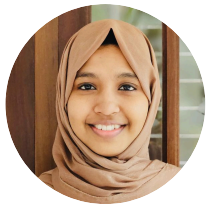
Sharmila Sherin *Data Engineer*

sharmilasherinpattu@gmail.com

+916238345162

linkedin.com/in/sharmila-sherin

github.com/SharmilaSherinRYZ



Career Journey

Passionate about transforming raw data into meaningful insights, I discovered my true calling in data engineering by leveraging the statistical foundations of my mathematics degree. Driven by curiosity and a love for solving complex problems, I am committed to building innovative data solutions that drive business intelligence.

Education

Executive Post Graduate Program in Data Science with Specialization in Data Engineering 12/2023 – 12/2024
Indian Institute of Information Technology (IIIT) Bangalore, India

BSc in Mathematics 07/2015 – 04/2018
NAM College Kallikkandy, Kannur Kannur, India

Technical Skills

- Programming Languages: Python, SQL
- Big Data Ecosystem: Hadoop MapReduce, Apache Spark, Kafka, Sqoop, Flume, NoSQL (HBase)
- Databases: MySQL, Apache HBase
- Cloud Computing: AWS (EC2, S3, EMR, Redshift, Glue, Lambda, Athena)
- Machine Learning: Linear Regression, Logistic Regression, K-Means Clustering
- Visualization Tools: Power BI

Professional Summary

- A data engineering enthusiast with a strong mathematical background, turning complex data into insights.
- Skilled in building efficient data pipelines with tools like Apache Spark and Hadoop.
- Proficient in cloud technologies, leveraging AWS services to design scalable and robust data architectures.
- Experienced in transforming raw data into meaningful business insights through ingestion and visualization.
- Passionate about using statistical methods and programming skills to solve real-world data challenges.

Projects

Data-Driven ATM Performance Analytics

Developed a comprehensive data pipeline to transform raw ATM transaction data into meaningful business insights for Spar Nord Bank. The project enabled detailed analysis of transaction volumes, performance metrics, and operational efficiency across the bank's ATM network. By leveraging advanced data engineering techniques, we provided actionable intelligence to optimize ATM operations and customer service.

Technologies Used:

- Sqoop
- PySpark
- AWS S3
- Amazon Redshift
- MySQL

Role & Responsibilities:

- Designed and implemented end-to-end ETL pipeline for ATM transaction data
- Extracted transaction data from MySQL database using Sqoop
- Performed complex data transformations using PySpark
- Loaded processed data into AWS S3 and Redshift for advanced analytics
- Created data models to enable comprehensive ATM performance reporting
- Developed scalable data processing solution to handle large volumes of transaction data

Musical Insights: Spotify Data Exploration

Created an innovative data pipeline to extract and analyze music data from Spotify's extensive platform. The project transformed raw streaming data into meaningful insights, providing a comprehensive view of music trends, listener behaviors, and content performance. By leveraging cloud technologies, we built a robust system for deep musical data analysis.

Technologies Used:

- Spotify API
- AWS Lambda
- AWS S3
- AWS Glue
- Amazon Athena

Role & Responsibilities:

- Designed and implemented end-to-end ETL pipeline using Spotify API
- Developed serverless data extraction process using AWS Lambda
- Transformed and processed streaming data using AWS Glue
- Created data storage and query infrastructure in AWS S3 and Athena
- Implemented data processing workflows for efficient music data analysis
- Ensured data quality and reliability throughout the pipeline

Languages

English, Hindi, Malayalam