## **Role Exploration Report: Data Scientist**

A **Data Scientist** is a multidisciplinary expert who transforms raw data into actionable insights. In healthcare, they play a pivotal role in building predictive models, analysing clinical trends, and optimizing patient outcomes using AI-driven tools.

## Responsibilities

- Collect, clean, and preprocess large datasets
- Perform exploratory data analysis and statistical modelling
- Build and evaluate machine learning models
- Communicate findings through visualizations and reports
- Collaborate with domain experts to guide decision-making

## **Skills & Tools**

- **Programming**: Python, R, SQL
- Libraries: Pandas, Scikit-learn, TensorFlow, ggplot2
- Visualization: Matplotlib, Seaborn, Power BI
- ML Techniques: Regression, Clustering, Decision Trees, Neural Networks
- **GenAI Tools**: Azure OpenAI, Llama Index, Vector DB, Prompt Engineering
- Soft Skills: Communication, critical thinking, domain knowledge

## **Real Job Post Snapshot**

Below is a real LinkedIn job post from LTI Mindtree for a GenAI Data Scientist role. Five key requirements highlighted:

- 1. Experience developing production-level GenAI solutions using Azure OpenAI
- 2. Familiarity with prompt engineering, RAG, and fine-tuning opensource LLMs

- 3. Strong foundation in ML lifecycle and model governance
- 4. Proficiency in Python, R, and statistical modelling techniques
- 5. Ability to collaborate across teams and present insights using visualization tools







