

Project: Schedule Delay Risk Analyzer

Project Overview

Construction schedules are full of uncertainty—from subcontractor issues to weather delays and coordination challenges. In this project, you'll analyze a **synthetic dataset** that simulates construction activity timelines under different conditions. Your task is to identify patterns in delays, explore risk factors, and evaluate which project activities or conditions are most prone to timeline overruns.

Objective

You will work with a **synthetic dataset** containing planned and actual durations for key construction activities across projects of different sizes. Each entry includes risk factors like **weather impacts**, **subcontractor availability**, and **project size**, allowing you to explore how each of these elements influences the likelihood and severity of delays.

Major Questions for Your Proposal

1. **Delay Drivers**

Which inputs (e.g., weather impact, subcontractor availability, project size) do you expect to be the **strongest predictors** of delay? Which activities do you think are most at risk?

2. **Analysis Strategy**

How will you identify patterns in delays? Will you compare groups (e.g., activities under "very late" subcontractor conditions)? Will you build a simple classifier or try to visualize delay trends across activity types?

3. **Visualization Plan**

What graphs or plots will best support your investigation? Will you show distributions of delay days, frequency of delay reasons by activity type, or project size breakdowns?

Note: Your proposal serves as a starting point—you can refine your strategy as your analysis progresses.

Project Definitions

Input Variables

- **activity_type:** Type of construction activity (e.g., foundation, framing, roofing, electrical, plumbing)
- **planned_duration_days:** The number of days allocated to complete the activity
- **weather_impact:** Degree to which weather affected the activity timeline (none, moderate, high)
- **subcontractor_availability:** Availability status of the subcontractor assigned to the task (on-time, late, very late)
- **project_size_category:** The size category of the overall construction project (small, medium, large)
- **crew_size:** Number of workers assigned to the activity
- **material_delivery_delay_days:** Number of days that material delivery was delayed beyond the planned date

- **inspection_required**: Whether the activity required a formal inspection before it could be closed out (yes or no)
- **activity_complexity**: Subjective rating of the complexity or coordination difficulty of the activity (low, medium, high)

Output Variables

- **actual_duration_days**: The actual number of days the activity took to complete
- **delay_days**: The number of days the activity exceeded its planned duration (0 if completed on time or early)
- **delay_reason**: The primary reason for the delay (weather, labor, material, unknown)

Data Source

You will be provided with a **synthetic CSV file** representing simulated activity logs across a variety of projects. The dataset includes a blend of categorical and numerical data, with relationships built in to reflect realistic delay trends based on weather, subcontractor performance, and activity complexity.

Project Requirements

1. Data Cleaning and Setup

- Load the dataset and inspect for missing values or inconsistencies.
- Confirm that delay calculations (actual – planned) make sense and are non-negative.
- Group and summarize data to prepare for analysis by activity type, project size, or delay reason.

2. Analysis

- Explore the frequency and severity of delays across different activity types.
- Investigate how inputs like weather impact, subcontractor availability, and project size affect actual durations.
- Look for patterns in delay reasons—are some more common for specific project types or sizes?

3. Visualization

- Use charts to reveal delay patterns, such as:
 - Bar charts showing average delay days per activity type
 - Stacked bars or heatmaps for delay reasons by project size
 - Box plots comparing planned vs. actual durations across categories
 - Scatter plots showing relationships between inputs and delay days

4. Discussion & Reflection

- Summarize your findings: which factors most influenced delays?
- Were your expectations from the proposal accurate?
- What recommendations would you give a project manager to mitigate these kinds of risks?