

**Caso de Estudio - Proyecto de Grado  
2022-2023**

***Datos Abiertos en la Contratación Pública***

Los estados, para llevar a cabo su misión, compran y contratan toda clase de bienes y servicios; desde aquellos relacionados con su operación cotidiana hasta las grandes inversiones orientadas a suplir necesidades básicas de la población y perseguir objetivos de desarrollo del país. Por tratarse del uso de recursos públicos, la pertinencia, competencia e integridad en la ejecución de estas compras y contratos es del interés de la ciudadanía, así como de agencias del gobierno, entes de control, o sectores relacionados con la provisión de bienes y servicios específicos (ej., medicamentos, construcción). La existencia de datos abiertos sobre compra y contratación pública sugiere posibilidades interesantes, como que la sociedad civil tenga insumos para evaluar el gasto de un gobierno (ej., frente al programa con el que fue elegido), o que agencias del gobierno, entes de control, periodismo investigativo u otras organizaciones puedan identificar fallas, ineficiencias o irregularidades en el uso de recursos públicos y, eventualmente, determinar mecanismos para reducirlas. En la práctica, sin embargo, es preciso resolver retos tecnológicos y organizacionales, entre otros, para lograr el nivel de madurez analítica que requeriría abordar preguntas de alto impacto e interés desde los datos. Este caso de estudio se plantea como un ejercicio participativo para abordar preguntas de interés sobre contratación pública en Colombia, buscando determinar en qué medida los datos abiertos permiten resolverlas, o qué obstáculos se deben solventar para lograrlo.

## Introducción

Los datos abiertos tienen implícita la idea de que existe un público para el cual sería relevante consumir información sobre las actividades que generan dichos datos. La contratación pública no es la excepción, pues diversos actores tienen interés en diferentes aspectos del uso de recursos públicos; por ejemplo, interesa saber sobre el propósito del gasto (pertinencia), la eficiencia económica y administrativa con que se hace (competencia), así como la transparencia, legalidad y ética en el proceso (integridad). Abordar preguntas sobre estos intereses a partir de los datos, sin embargo, no es una tarea trivial, ya que el sistema de compra y contratación pública es complejo, tanto a nivel jurídico y administrativo, como en relación con las plataformas tecnológicas y organizaciones involucradas en su funcionamiento. Entender qué procesos analizar, bajo qué lógica y métricas, y con cuáles datos, son preguntas de fondo que requieren entender el sistema y que, por lo general, implican resolver retos técnicos relacionados con el acceso, calidad, e interoperabilidad de los datos, antes de pensar en modelos y análisis. El objetivo de este caso de estudio es evaluar en qué medida los datos abiertos sobre contratación y compra pública en Colombia permiten enriquecer la discusión sobre los intereses mencionados (más que generar “análisis definitivos”).

Colombia cuenta con notables avances en plataformas de [datos abiertos](#), particularmente sobre [contratación pública](#). La [Agencia Nacional de Contratación Pública](#) (ANCP) administra las plataformas de contratación y define [manuales, guías y documentos tipo](#) para mejorar las prácticas de contratación y compra de las entidades del Estado. Aun así, existen escándalos de [corrupción](#), malas prácticas en gerencia de proyectos, o ineficiencias que contribuyen a una percepción de desconfianza en las instituciones. La existencia de datos abiertos debe permitir, por lo menos, proveer evidencia para enriquecer las complejas discusiones que pueden generar estos problemas. Como fuentes de consulta sobre el uso de datos abiertos en asuntos públicos, la organización *Open Contracting Partnership* ([OCP](#)) define [principios, estándares y prácticas](#) para el logro de una contratación pública abierta, justa y eficiente. Con mayor foco en el aspecto de integridad, el *Programa Interamericano de Datos Abiertos Anticorrupción* ([PIDA](#)) establece una [hoja de ruta](#) respecto al uso de [conjuntos de datos](#) para: (i) promover una cultura de la transparencia; y (ii) detectar y prevenir la corrupción. Por su parte, el [Instituto Anticorrupción](#) desarrolla investigaciones sobre transparencia, integridad y rendición de cuentas, apoyando la discusión y política pública sobre corrupción.

El sector de la infraestructura, por su importancia para el desarrollo y competitividad del país, y debido a los [complejos proyectos y grandes inversiones](#) que requiere, está expuesto a riesgos que lo hacen vulnerable a malas gerenciales (que se reflejan en problemas de calidad, demoras y sobrecostos), así como a prácticas cuestionables por parte de contratantes y contratistas (favorecimiento a un contratista, mala distribución de riesgos). Iniciativas como [Infraestructura Visible](#) (*InfraViz*) o el mismo [OCP](#) persiguen la transparencia y eficiencia en proyectos de provisión de infraestructura que utilizan grandes recursos públicos, a menudo ejecutados por privados que *deben* balancear el beneficio común con sus propios intereses. Actualmente, *InfraViz* reporta datos sobre el estado de la infraestructura en el país, pero no refleja actividad de contratación al respecto, entre otras, debido a las dificultades antes mencionadas. Los datos abiertos sobre contratación pública ofrecen la oportunidad de complementar el panorama que presenta *InfraViz* con información respecto a cómo se está haciendo el gasto público con respecto a ese importante sector.

## Actores involucrados

Desde el punto de vista del ciudadano, algunas preguntas fundamentales apuntan a si las inversiones del gobierno son las que “deberían ser” (ej., acorde a un programa de gobierno, o criterio de quien consulta), en qué medida las inversiones logran su cometido adecuadamente (según alguna métrica de desempeño o beneficio), o si hay indicios de ineficiencias o irregularidades (intencionales o no). Para el sector de la infraestructura (que pasa por temas de desarrollo sostenible, manejo de riesgo, gerencia de la construcción, entre otros), interesa poder tener un panorama claro del estado y desempeño de la contratación de proyectos de infraestructura del país. Por ejemplo, los datos abiertos *deberían* permitir examinar fácilmente la composición de la contratación en infraestructura<sup>1</sup>. Asimismo, un siguiente grado de procesamiento de los datos *debería* permitir explorar qué aspectos<sup>2</sup> de los proyectos de infraestructura se relacionan con mejores o peores indicadores de desempeño (ej., sobrecostos, demoras), o cuáles están relacionados con factores de alerta (ej., índices de concentración, riesgo de corrupción).

La ANCP, en su misión de velar por el uso eficiente de los recursos públicos y la transparencia en los procesos de contratación, ha expuesto como preguntas de su interés: ¿cómo es el comportamiento de los contratos (de prestación de servicios) por departamento y municipio, descubriendo dinámicas de los proveedores, como es el caso de adiciones, contratos recurrentes, varios contratos de corta duración, modificaciones de los contratos y contratistas con varios contratos?, ¿cómo es el comportamiento de proveedores en modalidades competitivas o sectores específicos para identificar dinámicas de oferta por regiones?, ¿cuál es el comportamiento, oferta y demanda del Régimen Especial en Colombia?, ¿cuál es la probabilidad de selección de un proveedor para ejecutar Órdenes de Compra en Acuerdos Marco de Precios de la TVEC, de acuerdo con el comportamiento histórico? En particular, contratistas y contratantes, como jugadores directos en el sistema, pueden estar interesados en qué tan buen “desempeño” o reputación tienen sus contrapartes (contratantes y contratistas), por ejemplo, en términos de eficiencia, incumplimiento, o concentración de actores con quienes han trabajado previamente, entre otros.

Desde el punto de vista de los entes control y organizaciones veedoras o activistas el foco del uso de datos abiertos está en la detección y prevención de actos irregulares que comprometan recursos públicos. Este es un tema delicado, tanto por la limitación en el acceso a datos de tipo judicial o sancionatorio, como por la responsabilidad de mantener la presunción de inocencia como parte del análisis y comunicación. En el mejor de los casos, se busca contar con banderas rojas que permitan priorizar la revisión manual de casos en los que algún índice (ej., riesgo de corrupción, o concentración) genere indicios de anomalías.

Como problemas comunes, relacionados con la calidad y el acceso a los datos, se tiene el reto de cómo llevar a la práctica: algoritmos de identificación de valores atípicos y perfilado de datos; métodos de OCR y modelos de lenguaje para contrastar con contrato en digital; y modelos de agrupamiento para la identificación de duplicados.

---

<sup>1</sup> Por ejemplo, en términos de montos totales, participación de regiones, sectores, entidades, contratistas, tipos de obra, fuentes de financiación, modalidad, u otros factores de interés accesibles desde los datos.

<sup>2</sup> Por ejemplo, modalidad de selección, contexto electoral, características de los contratistas, del contratante, del proyecto.

## Alcance del proyecto

El propósito de este proyecto es poner en práctica las competencias adquiridas en análisis de datos en el contexto planteado, buscando pasar de simplemente tener datos disponibles sobre contratación a poder abordar preguntas de interés para al menos un actor relevante del problema. Para esto, cada equipo debe formular uno de los varios *problemas de negocio* que pueden plantearse en este contexto y, alrededor de éste, identificar por lo menos un producto analítico (i.e., un artefacto que responda a una *pregunta analítica*) cuya implementación sea viable.

Hacer efectivo el potencial de los datos abiertos en contratación implica retos complejos, pues a menudo es considerable la distancia entre las preguntas que interesan a los involucrados y las preguntas que es posible abordar desde los datos (ej., por problemas de interoperabilidad entre conjuntos de datos o no disponibilidad de ciertos datos estructurados). Adicionalmente, el problema incluye varios actores cuyas necesidades no necesariamente se pueden abordar con una única solución. En ese sentido, planteamos las siguientes pautas para acotar el problema, y hacer explícito el alcance esperado del proyecto:

- La existencia o acceso a los datos no garantiza transparencia, pues es necesario contar con habilidades técnicas para acceder a la información y procesarla para que sea consumible. Una pregunta común a todos los actores es *¿qué hay en los datos?*, pues para todos es de interés un análisis descriptivo global que dé cuenta de la información que se puede conocer a partir de los datos abiertos de contratación. Para ello, es preciso entender los procesos de interés, buscando análisis y visualizaciones que tengan valor semántico para los usuarios previstos.
- El reto descriptivo de la pregunta anterior se puede especializar, permitiendo a un usuario explorar datos sobre un sector, modalidad o territorio específico, según su interés. Por ejemplo: explorar diferencias en licitaciones de concesión y de obra pública puede ser útil para *InfraViz*; un enfoque en régimen especial o prestación de servicios es un interés manifiesto de la ANCP; mientras que la caracterización de actores en la TVEC puede ser de interés de contratantes y contratistas.
- La exploración anterior debe permitir identificar qué análisis tendría sentido plantear en función de los datos disponibles (o que se puedan adquirir oportunamente). Por ejemplo, aunque pueden plantearse diversas métricas de desempeño para contratos y proyectos, una elección estándar accesible desde los datos es la desviación en duración o costo respecto a lo pactado originalmente. Otras medidas, como índices de riesgo de corrupción, re-negociaciones, o beneficio alcanzado del proyecto suelen ser más difíciles de lograr, aunque son atractivos y existe literatura al respecto.
- Una utilidad clave del análisis de datos es identificar factores que puedan estar relacionados con la idoneidad de los procesos de contratación (i.e., qué caracteriza los contratos que tienen buen o mal desempeño, y cómo definir ese desempeño), lo cual puede conducir a aprendizajes que mejoren los procesos de contratación.
- Por último, un usuario rara vez necesita ver el código o todos los análisis intermedios que conducen a una solución. Por esto, es importante identificar cuál es el producto o servicio que se quiere brindar al cliente (ej., infografía, reporte, tablero interactivo, aplicación de *software-as-a-service*) según su necesidad y características (ej., presupuesto, habilidad analítica del personal). El entregable al final del curso *Proyecto Aplicado en Analítica de Datos* será un prototipo de dicho producto.

El problema planteado es complejo y todos los actores son relevantes. La recomendación es que, aunque es necesario pensar en el problema de forma amplia para estructurar el proyecto, el foco debe estar en identificar un problema analítico puntual que sea viable dentro del plazo y dedicación horaria de los cursos en los que se desarrolla. Para acotar el alcance del ejercicio, se propone como caso base el abordaje de dos de las necesidades expuestas previamente para el sector de la infraestructura: ¿qué dicen de interesante los datos abiertos sobre la composición de la contratación en infraestructura?; ¿qué aspectos de los proyectos de infraestructura, y los procesos de contratación correspondientes, pueden explicar el buen o mal desempeño de los mismos?

Una vez completado el caso base, los equipos pueden solicitar el aval del profesor encargado del caso para abordar retos más ambiciosos si, como parte de su ejercicio de estructuración del problema, logran demostrar la viabilidad de los mismos.