# Executive Summary

This executive summary outlines the outcomes of a data engineering project analyzing hospital readmissions using the 'diabetic\_data.csv' dataset (~100,000 records). Implemented in PySpark on Google Colab, the objective was to build a scalable data pipeline that enhances patient outcomes and reduces readmission rates.

## Key Objectives

* - Ingest, clean, and transform raw hospital data
* - Identify and analyze patient readmission patterns
* - Produce analytical summaries and visual insights

## Findings

* - Older patients had longer hospital stays.
* - Certain diagnoses (e.g., 250.83 for diabetes) showed high readmission.
* - Patients with rising insulin needs had increased readmission risk.
* - Abnormal A1C levels (especially >8) correlated with higher readmission.

## Recommendations

* - Enhance discharge planning for high-risk diagnoses.
* - Provide targeted follow-up for older patients and those with insulin/A1C concerns.
* - Consider real-time monitoring using extended data pipelines.