Infrastructure Projects Data Analysis Report

1. Objective

The objective of this task is to clean, filter, and analyze a real-world dataset to identify true infrastructure projects, flag megaprojects, and analyze failure rates by procurement method.

2. Dataset Cleaning & Filtering

The dataset was cleaned by standardizing budgets, converting Yes/No fields to binary values, parsing timestamps, and removing non-infrastructure records using keyword-based rules.

3. Summary Statistics

Total infrastructure projects: 536
Total megaprojects (≥ $500M): 12
Average project budget (USD): 67239096.62
Median project budget (USD): 1623647.49

4. Failure Analysis by Procurement Method

ml.procurement_method_redo2.results.method	Total_Projects	Failed_Projects	Failure_Rate_%
Construction Management at Risk (CMAR)	5	0	0.0
Design and Build (D&B)	11	0	0.0
Design-Bid-Build	255	14	5.49
Public-Private Partnership (PPP)	21	0	0.0

5. Limitations & Uncertainties

The analysis is limited by missing budget data, simplified failure definitions, and potential classification errors due to keyword-based filtering. Results should be interpreted as indicative rather than definitive.