

## CONCLUSIONES GENERALES DEL PROYECTO

Elaborado por:  
Holman Sanchez  
Esteban Oliveros  
Jesús Córdoba

Con la base contratos–facturas se construyó la variable porcentaje (pagos/valor estimado $\times$ 100) y se depuraron fechas, NA y un outlier extremo. Tras la limpieza ( $n=581$ ), la mediana=100% y la media $\approx$ 85% evidencian ejecuciones completas pero con alta dispersión (0%–273%).

En el análisis bivariado, valor\_contrato y dias\_fin\_ejec muestran relación positiva moderada ( $r\approx 0.31$ ), mientras que la bandera de sobrecosto no guarda relación relevante con ninguna de ellas (correlaciones  $\sim 0$ ). Existen diferencias territoriales en la variabilidad de porcentaje. Para predecir sobrecostos  $>30\%$  se requieren variables de proceso y gobernanza no presentes en el dataset actual.

### 1) Calidad y preparación de datos

1. La creación de porcentaje al cruzar contratos con facturas permitió medir la ejecución financiera real y detectar sub/sobre ejecuciones.
2. El tratamiento de fechas y NA, y la remoción de un outlier por criterio de negocio, mejoró la confiabilidad estadística y visual.
3. El dataset mantiene pocas variables numéricas y de proceso, lo que limita la capacidad predictiva sobre sobrecostos.

### 2) Comportamiento univariado

4. **Distribución de porcentaje:** mediana en 100%, media 85%, con colas superiores  $>200\%$ ; coexisten subejecuciones ( $\approx 0\%$ ) y sobreejecuciones ( $>100\%$ ).
5. La alta dispersión ( $DE\approx 44$  pp) sugiere heterogeneidad operativa entre proyectos y posibles diferencias en gestión o registro.

### 3) Relaciones bivariadas

6. **Costo–tiempo:** valor\_contrato vs dias\_fin\_ejec exhibe tendencia positiva moderada ( $r\approx 0.31$ ). Proyectos más costosos tienden a durar más, pero con gran variabilidad: el valor no determina por sí solo la duración.
7. **Sobrecostos (bandera):** no se observan asociaciones significativas con valor\_contrato ( $r\approx -0.03$ ) ni con dias\_fin\_ejec ( $r\approx -0.06$ ). Los sobrecostos aparecen transversalmente en distintos tamaños y duraciones.
8. **Territorio:** los boxplots de porcentaje por municipio\_entidad muestran dispersión muy alta en municipios como Tuluá y Palmira, y mayor estabilidad en otros (p. ej., Andalucía, Calima). Esto indica brechas territoriales de ejecución.

### 4) Implicaciones para gestión y control

9. El riesgo de sobrecostos no se explica por el tamaño ni por el tiempo del contrato; es más probable que dependa de cómo se contrata y ejecuta (modalidad, interventoría, cambios de alcance, desempeño del contratista, gobernanza local).
10. La variabilidad territorial sugiere focalizar vigilancia y acompañamiento técnico en municipios con mayor dispersión de porcentaje.

#### 5) Limitaciones

11. **Información insuficiente** para un modelo predictivo robusto de bandera >30%: faltan variables de proceso, actor, régimen, hitos y cambios contractuales.
12. Posibles sesgos de medición (registros atípicos, heterogeneidad en reportes) y tamaño muestral acotado tras la depuración.

#### 6) Recomendaciones y próximos pasos

13. Enriquecer el dataset con: modalidad de contratación, sector, nivel\_entidad, tipo/alcance de interventoría, ID del contratista, adiciones y prórrogas, hitos y pagos parciales, fuente de recursos, año/mes, y variables territoriales (capacidad institucional, complejidad de obra).
14. **Análítica adicional:**
  - Correlación **Spearman** y **regresiones** (lineal log-log y cuantílica) para costo-tiempo.
  - **ANOVA** de porcentaje por municipio/sector.
  - **Modelos logísticos/árboles** para bandera con variables de proceso y territorio.
  - **Detección de anomalías** para pagos y variaciones contractuales.
15. **Gobernanza de datos:** políticas de estandarización de campos, reglas de outliers reproducibles y trazabilidad del cruce contratos-facturas.

**Conclusión final.** El proyecto establece una línea base: costo y tiempo se relacionan moderadamente, pero no explican los sobrecostos; estos requieren información de proceso y contexto. Con datos enriquecidos y pruebas adicionales, puede construirse un sistema de vigilancia predictiva que priorice auditorías por municipio, modalidad y sector, elevando la transparencia y la efectividad en la ejecución de obras públicas.