

# SANYA SYED

📍 Houston, TX | 📩 [sanyashireen@gmail.com](mailto:sanyashireen@gmail.com) | 💬 <https://www.linkedin.com/in/sanyasy/> | 🌐 [github.com/sanyassyyed](https://github.com/sanyassyyed)

## PROFESSIONAL SUMMARY

**Data Engineer** with experience designing ETL pipelines, integrating operational and analytical data, and supporting cloud-based analytics platforms. Strong foundation in Python, SQL, and data modeling, with experience across AWS and GCP. Brings a Data Science background and hands-on project work to build reliable, well-structured data pipelines with a focus on quality, automation, and maintainability.

## EDUCATION

**Msc Data Science** | Robert Gordon University | UK | 2018

**Bachelors In Computer Applications** | Madras University | India | 2008

## EXPERIENCE

**Data Engineer & Analytics Consultant** | Aqua Source Inc. | Sarasota, USA | June 2022 – Present

- Designed end-to-end data pipelines to track the order-to-fulfillment and billing lifecycle across clients, vendors, shipments, and payments.
- Built automated ETL workflows using Python and SQL (PostgreSQL) to consolidate order schedules, shipment updates, inspection reports, and billing data from multiple operational sources.
- Modeled relational datasets covering orders, shipments, inspections, invoices, and payments, improving data consistency across teams.
- Standardized inspection status data from 10+ vendor facilities, eliminating manual consolidation and reducing reporting discrepancies by ~70%.
- Created a centralized master dataset as a single source of truth for order status and fulfillment timelines, used daily by operations and finance.
- Deployed scalable, secure pipelines on AWS (S3, Lambda, IAM, Secrets Manager) and containerized workflows using Docker for reliable deployments.
- Integrated curated datasets into Snowflake and built Power BI dashboards to monitor shipment delays, billing cycles, and receivables, reducing ad-hoc reporting by ~60%.
- Partnered with sales, operations, finance, and quality teams to translate business processes into structured data models and actionable KPIs.

**Tools & Technologies:**

Python, SQL (PostgreSQL), AWS (S3, Lambda, IAM, Secrets Manager), Snowflake, Docker, Power BI, Pandas, Google Cloud Functions, Google Sheets API, Excel, CSV, SFTP

**Data Scientist / Data Engineer** | DeepMiner | Aberdeen, UK | Apr 2018 – Oct 2021

- Automated ingestion and transformation of structured and unstructured data by building scalable ETL pipelines using Python and SQL, reducing processing time by 40%.
- Enabled unified reporting by consolidating REST APIs, XML, and flat files into a centralized system.
- Improved analytics by optimizing data models, enhancing pipeline speed and downstream reporting accuracy.
- Delivered 10+ Power BI and Tableau dashboards, increasing operational efficiency by 25%.
- Reduced data quality issues by 50% through validation, integrity checks, and governance practices.
- Developed and deployed AI-driven NLP pipelines using SpaCy, CoreNLP, and BERT, improving text classification and entity recognition accuracy by 30%.

**Tools & Technologies:** Python, SQL, R, SpaCy, Stanford CoreNLP, scikit-learn, BERT, Power BI, Tableau, Pandas, REST APIs, XML, Neo4j

**Early Childhood Professional** | Honor Roll School | Houston, USA | Feb 2014 – Mar 2015

**System Support Executive** | Premium Lifestyles | Chennai, India | Jun 2008 – Jul 2009

## PROJECTS

- **Comail AI** – Agentic AI CV Tailoring & Outreach ([Kaggle](#)) – Developed an AI-driven, multi-agent system to automate resume customization and cold email generation from job descriptions, reducing manual effort and improving alignment with role requirements.
- **San Francisco Eviction Data ELT** ([GitHub](#)) - Designed and deployed an end-to-end pipeline using GCP, Docker, SQL, and PySpark, enabling 50% more accurate trend analysis.
- **Employee Hiring and History ETL** ([GitHub](#)) - Developed ETL pipeline with SQL & Python to analyze hiring trends and build Power BI dashboards, uncovering 46% turnover growth and diversity insights.
- **Identifying Sales Prospects from Textual Data - A Machine Learning Approach** - Built AI-driven text classification models using Scikit-learn, achieving 90%+ accuracy to automate prospect identification.

## PROFESSIONAL TRAINING & CERTIFICATIONS

**AI Agent Intensive Course** | Google X Kaggle | 2025

**Data Fundamentals Bootcamp** | WeCloudData | 2023

**Data Engineering Bootcamp** | DataTalksClub | 2023

**Professional Certificate For Data Visualisation With Python** | University Of Aberdeen, Scotland | 2022

**Diploma Software Engineering** | National Institute of Information Technology | India | 2007 – 2008

---

## TECHNICAL SKILLS

**Languages & Tools:** Python, SQL, Scala, Spark, PySpark, Hadoop, dbt

**Cloud Platforms:** AWS (Lambda, S3, EC2), GCP (Compute Engine, Functions, Data Lake), Azure

**Data Integration & Pipelines:** Docker, Prefect, CI/CD

**Databases & Data-warehouses:** PostgreSQL, MS SQL Server, BigQuery, Snowflake

**Reporting:** Power BI, Tableau

**Infrastructure & DevOps:** Terraform, GitHub, GitHub Actions

**AI / Machine Learning:** NLP, Transformers (BERT), Scikit-learn, SpaCy, CoreNLP, Agentic AI (Multi-Agent Systems)

**File Formats & APIs:** XML, CSV, REST APIs

**Other:** Data Modeling, Data Warehousing (Star/ Snowflake schema), Data Governance

---

## VOLUNTEERING WORK

**Data & Web Coordinator** | Slb Spouse Association | Global | 2013 – Present

**Tutor** | Westhill Library | Aberdeen, UK | 2017