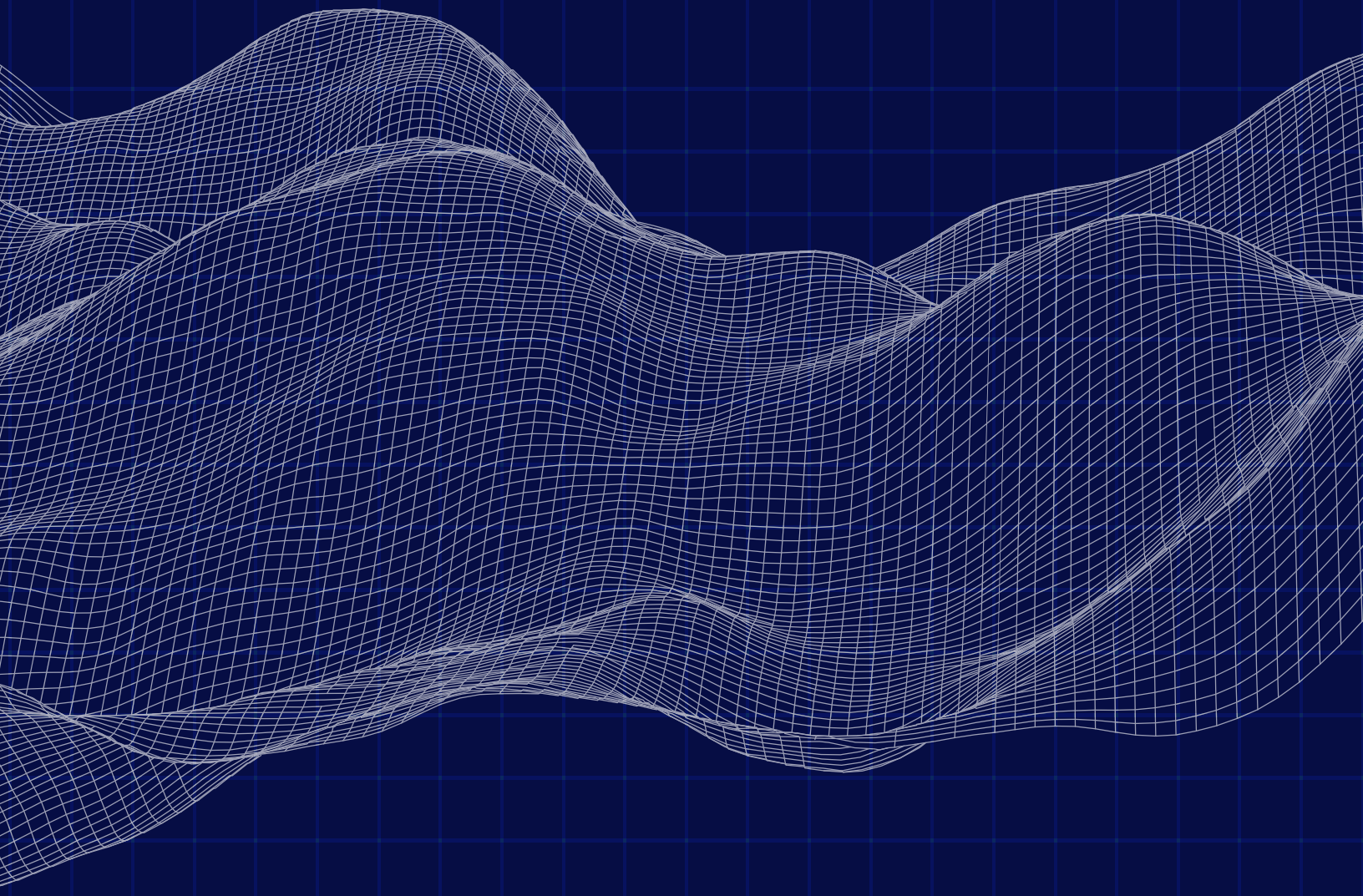
Gobierno de Colombia

IGAC
INSTITUTO GEOGRÁFICO AGUSTÍN CODAZZI

Boletín de Analítica N°3



Dirección de
INVESTIGACIÓN
y Prospectiva

Autores:
Oscar Romero
Mariana Ríos
Juan Pablo Gómez
Luis Andrés Campos

Construcción de un instrumento de visualización de los registros de transacciones inmobiliarias en Colombia.

Resumen

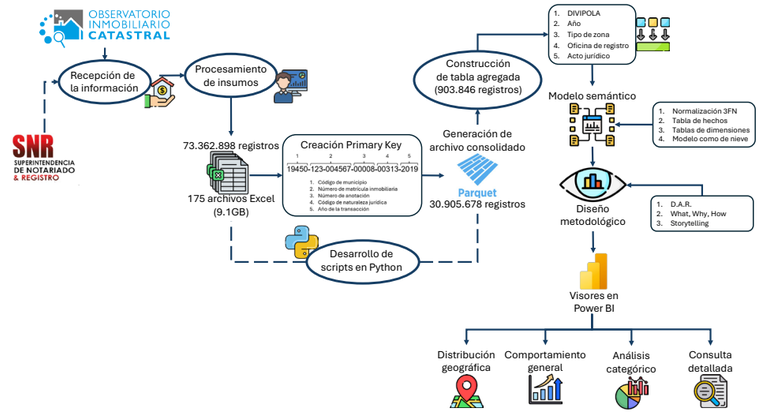
La información relativa a las transacciones registrales es fundamental para orientar procesos de análisis y toma de decisiones en materia de mercado inmobiliario en el país por parte de entidades públicas, particularmente aquellas que tienen funciones misionales de apoyo a la gestión catastral, organizaciones privadas, especialmente las vinculadas al sector inmobiliario, investigadores y ciudadanía en general interesada en compra o venta de vivienda en los diferentes municipios. Teniendo en cuenta lo anterior, la tercera entrega del Boletín de Analítica del Observatorio Inmobiliario Catastral – OIC del IGAC describe el proceso de construcción del instrumento de visualización de los registros de las transacciones jurídicas registradas en los folios de matrícula inmobiliaria, que ha implementado el OIC como una herramienta estratégica para apoyar estos procesos. De esta manera se continúa con la ruta de aprovechamiento de los datos de la Superintendencia de Notariado y Registro – SNR.

Introducción

Este boletín describe el proceso de consolidación y agregación aplicado a la información de transacciones inmobiliarias, con el objetivo de construir una herramienta para facilitar el análisis de la dinámica del mercado inmobiliario en el territorio nacional. Así las cosas, se integró y organizó una base de datos de aproximadamente 31 millones de registros, optimizando la consulta de esta información mediante filtros de interés específicos utilizando la herramienta Power BI.

Durante el desarrollo, se diseñó e implementó un modelo semántico que no solo organiza la información, sino que incrementa la eficiencia del visor de datos. En la Ilustración 1, se presenta el flujo del proceso para generar este instrumento de visualización y consulta. A continuación, se aborda la creación del tablero, la tabla agregada de datos, la estructura del modelo semántico y el diseño metodológico. Finalmente, se detallan las capas principales de visualización que conforman esta herramienta, disponible en el sitio web del OIC. Lo anterior proporciona un recurso valioso para el análisis de la dinámica inmobiliaria a partir de las transacciones, contribuyendo así al conocimiento y la toma de decisiones en el sector.

Ilustración 1 *Ruta de preprocesamiento, consolidación, agregación y generación de visualización de transacciones registrales.*



Fuente: *Observatorio Inmobiliario Catastral (2024).*

Construcción de la tabla agregada

El uso de una tabla agregada permite que los tableros de control sean más eficientes al reducir la cantidad de datos que se deben procesar en tiempo real. Al realizar las operaciones de agrupación y agregación previamente, los tableros pueden centrarse en la visualización de los datos y en ofrecer insights rápidos sin necesidad de cálculos complejos en cada consulta.

El objetivo de la construcción de esta tabla agregada es sintetizar el gran volumen de datos correspondientes a las transacciones registrales el cual asciende a más de 30 millones de datos, en un formato compacto que alcanza los 903.846 registros, lo que permite la creación de tableros analíticos ligeros y eficientes, mediante la consulta rápida de información relevante sobre las dinámicas de los predios en el tiempo y por diferentes categorías.

El proceso de construcción del modelo de datos para el análisis de transacciones inmobiliarias en Colombia utiliza PySpark, una herramienta poderosa para el procesamiento de grandes conjuntos de datos distribuidos de manera eficiente. La clave de este proceso es la función “GroupBy()”, que permite agrupar los datos según varios campos clave, tales como:

1. DIVIPOLA: indicador del municipio.
2. Año: año de radicación.
3. ORIP: Oficina de Registro de Instrumentos Públicos.
4. Acto jurídico: nombre del acto jurídico registral.
5. Tipo de zona: rural, urbano, sin información.
6. Categoría de ruralidad: Clasificación del DNP que diferencia las zonas del país.
7. Indicios de dinámica: actos jurídicos que representan cambios del mercado inmobiliario.
8. Transacciones con valor: indica si la transacción tiene valor asociado.
9. Folios nuevos: indica si la transacción es la primera anotación del folio de matrícula.

Una vez agrupados los datos, se aplican funciones que permiten sintetizar y resumir la información relativa al número de transacciones, la cual se obtiene a partir del conteo de los registros únicos, y su valor estimado, que se obtiene de la suma del valor promedio de las transacciones en cada nivel territorial. Esto es relevante para facilitar la interpretación y el análisis posterior de estas variables según los diferentes niveles de agregación.

Finalmente, los resultados se almacenan en un archivo Parquet para su fácil acceso y utilización en tableros de análisis, ya que son ideales por su cualidad de optimización para almacenamiento y lectura de datos, permitiendo que las consultas sean mucho más rápidas en comparación con otros formatos como CSV o JSON. Este enfoque facilita la creación de tableros ligeros que pueden responder rápidamente a las consultas del usuario, incluso cuando se manejan volúmenes de datos considerables.

Estructura del modelo semántico

La construcción de la tabla agregada permite una manipulación más sencilla, lo que facilitó el diseño de un modelo semántico que define cómo se relacionan y organizan los datos (Microsoft Learn, 2023). Este modelo ayuda a optimizar el rendimiento de las consultas a través de la compresión de datos y técnicas de carga diferida, asegurando que las visualizaciones y reportes sean más rápidos y eficientes.

Para tal fin se construyeron tablas de hechos y dimensiones, donde se identifican las columnas como atributos dimensionales y variables numéricas. Se crearon claves únicas para cada uno de los atributos dimensionales, conocidas como claves foráneas. Además, las tablas de dimensiones fueron normalizadas para evitar la redundancia de datos, facilitando así el análisis multidimensional.

La fase de normalización en el proyecto fue un proceso clave para asegurar la eficiencia y la precisión del modelo semántico, la estructuración de las tablas se desarrolló en tres etapas: la primera forma normal, busca asegurar que todos los campos en las tablas contengan valores únicos, esto implicó descomponer cualquier grupo de datos repetitivos o listas en valores simples; la segunda forma normal, elimina las dependencias parciales dentro de las tablas, cada atributo en una tabla debe depender completamente de la clave asignada; y, en la tercera forma normal se eliminan las dependencias transitivas, esto ayuda a evitar anomalías en la inserción, actualización o eliminación de datos en las tablas.

Finalmente, las tablas de hechos y dimensiones se guardan en nuevos archivos Excel para su posterior análisis o uso en herramientas de BI, en este caso, el modelo fue implementado en Power BI, donde se establecieron relaciones entre las tablas de hechos y dimensiones utilizando claves foráneas y primarias. Esta estructuración no solo permite realizar análisis profundos y dinámicos, sino que también mejora significativamente la velocidad de las consultas gracias al uso eficiente de tablas agregadas.

Diseño metodológico

El desarrollo del visor para la consulta del Registro de transacciones inmobiliarias en Colombia atiende las propuestas adelantadas por el OIC en materia de visualización y difusión, la cual contempla la adopción de una metodología de visualización robusta y dinámica, la metodología Dashboard, Analysis, Report –D.A.R.-, que fundamenta su enfoque de representación de datos (QlikView, 2013), en el que se exploran, analizan y se detallan, según se describe a continuación:

I. Exploración de Datos –Dashboard-: consiste en la generación de ayudas gráficas que permitan comprender el panorama general, así como identificar tendencias iniciales y posibles áreas de interés que requieren un análisis más profundo.

II. Análisis Específico -Analysis-: se trata de disponer ayudas gráficas que aporten valor y contexto para que el usuario pueda centrarse en los datos específicos que influyen en las decisiones de interés.

III. Resultado Detallado -Report-: Se presenta de manera granular los datos al nivel de detalle que se requiera, lo que permite una inspección y comprensión más completa de los datos.

Integrado con la Metodología D-A-R, se considera la Teoría de la Información de Claude Shannon para estructurar la transmisión eficiente y precisa de la información (QlikView, 2013), de tal manera que se minimice la ambigüedad y maximice la relevancia del mensaje que pretende transmitirse, teniendo en cuenta aspectos como la Percepción, Cognición y Acción. Así mismo, se integra el marco analítico de Tamara Munzner el cual se basa en niveles de análisis, centrados en tres preguntas esenciales (“What-Why-How”) que guían el proceso de visualización de datos desde la comprensión del contexto hasta la representación gráfica efectiva (Munzner, 2014).

Ilustración 2 Marco de trabajo para el diseño de visualizaciones de Tamara Munzer.

What?	Why?	How?
Instrumento de visualización de los datos de las transacciones jurídicas registrados en folios de matrícula inmobiliaria para consulta y descarga de datos por parte de los diferentes usuarios, entre los que se encuentran tomadores de decisiones a nivel gubernamental, privado y ciudadanía en general.	El objetivo de la visualización es el de identificar los territorios con mayor intensidad en el número de transacciones inmobiliarias que dan cuenta de la dinámica inmobiliaria en los departamentos y municipios del país, por zona y año, para el periodo 2015 a 2023.	El preprocesamiento de los datos se realiza mediante Python y la implementación de las visualizaciones se desarrolla previo diseño en PowerBI, para lo cual se estructura la información en capas que reflejan la evolución y la distribución de las transacciones inmobiliarias en los territorios.

Fuente: *Elaboración propia a partir de (Munzner, 2014)*

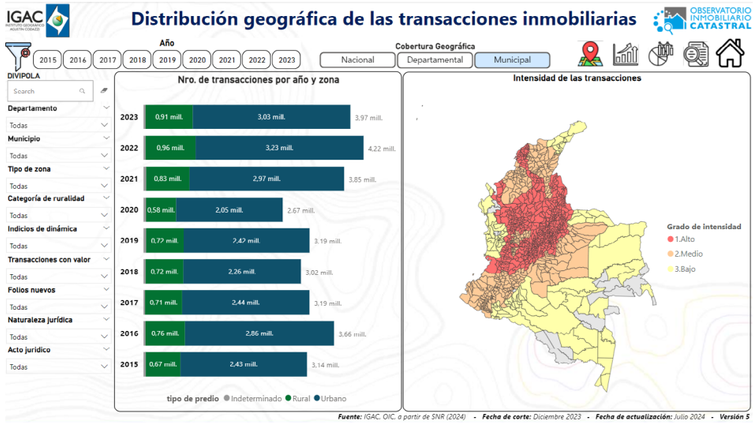
La adopción del marco estructurado por el OIC enfatiza la importancia de crear visualizaciones de datos que no solo sean sólidas y estéticamente atractivas, sino que se alineen con las necesidades y tareas de los usuarios finales. Así se garantiza que sus visualizaciones son herramientas poderosas para la transmisión de conocimientos, facilitando la toma de decisiones informadas y mejorando la comunicación de información crítica dentro del mercado inmobiliario.

Creación de capas

El proceso de creación de capas para el instrumento de visualización de las transacciones inmobiliarias en Colombia fue fundamental para asegurar que los usuarios pudieran interactuar con los datos de manera efectiva y comprensible. Este proceso comenzó con la identificación de las variables clave que permitirían analizar las transacciones inmobiliarias en diferentes niveles de detalle, tales como la cobertura geográfica, el tipo de zona, la categoría de ruralidad, y la naturaleza jurídica de los actos registrados. Una vez definidas estas variables, se procedió a estructurar la información en capas que representaran visualmente las dinámicas inmobiliarias en el territorio nacional. Estas capas se diseñaron para facilitar la consulta de los datos para el periodo abarcado desde el año 2015 al 2023, permitiendo así el análisis temporal y geográfico de las transacciones.

La primera capa creada corresponde a la distribución geográfica (Ilustración 3), que muestra la concentración de transacciones a nivel nacional, departamental y municipal, diferenciando entre zonas urbanas y rurales. Esta capa permite a los usuarios identificar fácilmente las regiones con mayor actividad inmobiliaria, así como las vinculadas a cada Oficina de Registro de Instrumentos Públicos -ORIP.

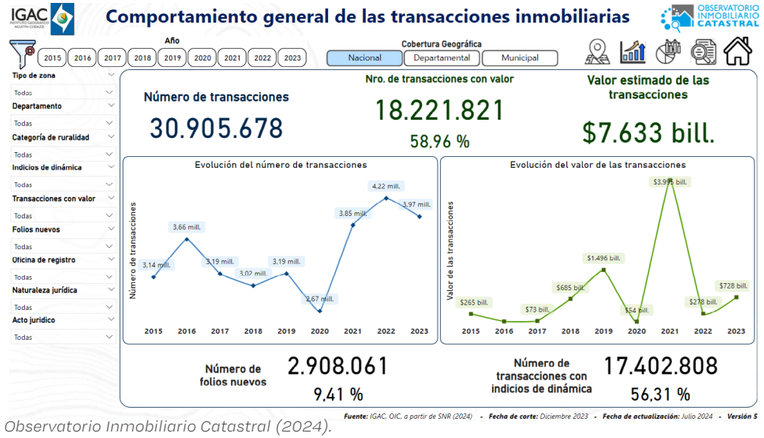
Ilustración 3 *Primera capa: Distribución geográfica*



Fuente: *Observatorio Inmobiliario Catastral (2024).*

Posteriormente, se desarrolló una capa enfocada en el comportamiento general de las transacciones (Ilustración 4) que muestra, las tendencias en el número de transacciones y los valores asociados, facilitando el análisis comparativo entre departamentos o municipios y categorías de ruralidad. La información fue presentada de manera que los usuarios pudieran reconocer patrones de dinámica inmobiliaria y la creación de nuevos folios de matrícula.

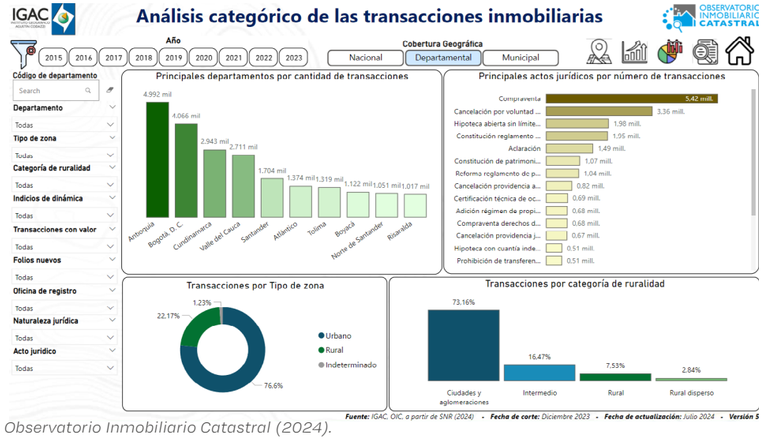
Ilustración 4 *Segunda capa: Comportamiento general de las transacciones*



Fuente: *Observatorio Inmobiliario Catastral (2024).*

Otra capa importante fue la dedicada al análisis categórico de las transacciones inmobiliarias (Ilustración 5). Aquí se destacó el rendimiento de las oficinas de registro, departamentos y municipios con mayor número de transacciones, permitiendo una segmentación detallada de los datos por zona y naturaleza jurídica de los actos. Esta capa proporciona información valiosa sobre cómo diferentes tipos de actos jurídicos influyen en el mercado inmobiliario de diversas áreas del país.

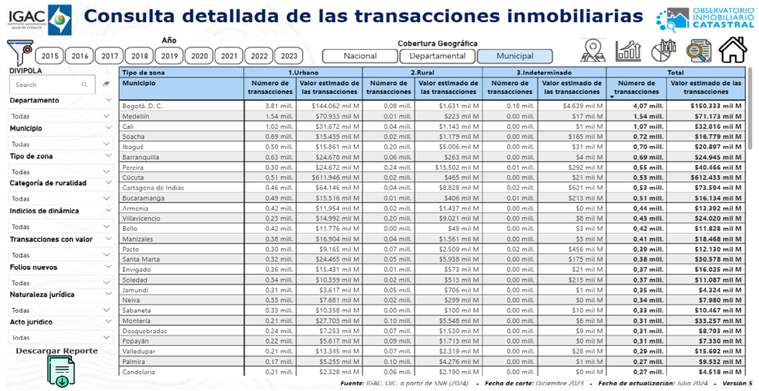
Ilustración 5 *Tercera capa: Análisis categórico*



Fuente: *Observatorio Inmobiliario Catastral (2024).*

Finalmente, se diseñó una capa para la consulta detallada de las transacciones, que ofrece un desglose minucioso del número y valor de las mismas, clasificadas según el tipo de zona y acto jurídico (Ilustración 6). Esta capa también facilita la descarga de reportes personalizados, permitiendo a los usuarios obtener datos específicos para su análisis o para la generación de informes. En conjunto, no solo enriquecen la experiencia del usuario, sino que también permiten un análisis profundo y flexible de las transacciones inmobiliarias en Colombia.

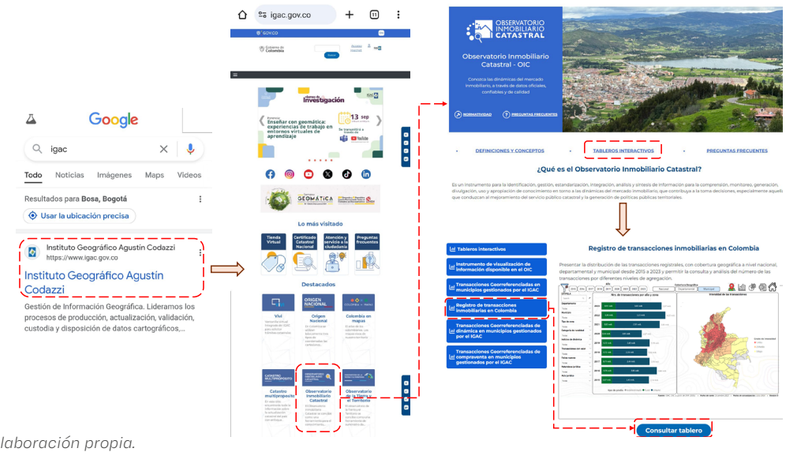
Ilustración 6 *Cuarta capa: Reporte detallado*



Fuente: *Observatorio Inmobiliario Catastral (2024).*

Los usuarios pueden acceder e interactuar con los distintos instrumentos de visualización como el de los registros de transacciones inmobiliarias en Colombia a través de la página web oficial del IGAC.

Ilustración 7 *Ruta de acceso a la herramienta de visualización del registro de transacciones*



Fuente: *Elaboración propia.*

De esta manera, el OIC avanza en la difusión de los resultados más destacados en cuanto a procesamiento y visualización de datos para aprovechamiento de diferentes usuarios, por medio del siguiente enlace:
https://www.igac.gov.co/el-igac/areas-estrategicas/direccion-de-investigacion-prospectiva/observatorio-inmobiliario-catastral

Bibliografía

Microsoft Learn. (2023, November 22). Modelos semánticos en el servicio Power BI - Power BI. Microsoft Learn. https://learn.microsoft.com/es-es/power-bi/connect-data/service-datasets-understand

Munzner, T. (2014). Visualization Analysis and Design. CRC Press. ISBN 978-1-4665-0893-4

QlikView. (2013, 11). DASHBOARD, ANALYSIS, REPORTING (DAR). Qlik Community. https://community.qlik.com/cyjdu72974/attachments/cyjdu72974/qlik-design-blog/2580/1/Technical%20Paper%20-%20DAR%20-%20US%20LETTER.pdf

Requísitos para difundir aportes en el boletín de analítica del IGAC

La DIP del IGAC, a través del OIC invita a los diferentes equipos de trabajo a difundir sus iniciativas en analítica de datos, para lo que se solicita el cumplimiento de los siguientes requisitos:

1. Identificar y describir el problema que se enfrenta y su relación con la analítica, ciencia de datos o inteligencia artificial.
2. Definir y presentar claramente la necesidad de tratamiento de datos en ese contexto.
3. La solución al problema se atiende usando: Inteligencia Artificial – IA, Modelos de Lenguaje – LLM, Analítica de datos mediante lenguaje Python, y visualización de datos.
4. Compartir el código empleado.
5. Describir de manera clara y concisa el proceso.

OBSERVATORIO INMOBILIARIO CATASTRAL

OIC

Si tienes dudas o quieres ampliar sobre el contenido presentado en esta entrega del Boletín de analítica escríbenos
obs_inmobiliario@igac.gov.co
visita nuestro sito web
https://www.igac.gov.co/index.php/el-igac/areas-estrategicas/direccion-de-investigacion-y-prospectiva/observatorio-inmobiliario-catastral

Para participar en el Boletín escríbenos al correo
obs_inmobiliario@igac.gov.co
indicando:
Nombre, Área, Tema y entregando la propuesta