

ANÁLISIS DE PROCESOS DE CONTRATACIÓN: SOBRECOSTOS Y SUS CAUSAS.



"Análisis de Procesos de Contratación para el Mantenimiento de Infraestructura en la Dirección Ejecutiva de la Administración Judicial Seccional Medellín: Sobrecostos y sus Causas"

Daniel Otálvaro Gallego

Corporación Universitaria Minuto de Dios

Rectoría Virtual

Programa Especialización en Gerencia de Proyectos

Diciembre de 2024

ANÁLISIS DE PROCESOS DE CONTRATACIÓN: SOBRECOSTOS Y SUS CAUSAS.

"Análisis de Procesos de Contratación para el Mantenimiento de Infraestructura en la Dirección Ejecutiva de la Administración Judicial Seccional Medellín: Sobrecostos y sus Causas"

Daniel Otálvaro Gallego

Trabajo de Grado presentado como requisito para optar al título de Especialista en Gerencia de Proyectos

Asesora
Doris Amanda Rosero García
Microbióloga, M.Sc., PhD.
Posdoctorado en Microbiología Ambiental

Corporación Universitaria Minuto de Dios
Rectoría Virtual
Programa Especialización en Gerencia de Proyectos
Diciembre de 2024

ANÁLISIS DE PROCESOS DE CONTRATACIÓN: SOBRECOSTOS Y SUS CAUSAS.

Contenido

Lista de tablas	5
Lista de figuras	6
Lista de anexos.....	7
Resumen	8
Abstract.....	9
Introducción.....	10
1 PLANTEAMIENTO DEL PROBLEMA.....	12
1.1 Descripción del problema.....	12
1.2 La pregunta de investigación.....	13
1.3 Los objetivos de investigación	13
1.3.1 Objetivo general.....	13
1.3.2 Objetivos específicos.....	13
1.4 Justificación de la investigación.....	14
2 MARCO DE REFERENCIA.....	16
2.1 Marco de Antecedentes.....	16
2.2 Marco Teórico	17
2.3. Marco normativo	17
3 METODOLOGÍA	18
3.1 Enfoque y alcance de la investigación	18
3.2 Población y muestra.....	19
3.2.1 Definición de la población.....	19
3.2.2 Cálculo y selección de la muestra	19
3.3 Instrumento(s)	20
3.3.1 Fuente de datos	20
3.3.2 Variables de interés y codificación	20
3.4 Descripción de procedimientos	21
3.5 Análisis de información.....	21
3.5.1 Herramientas de Software.....	21

ANÁLISIS DE PROCESOS DE CONTRATACIÓN: SOBRECOSTOS Y SUS CAUSAS.

3.5.2	Análisis Estadístico	22
3.5.3	Planeamiento del Análisis estadístico.....	22
3.6	Consideraciones éticas.....	22
3.6.1	Análisis de consideraciones éticas	22
3.6.2	Instrumentos de aceptación y autorización	22
4	HIPÓTESIS.....	23
4.1	Las variables	23
4.1.1	Variable(s) independiente(s)	23
4.1.2	Variable(s) dependiente(s).....	23
4.2	Planteamiento de hipótesis	23
5	RESULTADOS	24
5.1	Recolección de datos	24
5.1.1	Datos obtenidos	24
5.2	Análisis de datos	26
5.2.1	Análisis estadístico	26
5.2.2	Análisis de correlación	27
5.2.3	Análisis de regresión a través de técnicas de Machine Learning	30
6	DISCUSIÓN	33
7	CONCLUSIONES.....	35
8	RECOMENDACIONES.....	36
	Referencias.....	37
	Anexos.....	39

ANÁLISIS DE PROCESOS DE CONTRATACIÓN: SOBRECOSTOS Y SUS CAUSAS.

Lista de tablas

Tabla 1. Tabla Información Contratos..... 24

ANÁLISIS DE PROCESOS DE CONTRATACIÓN: SOBRECOSTOS Y SUS CAUSAS.

Lista de figuras

Figura 1 Estimación de la función de densidad de probabilidad del porcentaje de sobrecosto	26
Figura 2 Descripción estadística de los datos extraídos, el sufijo _mp indica que el valor se encuentra expresado en millones de pesos	27
Figura 3. Mapa de calor de los diferentes descriptores en el set de datos.	28
Figura 4 Scatter plot de los diferentes descriptores del set de datos.	29
Figura 5 Gráfico observado vs predicho del porcentaje de sobrecosto usando regresión lineal	30
Figura 6. Gráfico observado vs predicho del porcentaje de sobrecosto usando árbol de decisión (overfitting)	31
Figura 7. Gráfico observado vs predicho del porcentaje de sobrecosto usando bosques aleatorios.	32

ANÁLISIS DE PROCESOS DE CONTRATACIÓN: SOBRECOSTOS Y SUS CAUSAS.

Lista de anexos

Anexo 1 39

ANÁLISIS DE PROCESOS DE CONTRATACIÓN: SOBRECOSTOS Y SUS CAUSAS.

Resumen

Este estudio analizó la problemática de los sobrecostos y atrasos en la gestión de proyectos de infraestructura, enfocándose en los proyectos de mantenimiento realizados por la Dirección Ejecutiva de Administración Judicial Seccional Medellín. Mediante un enfoque mixto que integró análisis cuantitativo y cualitativo, se identificaron las causas subyacentes de estas desviaciones, evaluando variables clave como el valor inicial de los contratos, las prórrogas en los plazos y la calidad del servicio entregado. Se encontró una correlación significativa entre estas variables y los sobrecostos observados, lo que permitió comprender mejor las dinámicas que afectan la eficiencia de los proyectos. Paralelamente, se aplicaron técnicas avanzadas de machine learning para modelar y predecir los sobrecostos en función de los descriptores evaluados. Los resultados indicaron que los bosques aleatorios fueron la herramienta más eficaz, con un coeficiente de determinación (R^2) de 0.78, demostrando su capacidad para capturar relaciones no lineales y patrones complejos en los datos. Sin embargo, el tamaño reducido de la muestra representó una limitación que restringe la generalización de los hallazgos a otros contextos. El estudio ofrece una base sólida para optimizar la gestión de proyectos futuros, reduciendo desviaciones presupuestales y mejorando la asignación de recursos. Se plantearon recomendaciones orientadas a perfeccionar los procesos de contratación, fortalecer la supervisión y emplear modelos predictivos como herramientas estratégicas en la toma de decisiones.

Palabras clave: sobrecostos, mantenimiento de infraestructura, contratación pública, machine learning, bosques aleatorios.

ANÁLISIS DE PROCESOS DE CONTRATACIÓN: SOBRECOSTOS Y SUS CAUSAS.

Abstract

This study analyzed the issue of cost overruns and delays in infrastructure project management, focusing on maintenance projects carried out by the Executive Directorate of Judicial Administration, Medellín Section. Through a mixed-methods approach that integrated quantitative and qualitative analysis, the underlying causes of these deviations were identified, evaluating key variables such as the initial value of contracts, extensions of deadlines, and the quality of the delivered service. A significant correlation was found between these variables and the observed cost overruns, providing a better understanding of the dynamics affecting project efficiency. Additionally, advanced machine learning techniques were applied to model and predict cost overruns based on the evaluated descriptors. The results indicated that random forests were the most effective tool, achieving a coefficient of determination (R^2) of 0.78, demonstrating their ability to capture non-linear relationships and complex patterns in the data. However, the small sample size represented a limitation, restricting the generalizability of the findings to other contexts. The study provides a solid foundation for optimizing the management of future projects, reducing budget deviations, and improving resource allocation. Recommendations were proposed to enhance contracting processes, strengthen oversight, and use predictive models as strategic tools in decision-making.

Keywords: cost overruns, infrastructure maintenance, public procurement, machine learning, random forests.

Introducción

La Dirección Ejecutiva de Administración Judicial Seccional Medellín, como toda organización de alto nivel, se enfrentaba periódicamente a la necesidad de realizar mantenimiento a su infraestructura. La adecuada conservación y mejora de los edificios e instalaciones judiciales resultaba fundamental para garantizar un servicio eficiente y de calidad a la ciudadanía (Guatavita & Montoya, 2024). Estos procesos de mantenimiento y renovación eran cruciales para asegurar que las actividades judiciales se desarrollaran en entornos seguros y funcionales, facilitando el acceso a la justicia (William et al., 2019).

En las instituciones públicas, particularmente las judiciales, la transparencia y eficiencia en el uso de los recursos eran esenciales (Grau, 2006). Surgió la preocupación por los sobrecostos en los proyectos de mantenimiento y mejora de infraestructura (Medina, 2023). Estos sobrecostos representaban un desvío financiero que afectaba el presupuesto disponible, derivando también en retrasos significativos y una percepción negativa por parte de la ciudadanía y otros actores interesados (El Colombiano, 2020).

Estos desafíos impactaban la gestión financiera y tenían repercusiones operativas y reputacionales (Consejo Superior de la Judicatura, 2015). La responsabilidad de asegurar el adecuado desarrollo de las actividades judiciales recaía en gran medida en la capacidad de la entidad para gestionar eficientemente sus proyectos de infraestructura (Álvarez, 2020). Por lo tanto, era crucial implementar un análisis riguroso que permitiera identificar y abordar las causas de los sobrecostos y retrasos (Urrunaga Pascó-Font, 2009).

La presente investigación tuvo como objetivo principal realizar un análisis exhaustivo de los procesos de contratación y mantenimiento de infraestructura en la Dirección Ejecutiva de

Administración Judicial Seccional Medellín. A través de este análisis, se buscó identificar las causas de los sobrecostos y desarrollar estrategias que ayudaran a mitigarlos (Cadavid Rojas & Almanza Castillo, 2021). Al mejorar la gestión de estos proyectos, se esperaba optimizar los recursos y reducir los tiempos de entrega (Díaz, 2019).

El contexto económico y social en el que se desarrollaron estos proyectos también influyó. La inflación, las fluctuaciones en los costos de materiales y mano de obra, así como los factores socio-políticos, afectaron los presupuestos y cronogramas previstos (Pegnalver, 2013). La falta de una planificación y previsión adecuadas para estos factores externos resultó en ajustes y modificaciones durante la ejecución del proyecto, incrementando los costos y (Rosario Villarreal et al., 2019).

La coordinación y comunicación entre departamentos y actores involucrados fue esencial (Consejo Superior de la Judicatura, 2015). Los proyectos de infraestructura generalmente requirieron la colaboración de múltiples áreas dentro de la institución, así como la interacción con contratistas externos, proveedores y entidades reguladoras (Prats & Jacobs, 2014). Una gestión ineficaz de estas interacciones condujo a malentendidos, duplicación de esfuerzos y errores que afectaron negativamente el desarrollo del proyecto (Gaviria Gaviria, 2013).

Por lo tanto, este estudio no solo se centró en identificar y analizar las causas internas de los sobrecostos, sino también en entender cómo los factores externos y las dinámicas organizacionales contribuyeron a estos problemas. Al hacerlo, se proporcionó una visión integral que permitió desarrollar soluciones más efectivas y sostenibles para mejorar la gestión de los proyectos de infraestructura en la Dirección Ejecutiva de Administración Judicial Seccional Medellín (Sánchez I. 2007)

1 PLANTEAMIENTO DEL PROBLEMA

1.1 Descripción del problema

Colombia sufre un problema recurrente en sobrecostos de obras públicas. Al 25 de febrero de 2023, la Procuraduría emitió una alerta por sobrecostos que rondaban los 22 mil millones de pesos en la construcción de unas vías y un mirador turístico en Ciudad Bolívar, Bogotá (Medina, 2023). Noticias como esta se ven todos los días. Según la Cámara Colombiana de Infraestructura, en 2015, el 73% de los proyectos (unos 494 a la fecha) presentaban atrasos y se estimaba que para 2019 cada uno de estos proyectos terminaría costando un 20% más de lo presupuestado inicialmente (El Colombiano, 2020). Sin embargo, hay contratos que han tenido hasta un 100% de sobrecostos (El Colombiano, 2020).

En ambos casos, se mencionan como posibles causas de estos sobrecostos y atrasos la mala estructuración del proyecto, ya que no cuentan con los diseños y estudios iniciales necesarios que permitan plantear de forma correcta el problema y su solución (Cadavid Rojas & Almanza Castillo, 2021). Esto lleva a que, posteriormente, en la ejecución del proyecto se encuentren inconsistencias que generan sobrecostos y atrasos en el mismo (Pegnalver, 2013).

La Dirección Ejecutiva de Administración Judicial Seccional Medellín tiene como responsabilidad primordial asegurar el adecuado desarrollo de las actividades judiciales, garantizando la accesibilidad de toda la población a estos servicios a través de espacios e infraestructura adecuada (Consejo Superior de la Judicatura, 2015). Como parte de sus funciones, la entidad se encarga de la supervisión y ejecución de proyectos de infraestructura en todas las sedes adscritas, en estricto cumplimiento de la normativa vigente (Álvarez, 2020). Sin embargo, la Dirección Ejecutiva enfrenta desafíos significativos que impactan la eficiencia y efectividad en la ejecución de estos proyectos. Entre los desafíos más destacados se encuentran los sobrecostos y los retrasos en la entrega, los cuales generan repercusiones negativas en términos financieros, operativos y reputacionales para la entidad (Urrunaga Pascó-Font, 2009).

La identificación de las causas subyacentes de estos problemas se presenta como una necesidad prioritaria para la mejora continua de la gestión de proyectos de infraestructura en la Dirección Ejecutiva de Administración Judicial Seccional Medellín (Grau, 2006). Para mitigar los posibles sobrecostos y atrasos, surge la necesidad de realizar una investigación rigurosa que identifique y analice en profundidad las causas de los mismos en la entrega de proyectos de infraestructura en la Dirección Ejecutiva de Administración Judicial Seccional Medellín (William et al., 2019). Esta investigación pretende no solo mejorar la gestión de proyectos en la entidad, sino que también podría proporcionar insights y recomendaciones útiles para otras organizaciones que enfrentan desafíos similares en la gestión de su infraestructura (Consejo Superior de la Judicatura, 2015).

1.2 La pregunta de investigación

¿Es posible reducir los sobrecostos y tiempos en los proyectos de infraestructura de la Dirección Ejecutiva de Administración Judicial Seccional Medellín mediante un análisis riguroso de estos en la etapa de estructuración del proyecto?

1.3 Los objetivos de investigación

1.3.1 Objetivo general

Analizar las causas de los sobrecostos y atrasos en los proyectos de mantenimiento de infraestructura en la Dirección Ejecutiva de Administración Judicial Seccional Medellín, con el fin de desarrollar un modelo que permita describir y predecir los posibles sobrecostos de proyectos futuros y con esto mejorar la eficiencia la gestión de los mismos.

1.3.2 Objetivos específicos

Examinar los procesos de contratación y mantenimiento de infraestructura actualmente utilizados por la Dirección Ejecutiva de Administración Judicial Seccional Medellín, con el fin de variables de interés y describir los sobrecostos en los mismos.

Identificar la relevancia de cada una de las variables de interés con respecto a su capacidad de describir el sobrecosto de un proyecto de infraestructura.

Evaluar modelos predictivos que permitan predecir el posible sobrecosto de un proyecto de infraestructura con el fin de hacer una detección temprana del mismo y que se tomen acciones oportunas para minimizar el mismo.

1.4 Justificación de la investigación

La investigación presentada tiene como objetivo analizar en profundidad los procesos de contratación para el mantenimiento de infraestructura en la Dirección Ejecutiva de la Administración Judicial Seccional Medellín, con un enfoque específico en los sobrecostos y sus causas subyacentes. Este análisis es fundamental no solo para mejorar la gestión de recursos y la transparencia en los proyectos de infraestructura, sino también para fortalecer la eficiencia operativa y la calidad del servicio público en el sector judicial.

En el contexto actual, la gestión de proyectos de infraestructura en el ámbito público enfrenta desafíos significativos, especialmente en términos de sobrecostos y demoras. Estos problemas no solo afectan el presupuesto de las instituciones encargadas, sino que también tienen repercusiones negativas en la percepción pública de la eficiencia del Estado, lo que deteriora la confianza de los ciudadanos en las instituciones gubernamentales. En particular, la Dirección Ejecutiva de la Administración Judicial Seccional Medellín se enfrenta a este tipo de dificultades, lo cual pone en riesgo la sostenibilidad y el buen funcionamiento de la infraestructura judicial, fundamental para el acceso a la justicia.

El estudio de los sobrecostos en los proyectos de infraestructura es crucial, ya que permite identificar no solo las deficiencias en la planificación y ejecución de los contratos, sino también los factores externos que pueden influir en los costos, como la inflación, fluctuaciones en los costos de materiales y mano de obra, y cambios en la normativa. Al abordar las causas de los sobrecostos, la investigación proporcionará recomendaciones basadas en evidencia para mitigar estos problemas, optimizando el uso de los recursos públicos y reduciendo los tiempos de ejecución.

Además, esta investigación contribuye al desarrollo de buenas prácticas en la gestión de proyectos de infraestructura, un campo que aún presenta desafíos en términos de eficiencia en muchas instituciones del sector público. La adopción de estrategias eficaces para prevenir sobrecostos podría servir de modelo para otras entidades gubernamentales, no solo en Colombia, sino también en países de América Latina que enfrentan problemáticas similares en sus proyectos de infraestructura.

El impacto de esta investigación será relevante para diversos actores clave. Primero, beneficiará directamente a la Dirección Ejecutiva de Administración Judicial Seccional Medellín, al proporcionar un análisis detallado de sus procesos de contratación y estrategias para mejorar la eficiencia en la gestión de sus proyectos de infraestructura. Esto permitirá una asignación más adecuada de los recursos y una gestión más efectiva de los tiempos, garantizando así la entrega de proyectos en condiciones óptimas.

En segundo lugar, la comunidad se verá beneficiada, ya que una gestión más eficiente y menos costosa de los proyectos de infraestructura judicial resultará en un mejor servicio público, asegurando que los ciudadanos puedan acceder a espacios adecuados para realizar sus trámites judiciales sin demoras ni problemas estructurales. De esta forma, se contribuye al fortalecimiento del acceso a la justicia y al bienestar de la población en general.

Finalmente, esta investigación será útil para los estudiantes y profesionales en el campo de la administración pública y la gestión de proyectos, ya que les proporcionará herramientas y enfoques prácticos para abordar problemas de eficiencia y gestión de recursos en proyectos de infraestructura pública. Este estudio también contribuirá al cuerpo académico existente en la gerencia de proyectos, especialmente en el sector público, y ayudará a posicionar la investigación como una herramienta valiosa para la toma de decisiones informadas.

2 MARCO DE REFERENCIA

2.1 Marco de Antecedentes

Los sobrecostos y retrasos en proyectos de infraestructura han sido problemas recurrentes en el ámbito de la administración pública en Colombia y otros países de América Latina. Estos inconvenientes han generado un impacto significativo tanto en la asignación de recursos públicos como en la percepción de la eficiencia y transparencia de las instituciones gubernamentales. Estudios previos destacan que hasta el 73% de los proyectos de infraestructura pública en Colombia han enfrentado retrasos, y se estima que un porcentaje significativo de estos ha registrado sobrecostos de hasta el 20% o más con respecto a los presupuestos iniciales (El Colombiano, 2020). Sin embargo, en algunos casos, los sobrecostos pueden superar el 100%, evidenciando problemas estructurales en la planeación, adjudicación y ejecución de estos proyectos.

Diversos factores han sido identificados como posibles causas de los sobrecostos en proyectos de infraestructura pública. Por un lado, la falta de diseños y estudios técnicos adecuados durante la etapa de estructuración ha sido señalada como una causa clave, ya que dificulta la correcta definición de los requerimientos técnicos y financieros del proyecto (Cadavid Rojas & Almanza Castillo, 2021). Por otro lado, la ineficiencia en los procesos de contratación, la limitada supervisión técnica, y la falta de coordinación entre las diferentes entidades involucradas contribuyen significativamente a los retrasos y sobrecostos (Urrunaga Pascó-Font, 2009).

En el contexto colombiano, el impacto de estos problemas se ha intensificado debido a factores externos como la inflación y las fluctuaciones en los precios de los materiales y la mano de obra, que afectan directamente los costos estimados de los proyectos (Pegnalver, 2013). Asimismo, la implementación de normativas y regulaciones poco flexibles ha generado desafíos adicionales, dificultando la capacidad de las instituciones para adaptarse a las condiciones cambiantes del mercado y los requerimientos específicos de cada proyecto.

En la Dirección Ejecutiva de la Administración Judicial Seccional Medellín, los desafíos asociados con los sobrecostos y retrasos en proyectos de mantenimiento de infraestructura han sido evidentes. La necesidad de garantizar espacios funcionales y seguros para el desarrollo de actividades judiciales requiere una gestión eficiente de los recursos disponibles. Sin embargo, la falta de planeación adecuada, combinada con problemas administrativos y técnicos, ha resultado en costos adicionales significativos y en una percepción negativa por parte de la ciudadanía (Consejo Superior de la Judicatura, 2015).

A nivel global, investigaciones han destacado la importancia de adoptar estrategias basadas en la gestión de riesgos y la optimización de procesos para mitigar los sobrecostos en proyectos de infraestructura pública (Kerzner, 2019). En este sentido, el uso de herramientas avanzadas de análisis de datos y metodologías como el machine learning han demostrado ser útiles para identificar patrones y predecir riesgos, permitiendo la toma de decisiones más informadas y oportunas (Vela Moreno &

Luévanos Rojas, 2020). Estas estrategias aún no han sido aplicadas ampliamente en el contexto colombiano, lo que representa una oportunidad significativa para la innovación y el desarrollo de mejores prácticas.

Este marco de antecedentes resalta la necesidad de realizar un análisis integral de los procesos de contratación y ejecución de proyectos en la Dirección Ejecutiva de la Administración Judicial Seccional Medellín. Al identificar y abordar las causas de los sobrecostos y retrasos, esta investigación contribuirá al desarrollo de estrategias efectivas para mejorar la eficiencia y transparencia en la gestión de infraestructura pública, beneficiando tanto a la institución como a la comunidad en general.

2.2 Marco Teórico

Teoría de la Gestión de Proyectos: Proporciona el marco para planificar, ejecutar y controlar proyectos de manera eficiente y efectiva. Una gestión adecuada de los proyectos implica el uso de metodologías y prácticas estandarizadas que aseguren la calidad y el cumplimiento de los objetivos (PMI, 2017).

Teoría de la Gestión del Riesgo: Enfocada en identificar y mitigar los riesgos asociados con los sobrecostos y retrasos en proyectos de infraestructura. Una gestión proactiva de los riesgos es fundamental para minimizar su impacto en el proyecto (Klojčnik et al., 2018).

Teoría de la Eficiencia en la Administración Pública: Centrada en la utilización eficiente y transparente de los recursos públicos. La innovación y la adopción de prácticas empresariales pueden mejorar la eficiencia en el sector público (Hitchner, 1992).

2.3. Marco normativo

El marco normativo de esta investigación se basa en las regulaciones y normativas vigentes relacionadas con la contratación y ejecución de proyectos de infraestructura en el sector público. Estas normativas establecen los lineamientos y procedimientos que deben seguirse para garantizar la transparencia y eficiencia en la gestión de proyectos (Álvarez, 2020).

3 METODOLOGÍA

3.1 Enfoque y alcance de la investigación

El enfoque adoptado para esta investigación es **mixto**, combinando metodologías cuantitativas y cualitativas. Este enfoque permite abordar de manera integral las causas de los sobrecostos y retrasos en los proyectos de mantenimiento de infraestructura de la Dirección Ejecutiva de la Administración Judicial Seccional Medellín. La combinación de datos objetivos y análisis interpretativo proporciona una comprensión más profunda y detallada de los factores que influyen en la problemática estudiada, posibilitando la formulación de estrategias efectivas para su mitigación.

Desde el enfoque **cuantitativo**, se realizará un análisis estadístico de los datos recolectados, como valores iniciales de los contratos, adiciones presupuestarias, plazos originales y prorrogados, y el porcentaje de sobrecostos. Esta parte del análisis busca identificar correlaciones significativas, patrones recurrentes y factores de riesgo que contribuyen a los sobrecostos y retrasos. Adicionalmente, se evaluará la calidad del servicio a través de variables categóricas cuantificadas, lo que permitirá explorar posibles relaciones entre los costos y los resultados del proyecto.

Por otro lado, el enfoque **cualitativo** se centrará en el análisis documental de manuales, actas de contratación y reportes técnicos, así como en la realización de entrevistas semiestructuradas a expertos y funcionarios relacionados con los procesos de contratación. Este análisis cualitativo facilitará la identificación de factores organizacionales, normativos y administrativos que influyen en la eficiencia y eficacia de los proyectos.

En cuanto al **alcance de la investigación**, se clasifica como **descriptivo y exploratorio**. Descriptivo, porque se enfoca en caracterizar y documentar los procesos de contratación, las prácticas administrativas y las variables relacionadas con los sobrecostos en los proyectos analizados. Exploratorio, debido a la utilización de técnicas de modelado predictivo, como machine learning, para identificar patrones y generar modelos que permitan predecir el comportamiento futuro de los sobrecostos en proyectos similares. Este enfoque permite no solo comprender las causas actuales, sino también proponer soluciones basadas en evidencia para prevenir problemáticas futuras.

La temporalidad del estudio comprende los proyectos ejecutados entre los años 2021 y 2023, lo que proporciona un marco reciente y representativo del contexto actual de la Dirección Ejecutiva. Los hallazgos de esta investigación no solo aportarán valor a la institución objeto de estudio, sino que también ofrecerán una base para el desarrollo de mejores prácticas aplicables en otros contextos similares dentro del sector público.

3.2 Población y muestra

3.2.1 Definición de la población

La población de interés para esta investigación incluye todos los proyectos de mantenimiento de infraestructura realizados por la Dirección Ejecutiva de Administración Judicial Seccional Medellín durante el periodo comprendido entre los años 2021 y 2023.

La población objeto de esta investigación incluye todos los proyectos de mantenimiento y mejoramiento de infraestructura física ejecutados por la Dirección Ejecutiva de la Administración Judicial Seccional Medellín durante el período comprendido entre los años 2021 y 2023. Estos proyectos abarcan actividades relacionadas con la reparación, adecuación y modernización de las instalaciones judiciales, que son esenciales para garantizar un servicio eficiente y de calidad a la ciudadanía.

La selección de esta población responde a la necesidad de analizar los procesos de contratación y ejecución en un marco temporal reciente, que permita identificar patrones actuales de sobrecostos y demoras, así como su relación con factores internos y externos. La definición temporal también asegura la relevancia de los datos en el contexto económico y administrativo vigente.

- Dentro de esta población, se incluyen proyectos que cumplen con los siguientes criterios:
- Contratos formalizados y ejecutados bajo la responsabilidad de la Dirección Ejecutiva.
- Proyectos cuyo alcance haya sido documentado en términos de costos iniciales, costos finales, plazos establecidos y prorrogados, y calidad del servicio entregado.
- Infraestructura de uso judicial y administrativo vinculada directamente a las actividades de la entidad.

3.2.2 Cálculo y selección de la muestra

Dado que la población total incluye un número limitado de proyectos (27), se opta por un muestreo no probabilístico de tipo intencional o por conveniencia, que permite garantizar la inclusión de aquellos contratos con información completa y pertinente para el análisis. Este enfoque asegura la representatividad de los datos y maximiza su utilidad para el desarrollo de modelos predictivos y análisis estadísticos.

Los criterios de selección para la muestra incluyen:

- **Relevancia del proyecto:** Se priorizan contratos de mayor impacto presupuestal y operativo para garantizar la representatividad de los casos analizados.

- **Disponibilidad de información:** Solo se incluyen proyectos con datos completos sobre costos, tiempos, adiciones presupuestarias, calidad del servicio y otros descriptores clave.
- **Diversidad de casos:** Se busca cubrir una variedad de proyectos con distintos niveles de sobrecostos, calidad del servicio y plazos prorrogados para explorar patrones y variaciones significativas.

Aunque el tamaño de la muestra coincide con la población total (27 proyectos), el enfoque intencional asegura que los datos seleccionados cumplan con los requisitos necesarios para un análisis riguroso. Este enfoque también minimiza el impacto de posibles sesgos derivados de datos incompletos o inconsistentes en algunos contratos.

La elección de este diseño muestral es particularmente adecuada en estudios exploratorios y descriptivos como el presente, donde la disponibilidad y calidad de los datos son factores determinantes. Además, el análisis exhaustivo de esta muestra permitirá construir un modelo predictivo aplicable para futuros proyectos de infraestructura, optimizando la gestión de recursos en la Dirección Ejecutiva de la Administración Judicial Seccional Medellín.

3.3 Instrumento(s)

3.3.1 Fuente de datos

Los datos se recolectarán a partir de la base de datos de la Dirección Seccional de Administración Judicial de Medellín, Antioquia. Los criterios de búsqueda incluyen proyectos de infraestructura enfocados en el mejoramiento y mantenimiento de la infraestructura física de los despachos judiciales bajo la gestión de la Dirección, correspondientes al periodo comprendido entre los años 2021 y 2023.

3.3.2 Variables de interés y codificación

De los proyectos identificados, se extraerán las siguientes variables:

- Año del contrato.
- Número del contrato.
- Valor inicial del contrato.
- Plazo inicial en días.

- Valor de las adiciones presupuestarias.
- Plazo de las prórrogas en días.
- Calidad del servicio, clasificada como: mala, regular o buena.

En total, se recopilaron datos de 27 proyectos.

La variable calidad del servicio se categorizará numéricamente como se muestra:

- 0 = Mala.
- 1 = Regular.
- 2 = Buena.

Adicionalmente, se creará una variable auxiliar denominada porcentaje de sobrecosto, calculada como la relación entre el sobrecosto de cada proyecto y el valor inicial del mismo. Esta variable proporciona una medida adimensional para analizar los sobrecostos de manera uniforme.

3.4 Descripción de procedimientos

El acceso a los datos ya había sido autorizado previamente, y estos se encuentran en la base de datos de proyectos de la Rama Judicial. Por lo tanto, no se requiere ningún paso adicional para acceder a ellos.

3.5 Análisis de información

3.5.1 Herramientas de Software

El análisis de la información recolectada en el proyecto de investigación se realizará utilizando Python 3.10, junto con las librerías NumPy, Pandas, Matplotlib, Seaborn y Scikit-Learn. Estas herramientas serán empleadas para procesar y analizar los datos, permitiendo realizar cálculos estadísticos descriptivos, explorar correlaciones entre variables y desarrollar modelos predictivos de acuerdo con los objetivos del estudio.

3.5.2 Análisis Estadístico

3.5.3 Planeamiento del Análisis estadístico

En primer lugar, se procesarán los datos recolectados para obtener una descripción estadística que incluirá medidas como promedios, desviaciones estándar y distribuciones de probabilidad. Adicionalmente, se generarán gráficos visuales, como mapas de calor y diagramas de dispersión, para identificar patrones y relaciones preliminares entre las variables analizadas. Si se considera necesario, se implementarán técnicas avanzadas de análisis, como modelos de regresión y algoritmos de machine learning, con el objetivo de profundizar en la comprensión de los datos y explorar relaciones no lineales.

El procedimiento incluirá una validación de los modelos para garantizar su aplicabilidad y evitar problemas como el overfitting. Los resultados obtenidos serán interpretados para identificar factores críticos y proporcionar información relevante que contribuya a los objetivos del proyecto, permitiendo una base sólida para la toma de decisiones informadas en el ámbito de la investigación. Este enfoque asegurará un análisis riguroso y alineado con las metas propuestas.

3.6 Consideraciones éticas

3.6.1 Análisis de consideraciones éticas

No aplica

3.6.2 Instrumentos de aceptación y autorización

No aplica

4 HIPÓTESIS

4.1 Las variables

4.1.1 Variable(s) independiente(s)

En la sección 3.3.2. se hace una recolección exhaustiva de las variables independientes a analizar en el proyecto.

4.1.2 Variable(s) dependiente(s)

La variable dependiente es “Sobrecosto del Proyecto”.

4.2 Planteamiento de hipótesis

Dado que la investigación se centra en analizar y modelar las causas de los sobrecostos y atrasos en los proyectos de infraestructura, se propone la siguiente hipótesis para su validación empírica:

Hipótesis general:

El análisis detallado de los procesos de contratación, ejecución y variables asociadas a los proyectos de infraestructura en la Dirección Ejecutiva de Administración Judicial Seccional Medellín permite identificar patrones que explican los sobrecostos y retrasos, y desarrollar un modelo predictivo capaz de reducir estos problemas en proyectos futuros.

Hipótesis específicas:

1. Las variables relacionadas con los valores iniciales de los contratos, adiciones presupuestarias y plazos prorrogados tienen una relación significativa con los sobrecostos y pueden ser utilizadas para predecirlos.
2. La aplicación de modelos de machine learning permite predecir de manera efectiva el porcentaje de sobrecosto en función de las variables identificadas, contribuyendo a la gestión eficiente de los proyectos de infraestructura.

5 RESULTADOS

5.1 Recolección de datos

5.1.1 Datos obtenidos

Los datos obtenidos del proceso de análisis se relacionan en la siguiente tabla:

Tabla 1. Tabla Información Contratos

AÑO CONTRATO	CONTRATO N°	VALOR INICIAL	PLAZO INICIAL (Días)	VALOR ADICION	PLAZO PRORROGA (Días)	CALIDAD DEL SERVICIO	SOBRECOS TO %
2020	2020-021	\$ 411.934.360	90	\$ 75.119.076	45	BUENA	18,24%
2020	2020-022	\$ 29.988.000	90	\$ 14.994.000	45	BUENA	50,00%
2020	2020-055	\$ 1.036.770.130	100	\$ 92.501.438	90	REGULAR	8,92%
2021	2021-013	\$ 41.650.000	150	\$ -	0	BUENA	0,00%
2021	2021-019	\$ 55.230.280	60	\$ -	60	MALA	0,00%
2021	2021-021	\$ 521.147.842	150	\$ 83.748.826		BUENA	16,07%
2021	2021-026	\$ 28.084.000	120	\$ -		BUENA	0,00%
2021	2021-028	\$ 30.321.296,0	120	\$ -		BUENA	0,00%
2021	2021-029	\$ 291.838.798	120	\$ -		BUENA	0,00%
2021	2021-033	\$ 306.817.781	120	\$ -		BUENA	0,00%

ANÁLISIS DE PROCESOS DE CONTRATACIÓN: SOBRECOSTOS Y SUS CAUSAS.

25

2021	2021-034	\$ 262.717.407,00	120	\$ 88.189.564		REGULAR	33,57%
2021	2021-047	\$ 107.309.623	60	\$ -	15	BUENA	0,00%
2021	2021-016	\$ 15.291.381	30	\$ -		MALA	0,00%
2022	2022-030	\$ 20.230.000	120	\$ 5.057.500		BUENA	25,00%
2022	2022-031	\$ 220.304.471	120	\$ 74.602.327		BUENA	33,86%
2022	2022-035	\$ 23.512.020	120	\$ 8.817.008	30	BUENA	37,50%
2022	2022-042	\$ 661.669.311	120	\$ 60.955.863	68	REGULAR	9,21%
2022	2022-081	\$ 15.499.750	60	\$ -		BUENA	0,00%
2022	2022-082	\$ 12.713.960	60	\$ -		BUENA	0,00%
2023	2023-038	\$ 21.311.671	75	\$ 8.524.668	30	REGULAR	40,00%
2023	2023-027	\$ 83.585.600	60	\$ -		REGULAR	0,00%
2023	2023-037	\$ 20.788.003	75	\$ -		REGULAR	0,00%
2023	2023-040	\$ 797.915.153	75	\$ 99.081.818	30	REGULAR	12,42%
2023	2023-041	\$ 102.222.974	30	\$ 14.767.720		BUENA	14,45%
2023	2023-042	\$ 544.123.518	60	\$ 186.963.962		BUENA	34,36%
2023	2023-084	\$ 315.079.745	40	\$ -	30	REGULAR	0,00%
2023	2023-087	\$ 9.984.772	37	\$ 12.143.642	30	REGULAR	121,62%

Con esto, se da cumplimiento al primer objetivo específico.

5.2 Análisis de datos

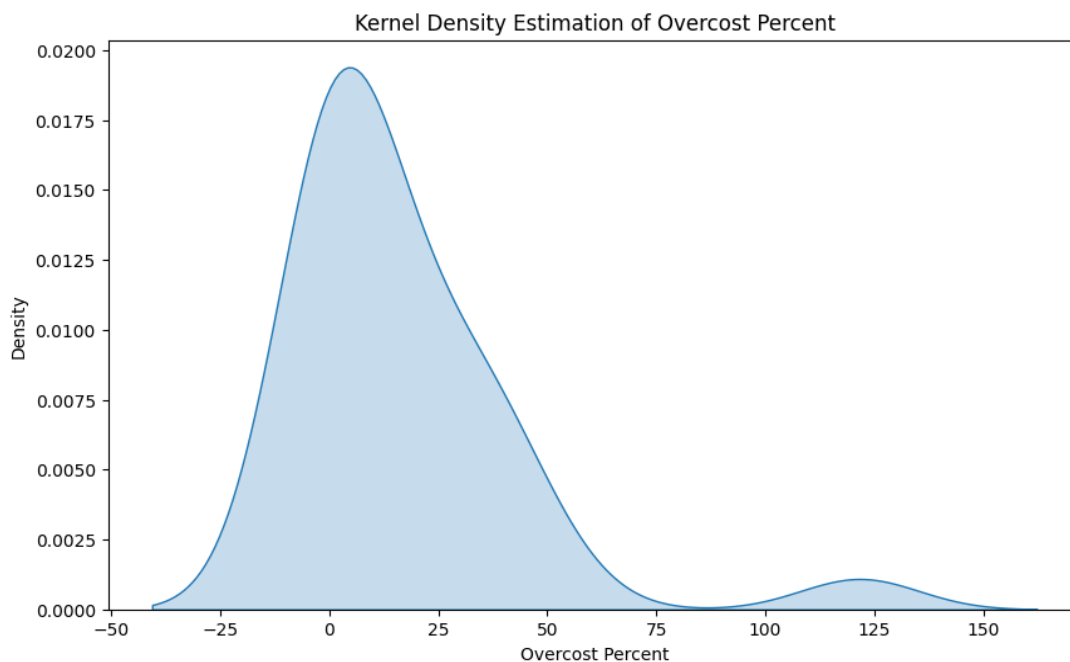
5.2.1 Análisis estadístico

El análisis estadístico de los datos se llevó a cabo utilizando el lenguaje de programación Python en su versión 3.10. Para este proceso, se emplearon las librerías **NumPy**, **Seaborn**, **Pandas**, **Matplotlib** y **Scikit-Learn**, las cuales facilitaron la manipulación matemática, el análisis descriptivo, la creación de gráficas y el desarrollo de modelos de regresión, respectivamente.

Dado que el objetivo principal del proyecto es identificar las causas de los sobrecostos en los proyectos analizados, y considerando que estos sobrecostos están estrechamente correlacionados con los valores iniciales de los contratos, se creó una variable intermedia denominada **porcentaje de sobrecosto** (`overcost_percent`). Esta variable, de carácter adimensional, indica el sobrecosto de cada proyecto en relación con su valor inicial, permitiendo analizar los sobrecostos de forma independiente al valor total del contrato. Además, proporciona una base para intentar explicarlos en función de los demás elementos presentes en el conjunto de datos.

El primer paso del análisis consistió en describir la función de densidad de probabilidad del porcentaje de sobrecosto. Aunque el número de muestras es relativamente bajo (27), se observó una distribución que recuerda a una curva gaussiana, centrada en torno al 17% y con una desviación estándar de 26%. **La Figura 1** ilustra esta distribución de probabilidad, obtenida mediante el método de estimación de kernel.

Figura 1 Estimación de la función de densidad de probabilidad del porcentaje de sobrecosto



El primer paso consistió en obtener la descripción estadística del conjunto de datos recolectados, como se muestra en la **Figura 2**. Entre los 27 contratos analizados, se encontró que el valor promedio de los contratos es de aproximadamente 222 millones de pesos, mientras que la duración media de estos proyectos es de alrededor de 88 días.

Figura 2 Descripción estadística de los datos extraídos, el sufijo `_mp` indica que el valor se encuentra expresado en millones de pesos

Descriptive Statistics:

	project_val_mp	project_len	project_overcost_mp	project_overlen	\
count	27.000000	27.000000	27.000000	27.000000	
mean	221.779328	88.222222	30.572867	17.518519	
std	276.917314	36.108100	47.464143	25.429193	
min	9.984772	30.000000	0.000000	0.000000	
25%	22.411845	60.000000	0.000000	0.000000	
50%	83.585600	90.000000	5.057500	0.000000	
75%	310.948763	120.000000	67.779095	30.000000	
max	1036.770130	150.000000	186.963962	90.000000	

	project_score	overcost_percent
count	27.000000	27.000000
mean	1.518519	16.859931
std	0.642733	26.180204
min	0.000000	0.000000
25%	1.000000	0.000000
50%	2.000000	8.922078
75%	2.000000	29.284111
max	2.000000	121.621625

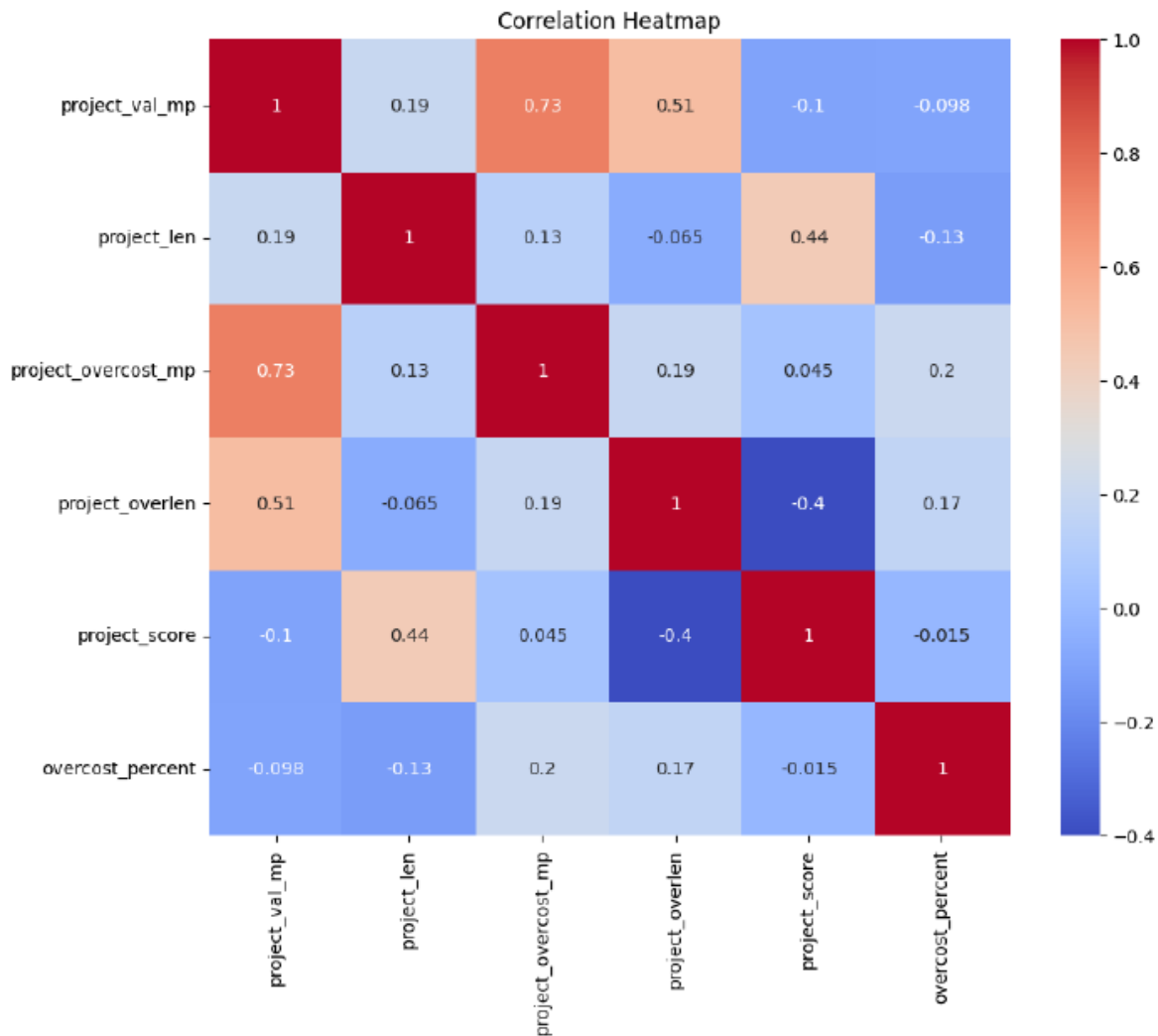
5.2.2 Análisis de correlación

Se realizó un análisis de correlación entre los descriptores de los datos. Debido a limitaciones de espacio en el presente documento, una versión completa de este análisis se encuentra en el Anexo 1. Sin embargo, las correlaciones principales pueden visualizarse en la **Figura 3**.

Los resultados preliminares indican, como era de esperarse, una alta correlación (0.73) entre los valores de los proyectos y los sobrecostos asociados. Además, se identificaron otras relaciones de interés, como el hecho de que los proyectos de mayor valor tienden a requerir prórrogas más extensas (0.51), y que aquellos con mayores prórrogas suelen obtener calificaciones más bajas en la calidad del servicio.

No obstante, ninguna de las categorías analizadas mostró una relación directa significativa con el porcentaje de sobrecostos, con valores de correlación que oscilan entre -0.2 y 0.2 .

Figura 3. Mapa de calor de los diferentes descriptores en el set de datos.

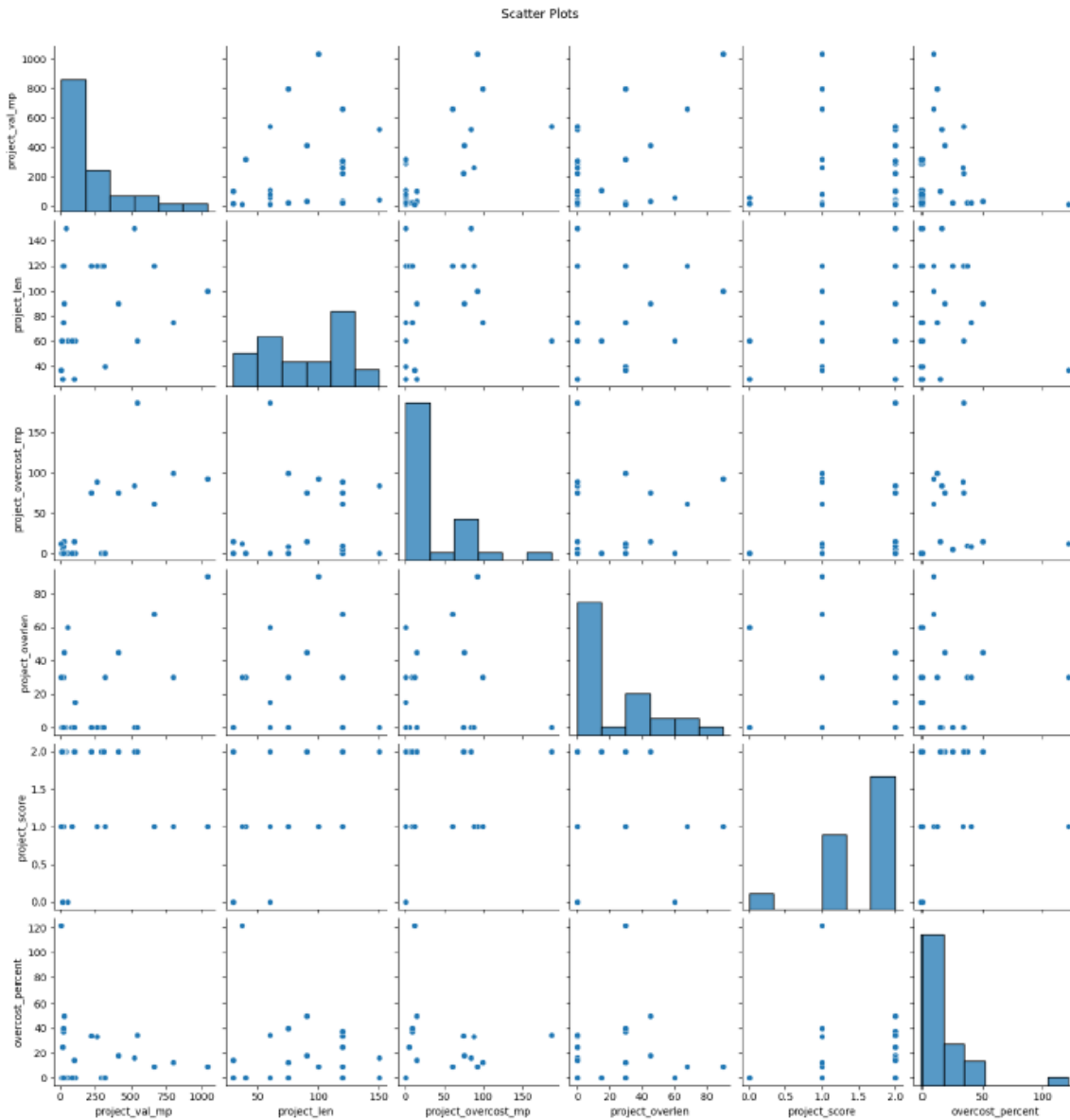


Continuando con el análisis, y con el objetivo de visualizar posibles relaciones no lineales que pudieran explicar el porcentaje de sobrecostos, se elaboró un gráfico de dispersión (**scatter plot**), presentado en la **Figura 4** (el código utilizado se encuentra en el **Anexo 1**).

En este gráfico se pueden observar con claridad algunas relaciones previamente indicadas en el mapa de calor, como la correlación lineal positiva entre el valor del proyecto y el sobrecosto. No obstante, también se identifican ciertas relaciones no lineales que no fueron evidentes en el análisis de correlación inicial.

Sin embargo, y a pesar de las observaciones gráficas, no se logró identificar ninguna relación lineal significativa entre el porcentaje de sobrecostos y los demás descriptores incluidos en el conjunto de datos.

Figura 4 Scatter plot de los diferentes descriptores del set de datos.



Con esto, se da cumplimiento al segundo objetivo específico.

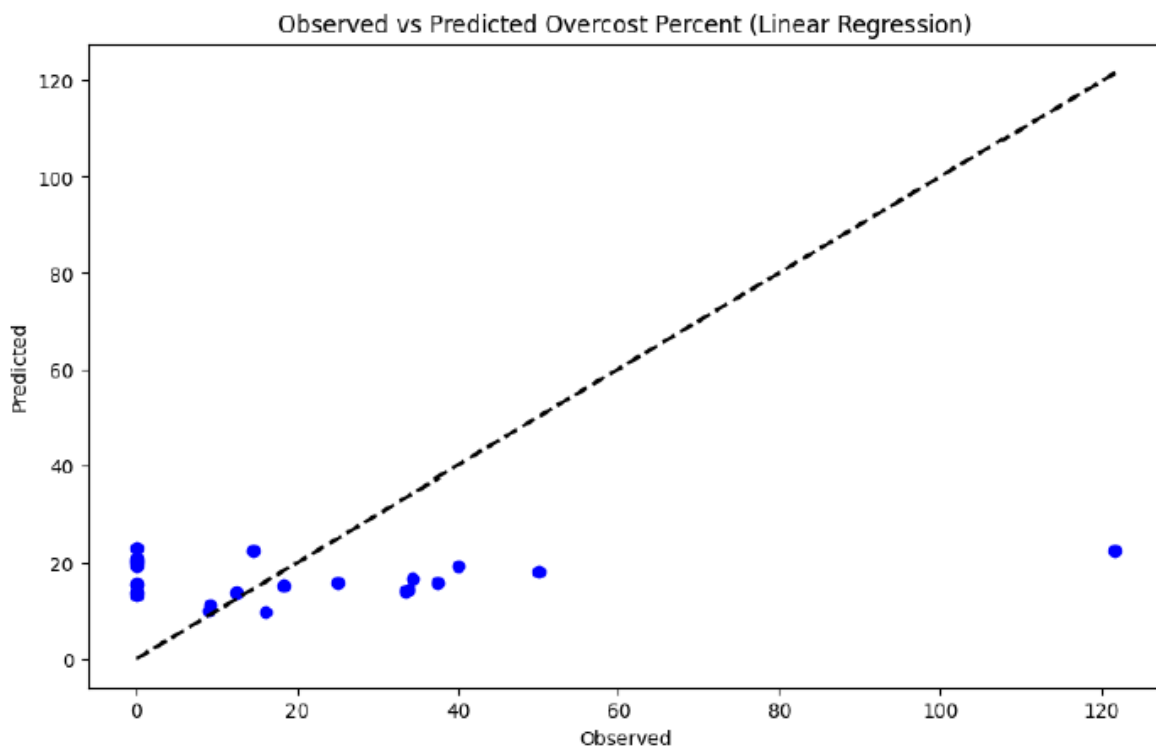
5.2.3 Análisis de regresión a través de técnicas de Machine Learning

En este punto, resulta evidente que ningún descriptor o combinación lineal de descriptores es suficiente para explicar el porcentaje de sobrecosto. Por lo tanto, se decidió cambiar de las técnicas de modelado estadístico tradicionales a enfoques de **machine learning**. Dado el tamaño reducido de la muestra, se optó por aplicar tres técnicas diferentes: **Regresión Lineal, Árboles de Decisión y Bosques Aleatorios**.

La regresión lineal se incluyó nuevamente para corroborar los resultados obtenidos en el análisis previo. Las **Figuras 5, 6 y 7** muestran los gráficos de valores observados frente a los predichos para cada uno de los modelos de regresión utilizados.

El modelo obtenido mediante **regresión lineal, Figura 5**, muestra resultados claramente erróneos. La mayoría de los valores predichos están por debajo de los valores observados, y además, el modelo presenta una limitación significativa al predecir valores superiores al 25% de sobrecosto. Este modelo obtuvo un error medio cuadrado de **624,891** y un coeficiente de determinación (R^2) de **0.053**, lo que confirma que no existe una combinación lineal de las variables predictoras capaz de describir de manera adecuada el porcentaje de sobrecosto.

Figura 5 Gráfico observado vs predicho del porcentaje de sobrecosto usando regresión lineal

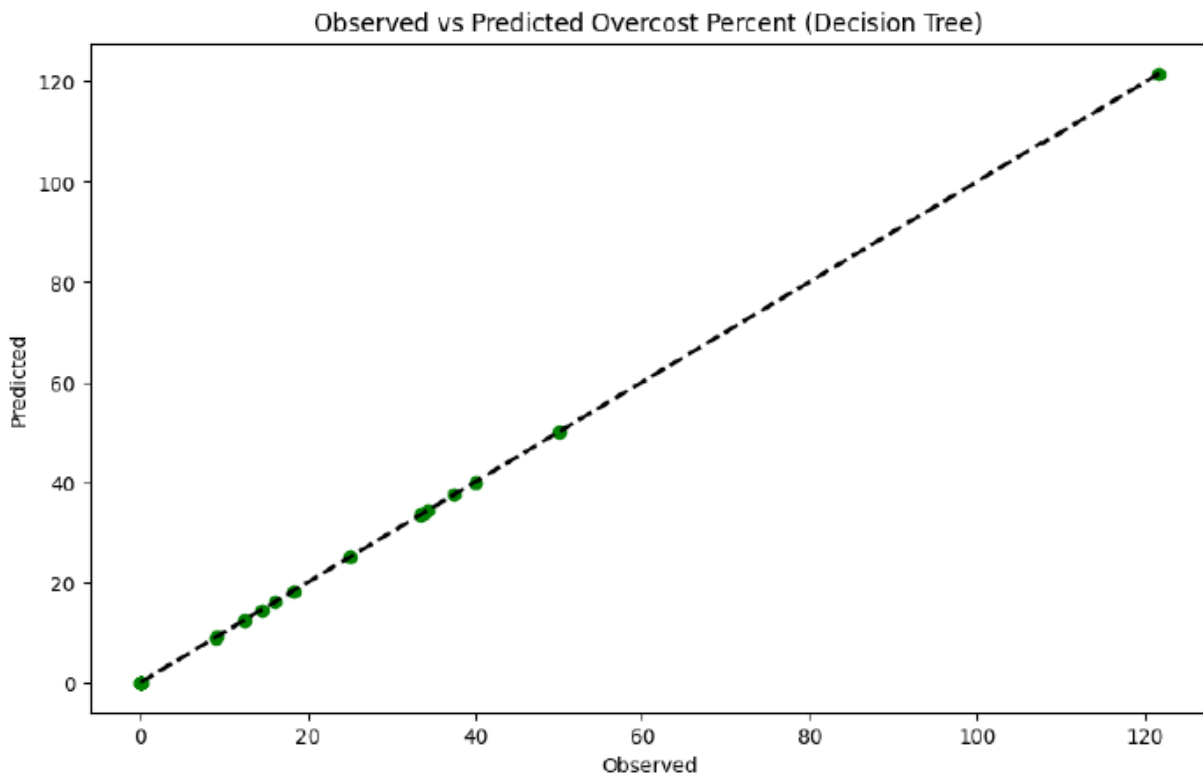


En el caso del modelo obtenido mediante el método de **árbol de decisión** (no lineal), los resultados son sorprendentemente precisos. Como se observa en la **Figura 6**, el modelo predice con exactitud cada una de las observaciones, lo que genera sospechas razonables. Este comportamiento sugiere que, debido al tamaño reducido del conjunto de datos, el modelo incurrió en **overfitting** del set de entrenamiento.

El overfitting implica que el modelo ha memorizado los datos en lugar de aprender patrones generalizables, lo que limita severamente su capacidad para realizar predicciones confiables sobre datos nuevos. Por esta razón, no se recomienda el uso de este modelo fuera del conjunto de datos analizado.

Cabe destacar que, en este caso, el modelo presenta un error medio cuadrado de **0** y un coeficiente de determinación (R^2) igual a **1**, lo cual confirma su desempeño perfecto en el set de entrenamiento, pero también refuerza las dudas sobre su aplicabilidad práctica en otros contextos.

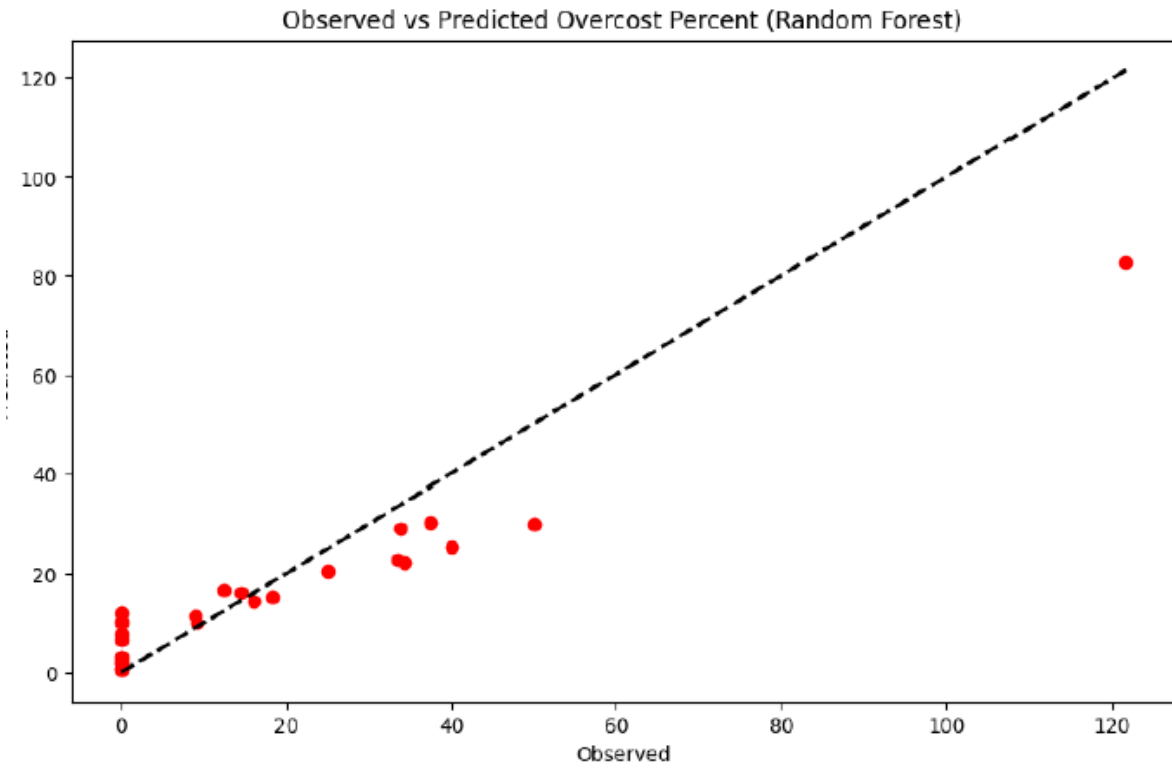
Figura 6. Gráfico observado vs predicho del porcentaje de sobrecosto usando árbol de decisión (overfitting)



Finalmente, se optó por utilizar el método de bosques aleatorios, cuyos resultados se presentan en la Figura 7. Este modelo obtuvo un error medio cuadrado de **145.71** y un coeficiente de determinación (R^2) de **0.78**, lo que indica un desempeño razonablemente bueno.

Con esto, se da cumplimiento al tercer objetivo específico.

Figura 7. Gráfico observado vs predicho del porcentaje de sobrecosto usando bosques aleatorios.



6 DISCUSIÓN

Los resultados de este estudio proporcionan una visión detallada de las causas de los sobrecostos en los proyectos de mantenimiento de infraestructura gestionados por la Dirección Ejecutiva de Administración Judicial Seccional Medellín. Se identificó una correlación significativa ($r = 0.73$) entre el valor inicial de los contratos y el porcentaje de sobrecostos, lo que sugiere que los proyectos de mayor envergadura están más expuestos a riesgos financieros y operativos que incrementan los costos. Este hallazgo es consistente con el estudio de Paltín Saraguro (2019), quien encontró que en el sector público de Ecuador, los proyectos de mayor tamaño presentaban sobrecostos más elevados debido a licitaciones subestimadas y una planificación insuficiente. Sin embargo, mientras Paltín destaca que los "sobrecostos relativos y absolutos sobresalen en proyectos de infraestructura sanitaria", nuestro estudio se enfoca en proyectos de mantenimiento de infraestructura judicial, donde las variables externas pueden diferir.

Además, se observó que las extensiones de plazo en la ejecución de los proyectos están asociadas con una disminución en las calificaciones de calidad del servicio. Este resultado coincide con las conclusiones de Nieto Camacho (2017), quien argumenta que las deficiencias en la planificación inicial generan una serie de problemas que afectan tanto los tiempos como la calidad percibida en proyectos de infraestructura pública en Bogotá. No obstante, mientras Nieto atribuye gran parte de los retrasos a factores de corrupción y politiquería, en nuestro estudio no se encontraron evidencias claras de tales factores, sugiriendo que la falta de recursos y las contingencias operativas desempeñan un rol más predominante.

En cuanto a la capacidad de los modelos predictivos para anticipar sobrecostos, se encontró que las combinaciones lineales de descriptores, evaluadas mediante regresión lineal, no fueron suficientes para modelar adecuadamente el porcentaje de sobrecostos. Los árboles de decisión mostraron una precisión notable en el conjunto de datos analizado, pero evidenciaron signos de sobreajuste (overfitting), lo que los hace inapropiados para datos no incluidos en el entrenamiento. Por otro lado, los bosques aleatorios demostraron ser el modelo más efectivo, logrando un coeficiente de determinación (R^2) de 0.78. Este resultado subraya la capacidad de los bosques aleatorios para capturar relaciones no lineales y patrones complejos en los datos, posicionándolos como una herramienta prometedora para predecir sobrecostos en proyectos futuros. Este hallazgo se alinea con Folleco David (2015), quien también identificó la utilidad de técnicas de machine learning en la valoración de riesgos en proyectos de infraestructura hospitalaria, destacando que "los modelos predictivos basados en machine learning permiten identificar patrones no evidentes con enfoques tradicionales".

El tamaño reducido del conjunto de datos, que incluyó solo 27 contratos, representó una limitación significativa para la generalización de los resultados. Esta restricción no solo afectó la robustez de los modelos predictivos, sino que también dificultó la validación de patrones no lineales identificados en los gráficos de dispersión. Además, la falta de inclusión de variables adicionales, como características específicas de los contratistas, condiciones iniciales de la infraestructura y factores externos como las condiciones climáticas, limitó la identificación de relaciones más significativas y completas en el análisis.

Paltín (2019) también subraya la importancia de contar con bases de datos más amplias y variables mejor definidas para mejorar la precisión de los análisis.

A pesar de estas limitaciones, el uso de técnicas de machine learning permitió explorar nuevas formas de abordar los sobrecostos, superando las limitaciones de los enfoques tradicionales. Los bosques aleatorios, en particular, destacaron por su capacidad para modelar los datos de manera efectiva, demostrando que es posible identificar patrones complejos y relaciones no lineales que no son evidentes en modelos lineales. Esta capacidad permite comprender mejor cómo interactúan variables clave como el valor inicial del contrato, las prórrogas y la calidad del servicio para generar sobrecostos en los proyectos.

Por último, los hallazgos resaltan la importancia de optimizar los procesos de contratación y seguimiento de proyectos. La identificación temprana de riesgos mediante modelos predictivos como los bosques aleatorios podría facilitar la implementación de medidas correctivas antes de que los problemas se traduzcan en desviaciones presupuestales significativas. Aunque los resultados son prometedores, la efectividad de estos modelos dependerá de la disponibilidad de datos más amplios y variados, así como de su validación en contextos operativos reales, como también lo enfatizan Folleco David (2015) y Nieto Camacho (2017) en sus respectivos estudios.

Es importante destacar que los sobrecostos en proyectos de infraestructura no son exclusivos de un sector o región específica. A nivel global, diversos estudios han identificado factores comunes que contribuyen a estas desviaciones presupuestales. Por ejemplo, un diseño inadecuado puede contribuir significativamente a los sobrecostos, afectando entre el 46% y el 85% de los proyectos evaluados (Paltín Saraguro, 2019). Además, la fluctuación en los precios de los insumos de construcción, como el asfalto y el concreto, ha mostrado variaciones anuales significativas, lo que impacta directamente en la viabilidad y ejecución de las obras.

La aplicación de inteligencia artificial (IA) en la construcción ha demostrado ser una herramienta eficaz para abordar estos desafíos. Los modelos de regresión, por ejemplo, se utilizan para predecir sobrecostos en función de factores como el tamaño del proyecto, la complejidad de las tareas y la experiencia del equipo, permitiendo una planificación más precisa y la implementación de medidas preventivas. Además, la IA facilita la optimización de la cadena de suministro y la gestión de riesgos, anticipando problemas potenciales y mejorando la eficiencia en la ejecución de los proyectos.

7 CONCLUSIONES

El análisis de los procesos de contratación y mantenimiento de infraestructura realizados por la Dirección Ejecutiva de Administración Judicial Seccional Medellín permitió identificar áreas clave donde los sobrecostos se presentan con mayor frecuencia. Se observó que la falta de estudios preliminares adecuados, las deficiencias en la planificación inicial y las inconsistencias en la ejecución contractual son factores que contribuyen significativamente a estos desvíos presupuestales. Además, la supervisión insuficiente y la limitada coordinación entre los actores involucrados agravan el problema. Este examen no solo permitió describir los sobrecostos, sino también entender cómo los procedimientos actuales afectan la eficiencia y sostenibilidad de los proyectos, proporcionando un punto de partida para proponer mejoras.

El análisis de las variables de interés evidenció que el valor inicial del contrato, las prórrogas en los tiempos de ejecución y la calidad del servicio tienen una relación directa y significativa con los sobrecostos. Variables auxiliares, como el porcentaje de sobrecosto calculado en relación con el valor inicial, facilitaron la comprensión de cómo estos factores influyen en el presupuesto de los proyectos. Además, se identificó que no todas las variables poseen la misma capacidad explicativa, lo que permitió priorizar aquellas con mayor relevancia. Este enfoque no solo ayuda a comprender las causas de los sobrecostos, sino que también sirve como base para optimizar la asignación de recursos y el diseño de estrategias de mitigación.

La evaluación de modelos predictivos, como la regresión lineal, los árboles de decisión y los bosques aleatorios, demostró que estos últimos ofrecen un desempeño superior en la predicción de sobrecostos. Aunque algunos modelos lineales resultaron limitados, los modelos no lineales, como los bosques aleatorios, lograron capturar mejor las relaciones complejas entre las variables. Estos modelos permiten anticipar con precisión razonable posibles desviaciones presupuestales, proporcionando una herramienta poderosa para la detección temprana de riesgos. Su implementación en los procesos de planificación y gestión no solo facilitaría la toma de decisiones oportunas, sino que también permitiría diseñar estrategias correctivas que minimicen el impacto financiero y operativo de los sobrecostos en futuros proyectos de infraestructura.

8 RECOMENDACIONES

A partir de los resultados obtenidos, se proponen las siguientes recomendaciones para optimizar la gestión de proyectos de infraestructura en la Dirección Ejecutiva de Administración Judicial Seccional Medellín:

8.1 **Modelos Predictivos para Estimación de Costos y Tiempos:**

Desarrollar modelos predictivos que utilicen datos históricos para estimar de manera más precisa los costos y plazos de ejecución de los proyectos. Esto permitirá anticipar desviaciones y tomar decisiones informadas durante la ejecución de los proyectos.

8.2 **Análisis de Riesgos y Detección Temprana de Desviaciones:**

Implementar herramientas que analicen patrones históricos para identificar riesgos potenciales, como sobrecostos o retrasos. Estos modelos pueden predecir problemas antes de que ocurran, lo que permite a los gestores tomar medidas correctivas a tiempo.

8.3 **Optimización de la Asignación de Recursos:**

Utilizar algoritmos que optimicen la distribución de recursos en los proyectos, ajustando las necesidades de personal, materiales y presupuesto de acuerdo con las características específicas de cada obra. Esto mejora la eficiencia y reduce costos operativos.

8.4 **Mejoramiento de la Gestión de Contratos:**

Desarrollar modelos que ayuden a evaluar la calidad y los riesgos de los contratos mediante el análisis de datos previos. Esto contribuirá a la identificación de cláusulas problemáticas o posibles riesgos, mejorando la toma de decisiones en la contratación.

8.5 **Ampliar el tamaño y diversidad de la muestra:**

Recopilar datos de contratos en otras regiones o periodos podría aumentar la representatividad del análisis y fortalecer las conclusiones.

8.6 **Incorporar más variables explicativas: Factores** externos como el clima, las condiciones económicas o indicadores de desempeño histórico de los contratistas deben integrarse para mejorar la capacidad de los modelos predictivos.

Estas recomendaciones permitirán mejorar la eficiencia y efectividad en la gestión de proyectos de infraestructura, optimizando el uso de recursos y minimizando los riesgos asociados a los plazos y costos.

Referencias

- El Colombiano. (2020, enero 27). Sobrecostos y retrasos, pecados usuales en obras públicas locales. El Colombiano. Recuperado de <https://www.elcolombiano.com/antioquia/sobrecostos-y-retrasos-pecados-usuales-en-obras-publicas-locales-GF12609483>
- García Sánchez M. (2007). *La nueva gestión pública: evolución y tendencias*. Recuperado de [https://www2.congreso.gob.pe/sicr/cendocbib/con4_uibd.nsf/B1168625381F1AE705257BCA00165978/\\$FILE/47_GarciaSanchez.pdf](https://www2.congreso.gob.pe/sicr/cendocbib/con4_uibd.nsf/B1168625381F1AE705257BCA00165978/$FILE/47_GarciaSanchez.pdf)
- Cadavid Rojas, R. A., & Almanza Castillo, L. V. (2021). Principales factores causales del sobrecosto en proyectos de construcción colombianos: una consulta a profesionales del sector. Trabajo de grado para optar al título de Ingeniero civil, Universidad Católica de Colombia. Recuperado de <https://repository.ucatolica.edu.co/server/api/core/bitstreams/6af4d496-97a8-403d-ab6a-0ff8a5293389/content>
- Consejo Superior de la Judicatura. (2015). *Plan sectorial de desarrollo rama judicial 2015–2018*. Recuperado de <https://disajcucuta.files.wordpress.com/2015/09/1-plan-sectorial-de-desarrollo-rama-judicial-2015-2018.pdf>
- Folleco David, J. (2015). Aplicación de técnicas de machine learning en la valoración de riesgos en proyectos de infraestructura hospitalaria. Universidad de Cuenca. Recuperado de <https://dspace.ucuenca.edu.ec/handle/123456789/21815>
- Gaviria Gaviria, P. A. (2013). *Diseño de un sistema de indicadores de sostenibilidad como herramienta en la toma de decisiones para la gestión de proyectos de infraestructura en Colombia* (Tesis de maestría). Universidad EAFIT. Recuperado de <https://repository.eafit.edu.co/items/f9624a18-2d10-48dd-b152-fd00d486ad17>
- Guatavita, C., & Montoya, G. (2024). *Conformación de documentos técnicos para el contrato de mantenimiento de infraestructura 2024 de las sedes de la Fiscalía General de la Nación Seccional Orinoquía*. Repositorio Institucional Universidad Cooperativa de Colombia; Universidad Cooperativa de Colombia, Facultad de Ingenierías, Ingeniería Civil, Villavicencio. Recuperado de <https://repository.ucc.edu.co/entities/publication/3d0f2539-ae09-440c-aeed-6388882ac658>
- Grau, N. C. (2006). La Transparencia en la Gestión Pública: ¿Cómo construirle viabilidad? *Estado, Gobierno, Gestión Pública: Revista Chilena de Administración Pública*, 8, 22–44. Recuperado de <https://dialnet.unirioja.es/descarga/articulo/2315250.pdf>
- Hitchner, E. (1992). *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector* [Reseña de libro]. *National Productivity Review*, 11(3), 71–73. Recuperado de <https://doi.org/10.1002/npr.4040110309>
- Klojčnik, T., Sagadin, T., & Kralj, D. (2018). *Project Management: A Systematic Approach to Planning, Scheduling, and Controlling Sustainable Transformation*. Recuperado de [https://www.iaras.org/iaras/filedownloads/ijems/2018/007-0004\(2018\).pdf](https://www.iaras.org/iaras/filedownloads/ijems/2018/007-0004(2018).pdf)

- Medina, M. (2023, febrero 25). Procuraduría emitió alerta por multimillonarios recursos en proyectos de infraestructura en Bogotá. Infobae. Recuperado de <https://www.infobae.com/colombia/2023/02/25/procuraduria-emitio-alerta-por-multimillonarios-recursos-en-proyectos-de-infraestructura-en-bogota/>
- Nieto Camacho, J. A. (2017). Análisis de los factores que inciden en los retrasos de los proyectos de infraestructura pública en Bogotá. Universidad Nacional de Colombia. Recuperado de <http://bdigital.unal.edu.co/58639/1/1013885403.2017.pdf>
- Paltín Saraguro, J. L. (2019). Sobrecostes en proyectos de infraestructura sanitaria en Ecuador. Universidad de Cuenca. Recuperado de <https://dspace.ucuenca.edu.ec/handle/123456789/31511>
- Pegnalver, D. (2013). Análisis del impacto social de las inversiones públicas en infraestructuras. La huella social Munich Personal RePEc Archive. *Uni-Muenchen.de*. Recuperado de https://mpra.ub.uni-muenchen.de/47647/8/MPRA_paper_47647.pdf
- PMI. (2017). A Guide to the Project Management Body of Knowledge (PMBOK Guide) (6th ed.). Project Management Institute.
- Prats, R., & Jacobs, J. (2014). Infraestructuras urbanas: entorno de colaboración e innovación SMART CITIES. Recuperado de https://quickclick.es/rop/pdf/publico/2014/2014_enero_3550_14.pdf
- Rosario Villarreal, M. A., Ocaña Fernández, Y., Capillo Chávez, C. H., Lavado Rodríguez, A. B., El Homrani, M., & Arias Romero, S. M. (2019). Factores que inciden en la gestión de proyectos de investigación científica. *Apuntes Universitarios*, 9(1), 46–67. Recuperado de <https://doi.org/10.17162/au.v9i1.349>
- Urrunaga Pascó-Font, R. (2009). Relevancia de la infraestructura y análisis de los sobrecostos que genera su déficit. *Apuntes: Revista de Ciencias Sociales*, 65, 61–94. Recuperado de <https://doi.org/10.21678/apuntes.65.591>
- Vela Moreno, V. B., & Luévanos Rojas, A. (2020). Optimización de recursos en los proyectos de construcción de la Comarca Lagunera de México mediante la programación lineal. En *13° Simposio de Investigación en Sistemas Constructivos Computacionales y Arquitectónicos (SISCCA 2020)*. Facultad de Ingeniería, Ciencias y Arquitectura, Universidad Juárez del Estado de Durango. Gómez Palacio, Durango, México. Recuperado de <http://fica.ujed.mx/Optimizaci%C3%B3n%20de%20recursos%20en%20los%20proyectos%20de%20construcci%C3%B3n.pdf>
- William, R., David, J., & Pereyra, T. (2019). Propuesta de mejora del procedimiento de administración y control de contratos de obras públicas en la construcción y mantenimiento de infraestructura penitenciaria, 2019. Continental.edu.pe. Recuperado de <https://hdl.handle.net/20.500.12394/10436>

Anexos

Anexo 1. Código fuente

script file:///Users/jotalvaro/tmp/daniel/script.html

```

In [1]: import numpy as np

In [16]: # Data ingestion
project_val= np.array([411934360, 29988000, 1036770130, 41650000, 55230280,
project_len = np.array([90,90,100,150,60,150,120,120,120,120,60,30,120,1
project_overcost = np.array([75119076,14994000,92501438,0 ,0 ,83748826,0 ,0
project_overlen = np.array([45,45,90,0,60,0,0,0,0,0,15,0,0,0,30,68,0,0,30,
project_score_labels = ["BUENA","BUENA","REGULAR","BUENA","MALA","BUENA","BU
score_mapping = {'MALA': 0, 'REGULAR': 1, 'BUENA': 2}
project_score = np.array([score_mapping[score] for score in project_score_la
overcost_percent = project_overcost/project_val

In [ ]: import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt

# Create a DataFrame from the Lists
data = {
    'project_val_mp': project_val / 1000000.0,
    'project_len': project_len,
    'project_overcost_mp': project_overcost / 1000000.0,
    'project_overlen': project_overlen,
    'project_score': project_score,
}

df = pd.DataFrame(data)

# Calculate overcost percentage
df['overcost_percent'] = df['project_overcost_mp'] / df['project_val_mp'] *

# Descriptive Statistics
print("Descriptive Statistics:")
print(df.describe())

# Correlation Analysis
print("\nCorrelation Matrix:")
print(df.corr())

# Histograms
df.hist(bins=10, figsize=(10, 8))
plt.suptitle('Histograms')
plt.show()

# Box Plots
plt.figure(figsize=(10, 8))
sns.boxplot(data=df)
plt.title('Box Plots')
plt.show()

# Scatter Plots
sns.pairplot(df)
plt.suptitle('Scatter Plots', y=1.02)

```

script

file:///Users/jotalvaro/tmp/daniel/script.html

```
plt.show()

# Bar Plots
df.plot(kind='bar', figsize=(15, 8))
plt.title('Bar Plots')
plt.show()
```

Descriptive Statistics:

	project_val_mp	project_len	project_overcost_mp	project_overlen \
count	27.000000	27.000000	27.000000	27.000000
mean	221.779328	88.222222	30.572867	17.518519
std	276.917314	36.108100	47.464143	25.429193
min	9.984772	30.000000	0.000000	0.000000
25%	22.411845	60.000000	0.000000	0.000000
50%	83.585600	90.000000	5.057500	0.000000
75%	310.948763	120.000000	67.779095	30.000000
max	1036.770130	150.000000	186.963962	90.000000

	project_score	overcost_percent
count	27.000000	27.000000
mean	1.518519	16.859931
std	0.642733	26.180204
min	0.000000	0.000000
25%	1.000000	0.000000
50%	2.000000	8.922078
75%	2.000000	29.284111
max	2.000000	121.621625

Correlation Matrix:

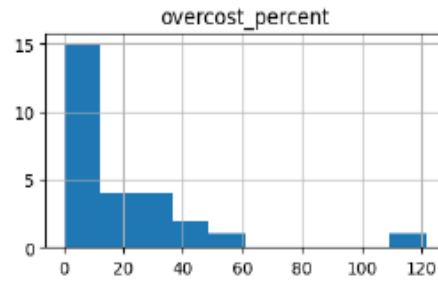
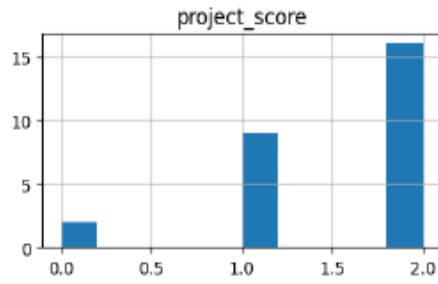
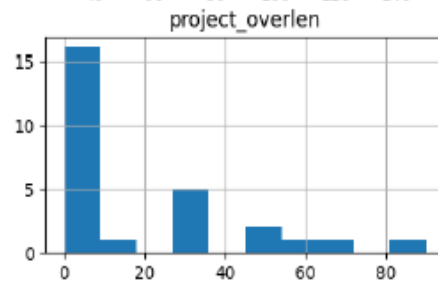
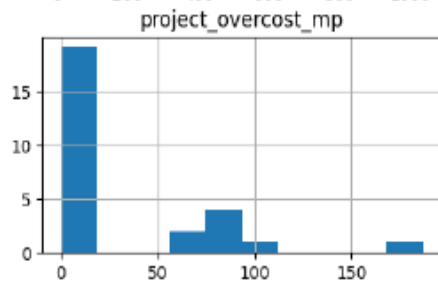
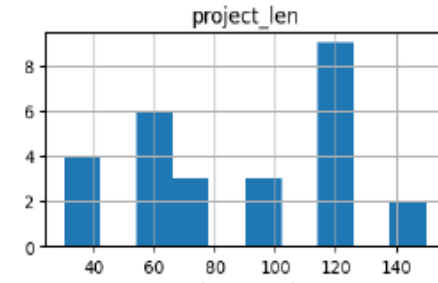
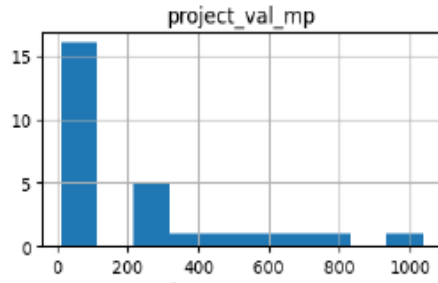
	project_val_mp	project_len	project_overcost_mp	\
project_val_mp	1.000000	0.191011	0.733668	
project_len	0.191011	1.000000	0.130014	
project_overcost_mp	0.733668	0.130014	1.000000	
project_overlen	0.509297	-0.065308	0.188687	
project_score	-0.101075	0.438991	0.045451	
overcost_percent	-0.097676	-0.125835	0.203183	

	project_overlen	project_score	overcost_percent
project_val_mp	0.509297	-0.101075	-0.097676
project_len	-0.065308	0.438991	-0.125835
project_overcost_mp	0.188687	0.045451	0.203183
project_overlen	1.000000	-0.400659	0.165892
project_score	-0.400659	1.000000	-0.015001
overcost_percent	0.165892	-0.015001	1.000000

script

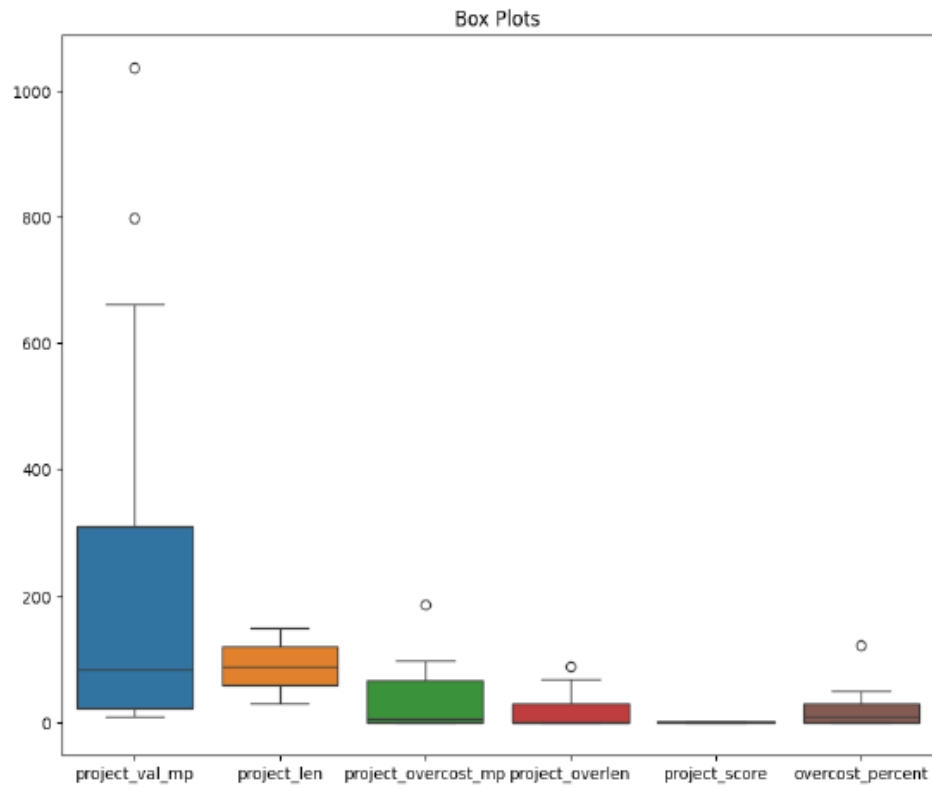
file:///Users/jotalvaro/tmp/daniel/script.html

Histograms



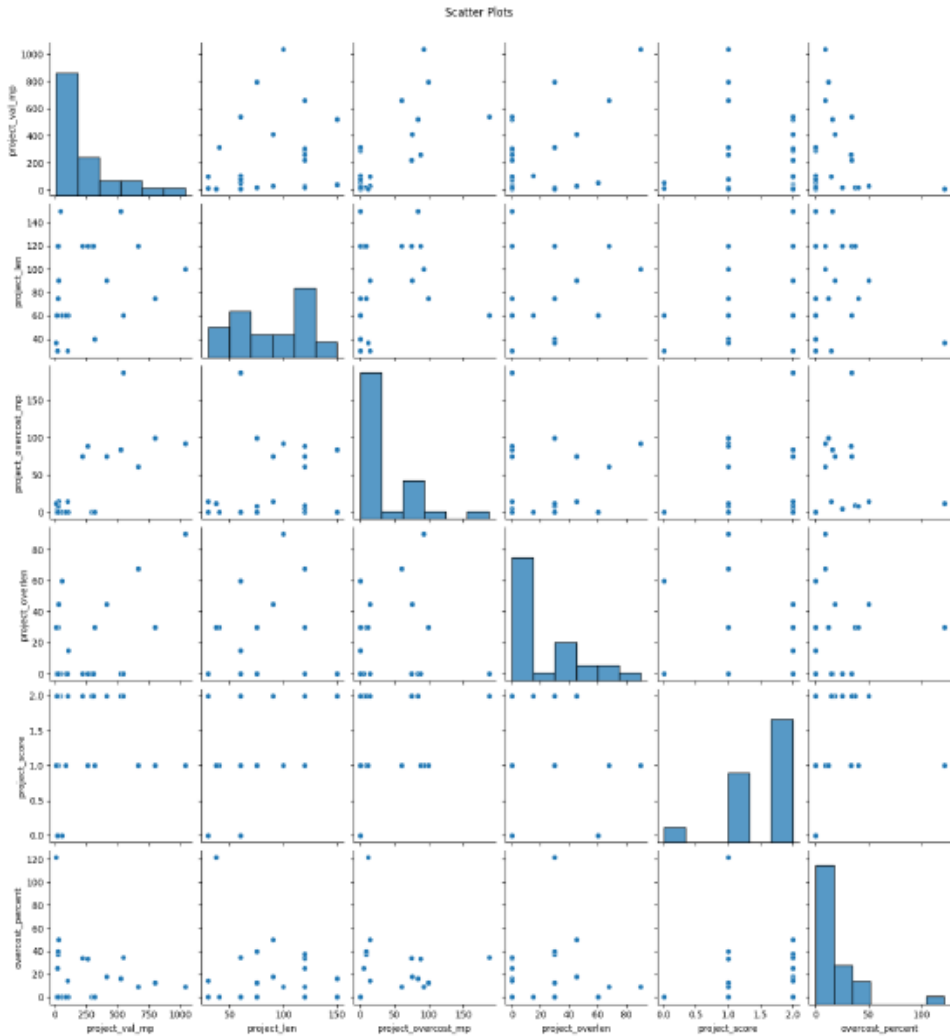
script

file:///Users/jotalvaro/tmp/daniel/script.html



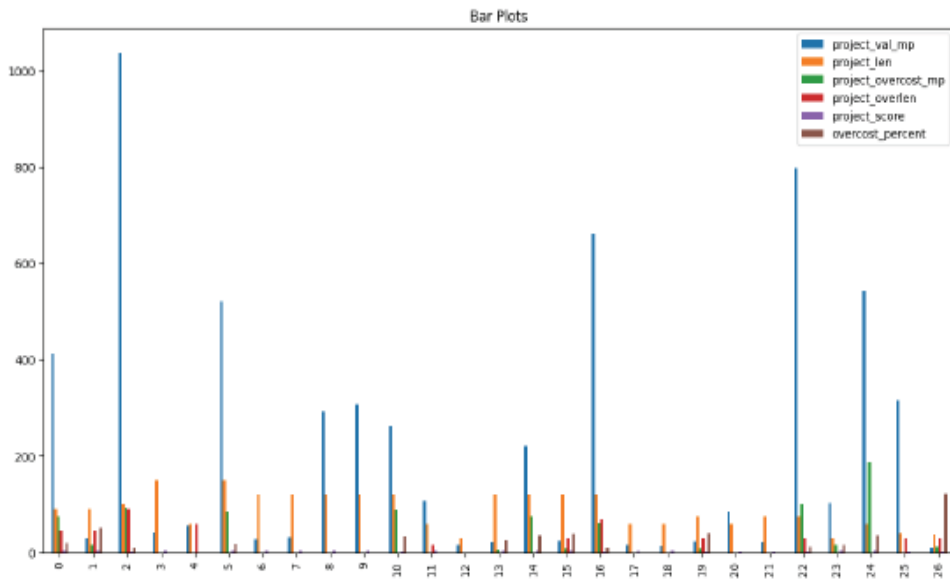
script

file:///Users/jotalvaro/tmp/daniel/script.html

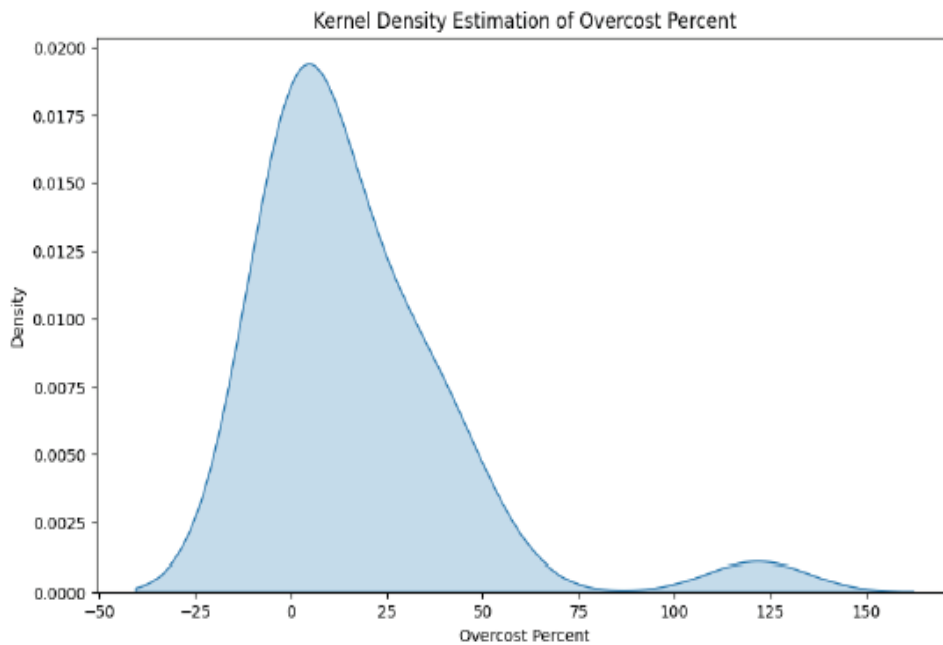


script

file:///Users/jotalvaro/tmp/daniel/script.html



```
In [ ]: # Kernel Density Estimation (KDE) for overcost_percent
plt.figure(figsize=(10, 6))
sns.kdeplot(df['overcost_percent'], fill=True)
plt.title('Kernel Density Estimation of Overcost Percent')
plt.xlabel('Overcost Percent')
plt.ylabel('Density')
plt.show()
```



```
In [40]: from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.tree import DecisionTreeRegressor
```

scripr

me:///Users/jotaivarom/imp/daniel/scripr.nmml

```
from sklearn.ensemble import RandomForestRegressor

# Define the features and target variable
X = df[['project_val_mp', 'project_len']]
y = df['overcost_percent']

# Create and train the model
model = LinearRegression()
model.fit(X, y)

# Make predictions
y_pred = model.predict(X)

# Display the coefficients
coefficients = pd.DataFrame(model.coef_, X.columns, columns=['Coefficient'])
print(coefficients)

# Evaluate the model
mse = mean_squared_error(y, y_pred)
r2 = r2_score(y, y_pred)

print(f'Linear Regression Mean Squared Error: {mse}')
print(f'Linear Regression R-squared: {r2}')

# Feature importance for Linear Regression
print("\nLinear Regression Feature Importance:")
print(coefficients)
print("=====")

# Decision Tree Regression
tree_model = DecisionTreeRegressor()
tree_model.fit(X, y)
y_tree_pred = tree_model.predict(X)
tree_mse = mean_squared_error(y, y_tree_pred)
tree_r2 = r2_score(y, y_tree_pred)
print(f'Decision Tree Mean Squared Error: {tree_mse}')
print(f'Decision Tree R-squared: {tree_r2}')

# Feature importance for Decision Tree
tree_importances = pd.DataFrame(tree_model.feature_importances_, X.columns,
print("\nDecision Tree Feature Importance:")
print(tree_importances)
print("=====")

# Random Forest Regression
forest_model = RandomForestRegressor()
forest_model.fit(X, y)
y_forest_pred = forest_model.predict(X)
forest_mse = mean_squared_error(y, y_forest_pred)
forest_r2 = r2_score(y, y_forest_pred)
print(f'Random Forest Mean Squared Error: {forest_mse}')
print(f'Random Forest R-squared: {forest_r2}')

# Feature importance for Random Forest
forest_importances = pd.DataFrame(forest_model.feature_importances_, X.columns,
print("\nRandom Forest Feature Importance:")
print(forest_importances)
```

script

file:///Users/jotalvaro/tmp/daniel/script.html

```

print("=====")

# Plotting the observed vs predicted values
plt.figure(figsize=(10, 6))
plt.scatter(y, y_pred, color='blue')
plt.plot([y.min(), y.max()], [y.min(), y.max()], 'k--', lw=2)
plt.xlabel('Observed')
plt.ylabel('Predicted')
plt.title('Observed vs Predicted Overcost Percent (Linear Regression)')
plt.show()

# Plotting the observed vs predicted values for Decision Tree
plt.figure(figsize=(10, 6))
plt.scatter(y, y_tree_pred, color='green')
plt.plot([y.min(), y.max()], [y.min(), y.max()], 'k--', lw=2)
plt.xlabel('Observed')
plt.ylabel('Predicted')
plt.title('Observed vs Predicted Overcost Percent (Decision Tree)')
plt.show()

# Plotting the observed vs predicted values for Random Forest
plt.figure(figsize=(10, 6))
plt.scatter(y, y_forest_pred, color='red')
plt.plot([y.min(), y.max()], [y.min(), y.max()], 'k--', lw=2)
plt.xlabel('Observed')
plt.ylabel('Predicted')
plt.title('Observed vs Predicted Overcost Percent (Random Forest)')
plt.show()

```

```

                Coefficient
project_val_mp  -0.007226
project_len     -0.080652
Linear Regression Mean Squared Error: 645.8520803455909
Linear Regression R-squared: 0.021462587204706174

```

Linear Regression Feature Importance:

```

                Coefficient
project_val_mp  -0.007226
project_len     -0.080652

```

Decision Tree Mean Squared Error: 0.0

Decision Tree R-squared: 1.0

Decision Tree Feature Importance:

```

                Importance
project_val_mp  0.93509
project_len     0.06491

```

Random Forest Mean Squared Error: 111.25242865768953

Random Forest R-squared: 0.8314402523134464

Random Forest Feature Importance:

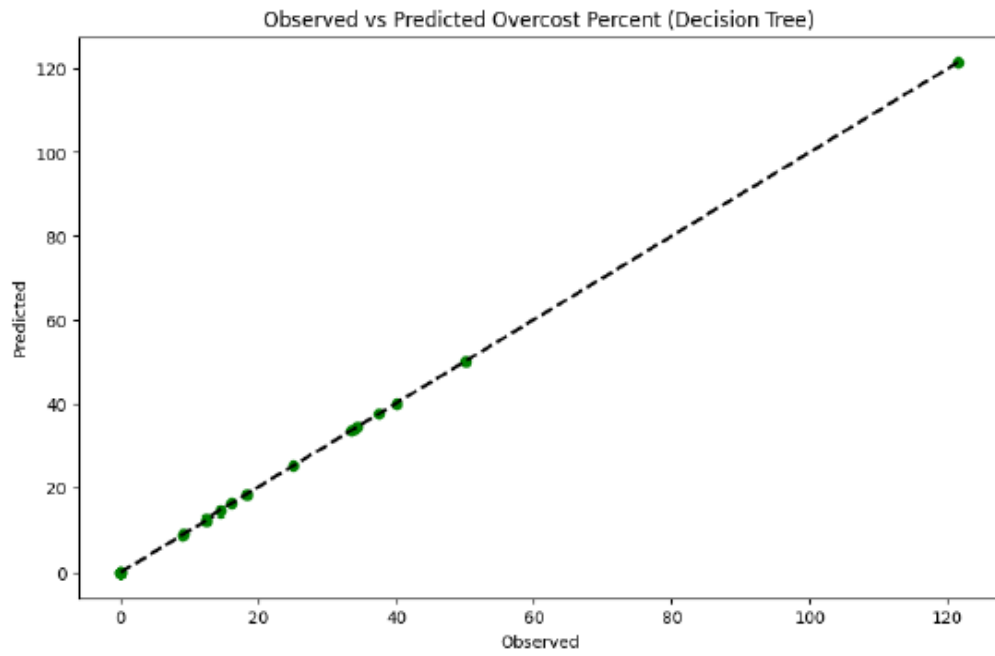
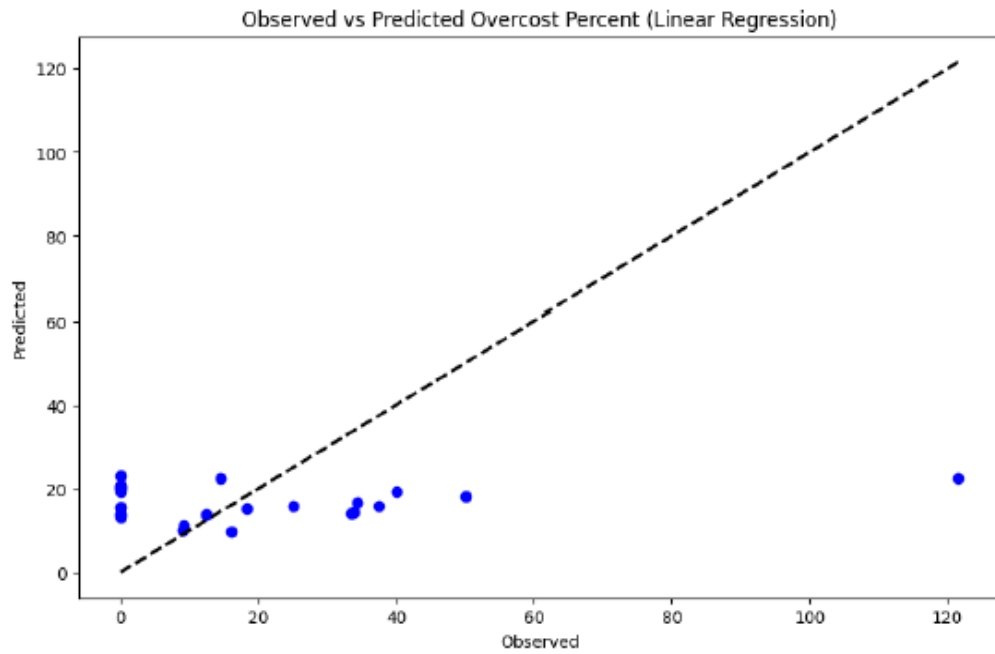
```

                Importance
project_val_mp  0.814234
project_len     0.185766

```

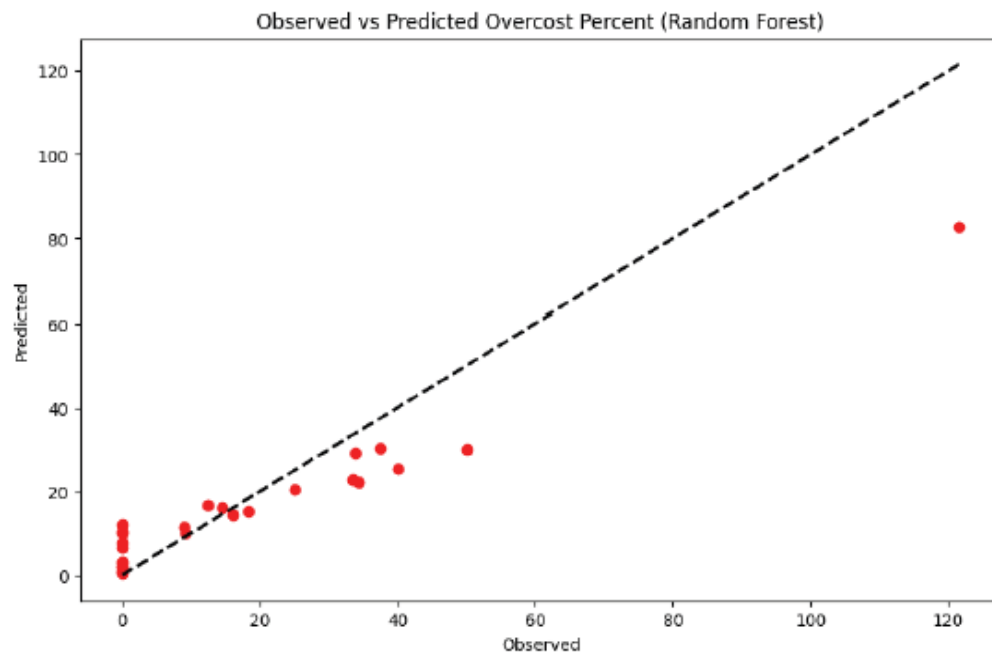
script

file:///Users/jotalvaro/tmp/daniel/script.html



script

file:///Users/jotalvaro/tmp/daniel/script.html



```
In [42]: from statsmodels.formula.api import ols
import statsmodels.api as sm

# Create a DataFrame with the relevant data
anova_data = df[['project_val_mp', 'project_len', 'project_score', 'overcost

# Fit the model
model = ols('overcost_percent ~ project_val_mp + project_len + C(project_score

# Perform ANOVA
anova_table = sm.stats.anova_lm(model, typ=2)

# Display the ANOVA table
print(anova_table)
```

	sum_sq	df	F	PR(>F)
C(project_score)	1867.975832	2.0	1.319698	0.287554
project_val_mp	561.267396	1.0	0.793055	0.382814
project_len	183.072213	1.0	0.258676	0.616094
Residual	15570.030338	22.0	NaN	NaN

```
In [44]: plt.figure(figsize=(10, 8))
sns.heatmap(df.corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```

script file:///Users/jotalvaro/tmp/daniel/script.html

