

# SHAHEEN RAHAMATULLAH

0743 501 0008 | [shaheen.rahamatullah@gmail.com](mailto:shaheen.rahamatullah@gmail.com) | Greater Manchester, U.K.

[linkedin.com/in/shaheen-rahamatullah](https://www.linkedin.com/in/shaheen-rahamatullah) | [github.com/ShahveenRahamatullah](https://github.com/ShahveenRahamatullah)

Data Science Portfolio: [datascienceportfol.io/Shahveen](https://datascienceportfol.io/Shahveen)

Work Authorisation: Graduate visa (PSW) valid until Jan 2026

---

A Data Science professional with a recent Master's degree in Data Science from the University of Sheffield, UK, and a proven track record across the retail, supply chain, and insurance sectors. Specialising in leveraging advanced analytics and machine learning techniques to drive business insights and outcomes. Proficient in SQL, Python, R, and a suite of data visualisation and analysis tools, including Tableau, Power BI, and Databricks. Demonstrated history of successful projects in data analysis, predictive modelling, and dashboard development. Eager to apply a blend of technical skills, strategic thinking, and commercial experience to solve complex challenges in Data Science, Analytics, and Engineering roles. Open to opportunities across the UK, offering flexibility and a keenness to contribute to innovative data-driven solutions.

## EDUCATION

---

MSc in Data Science (Distinction)                      The University of Sheffield, UK    Sep '22 - Sep '23  
(Data Analysis, Data Visualisation, Data Mining, Database design, Big data analytics and Business Intelligence)

BE in Computer Science and Engineering (Distinction)                      Anna University, India  
(Artificial intelligence, Object Oriented Programming, Software Testing, Software Engineering)

## SKILLS

---

- Programming Languages - SQL, Python, R
- Python Libraries - Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, PySpark, Streamlit, Langchain, BERT
- Visualisation Tools -Tableau, Power BI (DAX), MS Excel, MS PowerPoint, MS Visio
- Data Intelligence Platform - Databricks
- Interactive computing platform - Jupyter Notebook
- Data Analysis, Statistical Analysis, Predictive Analysis, Machine Learning, Data Mining, Logistic regression, Random forest, Artificial Neural Network (ANN), Decision tree, Linear regression, Large language Model (LLM), NLP, BERT
- Integrated Development Environment - PyCharm, Anaconda, R-Studio
- Version Control - Git
- Code Repositories - Github
- MS Office, SPSS, KNIME, Power Query
- Cloud Services - AWS, Azure
- RDBMS - MySQL, Oracle
- No SQL - MongoDB
- Collaboration - Jira, Confluence, Slack, Microsoft Teams, Microsoft 365, Trello, Zoom, Service Now
- Software Development Methodologies - Agile, Waterfall
- Soft Skills - Communication, Presentation, Storytelling, Project Management, Time Management, Leadership, Teamwork, Flexibility, Attention to detail

## WORK EXPERIENCE

---

Associate Manager, HTC Global Services, India, Feb 2013 – Aug 2022  
Technical Lead, Infosys Limited, USA & India, Aug 2005 – Jan 2013

## CERTIFICATE COURSES

---

- PL-300 | DA-100 Microsoft Power BI Certification Exam Prep - Udemy
- Ultimate AWS Certified Solutions Architect Associate SAA-C03 - Udemy
- Tableau Desktop Specialist Certification Exam Prep 2024 - Udemy
- Azure AI Fundamentals - Microsoft

## KEY PROJECTS

---

**Data Scientist Internship, jobsRmine, UK, Sep 2023 to Current**      **Industry:** Recruitment

In this pivotal role within a dynamic job portal environment, I spearheaded the development of an advanced interactive dashboard to scrutinise and interpret the portal's performance metrics comprehensively. This initiative enhanced the data's accuracy and discoverability and played a critical role in elevating the portal's user engagement and operational efficiency. Key contributions and learning outcomes include:

- **Data Analytics Excellence:** Employed sophisticated data analysis techniques using Python (Pandas, NumPy, Scikit-learn, BERT) and SQL, enabling the extraction of actionable insights from vast datasets, thereby directly supporting strategic decision-making processes.
- **Predictive Analytics Implementation:** Leveraged statistical models and machine learning algorithms to forecast job market trends, significantly improving the portal's predictive capabilities and enabling proactive adjustments to market shifts.
- **Text Classification:** Utilised sophisticated NLP models such as BERT and BERTopic for text classification, accurately categorising job descriptions into their respective job categories.
- **Similarity Score:** Developed a similarity score metric to assess the alignment between candidates' resumes and job descriptions through cosine similarity.
- **Dashboard Development:** Designed and implemented a state-of-the-art Power Bi dashboard, presenting complex data analyses through intuitive visualisations that facilitated quick insights for technical and non-technical stakeholders.
- **Data-Driven Improvements:** Conducted a deep dive into historical data, identifying and implementing quality enhancements that boosted the accuracy and usability of job market insights, thereby enhancing user satisfaction and engagement.
- **Technological Proficiency:** I gained hands-on experience with various tools and technologies, including Jupyter Notebook for interactive computing, Git for version control, and advanced data visualisation techniques, preparing me for a versatile role in data science and analytics.

This role provided a comprehensive platform for applying theoretical knowledge to practical challenges, reinforcing my analytical skills and contributing to my professional growth in data science and analytics.

**Technologies:** Python, Pandas, NumPy, BERT, BERTopic, Cosine Similarity, Transformers, Power Bi, MySQL, MS Excel, Jupyter Notebook, git, data analysis, data visualisation, data cleaning, reports, dashboard creation and presenting results

### **Data Science Projects, Full-time student at Univ. of Sheffield, UK, Sep 2022 - Sep 2023**

- **Stroke prediction:** Developed and implemented predictive models using Python to predict stroke from an imbalanced health dataset. Utilised logistic regression, random forest and ANN algorithms, addressing class imbalance through resampling techniques. Optimised model performance through cross-validation and hyperparameter tuning, resulting in accurate stroke predictions.
- **What's Cooking Streamlit app:** Developed a Streamlit web application called "what's cooking" to simplify meal planning for daily cooking using LLM, Langchain libraries in Python and openAI.
- **Big data analysis - NLP:** Analysed 50K unstructured job descriptions using big data solutions using PySpark, Pandas, Dataframes, and Databricks, incorporating Natural Language Processing (NLP) techniques like tokenization, lemmatization and stemming and found valuable insights.
- **Sales analysis:** Built an interactive Tableau dashboard to analyse the sales trends and marketing decisions for best and worst-selling products from a sales dataset.
- **Titanic survivor prediction:** Built a couple of supervised machine learning predictive models like decision tree and logistic regression to predict the survivors on the Titanic using KNIME. Conducted data cleaning, exploration, and feature engineering and evaluated the optimal model.
- **Fall in elderly prediction:** Conducted a comprehensive analysis by applying statistical methods like the chi-square test and predicted falls among the elderly using logistic regression in SPSS.
- **COVID-19 mortality prediction:** Evaluated the influence of population density and the stringency index on COVID-19 mortality among high-income European countries by applying multiple linear regression using R.
- **Domestic violence analysis:** Employed advanced data visualisation techniques and graphical representations using R to identify valuable insights on the domestic abuse issue in England and Wales.
- **Technologies:** Python, Pandas, NumPy, Scikit-learn, Matplotlib, Searborn, Logistic Regression, Random Forest, ANN, LLM, Langchain, Streamlit, openAI, Pycharm, Tableau, MS Excel, PySpark, Pandas, Dataframes, Databricks, NLP, tokenization, lemmatization and stemming, cross-validation, hyperparameter tuning, sampling technique, SMOTEEN, KNIME, SPSS, Git, multiple linear regression, R-studio, statistical analysis, chi-square, Anaconda, model evaluation

### **Result Based Service - HTC, State Farm, USA, Jan 2019 to Aug 2022    Industry: Insurance**

- Slashed 10% of overall incidents by identifying trends/patterns, recognising key issues and recommending actions on incidents reported for 40+ applications.
- Increased revenue by 5% by recommending compelling proposals for business problems.
- Enhanced 90% of data accuracy and efficiency in reporting by improving the data collection and cleaning process.

- Communicated clearly with senior management and stakeholders to gather and translate business needs for data analysis and dashboard creation.
- Created training plans and provided training sessions on tools and best practices to more than 300+ members, achieving a 100% compliance rate.
- **Technologies:** SQL, data analysis, data visualisation, document and process review, MS Excel, MS Powerpoint, ServiceNow, communication, project management, organisation, agile, presentation, training, software development and maintenance, working under pressure, attention to detail, teamwork

**Database Management, HTC, State Farm, USA,** Jan 2017 to Dec 2018 **Industry:** Insurance

- Optimised 100+ complex SQL queries with joins, subqueries, and stored procedures and developed 100+ database objects, including tables, indexes, and views, increasing application efficiency by 20%.
- Created process documents and trained new team members.
- **Technologies:** SQL, DB2, Cobol, JCL, MS Excel, HPSM, Mainframe, database management, communication, code review

**Data Migration, HTC, State Farm, USA,** Feb 2013 to Dec 2016

**Industry:** Insurance

- Reduced content retrieval time by 40% for 1000+ users, ensuring zero data loss and minimal downtime by orchestrating a seamless migration of a large and intricate content management database and its application using the ETL process.
- **Technologies:** SQL, DB2, ETL, MS Word, MS Excel, Cobol, JCL, Mainframe, data migration, waterfall, software development, teamwork, data engineering, code review, attention to detail, working under pressure

**Supply chain applications, Infosys, PepsiCo Fritolay, USA** Nov 2007 to Jan 2013 **Industry:** Retail

- Reduced 15% of application downtime by proactively collaborating with customers, analysing and resolving data issues for Pepsico Fritolay's supply chain applications across 30+ units and 200+ distribution centres in the USA.
- Conducted training sessions for newly onboarded team members.
- **Technologies:** SQL, DB2, MS Word, MS Excel, Cobol, JCL, Mainframe, HPSM, software maintenance, customer handling

**Financial reporting, Infosys,** Aug 2005 - Apr 2006

**Industry:** Retail

Revolutionised inventory and financial reporting for Hallmark Cards and Sysco by enhancing inventory management efficiency by 60% through predictive SQL-based sales planning and eradicating all data inconsistencies in financial reporting with custom BI solutions