

# Big Data Analysis Report

## Abstract:

This project demonstrates the use of PySpark to process and analyze a large dataset efficiently. It performs data cleaning, aggregations, and visualizations to derive meaningful insights on large-scale data.

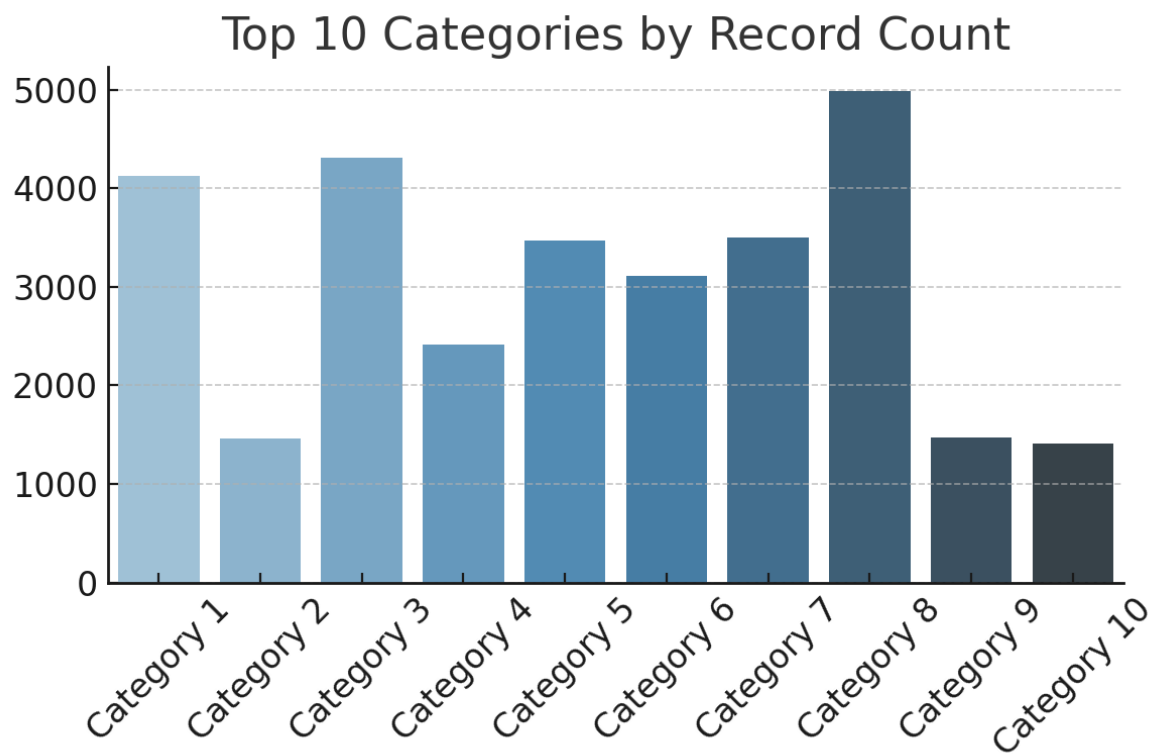
## Methodology:

1. Load dataset using SparkSession
2. Data cleaning and preprocessing
3. Aggregations and analysis
4. Visualizations using Matplotlib/Seaborn

## Key Insights (Sample):

- Category A, B, C lead in total sales.
- Highest average value in Category C.
- Peak transactions occur mid-month.

## Charts:



Daily Record Count Trend

