

Shourya Singh

London, UK | singhshr777@gmail.com | 07780 148148 | <https://www.linkedin.com/in/shouryasingh24>

PROFESSIONAL SUMMARY

Data Engineer with 4+ years' experience designing scalable data pipelines, curating high-quality datasets, and applying statistical analysis to evaluate data integrity and impact. Strong proficiency in Python (Pandas, NumPy) and SQL, with hands-on experience building reproducible workflows for data ingestion, transformation, and validation. Passionate about improving dataset quality, understanding model-relevant signals, and aligning data systems with responsible AI principles.

TECHNICAL SKILLS

- **Python & Data Libraries:** Pandas, NumPy, data preprocessing, EDA, data validation
- **SQL & Data Warehousing:** Advanced SQL, BigQuery, large-scale dataset transformation
- **Data Engineering:** ELT pipelines, ingestion frameworks, schema management, version-controlled workflows
- **Dataset Quality & Analysis:** Data profiling, anomaly detection, variance analysis, metric design
- **AI & Experimentation:** Structured evaluation frameworks, data impact analysis, AI-assisted workflows

WORK EXPERIENCE

Commercial Data Analyst | Department for Transport

May 2024 – Present | London, UK

- Designed and maintained scalable data ingestion and transformation pipelines integrating 5+ structured data sources into curated analytical datasets, improving data reliability by 25% across a £1.2bn+ portfolio.
- Applied systematic data profiling and preprocessing techniques (schema validation, anomaly detection, completeness checks) to improve dataset quality and integrity, reducing downstream reporting errors by 30%.
- Conducted exploratory data analysis on high-volume datasets (millions of records) to identify signal inconsistencies and refine metric definitions, improving analytical robustness.
- Built automated, version-controlled workflows for data preprocessing and validation using Python and SQL, reducing manual processing effort by 50% and increasing reproducibility of analytical outputs.

Data Analyst | NHS

August 2023 – May 2024 | Bristol, UK

- Processed and analysed 5M+ structured and semi-structured healthcare records, performing quality assessment and cleansing to ensure dataset consistency across 12 Trust environments.
- Standardised dataset definitions and transformation logic to improve cross-environment comparability, increasing reporting consistency by 40%.
- Designed repeatable preprocessing and validation workflows to assess data sufficiency and reduce measurement bias in operational analytics.
- Collaborated with governance stakeholders to ensure responsible handling of sensitive data, aligning datasets with privacy and compliance standards.

Business Analyst | Pollen Pharma Pt. Ltd

September 2020 – January 2022 | India

- Curated and transformed 100k+ transactional records into structured datasets for segmentation and performance evaluation, improving data accessibility for marketing decision-making.
- Conducted statistical and exploratory analysis to identify high-impact product-location combinations, contributing to a 15–20% uplift in annual regional sales performance.
- Built reusable reporting datasets and dashboards that reduced manual analytical workload by 30% and improved data discoverability for 8–10 stakeholders.
- Implemented structured validation and preprocessing checks to improve dataset completeness and reduce inconsistencies in campaign performance tracking.

PROJECTS

AI-Driven Evaluation Platform | Python, Streamlit, SQLite: [Github](#)

- Built a Streamlit-based analytics product to evaluate CVs against job descriptions, with persistent storage of analysis runs using SQLite.
- Designed an AI-ready evaluation workflow to generate candidate fit summaries, skill alignment, and gap analysis, with historical retrieval by run ID.
- Applied production-style practices including virtual environments, Git version control, modular architecture, and clean repository hygiene.
- Demonstrated application of advanced analytics concepts to real-world decision support and optimisation use cases.

EDUCATION

MSc Business Analytics, University of Exeter

Award: Merit | 2022 – 2023

BTech Mechanical Engineering, Vellore Institute of Technology

Award: 7.8 CGPA | 2016 – 2020

CERTIFICATIONS & AWARDS

- Microsoft – Power Platform Apps for Data Analysts (Power BI, Power Query, Power Automate)
- DataCamp – Data Engineering Certification (<https://www.datacamp.com/certificate/DEA0010994684473>)

KEY ACHIEVEMENTS

- Supported total £1.2B contract pipeline pricing and negotiation process, integrating financial risk analysis.
- Standardized KPI frameworks across 12 NHS Trusts, strengthening reporting comparability and efficiency (+40%).
- Improved customer acquisition by 25% via segmentation-driven pricing strategy at Pollen Pharma.