



## SUMMARY

Result-oriented Data Science Engineer with 2+ years of experience in creating and implementing automated data pipelines and machine learning models. Skilled in processing big data, identifying patterns, and extracting valuable insights. Excellent skills in ETL, DE, ML, Python, SQL, and Excel. Recognized for automating work processes and achieving significant cost reductions and accuracy.

## KEY SKILLS

• Automation • Stored Procedures • Big Data • Data Mining & Analytics • Data Visualization • Price Optimization • Portfolio Pricing • Extract, Transform, Load (ETL) • Predictive Modelling • Web Scraping • EDA • Gen Ai • Hadoop Ecosystem • SSIS packages

## TECHNICAL SKILLS

**Big Data Ecosystem:** Hadoop MR, NoSQL  
**Languages:** Python, SQL, VBA  
**Cloud Computing:** AWS, EC2, S3, EMR  
**Machine Learning:** XgBoost, Scikit-learn, Statsmodels, TensorFlow, NLTK  
**Database:** MySQL, MongoDB, HBase  
**Statistics/ML:** Linear/Logistic Regression, Classification, Ensemble Trees, KNN  
**Tools:** VS SSIS ETL, Advance Excel, Tableau, Visual Studio Code

## CERTIFICATIONS

- Advanced Learning Algorithms Professional Certificate | Deeplearning.ai | May '23
- Building LLM Models | ChatGPT | Fourth Brain | Nov'23

## ADDITIONAL ACHIEVEMENTS

- Felicitated with the **Genius Award** by EXL in May '23 for excellence in **process automation**.

## EDUCATION

Liverpool John Moores University + IIIT Bangalore -  
**M.S. Data Science**

Correspondence/Distance

- Data Engineering Specialization

**BITS PILANI, Pilani Campus - B.E. Civil Engineering**

Full Time

- Secured **7.7CGPA**

**12th Score: 90.6%**

**10th Score: 9.6 CGPA**

## PROFESSIONAL EXPERIENCE

**Data Science Engineer**

May '23 - Dec '24

EXL Services

Gurugram, IN

**Analytics & Machine Learning Methodologies**

- Done extensive data cleaning, manipulation and processing for analytical and pricing models.
- Train, test and validated **volume prediction model**. Implemented **Xgboost/Random Forest** architecture to achieve **MAPE of around 10-12%** on train and test data sets.
- Also, formulated a **price elasticity algorithm model** to predict volume based on current price elasticity trend.
- Optimized prices across product portfolio by developing a customer behavior-based pricing model.
- Implemented sanity checks at various points in the process to ensure data integrity.

**Automation & Algorithm Development**

- Automated all repetitive tasks using **Python** (Xlwings, Pandas, Numpy, Sqlalchemy, win32, NLTK, regex) and **SQL stored procedures** to create **end-to-end data pipelines** and automating final reports.
- Decreased various processes time consumption and saved 40 working hours (70%)/Wk. of resources.

**Chat GPT (Gen AI)**

**Worked on an automated agent based on Chat GPT**

- Made and implemented **GPT prompts** to analyze user data and provide trends and facts. Created web scrapers on python to scrape summary for explaining analysis using latest trends and facts, regarding to user's query.

## KEY DATA PROJECTS

- Domain: Hadoop MR | Tech Stack: AWS, Hadoop, MRJob, Sqoop, Flume, Linux Shell | Jan' 24**
  - Used **Apache Sqoop** and **Apache Hbase** to process data from AWS RDS and Hadoop to NoSQL HBase database.
  - Created **map reduce scripts** using mrjob library to process and aggregate big data stored on Hadoop distributed file system.
- Domain: Lead Classification | Tech Stack: Python, ML | Oct '23**
  - Objective: Classify the lead which have potential to convert in customers on basis of their inputs and choices.
  - Solution: Designed a machine learning model using logistic regression to classify leads with 82% accuracy.