insurance providers, leading unit acceptance testing through all 3 stages of implementation.
- Formulated 50+ scripts for NJIT's DegreeWorks platform, improving course enrollment efficiency and reducing scheduling errors.
- Engineered the ERD for the summer program students' database, improving data retrieval speed by 25% and reducing redundancy.

Financial Data Analyst (CSR Report and Financial Performance Analysis)   *March 2023-June 2023*
- Analyzed large-scale ESG data for 200+ companies, using statistical methods to identify trends and generate insights in CSR reporting.
- Categorized 1000+ filtered keywords and assessed the polarity, facilitated focused analysis and generated targeted insights.
- Designed 40+ interactive dashboards using Power BI and Tableau, monitoring KPIs and financial impacts for executive stakeholders.

Faces Canada   **India**
Supply Chain Data Analyst   *January 2021 – July 2021*
- Analyzed ERP data for 500+ orders, identified order inaccuracies and implemented a centralized inventory management system.
- Developed automation strategies that streamlined warehouse operations and cut processing time by over 50 hours monthly.
- Overhauled order fulfillment process reducing holding costs by 15% through precise packing procedures and inventory tracking.

EDUCATION

New Jersey Institute of Technology   **Newark, NJ**
MS in Information Systems, Teaching Assistant (GPA: 3.95/4.00)   *September 2022-May 2024*

PROJECTS (*Access projects - Github*)

- **PySpark - Pipeline for Flight Delay Prediction:** Built a distributed ETL pipeline in PySpark to process large-scale flight datasets and deployed an ML model achieving 85% accuracy. Implemented feature engineering, and automated ingestion for scalable forecasting.
- **National Science Foundation Website:** Developed a responsive EV charging station platform using HTML, CSS, and JavaScript, integrating accessibility standards and embedding analytics pipelines to capture and analyze user-behavior data.
- **SpaceAlpha- Distributive Native Space Data Platform:** Created a distributed data platform leveraging MongoDB, Redis, Rust, and Google Cloud, building ingestion pipelines and API-driven services to centralize multi-source space-tech information.

LEADERSHIP EXPERIENCE & ACHIEVEMENTS

Hackathon Victories: Successfully led 4 member teams to win multiple hackathons, showcasing technical expertise and strategic acumen.
Marketing Leadership: Served as Marketing Head for 4 years, driving successful sponsorships for the annual Techno-Cultural Fest.
Conference Organization: Spearheaded E-Summit 2022, collaborating with over 20 industry leaders to boost participant engagement.