

Summary

Data Engineer with 6+ years of experience delivering production-ready data solutions and scalable ETL pipelines. Expert in SQL, Python, and data visualization tools, with a strong track record of optimizing workflows in cloud environments. Demonstrated ability to collaborate with teams to maintain data quality and support analytic product development through effective documentation and project management.

Experience

Pavane Solutions Inc.	Data Engineer	Remote, NJ Jul 2024 – Present
<ul style="list-style-type: none">Designed scalable batch and real-time data pipelines to support business-critical ETL workflowsAutomated PySpark ETL with Apache Airflow, reducing manual effort by 40% and increasing reliability.Migrated on-prem systems to AWS and Snowflake, enabling scalable, cloud-native data infrastructure.Optimized data ingestion pipelines, improving performance by 25% while maintaining data integrity.		
Saint Peter’s University	Data Science Researcher	Jersey City, NJ Nov 2023 – Feb 2025
<ul style="list-style-type: none">Built ACO + Grid Search metaheuristic algorithm for logistic regression optimization, improving breast cancer model accuracy by 15%.Designed real time AQI-based routing with Azure Maps to optimize travel for respiratory patients, enhancing public health safety.Developed an NLP-driven news cloud tool using real-time APIs; both projects presented at NJBDA as part of university-led research.		
IT Nopal Technologies	Data Scientist II	New Delhi, India Jun 2021 – Jan 2023
<ul style="list-style-type: none">Maintained dynamic pipelines for high-volume data sports applications, boosting processing efficiency and reliability.Reduced AWS deployment costs by 15% by optimizing Lambda, EC2, and S3 usage across infrastructure.Integrated data pipelines with APIs and SQL-based ETL systems, increasing throughput while maintaining pipeline integrity.Worked with DevOps teams to containerize ML workflows using Docker and manage them via CI/CD pipelines on GitLab.Automated key SQL processes, delivering 20% efficiency gains through insight-driven database management.Achieved 100% data integrity by deploying Python scripts to process JSON data from 50K+ sports matches.		
IT Nopal Technologies	Data Scientist I	New Delhi, India Jan 2019 – May 2021
<ul style="list-style-type: none">Migrated NoSQL to PostgreSQL with PyODBC, reducing query time by 85% and EC2 cost by 70%.Developed real-time data access pipeline using AWS Lambda and EC2, improving data retrieval speed by 30%.Created interactive dashboards with Plotly and SQL for monitoring product KPIs and business health.Built interpretable ML models for classification and prediction, deployed as part of customer analytics workflows.		

Projects

Interactive Keyword-Based News Retrieval System – NLP & Real-Time Content Delivery NJBDA – Under Review	GitHub Live Demo
<ul style="list-style-type: none">Built a real-time keyword-based platform to extract and process 300+ articles using NLP and REST APIs (NYT, Bing).Automated API-to-UI pipeline using Python and Streamlit, reducing content delivery time for end users.Engineered keyword-driven WordClouds linking text to articles/videos, improving user engagement and clarity.Tools & Tech: Python, TextBlob, REST APIs (NYT, Bing), JSON, Web Scraping, Streamlit, WordCloud, Real-Time Data Engineering	
Optimization with Metaheuristic Algorithms – Breast Cancer Diagnosis	GitHub
<ul style="list-style-type: none">Developed ACO + Grid Search hybrid tuner, reducing logistic regression compute time by 20% and improving model efficiency.Increased predictive accuracy by 15% over traditional tuning methods and minimized error rates by 30% on breast cancer datasets.Applied advanced metaheuristic techniques to optimize model performance, enabling more precise and efficient healthcare diagnostics.Tools & Tech: Python, Scikit-learn, Ant Colony Optimization, Grid Search, Logistic Regression, Hyperparameter Tuning, Healthcare Analytics	
Health-Centric Navigation & Air Quality Management for Sensitive Populations NJBDA – Under Review	GitHub Live Demo
<ul style="list-style-type: none">Developed a smart routing system using Azure Maps and a custom AQI-based algorithm to guide users through least polluted paths.Integrated real-time AQI and weather data into a custom path-finding algorithm optimizing travel routes for respiratory health.Demonstrated reduced health risk in 100+ route simulations, showcasing potential for smart city and healthcare applications.Tools & Tech: Azure Maps, REST APIs, Python, JSON Parsing, Real-Time Data Processing, Geospatial Analysis, Environmental Data Integration, Route Optimization.	

Achievement & Honors

Data Science Club President (Sep 2024 – Apr 2025), Lead Presenter – NJBDA Research Conference & Symposium: (Presented 2 Research Paper), 1st Prize – Data Science Showcase, Alpha Sigma Nu Honor Society Inductee, Data Storyteller Award Winner

Education

Saint Peter’s University, Master of Science in Data Science (GPA - 3.95/4.00)	Jersey City, NJ Feb 2023 – Feb 2025
AKTU, Bachelor of Technology in Computer Science and Engineering: (GPA - 3.57/4.00)	India Aug 2016 – Aug 2020

Skills

- Programming Languages & Databases:** Python, SQL, C and C++, SQL Server, PostgreSQL, MongoDB
- Big Data & Data Mining:** Tableau, Statistical Modeling, Python, (e.g. Scikit-learn, Pandas, NumPy), Text Analysis (Structured & Unstructured), AWS (S3, Redshift, RDS, Lambda, EC2, S3), Azure (Databricks, SQL DB), Big Query
- Data Science & AI Technologies:** ETL, DevOps, End-to-End ML Methodology, Cloud Infrastructure (AWS, Azure), Time series analysis, Predictive Modeling (Supervised & Unsupervised Learning), Hypothesis testing, feature Selection, Data Manipulation, Data Visualization