

# Project: Equipment Productivity Tracker

---

## Project Overview

Construction equipment plays a critical role in site operations—but its productivity depends on many factors, including weather, operator skill, crew support, and site conditions. In this project, you will analyze a **synthetic dataset** that simulates field performance for a variety of heavy equipment types (e.g., excavators, cranes, bulldozers). Your goal is to investigate how different input conditions affect **daily productivity**, **fuel usage**, and **operating delays**, and to surface patterns that could help improve efficiency on real construction sites.

## Objective

You will work with a **synthetic dataset** that simulates equipment operation logs from different construction sites. Each entry includes inputs such as **equipment type**, **operator experience**, **crew size**, and **weather conditions**, and tracks performance metrics such as **volume moved**, **fuel usage**, and **number of delays**. Your goal is to identify which factors most impact productivity and whether any patterns emerge across different site conditions or machine types.

## Major Questions for Your Proposal

### 1. Focus Variables & Hypotheses

Which variables do you expect will have the **strongest influence** on daily productivity (volume moved)? How might material type or operator experience play a role? Are there any variable combinations you're especially curious about?

### 2. Exploration Strategy

How do you plan to **analyze relationships** between inputs and outputs (e.g., correlations, group comparisons, basic regression)? Will you focus on one specific equipment type or explore all types together?

### 3. Visualization Plan

What plots or graphics will you use to reveal patterns in the data? Do you expect to show relationships between variables (e.g., operator experience vs. fuel efficiency), categorical breakdowns (e.g., material type vs. delays), or comparisons across equipment types?

**Note:** Your answers don't lock you into a specific direction—this just helps guide your thinking and gives you a solid starting point.

## Project Definitions

### Input Variables

- **equipment\_type**: The type of heavy equipment used (e.g., excavator, crane, loader, bulldozer).
- **equipment\_age\_years**: Number of years since the equipment was manufactured or placed in service.
- **maintenance\_score**: A numeric rating from 1 to 10 representing the condition and upkeep of the equipment. Higher values mean better maintenance.

- **operator\_experience\_years**: Number of years the primary operator has been running this type of equipment.
- **crew\_size**: Number of workers assigned to assist the equipment operator (excluding the operator).
- **material\_type**: The material being handled (e.g., clay, sand, gravel, rebar).
- **workday\_weather**: Overall weather for the day (sunny, rainy, windy, etc.).
- **soil\_condition**: Site conditions where applicable (e.g., loose, compacted, wet).
- **daily\_distance\_moved\_m**: Total horizontal distance (in meters) the equipment moved during the day.
- **load\_cycles\_per\_day**: Number of complete cycles the equipment performed (e.g., scoop + dump, lift + place).

## Output Variables

- **volume\_moved\_per\_day\_cubic\_m**: Total material volume moved during the workday.
- **fuel\_used\_liters**: Amount of fuel consumed during operation.
- **operating\_hours**: Total machine run-time (excluding idle time).
- **number\_of\_delays**: Count of delays during operation due to mechanical, coordination, or external issues.

## Data Source

A **synthetic CSV dataset** will be provided containing simulated logs from various equipment types under varying site conditions. These values are generated using simplified models to reflect realistic field variation while avoiding real-world data complexity.

## Project Requirements

### 1. Data Cleaning and Setup

- Load the dataset and inspect the structure and types of each variable.
- Check for missing or anomalous values (e.g., zero operating hours, negative fuel use).
- Organize the data for analysis (e.g., filter by equipment type, create summary stats by category).

### 2. Analysis

- Investigate the impact of different inputs (e.g., weather, material type, experience) on key outputs like **volume moved** or **fuel used**.
- Look for strong predictors of **delays** or **low productivity** (e.g., is rain a common cause of downtime?).
- Identify any combinations (e.g., low crew size and high material resistance) that lead to inefficiency.

### 3. Visualization

- Create clear, compelling visualizations that help explain your findings. Some ideas:
  - Scatter plots (e.g., operator experience vs. fuel used)
  - Box plots or bar charts (e.g., productivity by material type)
  - Line or stacked plots if comparing multiple equipment types

#### 4. Discussion & Reflection

- Summarize the **most influential factors** on productivity or delays.
- Comment on any **unexpected findings** or insights.
- Suggest **strategies or considerations** for improving equipment efficiency on real jobsites based on your analysis.