# Cloud-Based Generalized Data Analysis Platform

# STAR

**S**
**Situation**

Traditional data analysis methods were slow and manual, taking 4-5 minutes per dataset, limiting accessibility and efficiency.

**T**
**Task**

Develop a cloud-based platform to streamline the data analysis process, making it faster, more efficient, and accessible to all users.

**A**
**Action**

Utilized Amazon Web Services for data storage and processing, employing Spark and SparkML for rapid descriptive statistics and predictive analysis. Simplified the user experience by enabling CSV file uploads for automatic analysis.

**R**
**Results**

The platform significantly reduced the time required for data analysis, providing descriptive statistics and predictive insights within minutes, and improved accessibility for users with varying technical expertise.

# Leveraged Knowledge

- AWS (S3, Lambda, Glue)
- Spark/SparkML for data processing
- HTML/JavaScript for UI development
- Machine Learning for predictive analysis

# Project Analysis

| | POSITIVE | NEGATIVE |
|---|---|---|
| INTERNAL | STRENGTH<br>› Rapid analysis, user-friendly interface, leverages cloud computing for scalability and efficiency, accessible to non-technical users. | WEAKNESS<br>Dependence on cloud services may raise concerns about data privacy and internet connectivity requirements. |
| EXTERNAL | OPPORTUNITY<br>› Growing demand for data analysis tools across industries, potential to integrate with other cloud services for enhanced functionalities. | THREAT<br>Competition from existing cloud-based analysis platforms, technological advancements may require constant updates to stay relevant. |

# Thank You