

# Piyush Pandey

[in piyush-pandey03](#) | [✉️ piyushpandey4263@gmail.com](mailto:piyushpandey4263@gmail.com) | [+9311562074](#)

## SUMMARY

Data Analytics and Cloud enthusiast with hands-on experience in Azure, Databricks, and PySpark. Skilled in building data pipelines, preprocessing large datasets, and creating BI dashboards. Strong problem-solving skills with experience across internships and end-to-end data projects.

## PROFESSIONAL EXPERIENCE

### Business Analytics, Qlik (Remote Internship)

April 2024 - June 2024

- Executed data analysis using advanced tools to derive meaningful insights.
- Developed dashboards and visualizations to communicate analytical findings.
- Tackled complex data challenges to enhance problem-solving abilities.

### Data Analyst Intern, Ernst & Young (EY)

November 2025 - January 2026

- Executed Parquet data preprocessing in Databricks using PySpark, ensuring schema consistency, data integrity, and readiness for downstream analytics.
- Built Databricks-managed tables and implemented automated data ingestion into MySQL using Python-based connectors and SQL-driven extraction.
- Tackled complex data challenges to enhance problem-solving abilities.

## PROJECTS

### Enterprise Parquet-to-Analytics Pipeline using Databricks

2025

- Built a PySpark-based data pipeline in Azure Databricks to preprocess and validate Parquet datasets, enforcing schema consistency and data quality checks.
- Created Databricks-managed tables and implemented automated ingestion into MySQL using Python connectors and SQL-based transformations.
- Designed an end-to-end analytics-ready workflow with logging and error handling to support reliable downstream reporting and analytics.

### Databricks Lakehouse with Auto Loader and Delta Live Tables

2025

- Built Azure Databricks lakehouse (bronze/silver/gold) with streaming Auto Loader for idempotent, exactly-once incremental ingestion governed by Unity Catalog.
- Implemented silver PySpark transforms and gold star schema; delivered SCD1 in PySpark and SCD2 via Delta Live Tables with expectations and lineage.
- Orchestrated ETL with Databricks Jobs and published BI-ready datasets via SQL Warehouse; provisioned ADLS Gen2, Workspace, and Access Connector for secure lake access.

## EDUCATION

### VIT Bhopal University

Oct.2022 - Apr.2027

- Integrated M.Tech - Data Science (CGPA:-8.07)

### Kendriya Vidyalaya No.2

May 2021 - Apr 2022

- Class XII(Precentage:79%)

## SKILLS

**Languages:** Python, Java, C++, SQL.

**Data Tools:** Pandas, NumPy, Scikit-learn, Power BI, Pyspark

**Cloud:** Microsoft Azure (ADF, Databricks, ADLS), Git, Azure DevOps