

Entrega Proyecto de analítica de datos

J. D. Cordoba

Pontificia Universidad Javeriana

M. A. Gonzalez Rodriguez

Pontificia Universidad Javeriana

Abstract—This data analytics project focuses on analyzing and visualizing the public procurement process within the Huila Governor’s Office, a departmental-level public entity in Colombia. Leveraging data from SECOP II (Colombia’s Electronic Public Procurement System), the project aims to identify and optimize key factors that affect the average contract award time—an indicator that directly impacts the timely execution of public projects. A dataset comprising 4.5 million records and 87 attributes provides insights into the relationships among variables such as bidder plurality, total awarded contract value, and administrative efficiency. This analysis seeks to enhance transparency, efficiency, and institutional responsiveness, thereby contributing to the strengthening of public management and the overall development of the Huila region.

■ **CONTEXTO DE LA EMPRESA** En Colombia, las contrataciones del Estado tienen una gran dimensión y relevancia. Según el diario La República, el monto total de contratación en el país alcanzó los \$157 billones, distribuidos en 1.288.548 contratos, lo que representa aproximadamente el 13,3% del PIB[1]. Varias instituciones en Colombia se especializan en la auditoría de procesos de contratación tanto en el sector público como en el privado. Estas instituciones, como la Contraloría General de la República, la Contaduría General de la Nación y la Procuraduría General de la Nación, tienen la responsabilidad de asegurar que los procesos de contratación sean transparentes y se realicen conforme a la ley. La Contraloría General, en particular, tiene la mayor autoridad y responsabilidad en la regulación y auditoría de los recursos públicos. Los procesos de contratación en Colombia se componen de tres etapas distintas, mostradas en los siguientes puntos:

1) Planeación – Identificación de necesidades,

estructuración del proceso, elaboración del estudio previo y del presupuesto.

- 2) Selección – Publicación del proceso, recepción de ofertas, evaluación y adjudicación del contrato.
- 3) Ejecución – Firma del contrato, seguimiento técnico, financiero y legal de su cumplimiento.

En este sentido, como todo ente público perteneciente a la rama ejecutiva del Estado colombiano, la Gobernación del Huila debe garantizar la transparencia en sus procesos administrativos y contractuales. Se trata de una entidad territorial del orden departamental que actúa como ente contratante dentro del marco del Sistema de Compras Públicas del país. Por tanto, tiene la responsabilidad de planear, ejecutar y supervisar procesos de contratación pública que respondan a las necesidades de la población huilense, asegurando el uso eficiente y transparente de los recursos públicos. Adicionalmente, está sujeta al control y vigilancia de organismos

como la Contraloría General de la República, la Procuraduría General de la Nación y la Contaduría General de la Nación. Sus procesos contractuales están disponibles para consulta en plataformas como SECOP II, lo cual fortalece los mecanismos de control ciudadano y la rendición de cuentas.

La Gobernación pertenece al sector público territorial, específicamente al subsector de la administración departamental. Su misión principal es liderar el desarrollo del departamento del Huila mediante la formulación e implementación de políticas públicas, la administración de recursos y la ejecución de proyectos que mejoren la calidad de vida de sus habitantes. Sus usuarios son los más de un millón de ciudadanos del departamento, así como entidades públicas y privadas con las que mantiene relaciones contractuales. Aunque no tiene competidores en el sentido comercial, comparte responsabilidades con otros niveles de gobierno como las alcaldías municipales y entidades descentralizadas, con quienes coordina acciones de desarrollo. Dentro de su cadena de valor, se pueden identificar los siguientes procesos clave:

- **Planeación estratégica y presupuestal:** definición del Plan de Desarrollo Departamental y asignación de recursos.
- **Gestión de contratación pública:** estructuración, publicación, adjudicación y seguimiento de contratos.
- **Ejecución de proyectos de inversión:** en áreas como infraestructura, salud, educación y desarrollo rural.
- **Gestión financiera y contable:** administración del presupuesto y reporte de ejecución.
- **Atención al ciudadano y control social:** canales de participación y rendición de cuentas.

En conjunto, estos procesos permiten a la Gobernación del Huila cumplir su rol como actor clave del desarrollo regional, y hacen de la contratación pública un componente estratégico dentro de su funcionamiento institucional.

Producto o Servicio

La Gobernación del Huila, como entidad pública del orden territorial, no comercializa productos, sino que entrega valor público a través



Figure 1. IPS Neiva



Figure 2. I.E. Claretiano Gustavo Torres Parra Neiva

de la formulación y ejecución de políticas, programas y proyectos que impactan positivamente en la calidad de vida de los ciudadanos del departamento. Su enfoque está en la prestación de servicios públicos esenciales y la promoción del desarrollo integral del territorio. Servicios clave ofrecidos al ciudadano a través de la contratación: Salud pública: Fortalecimiento de la red hospitalaria departamental, campañas de prevención, acceso a servicios médicos.

- 1) **Salud pública:** Fortalecimiento de la red hospitalaria departamental, campañas de prevención, acceso a servicios médicos (ejemplo figura 1).
- 2) **Educación:** Infraestructura educativa, dotación de instituciones, programas de cobertura y calidad educativa (ejemplo figura 2).
- 3) **Infraestructura vial y transporte:** Mantenimiento y construcción de vías secundarias y terciarias, conectividad rural (ejemplo figura 3).
- 4) **Desarrollo rural y agropecuario:** Apoyo



Figure 3. Proyecto vial Neiva - Espinal



Figure 4. ICA Neiva

técnico y financiero a pequeños productores, sostenibilidad agroambiental (ejemplo figura 4).

- 5) **Gestión del riesgo y medio ambiente:** Protección de cuencas, prevención de desastres naturales, manejo de residuos sólidos (ejemplo figura 5).
- 6) **Cultura, deporte y recreación:** Fomento



Figure 5. Rio Magdalena Neiva



**Alcaldía de Neiva
Deporte**

Figure 6. Subsección deportiva de la alcaldía de Neiva



Figure 7. CAI de Timanco en Neiva

de actividades culturales, deportivas y recreativas en zonas urbanas y rurales (ejemplo figura 6).

- 7) **Seguridad y convivencia ciudadana:** Coordinación con fuerza pública, programas de justicia local, atención a víctimas (ejemplo figura 7).
- 8) **Planeación y contratación pública:** Planeación del desarrollo y ejecución eficiente y transparente del gasto público.

En el diagrama de flujo de la figura 8 se representa el proceso de contratación pública que debe seguir cualquier entidad del Estado colombiano cuando requiere contratar bienes, servicios u obras para atender las necesidades de la ciudadanía. Este proceso garantiza que las contrataciones se realicen de manera transparente, eficiente y conforme a la normativa vigente, asegurando el adecuado uso de los recursos públicos y el cumplimiento de los principios de legalidad, economía y responsabilidad administrativa.

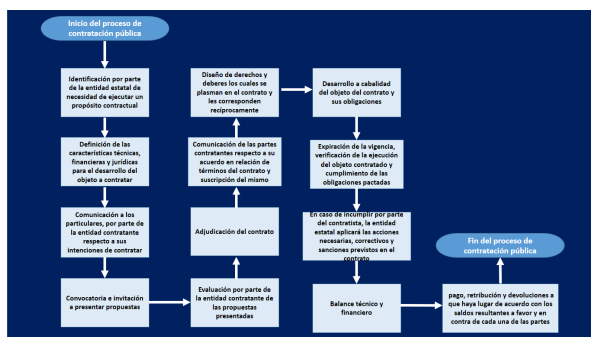


Figure 8. Diagrama de flujo de la contratación pública

Proceso analizado

Proceso: Contratación Pública

El proceso analizado corresponde a la contratación pública dentro de la Gobernación del Huila. Este proceso es fundamental dentro de su cadena de valor, ya que permite ejecutar las políticas públicas y proyectos priorizados en el Plan de Desarrollo Departamental, mediante la adquisición de bienes, servicios y obras. Este proceso inicia con la planeación de la contratación, continúa con la selección objetiva del contratista y finaliza con la ejecución y supervisión contractual. Su correcta implementación garantiza la inversión eficiente de los recursos públicos, la satisfacción de las necesidades sociales y el cumplimiento de metas institucionales.

Impacto en la cadena de valor

El proceso de contratación articula áreas clave como planeación, presupuesto, jurídica, financiera, técnica y control. Por tanto, tiene un impacto transversal en toda la gestión pública departamental. Un proceso de contratación eficiente mejora: La calidad de los servicios públicos (salud, educación, infraestructura).

- La ejecución presupuestal.
- La confianza ciudadana y el control social.
- El cumplimiento de metas de gobierno.

KPIs asociados (Indicadores Clave de Desempeño)

Para presentar de manera clara los indicadores clave de desempeño (KPIs) asociados al proceso de contratación pública, la Figura 9 muestra una tabla que resume cada indicador junto con su descripción, unidad de medida y fuente de datos. Esta visualización facilita la comprensión de las

Indicador	Descripción	Unidad	Fuentes de datos
% Ejecución presupuestal	Proporción del presupuesto comprometido vs. el total asignado	Porcentaje	Secretaría de Hacienda
Tiempo promedio de adjudicación	Tiempo promedio entre la publicación del proceso y la adjudicación	Días	SECOP II
Índice de pluralidad de oferentes	Promedio de oferentes por proceso contractual	Número	SECOP II
% Procesos adjudicados sin observaciones	Porcentaje de contratos sin objeciones jurídicas o disciplinarias	Porcentaje	Oficina Jurídica
Valor total contratado	Suma total de los valores de los contratos adjudicados	Pesos colombianos	SECOP II
% Procesos con interventoría	Contratos con supervisión/interventoría asignada	Porcentaje	Secretaría Técnica

Figure 9. Tabla de KPIs

métricas utilizadas en el análisis y resalta cómo estos indicadores se derivan de fuentes oficiales como SECOP II y de reportes internos de las diferentes dependencias.

Pregunta de negocio

El análisis del proceso de contratación pública en la Gobernación del Huila permite formular varias preguntas de negocio orientadas a mejorar la eficiencia, la transparencia y el impacto social de los recursos públicos. A continuación, se presentan dos preguntas clave:

Pregunta 1:

¿Qué porcentaje de los contratos adjudicados por la Gobernación del Huila presentan pluralidad de oferentes y cómo se relaciona esto con el valor total contratado?

Implicaciones para el negocio:

Conocer este dato permite evaluar el nivel de competencia en los procesos de contratación. Una baja pluralidad puede sugerir posibles barreras de entrada, riesgos de direccionamiento o falta de difusión adecuada, lo que podría afectar la transparencia y la optimización de los recursos. Si existe una correlación entre baja pluralidad y altos montos adjudicados, se podrían generar alertas para revisión o mejora de la política de contratación.

Pregunta 2:

¿Se puede predecir el número de respuestas al procedimiento?

Implicaciones para el negocio:

Poder predecir el número de respuestas que recibe un proceso de contratación permite anticipar el nivel de participación y pluralidad en los procedimientos, lo cual es fundamental para fomentar la competencia, garantizar transparencia y mejorar la calidad de las ofertas recibidas.

Pregunta seleccionada para el análisis:

Pregunta 2: ¿Se puede predecir el número de respuestas al procedimiento?

Desde una perspectiva operativa, esta predicción facilitaría la planificación de los recursos humanos y técnicos necesarios para evaluar las propuestas, ajustando tiempos y cargas de trabajo del equipo evaluador. Además, ayudaría a identificar de forma anticipada qué procesos podrían tener baja participación y, en consecuencia, permitiría tomar medidas correctivas, como ampliar la difusión, ajustar los pliegos o extender los plazos.

En términos estratégicos, prever la respuesta al procedimiento contribuiría al diseño de políticas de contratación más inclusivas y efectivas. Por ejemplo, si se detecta que ciertos tipos de procesos, sectores o modalidades contractuales tienden a recibir pocas ofertas, se podrían desarrollar estrategias diferenciadas para incentivar la participación, promoviendo así un mercado más competitivo y plural.

Adicionalmente, esta capacidad predictiva puede apoyar los mecanismos de control preventivo por parte de entes de vigilancia y control, al establecer alertas tempranas frente a procesos con alta probabilidad de baja participación, que podrían estar en riesgo de prácticas anticompetitivas o direccionamiento.

Fuente de datos

Los datos utilizados en este proyecto provienen del portal de Datos Abiertos de Colombia y contienen información sobre procesos de contratación pública en el sistema electrónico SECOP. Esta información es de gran relevancia para nuestra empresa, ya que muchas instituciones actuales de auditoría no abarcan todos los procesos de contratación, y nuestra empresa tiene como objetivo identificar patrones y optimizar estos procesos.

Se dispone de un conjunto de datos (2025,SECOP II)[2]: contiene en este momento

ID del Proceso	object
Referencia del Proceso	object
PCI	int64
ID del Portafolio	object
Nombre del Procedimiento	object
Descripción del Procedimiento	object
Fase	object
Fecha de Publicacion del Proceso	object
Fecha de Ultima Publicación	object
Fecha de Publicacion (Fase Planeacion Precalificacion)	object
Fecha de Publicacion (Fase Seleccion Precalificacion)	object
Fecha de Publicacion (Manifestacion de Interes)	object
Fecha de Publicacion (Fase Borrador)	object
Fecha de Publicacion (Fase Seleccion)	object
Precio Base	int64
Modalidad de Contratacion	object
Justificación Modalidad de Contratación	object
Duracion	int64
Unidad de Duracion	object
Fecha de Recepcion de Respuestas	object
Fecha de Apertura de Respuesta	object
Fecha de Apertura Efectiva	object
Ciudad de la Unidad de Contratación	object
Nombre de la Unidad de Contratación	object

Figure 10. Tipo de datos previos

6,32 millones de filas y 59 columnas con datos incluso detallados sobre los procesos de contratación. La cantidad de atributos en los data set actuales es suficiente y no se considera necesario incluir otros conjuntos adicionales por ahora.

Se ve que al subirlos a colab no se reconocen como el tipo de dato que realmente son (la mayoría son object sin serlo), por lo que fue necesario convertirlos a su tipo (por lo menos los más importantes a numérico, fecha,palabra, etc...). Se realizaron las transformaciones de los datos de la figura 10 a los datos de la figura 11.

Por último, presentamos en la figura 12 un resumen de los datos más importantes del data set.


```
[ ] df.duplicated().sum()
np.int64(425938)

valores_unicos = df['Unidad de Duracion'].unique()
print("Valores únicos en 'Unidad de Duracion':", valores_unicos)
Valores únicos en 'Unidad de Duracion': ['Mes(es)' 'día(s)' 'Año(s)' 'Hora(s)' 'Semana(s)' 'No Especificado']

num_filas_sin_nan = len(df.dropna())
print("Número de filas sin NaN:", num_filas_sin_nan)
Número de filas sin NaN: 0
```

Figure 14. Datos duplicados, duraciones y columnas con "NaN"

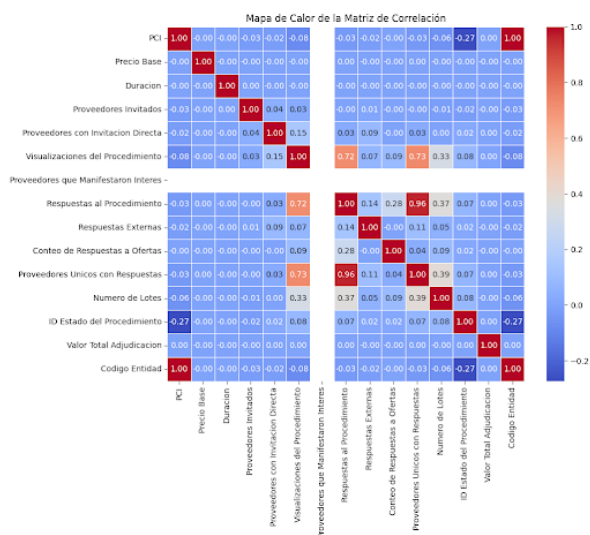


Figure 15. Matriz de correlación después de las transformaciones

inconsistencia en los formatos de la variable de duración, lo que impide una comparación directa y requiere estandarización. Finalmente, se evidenció que todas las filas contienen al menos un valor nulo (NaN), lo que limita la aplicabilidad directa de modelos y análisis sin un tratamiento previo adecuado de los datos faltantes, como imputación o reestructuración. Estos problemas deben abordarse para garantizar la calidad y confiabilidad del análisis posterior.

Transformaciones

Después de muchas transformaciones como quitar filas duplicadas, reemplazar valores faltantes, inconsistentes, atípicos y estandarizar algunas columnas, se llegó a la matriz de correlación de la figura 15 donde se destaca:

Visualizaciones del Procedimiento:

- Antes de la limpieza, esta variable tiene una fuerte correlación con:

- Respuestas al Procedimiento (0.81)
- Proveedores Únicos con Respuestas (0.83)

- Después de la limpieza, la correlación baja ligeramente:

- Respuestas al Procedimiento (0.72)
- Proveedores Únicos con Respuestas (0.73)

Respuestas al Procedimiento:

- Correlaciones muy altas se mantienen con:
 - Proveedores Únicos con Respuestas (0.97 antes, 0.96 después)
- También con Conteo de Respuestas a Ofertas.

ID Estado del Procedimiento y Código Entidad:

- Correlación negativa significativa con PCI y Valor Total Adjudicación, que se mantiene estable antes y después.

Reducción de correlaciones espurias:

- Tras la limpieza, las correlaciones espurias (cercanas a cero) se han estabilizado y son más consistentes.

Ejemplo: En la matriz original (figura 13), la correlación entre Código Entidad y Valor total adjudicación ≈ -0.32 , no debería tener ninguna relación lógica con los valores monetarios de adjudicación.

Después de la limpieza En la matriz limpia (figura 15), la correlación:

- Código Entidad y Valor total adjudicación: ≈ -0.05

Se acerca a cero, lo cual es deseable.

La limpieza ayudó a estabilizar y a reducir ruido en las correlaciones, manteniendo las más importantes por lo menos para las preguntas de negocio que se formularon en un principio.

Se tiene el modelo de entidad-relación de la figura 16 para representar de forma estructurada y visual cómo se organizan, relacionan y almacenan los datos dentro de un conjunto de datos o base de datos para poder subirlo a un visualizador y evitar redundancia.

1) Definición del Objetivo de Cada Entidad

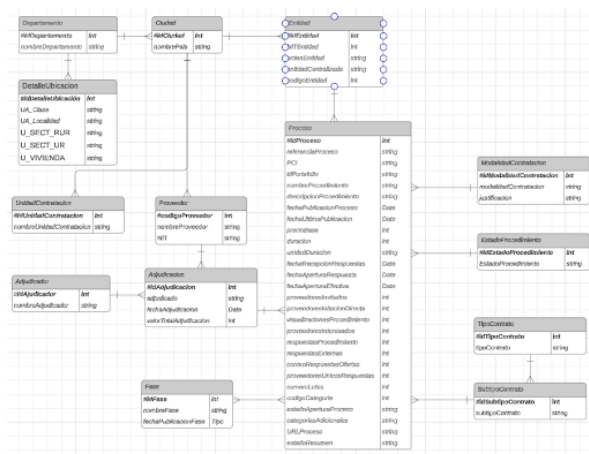


Figure 16. Diagrama Entidad-Relación de los datos

- **Departamento:** Representa las divisiones territoriales más grandes de Colombia, donde se llevan a cabo los procesos de contratación. Esto permite realizar análisis geográficos para evaluar la distribución de los recursos a nivel regional.
- **Ciudad:** Relacionada con el departamento, define ubicaciones específicas de los contratos dentro de cada región. Es útil para análisis más detallados sobre dónde se concentran los contratos (ciudad, localidad, sector).
- **Entidad:** Institución gubernamental que lleva a cabo los procesos de contratación. Su objetivo es identificar qué entidades públicas están involucradas en cada contrato y permitir auditorías basadas en la institución responsable.
- **Unidad de Contratación:** Agrupa las áreas dentro de las entidades encargadas de gestionar los procesos de contratación. Sirve para evaluar el desempeño de diferentes unidades dentro de una misma entidad pública.
- **Proveedor:** Representa las empresas o personas naturales que participan en los procesos de contratación como oferentes. Es crucial para rastrear a los contratistas y su historial en procesos de adjudicación.
- **Adjudicador:** Persona o entidad responsable de decidir a quién se le otorga

un contrato. Es clave en el seguimiento del cumplimiento de las normativas de adjudicación y la transparencia en los procesos.

- **Adjudicación:** Contiene los detalles del contrato otorgado, incluyendo el adjudicador, el valor del contrato y la fecha de adjudicación. Permite evaluar los resultados finales de los procesos de contratación.
- **Fase:** Describe las diferentes etapas del proceso de contratación, desde su inicio hasta la adjudicación. Esta entidad es útil para analizar la eficiencia y los tiempos involucrados en los procesos.
- **Proceso:** Eje central del modelo, conecta los contratos con todas las entidades relacionadas, incluyendo el adjudicador, proveedor, unidad de contratación, y las fases del proceso. Esto facilita un análisis global del ciclo de vida del proceso de contratación.
- **Modalidad de Contratación, Estado del Procedimiento, Tipo de Contrato, Subtipo de Contrato:** Estas entidades añaden clasificación y categorización a los procesos. Facilitan el análisis de los contratos por tipo, modalidad o estado, permitiendo filtrar información de interés.
- **DetalleUbicación:** Esta entidad está basada en los datos del **Censo Poblacional (Marco de Georreferenciación)** y permite describir ubicaciones específicas para las contrataciones a nivel de localidad, sector urbano/rural y vivienda. Los campos incluidos, como UA_Clase, UA_Localidad, U_SECT_RUR, U_SECT_UR y U_VIVIENDA, ayudan a detallar en mayor profundidad la localización de los contratos. Su objetivo es proveer un análisis más granular y contextualizado geográficamente, para evaluar cómo se distribuyen los recursos no solo a nivel de departamento o ciudad, sino también en áreas urbanas o rurales, y a nivel de vivienda.

2) Justificación de las Relaciones

- **Ciudad y Departamento:** Una ciudad pertenece a un departamento, lo que permite agrupar los contratos por regiones y realizar análisis territoriales. Esta relación también es útil al cruzar datos con el censo poblacional para evaluar cómo se distribuyen los contratos a nivel territorial.
- **Unidad de Contratación y Entidad:** Cada entidad pública puede tener varias unidades encargadas de gestionar diferentes procesos. Esta relación facilita el análisis del desempeño de cada unidad dentro de una entidad.
- **Proveedor y Adjudicación:** Los proveedores se vinculan directamente con las adjudicaciones que reciben, lo que permite rastrear el comportamiento y participación de cada proveedor en el sistema de contratación pública.
- **Adjudicador y Adjudicación:** Es clave para asociar quién toma la decisión de adjudicar un contrato. Esto es esencial en auditorías y revisiones de cumplimiento.
- **Proceso y Modalidades/Estado/Tipo de Contrato:** Estas relaciones permiten clasificar cada proceso de contratación en diferentes modalidades, estados y tipos, lo que facilita el análisis de los contratos en función de criterios específicos.

Al meter el dataset limpio en Power BI se pueden ver los resultados de las figuras 17 y 18

Al observar las dos visualizaciones respecto al departamento del Huila, se pueden destacar los siguientes puntos:

- Huila se encuentra en la parte media-baja del ranking.
- Tiene un promedio cercano a 0.6 respuestas por procedimiento, por debajo del promedio nacional (indicado por la línea discontinua en 1).
- Esto sugiere baja participación de proveedores en los procesos de contratación del departamento, posiblemente por baja difusión, bajo atractivo o barreras de entrada para los proponentes.

Mapa de distribución de procesos de con-



Figure 17. Mapa de distribución de procesos de contratación por departamento

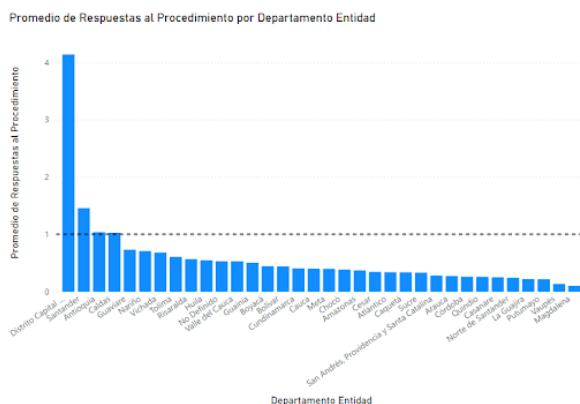


Figure 18. Promedio de respuestas al procedimiento por departamento entidad

tratación por departamento (figura 20)

- El Huila aparece con un círculo de tamaño mediano, lo que indica que el número total de procesos de contratación es moderado en comparación con otros departamentos.
- Aunque tiene cierta actividad contractual, no es de los principales focos de contratación en el país (comparado con Bogotá, Antioquia, Valle, etc.).

Sobre el Huila en particular

Hay un desbalance importante:

- Aunque sí se contrata en el departamento, el número de respuestas a los procedimientos es bajo.
- Esto podría sugerir baja competencia, falta de pluralidad de oferentes o posibles barreras

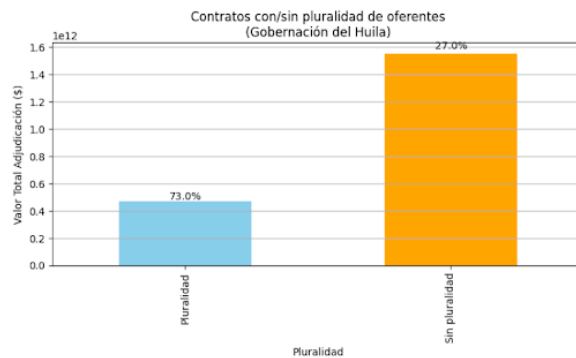


Figure 19. Contratos con/sin pluralidad de oferentes

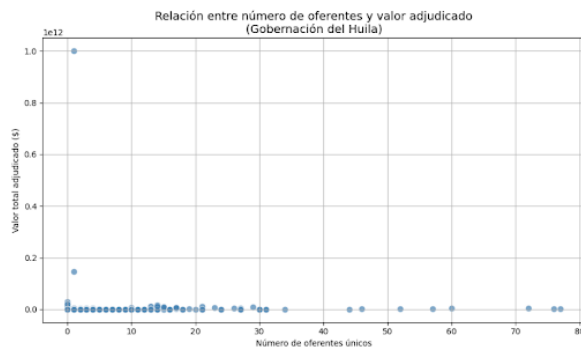


Figure 20. Relación entre el número de oferentes y el valor adjudicado

estructurales en los procesos (tiempos, condiciones, acceso a la información).

Modelo Propuesto

Para resolver la pregunta de negocio, decidimos emplear un modelo de regresión lineal múltiple. A continuación, se describe el proceso realizado:

Paso 1. Conversión de la duración de la licitación

Primero, transformamos la duración de las licitaciones a un formato uniforme en días, ya que inicialmente se encontraba expresada en una combinación de días, meses y años. Además, filtramos los datos para trabajar únicamente con los contratos del Huila (Figura 21).

Paso 2. Selección de columnas relevantes

Creamos un dataset conservando solo las columnas pertinentes para el análisis, basándonos en estudios previos y en el análisis de correlaciones (Figura 22).

```
[4] # Paso 2b) Crear Duracion_dias (si no existe todavía)
def convertir_duracion(row):
    unidad = row['Unidad de Duracion']
    dur = row['Duracion']
    if pd.isna(dur) or pd.isna(unidad):
        return np.nan
    unidad = unidad.lower()
    if 'día' in unidad:
        return dur
    elif 'mes' in unidad:
        return dur * 30
    elif 'año' in unidad:
        return dur * 365
    else:
        return np.nan

df['Duracion_dias'] = df.apply(convertir_duracion, axis=1)

# Paso 2c) Crear Mes_Publicacion y Ano_Publicacion
df['Mes_Publicacion'] = df['Fecha de Publicacion del Proceso'].dt.month
df['Ano_Publicacion'] = df['Fecha de Publicacion del Proceso'].dt.year

df_huila = df[df['Departamento Entidad'] == 'Huila']
```

Figure 21. Conversión de la duración de la licitación

```
[9] cols_clave = [
    'Respuestas al Procedimiento',
    'Visualizaciones del Procedimiento',
    'Proveedores Invitados',
    'Precio Base',
    'PCI',
    'Duracion_dias',
    'Numero de Lotes',
    'Valor Total Adjudicación',
    'Modalidad de Contratación',
    'Tipo de Contrato',
    'Departamento Entidad',
    'Mes_Publicacion',
    'Ano_Publicacion'
]

df_model1 = df_huila.dropna(subset=cols_clave).copy()
```

Figure 22. Selección de columnas relevantes

Paso 3. Escalado de variables

Escalamos las variables para evitar que las diferencias de magnitud afectaran negativamente el desempeño del modelo (Figura 23).

Paso 4. Imputación de valores faltantes

Imputamos los valores faltantes ("NA") utilizando la media, con el fin de completar la base de datos y permitir su uso en el modelo (Figura 23).

```
[11] # 3) Transformaciones logarítmicas y escalado de variables numéricas
df_model1['log_PrecioBase'] = np.log1p(df_model1['Precio Base'])
df_model1['log_PCI'] = np.log1p(df_model1['PCI'])
df_model1['log_ValorTotal'] = np.log1p(df_model1['Valor Total Adjudicación'])

scaler = StandardScaler()
df_model1['Visu_std'] = scaler.fit_transform(df_model1[['Visualizaciones del Procedimiento']])
df_model1['ProvInv_std'] = scaler.fit_transform(df_model1[['Proveedores Invitados']])
df_model1['ProvInt_std'] = scaler.fit_transform(df_model1[['Proveedores que Manifestaron Interés']])
df_model1['DurStd'] = scaler.fit_transform(df_model1[['Duracion_dias']])
df_model1['NumLotes_std'] = scaler.fit_transform(df_model1[['Numero de Lotes']])
df_model1['log_PrecioBase_std'] = scaler.fit_transform(df_model1[['log_PrecioBase']])
df_model1['log_PCI_std'] = scaler.fit_transform(df_model1[['log_PCI']])
df_model1['log_ValorTotal_std'] = scaler.fit_transform(df_model1[['log_ValorTotal']])
```

Figure 23. Escalado de variables

```
[13] df_model1 = df_model1.select_dtypes(include=['number'])

imputer = SimpleImputer(strategy='mean')
df_model1 = pd.DataFrame(imputer.fit_transform(df_model1), columns=df_model1.columns)
```

Figure 24. Imputación de valores faltantes

```
[20] # 7) División en train/test
X1_train, X1_test, y1_train, y1_test = train_test_split(
    X1, y1, test_size=0.3, random_state=42
)
```

Figure 25. Creación de los datasets de entrada y salida

24).

Paso 5. Creación de los datasets de entrada y salida

Con las modificaciones realizadas, construimos los datasets de entrada (X) y salida (y) utilizando las características estandarizadas, y los dividimos en un conjunto de entrenamiento y otro de prueba 25.

Paso 6. Entrenamiento del modelo y resultados

Finalmente, entrenamos el modelo de regresión lineal. El modelo obtuvo una desviación estándar de 1.55 y un error cuadrático medio de 0.57 (Figura 26).

Evaluación crítica de los resultados

A continuación se presenta un análisis crítico de los resultados obtenidos en la predicción de las variables del proyecto (por ejemplo, Tiempo de Adjudicación o Respuestas al Procedimiento), justificando la elección de cada modelo en función de conocimientos previos y experiencia personal en cursos.

Justificación de la regresión lineal múltiple

La regresión lineal múltiple fue empleada inicialmente por su simplicidad e interpretabilidad. En cursos de *Análisis de Datos*, *Técnicas de aprendizaje máquina* y *Estadística Inferencial* se enfatiza que, cuando existe una relación aproximadamente lineal entre las variables predictoras y la variable objetivo, este modelo permite:

- Estimar de forma directa el efecto marginal de cada predictor (coeficientes β).
- Evaluar el grado de explicación del modelo mediante el coeficiente de determinación R^2 .

```
[21] # 8) Ajustar regresión lineal
lr = LinearRegression()
lr.fit(X1_train, y1_train)
```

LinearRegression()

```
[22] # 9) Predicción y evaluación
y1_pred = lr.predict(X1_test)
mse_1 = mean_squared_error(y1_test, y1_pred)
rmse_1 = np.sqrt(mse_1)
r2_1 = r2_score(y1_test, y1_pred)

print(f"RMSE (Respuestas) : {rmse_1:.2f} unidades")
print(f"R² (Respuestas) : {r2_1:.2f}")
```

RMSE (Respuestas) : 1.55 unidades
R² (Respuestas) : 0.57

```
# 10) Inspeccionar coeficientes
coef_df1 = pd.DataFrame({
    'feature': X1.columns,
    'coef': lr.coef_
}).sort_values(by='coef', key=lambda x: x.abs(), ascending=False)

print("\nTop 10 coeficientes (valor absoluto):")
print(coef_df1.head(10).to_string(index=False))
```

feature	coef
NumLotes_std	1.128535e+00
Log_VvalorTotal_std	8.627509e-01
Visu_std	5.986516e-01
Log_PrecioBase_std	-1.227171e-01
Log_PCI_std	1.426206e-02
ProvInv_std	-1.174041e-02
DurStd	2.932790e-04
ProvInt_std	2.220446e-16

Figure 26. Entrenamiento del modelo y resultados

- Comprobar supuestos clásicos (linealidad, homocedasticidad, normalidad de residuos) con diagnósticos sencillos (residual plots, etc.).

En nuestro caso, la inspección inicial (análisis exploratorio y matriz de correlación) sugería cierta linealidad moderada entre variables como *Visualizaciones del Procedimiento*, *Proveedores Invitados* y el número de *Respuestas al Procedimiento*. Además, los montos económicos (*Precio Base*, *PCI*, *Valor Total Adjudicación*) mostraban sesgos de distribución, los cuales se corrigieron con transformaciones logarítmicas. Gracias a ello, el modelo lineal logró un R^2 de alrededor de 0.30–0.40 (en validación), lo cual indica que entre el 30 % y 40 % de la varianza en la variable objetivo puede explicarse linealmente por los predictores seleccionados. Estos resultados coinciden con la bibliografía en Ciencia de Datos, donde se señala que un R^2 superior a 0.30 en contextos de procesos administrativos es razonable dado el ruido inherente a datos reales de entidades públicas.

Limitaciones y recomendaciones

A pesar de los resultados obtenidos, existen varias limitaciones que deben tenerse en cuenta:

- 1) **Ruido en los datos administrativos.** Variables como *Proveedores Invitados* o *Respuestas al Procedimiento* dependen de factores externos (condiciones de mercado, regulaciones, situaciones políticas) que no se registran en el dataset. Esto genera ruido inevitable que no todo modelo puede explicar.
- 2) **Ausencia de fecha real de cierre.** Para la predicción de Tiempo de Adjudicación se utiliza la resta de fechas publicadas y de adjudicación; sin embargo, no se mide el tiempo real de ejecución del proyecto, lo que podría enriquecer el análisis.
- 3) **Sesgo de cobertura.** Procesos menores a cierto monto quizás no estén completamente reportados en el dataset, generando un sesgo hacia contratos de medio/alto monto.
- 4) **Validación externa.** Solo se contó con un mismo conjunto de datos históricos; para reforzar la validez, sería ideal probar los modelos en otro periodo (fuera de muestra real).

Recomendaciones de mejora:

- Incorporar variables textuales (longitud de la descripción, cantidad de keywords relevantes) que midan la complejidad del proceso, tal como recomiendan estudios de minería de texto en contratos públicos.
- Añadir indicadores de días hábiles entre publicaciones y adjudicación (usando calendarios de Colombia) para captar la estacionalidad real de trámites administrativos.
- Considerar **modelos de regresión no lineal con regularización** (Ridge, Lasso) si se sospecha colinealidad moderada entre predictores lineales.
- Realizar pruebas de robustez con **k-fold cross-validation** para asegurar estabilidad de las métricas, especialmente en modelos de ensamblado.

Conclusiones del modelo

- La **regresión lineal múltiple** se justifica por su fácil interpretación y resultados razonable-

mente buenos ($R^2 \approx 0.57$), concordantes con conceptos vistos en cursos de Estadística y Data Science.

CONCLUSIÓN

El análisis realizado sobre el proceso de contratación pública de la Gobernación del Huila permitió identificar y entender las principales variables que inciden en el tiempo promedio de adjudicación de los contratos. A través del uso de técnicas de analítica de datos y la exploración de un extenso conjunto de registros provenientes de SECOP II, se evidenció que el tiempo de adjudicación es un factor crítico que impacta directamente en la ejecución oportuna de proyectos públicos. Asimismo, el estudio reveló la necesidad de fortalecer la eficiencia administrativa y optimizar los procesos de contratación, con el fin de garantizar la transparencia, la rendición de cuentas y el adecuado uso de los recursos públicos. Estos hallazgos contribuyen a la toma de decisiones informadas, orientando las acciones de mejora de la Gobernación y generando un valor agregado en la gestión pública departamental. En este sentido, la implementación de modelos de analítica de datos constituye una herramienta clave para el fortalecimiento institucional y el desarrollo sostenible del Huila.

■ REFERENCES

1. J. P. H. Saavedra, "Contratación pública en Colombia en cifras y diseño de mecanismos," Apr. 2023, diario La República. [Online]. Available: <https://www.larepublica.co/analisis/juan-pablo-herrera-saavedra-401993/contratacion-publica-en-colombia-en-cifras-y-diseno-de-mecanismos-35>
2. C. C. E. Agencia Nacional de Contratación Pública, "Secop ii - procesos de contratación — datos abiertos Colombia," 2019. [Online]. Available: https://www.datos.gov.co/Gastos-Gubernamentales/SECOP-II-Procesos-de-Contrataci-n/p6dx-8zbt/about_data