

---

## PROFESSIONAL SUMMARY

**Data Engineer & Analyst** with 3+ years of experience designing ETL pipelines, integrating multi-source data, delivering analytics dashboards, and implementing AI/ML solutions. Skilled in Python, SQL, Power BI, and cloud platforms (AWS, GCP). Experienced across professional and pro bono projects, applying data governance, quality, and security best practices. Continuously upskilled through bootcamps and hands-on projects, demonstrating adaptability and immediate readiness to contribute.

---

## 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

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

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

---

## EDUCATION

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

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

---

## PROFESSIONAL TRAINING & CERTIFICATIONS

**Applied Data Engineering Certificate Program** | WeCloudData | Ongoing

**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

---

## EXPERIENCE

**Data Engineer & Analyst (Pro Bono Projects)** | Aqua Source Inc. | Sarasota, USA | Jan 2024 – Jun 2025

*Projects: Seafood Logistics Dashboard & ETL Automation*

- Built ETL pipelines using Python, SQL, GCP Functions, and Sheets API, reducing manual reporting by 60-80%.
- Cleaned and structured datasets, improving data consistency and enabling real-time monitoring of delivery and inventory.
- Developed interactive Power BI dashboards for tracking KPIs, operational performance, and forecasting.
- Ensured secure data handling via GCP Secret Manager, SFTP, and pipeline documentation.

**Tools & Technologies:** Python, SQL, Power BI, Google Cloud Functions, Sheets API, Secret Manager, CSV

**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

**Professional Development & Career Transition** | Houston, USA | Oct 2021 – Present

- Built and deployed production-grade pipelines using Python, SQL, Docker, and AWS Lambda, reducing latency by 40%.
- Published Power BI dashboards tracking KPIs from multi-source data, enhancing reporting efficiency.

- Applied data lineage, metadata documentation, and pipeline monitoring techniques to maintain transparency and traceability across project stages.
- Balanced full-time upskilling with substitute teaching and peer mentoring returners entering data careers.

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

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

---

## PROJECTS

- **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.
- 

## VOLUNTEERING WORK

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

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