**Modelo de búsqueda automática de apartamentos en venta en la zona sur de Cali a partir de la información de Fincaraiz.com.co**

**Angie Patricia Manzano Meléndez**

**Juan Camilo Vargas Vélez**

**Kevin Santiago Artunduaga Vivas**

**Samuel Pinzón Valderrutén**

**Universidad Autónoma de Occidente**

**Programación**

**Jack Daniels Márquez Franco**

**2022**

**Contexto**

Fincaraiz.com.co forma parte de Frontier Digital ventures, actualmente posee la mayor cantidad de ofertas inmobiliarias de Colombia, ubicándose como el portal inmobiliario líder del país con más de 30 años de experiencia en la publicación de anuncios clasificados de inmuebles en venta y arriendo, vivienda usada, inmuebles comerciales, proyectos nuevos de vivienda y otras propiedades. Posicionándose como el marketplace número uno en todo el país en este sector, con más de 370 mil inmuebles por todo el territorio.

Así mismo, Fincaraíz.com.co proporciona publicidad en forma de anuncios clasificados de constructoras, inmobiliarias y particulares en las ciudades principales del país, además de poder visualizarlas, compararlas y tener comunicación directa con el vendedor.

Por otra parte, Finca Raíz.com.co cuenta con más de 6 millones de visitas y 53 mil descargas mensuales en dispositivos móviles para Android y iOS, destacándose como una plataforma versátil para los medios de comunicación, no solo centrándose en la manera tradicional de compra y venta de inmuebles. Gracias a esto se ha reducido el tiempo de compra haciendo uso de las herramientas digitales para satisfacer la necesidad de los clientes que quieren comprar o vender bienes raíces.

Es importante tener en cuenta, que Fincaraiz.com.co trabaja con un sistema de búsqueda automática para filtrar los tipos de inmuebles en los que el cliente esté interesado, de manera que pueda consultar información de las diferentes ofertas de propiedades. Resulta conveniente mencionar que Fincaraiz no es más que una página intermediaria al momento de las negociaciones, es decir, no interviene en ninguna de las dos partes, y no se hace responsable del cumplimiento de los acuerdos pactados entre estas. Además de que no se da ninguna garantía sobre la veracidad de la información subida en la página por lo que las decisiones que tome el cliente serán bajo su propio riesgo y responsabilidad.

¿Por qué se escogió Fincaraiz? Porque Fincaraíz.com.co es el líder en términos de usuarios, clientes y contenido inmobiliario en Colombia. Asimismo, posee una gran variedad de inmuebles como casas, apartamentos, locales, entre otros. Por lo que es adecuado para el desarrollo de este proyecto de datos poder escoger cualquier tipo de inmueble y así extraer, tratar, transformar y presentar la información.

**Antecedentes**

Dentro de esta investigación al estilo desk research, se encontraron varios artículos. Sin embargo, para mejor clasificación, denominación y entendimiento, se decidió que sería mejor denominarlas de dos formas: Aquellos que sirven para conocer el manejo de la información; y aquellos que están para la valorización de los inmuebles.

En cuanto al manejo de la información, está la compañía de software Landlordstation basada en la administración de inmuebles, fundada en 2010 por la principal razón de que existían problemas al momento de obtener información sobre personas que podrían ser inquilinos de ciertos inmuebles, lo cual hacía de la publicación de bienes raíces una actividad poco segura. En respuesta a esta problemática la compañía trabajó con SmartMove, un software para proporcionar información verídica sobre los posibles inquilinos de los inmuebles, brindando un parte de seguridad a los propietarios. Demostrando de esta forma la importancia de la veracidad de la información publicada en estos tipos de páginas, lo cual hace de la búsqueda y arrendamiento de los inmuebles un proceso más ágil y eficaz.

De igual forma está el proyecto titulado “Formulación de un plan de mercadeo digital para una aplicación digital que promueve el consumo colaborativo de inmuebles físicos de uso privado y comercial para vehículos automotores” de la Universidad Autónoma de Occidente en el año 2018, el cual tiene como objetivo desarrollar un plan de marketing digital para Parkeep una empresa de tecnología ubicada en la ciudad de Cali que permita crear espacios de parqueo de vehículos que fomenten el consumo colaborativo de bienes y servicios para ambos usuarios y oferentes. Dicho plan de marketing digital utiliza el esquema de la metodología RACE y el cual tiene un enfoque en el crecimiento de las ventas de las compañías y se compone de 5 etapas: Oportunidad, Estrategia, Acción , Monitoria y Control y por último Presupuesto, con el objetivo de diseñar estrategias que permitan posicionar la app en el mercado teniendo en cuenta las fortalezas, debilidades y oportunidades de la empresa, y que además les permita transar espacios de parqueo. Esta información es relevante a la hora de generar soluciones mediante la elaboración y análisis de la información obtenida para tener la solución más ideal con respecto a la problemática que se tenga.

Así mismo está el trabajo de grado realizado en 2019 por parte de estudiantes de la Universidad Autónoma de Occidente, titulado: “Plataforma para el alquiler de inmuebles residenciales en la empresa SWAPPS” el cual consta de un aplicativo móvil y una página web para usuarios que poseen inmuebles y buscan una manera ágil y segura de arrendarlos, para ello pueden observar un informe personal del posible inquilino, que muestra información, del salario promedio, si se encuentra reportado en datacredito, entre otros. Todo esto debido a que el alquiler de inmuebles ha estado en su auge por medio de páginas especializadas en el tema, pero a su vez por medio de estas plataformas existen procesos que pueden ser tardados por la parte legal, por lo que se comienza a evidenciar falsificación de documentos para aparentar ser una persona apta para el inmueble, por esto muchas empresas han decidido desarrollar un software para impedir la falsificación de estos documentos y poder brindar un mejor servicio a sus clientes.

Finalmente en el año 2008 en la universidad Autónoma de Occidente se realizó un trabajo de grado titulado “Desarrollo de un aplicativo de inmuebles - inmuebles web” hecho con el fin resolver las nuevas necesidades de la administración de los inmuebles del área, para que no haya un retraso en las obras que son realizadas en el país y esperar a ser registradas en un lugar en específico, sino que sea desde cualquier sitio. Así que para ello se hizo el desarrollo del aplicativo inmueble, con el fin de realizar el manejo actual de la información en el área, brindando datos oportunos, ágiles y actualizados, al mismo tiempo beneficios a los empleados debido que no se debe esperar a estar en el sitio de trabajo. Además de facilitar la rapidez de la documentación requerida por los licitadores, estableciéndose en la vanguardia de la tecnología. Como resultado se obtuvieron los objetivos propuestos, haciendo uso del UML (lenguaje unificado), el cual optimizó la traducción a un lenguaje de programación de los resultados obtenidos en el análisis y diseño, en otras palabras el diseño de las interfaces y el software.Este trabajo muestra cómo realizar una solución óptima por medio del buen manejo de la información obtenida, para agilizar los procesos, que en este caso serán la búsqueda ágil de inmuebles.

Respecto a la valoración de los inmuebles está el proyecto titulado “Valoración de inmuebles comerciales mediante la aplicación del método de capitalización directa y el método de flujos de caja descontados” es presentado por dos estudiantes de la universidad privada Boliviana en el 2016, centrándose en la determinación del valor de los bienes raíces gracias al uso de dos métodos, el primero es la capitalización directa y el segundo son los flujos de caja de descontados o también llamado descuento de flujos, para ambos es necesario saber el beneficio económico efectivo que obtienen los propietarios del inmueble. En la cual para la tasa de capitalización se calculó mediante el arrendamiento por metro cuadrado y la comparación del valor de los inmuebles en el mercado de la ciudad de La Paz. En los dos métodos los resultados obtenidos fueron distintos, debido a que se usaron diferentes tasas de crecimiento en las utilidades esperadas, sin embargo se lograron encontrar semejanzas en ambas, en consecuencia se propone incorporar una ajuste aleatorio de la tasa de capitalización y un ajuste de expectativas en las tasas de interés. Este último antecedente sirve para aprender a darle el valor adecuado los inmuebles que están publicados en la plataforma por medio de diferentes métodos y así poder brindar información más precisa para facilitar los procesos mencionados en los otros antecedentes, que son la agilidad en la búsqueda y en el arrendamiento de estos inmuebles.

**Justificación**

Es bien sabido que la actividad inmobiliaria se ha afianzado en los últimos años en Colombia. Para el 2021 se registró el mejor año en venta de viviendas con más de 209 mil unidades VIS (Viviendas de Interés Social) y No VIS vendidas, según datos revelados por Galería Inmobiliaria. Lo que muestra el gran momento que está pasando la vivienda en el país.

La vivienda es importante para las personas no solo por su desarrollo y crecimiento sino también por las necesidades básicas que satisface como protección, sanidad y bienestar de las personas. Por lo que adquirir una vivienda es uno de los proyectos más comunes e importantes para las personas, principalmente las familias recién casadas.

El mercado de vivienda en Colombia, en el primer semestre del año 2021 se incrementó la venta de inmuebles en 53% y las de arriendo en un 47% en principales ciudades del país y municipios aledaños, según datos extraídos del estudio “Evolución del mercado de la vivienda en Colombia”, los apartamentos fueron los más ofertados en el periodo 2019-2021 con cerca de 68% en ventas y 77% en arriendo del total de inmuebles seguido por casas con un 13% y apartaestudios con un 10%, y en cuanto a venta y arriendo, los apartamentos fueron los más consultados por los Colombianos gracias a el gran crecimiento de los conjuntos residenciales por temas de seguridad en las ciudades principales del país.

La información anterior ponía en contexto la importancia de la vivienda en Colombia y cuáles eran los inmuebles más buscados en el mercado. Ahora se centrará en un contexto más regional como la ciudad de Santiago de Cali, donde el mercado de apartamentos, especialmente la venta de estos es la mayor oferta inmobiliaria, con un aumento del 76,9% en el 2021. En Fincaraiz en Mayo del 2022 se encontraban ofertados 9,142 apartamentos a la venta, esta cifra es una muestra representativa para realizar el proyecto. Pero para presentar un análisis más preciso se buscó un contexto más específico que cuente con la cantidad suficiente de datos para realizar el análisis, se definió enfocarse en la zona sur de la ciudad, que ofertaba más de la mitad de la muestra total.

Teniendo en cuenta todo lo anterior, el presente proyecto busca utilizar los datos de los inmuebles que se encuentran en Fincaraiz, específicamente apartamentos en venta ubicados en la zona sur de la ciudad de Cali. Puesto que la búsqueda de este tipo de inmueble es altamente demandada, se requiere desarrollar un modelo que facilite el proceso que en ocasiones puede llegar a ser complejo. A partir de estos factores; la alta demanda de este tipo de inmueble en los últimos años y la veracidad de la información que se brinda en las publicaciones de estos. De manera que responda a la siguiente pregunta:

¿Cómo desarrollar un modelo de búsqueda automática de apartamentos en venta en la zona sur de Cali, a partir de la información de la página web de Fincaraiz.com.co?

**Objetivo General**

Desarrollar un modelo de búsqueda automática de apartamentos en venta ubicados en la zona sur de Cali a partir de la información de Fincaraiz.com.co

**Objetivos Específicos**

* Identificar la información brindada por la página Fincaraiz.
* Diseñar un modelo de búsqueda a partir de la información para facilitar el arriendo de apartamentos basado en ETL’s.
* Analizar los datos arrojados por la búsqueda.

**Metodología**

La investigación de este proyecto es tipo descriptiva, debido a que busca realizar un informe detallado de cómo se va a realizar el modelo de búsqueda, con sus características y la información requerida que se va a usar de la página fincaraiz.com. Para la metodología se usará Thinking based learning (TBL), ya que se busca contextualizar, analizar, relacionar,argumentar, convertir los conocimientos ya aprendidos y desarrollarlos mediante las destrezas adquiridas.

**Etapa 1**

Identificar la información brindada por la página de Fincaraiz.

* Hacer una análisis visual de qué realiza la página de Fincaraiz.
* Escoger el inmueble sobre el cual se va a centrar el proyecto.
* Reconocer cuál es la información que presenta la página del inmueble escogido

**Etapa 2**

Diseñar un modelo de búsqueda a partir de la información para facilitar la venta de apartamentos basado en ETL

* Obtener toda la información necesaria por medio de web scraping.
* Filtrar la información obtenida para que sea usada.

**Etapa 3**

Analizar los datos arrojados por la búsqueda.

* Utilizar la herramienta power bi para el análisis de la información
* Realizar las estadísticas y presentarlas de una manera gráfica.
* Extraer la conclusiones de la gráficas.

**Análisis del archivo robots.txt**

Un archivo robots.txt indica a los rastreadores de los buscadores a qué URLs de tu sitio pueden o no pueden acceder. El robots.txt de Fincaraiz.com.co está hecho solo para unos User-agent (robots/buscadores) en específico, como: Googlebot, Alexabot, LinkedInbot, Twitterbot, entre otros. Y a estos User-agent mencionados anteriormente no les es permitido entrar y rastrear ciertas urls de la página como: InsertResult.aspx, Legal.aspx, comparar\_inmuebles.aspx, loginUser.aspx, entre otras más.

**Análisis del mapa del sitio web**

El mapa del sitio web es una representación visual de el, este te facilita la búsqueda del contenido de una manera un poco más organizada, es fundamental debido a que es una guía para los usuarios que deseen ingresar a una determinada página web, recurriendo a esta cuando no se haya encontrado la información deseada en la página principal, además de su navegación.

Existen dos tipos de sitemaps: esta el HTML que sirve de ayuda a los visitantes debido a que muestra una lista de páginas relacionadas, desde las particulares a las generales. Permitiendo una mejor estructuración de la web. Y el XML es el que ayuda a los motores de búsqueda para que más usuarios puedan conocer la página web, proporcionándole a los motores de búsqueda las URL para una visualización del mapa completo de todas las páginas. Por otra parte también favorece al trabajo de los rastreadores web o mejor conocidas como web spider (arañas web) que se encargan de la indexación, facilitando este proceso.

Dicho esto fincaraiz.com.co solo cuenta con sitemaps XML, y son los siguientes:

* Sitemap: http://www.fincaraiz.com.co/sitemaps/SiteMap\_Index.xml
* Sitemap:http://www.fincaraiz.com.co/sitemaps/SiteMapFCCOL\_Clients\_Index.xml
* Sitemap: http://www.fincaraiz.com.co/sitemaps/image\_sitemap.xml

De los tres XML que proporciona, el primero es el principal, debido a que muestra información relevante a todos los inmuebles que se encuentran registrados en la página, luego en el segundo se muestra lo que hace parte de la indexación de los sitemap, conteniendo en el dos XML, uno que indica que la ubicación de las URL cambian constantemente cada que se recargue la página y el otro sirve para adaptar el formato de fecha al W3C. El último te direcciona a un error 404, el cual significa que no es capaz de comunicarse con el servidor para este recurso.

**Cálculo de tamaño**

El tamaño de la página es el tamaño de archivo descargado de una página web determinada, basándose en la construcción de los archivos que conforman el sitio. En otras palabras, cada uno de los archivos tiene diferentes tamaños, que a la hora de juntarlos equivalen al tamaño de la página web.

La media del tamaño es de 3 MB(megabytes) , pero esto no es un punto positivo porque entre más MB consuma, más pesada será la página y su velocidad de carga más lenta será. Así que por medio de la página tools.pingdom.com podemos observar que el tamaño del sitio web de Fincaraiz.com.co es de 1,8 MB un tamaño no muy grande a diferencia de los sitios web actuales, considerándose como una página ligera.

**Descripción del propietario del sitio**

Gracias a la pagina co.godaddy.com se pudo ver cuales son los resultados del whois de Fincaraiz.com.co, debido a que esta pagina tiene un dominio privado hay algunas características de información que aparecen privadas, y los resultados permitidos son:

Domain Name: fincaraiz.com.co

Registry Domain ID: D607344-CO

Registrar WHOIS Server:

Registrar URL: http://mi.com.co/

Updated Date: 2021-07-23T05:28:20Z

Creation Date: 2004-05-18T00:00:00Z

Registry Expiry Date: 2024-05-27T23:59:59Z

Registrar: Central Comercializadora de Internet S.A.S

Registrar IANA ID: 88888

Domain Status: clientTransferProhibited https://icann.org/epp#clientTransferProhibited

Registrant Organization: 800229663 1

Registrant State/Province: Bogota

Registrant Country: CO

Name Server: ns-871.awsdns-44.net

Name Server: ns-1778.awsdns-30.co.uk

Name Server: ns-1.awsdns-00.com

Name Server: ns-1462.awsdns-54.org

DNSSEC: unsigned

**Descripción de la tecnología usada**

Por medio del build with se conoció que Fincaraiz.com.co fue construida con Microsoft Frontpage, el cual fue un editor de HTML para páginas web. Se descontinuó en 2007 y fue reemplazado por Microsoft Expression Web y Sharepoint Designer debido a que muchos consideraban que el código HTML generado por este era muy descuidado. También el build with menciona que debido al uso de ciertas funciones de Javascript la página contiene código generado, al menos inicialmente por Adobe Dreamweaver que es una aplicación informática destinada a la construcción, diseño y edición de sitios y aplicaciones Web basados en estándares.

Como en la mayoría de desarrollos de páginas web se evidencia el uso de frameworks, este también es el caso de nuestro sitio a trabajar, y encontramos que Fincaraiz.com.co fue desarrollada con los siguientes frameworks: ASP.NET, que consiste en poder crear sitios web dinámicos, aplicaciones web y servicios web XML, también utiliza la versión 4.0, la cual es una versión mejorada en cuestión de la plantillas de interfaz de usuario, nuevas capacidades de JavaScript y de enlace de datos para el Ajax (Ajax: tecnología que permite que un usuario de una aplicación web interactúe con la página web sin la interrupción que implica volver a cargar la página web).

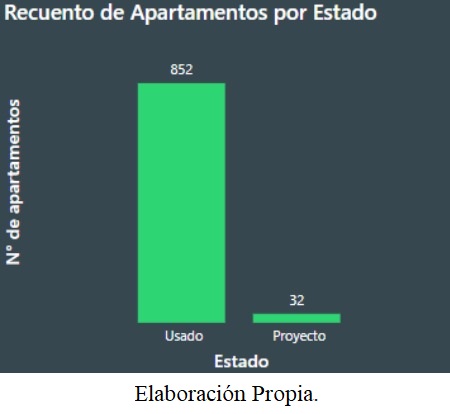
Trabaja con ASP.NET Ajax, que es un framework gratuito para crear rápidamente una nueva generación de experiencias web más eficientes, más interactivas y altamente personalizadas que funcionan en los navegadores más populares. Está también Next.js, el cual es un framework de trabajo que permite a los desarrolladores crear aplicaciones de página única y aplicaciones web de alto rendimiento a través de la renderización del lado del servidor. Y por último trabaja con Organization Schema (Schema.org) el cual es un proyecto común para unificar la forma de etiquetar y enriquecer la información de los contenidos de una página web para que los buscadores puedan indexarlo, clasificarlo y mostrarlo, con el fin de mejorar la experiencia de los usuarios.

**Descripción de Dataset**

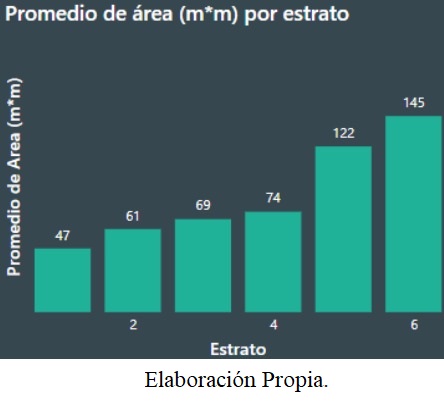
En el dataset se presenta toda la información extraída de la página de fincaraiz.com.co. La muestra es de 886 apartamentos, y cada uno de ellos muestra sus respectivos datos, como el nombre; el barrio donde se encuentra ubicado el apartamento; el estado, si es un proyecto, si ya ha sido usado o por el contrario es nuevo; el precio por apartamento; la constructora o inmobiliaria a la que pertenece; el código para identificar cada apartamento, ya que fincaraiz le asigna un código a cada anuncio que se publica en la página.

Por lo cual es importante tener tal código en los datos para que al momento de aplicar el modelo de búsqueda automático, en donde si la persona encuentra el apartamento indicado solo sea cuestión de ubicarse por medio del código de fincaraiz para buscar más información acerca de este; el número de habitaciones; el número de baños; el área total del apartamento; y el área privada del mismo, el estado del apartamento, el cual hace referencia sobre qué condiciones se encuentra, si es buena, excelente, está en construcción y demás.

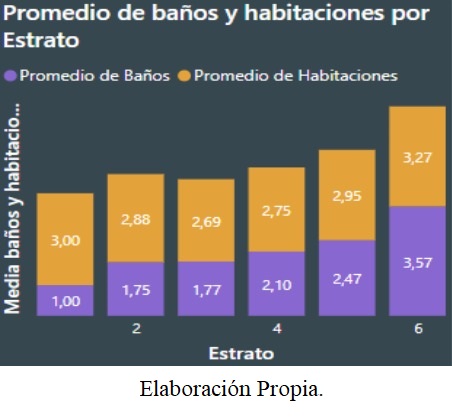
En el dataset se encuentran variables con diferentes formatos, como strings, enteros y decimales. De la variables con formato string se encuentra el nombre, barrio, estado, constructora, código (pese a que lo que se encuentra en dicha variable son números, se toma como string, ya que con esos números no se hacen operaciones, sólo es un indicador de los apartamentos), y el estado del apartamento. Para las variables con formato entero están el estrato, la cantidad de habitaciones y la cantidad de baños. Con formato de número decimal se tiene el área total y el área privada.

**Resumen básico estadístico de las variables (Gráficos y tablas)**

Gráfica 1- Recuento de apartamentos por estado: En esta gráfica de barras vertical se puede ver la comparación de la cantidad de apartamentos por estado; si es un proyecto o si ya ha sido usado. Concluyendo que los apartamentos que se encuentran en la plataforma son mayormente usados que recién construidos.



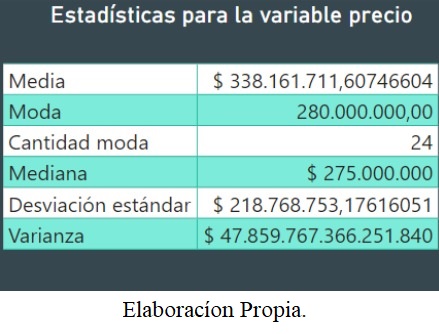
Gráfica 2 - Promedio de área (m\*m) por estrato: Esta gráfica está representada también con un gráfico de barras, y se puede observar el área media de los apartamentos que se encuentran en los diferentes estratos. A simple vista se puede analizar que los apartamentos de estrato más alto son los que tienen más área construida.

Gráfica 3 - Promedio de número de baños y habitaciones por estrato: Esta gráfica de barras agrupadas es una visualización del número de baños y de habitaciones promedio que se puede encontrar en los apartamentos clasificados por estrato. Una información que podemos concluir de la gráfica es que el número de habitaciones promedio para todos los estratos es de 3 habitaciones, es decir, que es muy común encontrar apartamentos con tres habitaciones.



Gráfica 4 - Promedio del precio de apartamentos por estrato: El siguiente es un gráfico de líneas que muestra el precio promedio de los apartamentos clasificados por estrato. Presenta una escala en el eje Y en miles de millones debido a que existen apartamentos que pasan por estos valores. Una conclusión de este gráfico es que al igual que ocurría en el gráfico 2, quienes tienen los valores más altos son los apartamentos de los estratos mayores (5 y 6)

Gráfica 5 - Top 5 constructoras con mayor número de apartamentos: Por último tenemos este gráfico de barras horizontales, en el cual se muestra un top 5 de las constructoras e inmobiliarias que están más presentes en los anuncios publicados de ventas de apartamentos. De la gráfica se puede deducir que la gran mayoría de apartamentos en venta no están sujetos a una inmobiliaria o constructora reconocida, sino que se publican de manera independiente.



Tablas: Se realizaron seis tablas entre todas las variables de formato numérico, todas presentan el mismo tipo de información, es decir, en todas se muestra la media, la moda, la mediana, la desviación estándar y la varianza pero de cada una de las variables escogidas (precio, área total, área privada, estrato, número de baños y número de habitaciones).

**Resolución del problema**

Teniendo en cuenta el objetivo del proyecto, el cual es agilizar los procesos de búsqueda de apartamentos publicados en la página de fincaraiz.com.co por medio de la misma información que brinda la página, la solución que se seleccionó fue crear un modelo de búsqueda automático para satisfacer los requerimientos. Dicho modelo consta de una parte de programación y otra parte de visualización.

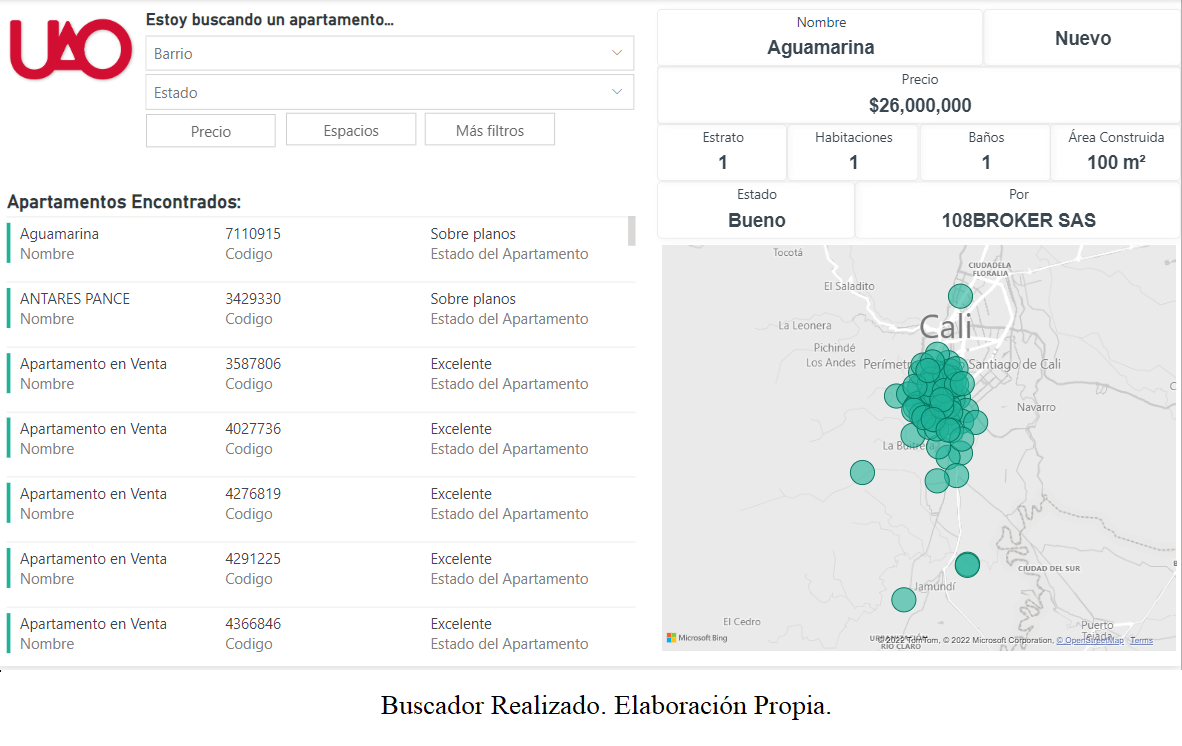
En la parte de programación se encuentra el código que permite conseguir la información de los apartamentos de Fincaraíz. Para esto, fue necesario primero identificar la información de interés de los inmuebles. Es importante tener en cuenta que en los anuncios de los apartamentos no todos cuentan con la misma estructura de información, debido a que hay algunos que cuentan con mucha más información que otros. Esto ocurre, porque entre la venta de apartamentos se encuentran proyectos de construcción, donde solo se brinda la información básica como: Nombre de la unidad, ubicación, estrato, estado de proyecto si se encuentra sobre planos o en construcción, y tipos de apartamentos donde se brinda información acerca del precio, habitaciones, baños y áreas. Mientras que en apartamentos usados, se pueden encontrar más características del inmueble además de la información mencionada anteriormente, como por ejemplo, parqueaderos, piso en el que se encuentra, precio de la administración, entre otros. Teniendo en cuenta lo anterior, se decidió enfocar en la información que todos los anuncios tuvieran sin importar si era usado o proyecto, es decir: Nombre, barrio, estado (Usado o Proyecto), precio, vendedor, código del apartamento en Fincaraíz, habitaciones, baños, estrato, estado del apartamento, área total y área privada.

Después de identificar la información de interés, era necesario conseguirla. Esto se realizó a través de web scraping, una técnica utilizada para extraer información de sitios web como lo es fincaraiz.com.co. Con la ayuda de Selenium, una librería de Python utilizada para realizar web scraping se pudo conseguir la información. Hay que tener en cuenta que para extraer cualquier información de un sitio web es necesario conocer el XPath del elemento en el html de la página. También hay que recalcar que mucha de la información necesaria se encontraba al abrir el anunció y no en la card del mismo, además de que los anuncios de apartamentos en estado de proyecto y usado, cuentan con una estructura de html diferente. Por lo que era necesario realizar un proceso de web scraping para cada uno.

Para los apartamentos en estado de proyecto se decidió utilizar el primer tipo de apartamento como muestra del anuncio. El código consiste en extraer la información a través de los XPaths de las variables de interés, previamente obtenidos al revisar el html de la página. Sí el código no encuentra un elemento con el XPath asignado a la variable le asigna el valor de None.

Por otro lado, para los apartamentos usados a diferencia de los proyectos, se utilizan dos arreglos unidimensionales. Uno para todos los títulos de las características del apartamento, mientras que en el otro se encuentran los valores respectivos de cada característica. Todos estos datos obtenidos a través de su respectivo por XPath. Posteriormente el código utiliza la función index() para buscar la información de interés en el arreglo de característica, y con esa posición encontrar el valor de esa característica en el otro arreglo. De esta manera se asegura de obtener solo la información necesaria. Así mismo, si no encuentra algún dato le asigna el valor de None.

Toda la información recolectada de cada apartamento se almacena en un diccionario para mejor organización. Con la ayuda de la librería CSV de Python se creó un archivo en formato csv codificado en Latin-1, que permite la visualización de caracteres especiales, con la información de 1714 apartamentos encontrados en Fincaraíz, dando por finalizada la parte de programación.

Una vez recolectada toda la información, se procedió con la limpieza y organización de los datos mediante PowerBi, debido a que no se contaban los conocimientos necesarios para llevarlo a cabo en otra aplicación. Primero, se eliminaron los apartamentos duplicados debido a que Fincaraiz fijaba algunos anuncios como destacados, que aparecían en cada una de las páginas de búsqueda. También se corrigieron los nombres de los apartamentos, barrios y vendedores, luego los datos errados o faltantes; aunque en la mayoría de apartamentos carecían de cierta información, sólo se eliminaron los apartamentos con más tres datos faltantes y por último, la eliminación de dos apartamentos, que estando dentro del perímetro urbano analizado sus precios no eran coherentes a la de la zona, con precios de $1.300.000 y $215.000.000.000 repectivamente. Lo que dificulta el proceso de filtrado de apartamentos a partir de su precio. Por último, con el fin de mejorar y brindar información visual de cada apartamento se agregaron dos columnas longitud y latitud; estas son las coordenadas de los barrios, que fueron agregadas manualmente para realizar el objeto visual de mapa en el buscador.

El buscador realizado cuenta con el registro de 884 apartamentos obtenidos de Fincaraiz, a partir del proceso mencionado anteriormente. Asimismo, el buscador cuenta con siete filtros: barrio, estado, precio, habitaciones, baños, estrato y área construída. Para facilitar la búsqueda del inmueble. Adicionalmente, cuenta con objetos visuales como: lista de los apartamentos y un mapa con la ubicación del barrio, para darle una mejor idea al comprador de la ubicación. Por último, incluye un espacio en el cual se muestran las características del apartamento seleccionado.

**Trabajo futuro y recomendaciones**

Para un futuro trabajo del proyecto sería digitalizar el modelo de búsqueda, que no sea solo desde una plataforma visual como Power Bi, sino una programación de toda la interfaz necesaria para realizar el objetivo de una manera correcta, facilitando la búsqueda e indagación en los apartamentos y si es capaz aplicarlo a todos los inmuebles (no tan solo apartamentos), claramente para su realización sería necesario unos mejores equipos, capaces de soportar toda la información recibida sin colapsar en el intento de recolectar la información de cada inmueble.

Otra propuesta para un trabajo futuro sería poder determinar los criterios que la página tiene para descartar ciertos anuncios, como por ejemplo: son poco visitados, cuentan con información engañosa o simplemente el que oferta se declina de la idea de vender. Teniendo en cuenta que no se encontraron los 4674 registros de los que se mencionan al inicio de la página.

Como recomendaciones todo sería más enfocado a la plataforma Fincaraiz, debido a la facilidad que tienen los dueños de los inmuebles para publicar cualquier tipo de información en un anuncio, sin pasar por un filtro o monitoreo de lo escrito. Esto se puede observar en que la mayoría de apartamentos en venta tienen el mismo nombre “Apartamento en Venta” , lo que genera poca precisión de lo que se está ofreciendo y/o confusión en el comprador. Lo que conlleva a un problema para los que quieran adquirir algún apartamento, casa, finca, entre otros. Debido a que pueden caer en una estafa, dejándose llevar por características que ni siquiera son las correctas.

Para el proyecto se debería realizar algunas correcciones en la extracción de los datos, para automatizar todo sin requerir ayuda manual para los errores expuestos, cosas como los metros cuadrados que al sacarlos se use un replace, que permite al csv no confundir el formato de string por tener letras y sea en un entero o decimal, a su vez errores básicos como las mayusculas o minusculas de los barrios, determinando de una sola manera cada elemento, que a la hora de querer hacer un procesado de estos no se confunda la información vista.

**Uso de librerías de visualización**

Para la visualización de la información se utilizó Power Bi, el cual es una herramienta para el análisis de datos y visualización interactiva de los mismos. En este se pueden hacer diferentes tipos de gráficos con las variables de la información recolectada. El uso que se le dio al Power Bi dentro del proyecto fue para cargar, transformar y hacer visualizaciones de los datos.

Hablando más específicamente de Power Bi como herramienta de visualización, se utilizó para la creación de tablas y gráficas que representan de mejor manera la información dada en el dataset. Para la creación de algunas de las estadísticas requeridas fue necesario la implementación de la función DAX para generar visualizaciones más específicas, y para lo demás, se utilizaron diferentes gráficos ofrecidos por la aplicación como: diagramas de barras, de líneas y demás funciones que permiten la visualización de la información de mejor manera.

**¿Cómo y en donde se publicarían los datos obtenidos?**

Los datos extraídos se podrían utilizar para realizar informes de la Secretaria de Vivienda Social de Cali para mostrar el panorama y/o comportamiento que vive la venta de apartamentos en la ciudad. También podrían ser de interés para las inmobiliarias tener una idea del mercado en la zona sur de la ciudad y realizar estrategias comerciales alrededor de esta información.

**Conclusiones**

Durante la realización del proyecto se tuvieron diferentes procesos y etapas necesarias para la extracción de la información, dentro de los cuales se tuvo la creación de un algoritmo de Python capaz de obtener la información de los correspondientes XPath, la creación de un archivo csv, limpieza y manipulación de los datos a través de PowerBI para una mejor visualización. Una de las dificultades encontradas en la realización de este proyecto fue la obtención de los XPath, debido a que Fincaraiz está clasificada como una página dinámica, es decir, las direcciones o XPath cambian constantemente por la inmensa cantidad de información que es agregada. La página necesitaba reestructurarse por lo que se optó por usar Css Selector, otro método para obtener un XPath preciso.

De acuerdo con la información de Fincaraiz, en total son aproximadamente más de 4000 apartamentos, pero de acuerdo a la suma de todos los cards en todas las páginas fueron encontrados cerca de 1200 apartamentos, una cifra mucho menor que la página menciona, se desconoce el motivo por el cual ocurre esta situación y se plantea como un trabajo futuro para conocer si fincaraiz engaña con su información o se pierden algunos anuncios. También durante la limpieza de los datos dentro se identificaron dos apartamentos con valores atípicos, cuyos precios eran desproporcionados en comparación a los demás, uno por ser extremadamente costoso ($215.000.000.000) y otro por ser muy economico ($1.300.000), lo que dificulta la busqueda de apartamentos en el buscador.

Con los datos obtenidos del proyecto se podría abordar desde la inteligencia artificial, un algoritmo de Deep Learning, capaz de identificar anuncios falsos de apartamentos en venta y de esta forma acabar con las estafas que se pueden presentar en la plataforma. Por ejemplo, si se publica un inmueble con un precio o área exagerado, este será eliminado y reportado al anunciante y a Fincaraiz. A partir de estos reportes, la IA aprendería automáticamente para identificar los distintos métodos de estafa en los anuncios, y así mejorar la experiencia de usuario en la página.

A la hora de hacer el web scraping, se lograron los fines de aprendizaje requeridos sobre cada una de las rutas que se recolectaron, teniendo en cuenta que método usar dependiendo de la plataforma con la que se encuentre, además de la elección de las librerías. Lográndose un resultado satisfactorio, que en caso de que se enfrenten páginas con información necesaria para su extracción de datos, se sepa qué método usar y por qué proceso llevar a cabo la programación.

**Bibliografía**

CatastroDistrital, Fincaraíz.com.co. (2021). Evolución del mercado de la vivienda en Colombia. *Observatorio Técnico Catastral*.

Clavijo López, S, Patiño Molano, J y . (2018-10-23.). *Formulación de un plan de mercadeo digital para una aplicación digital que promueve el consumo colaborativo de inmuebles físicos de uso privado y comercial para vehículos automotores.* Universidad Autónoma de Occidente.

Cobo Ospina, A y . (2008.). *Desarrollo de un Aplicativo de Inmuebles - Inmuebles Web.* Universidad Autónoma de Occidente.

LANDLORDSTATION [en línea]. About. [Consultado el 20 de enero de 2018] Dallas, Texas, Estados Unidos. Disponible en internet: <https://goo.gl/TxkeHm>

Ruiz Curacá, W. S., & Valencia Marín, J. F. (2019). *Plataforma para el alquiler de inmuebles residenciales en la empresa SWAPPS*. Universidad Autónoma de Occidente.

Vargas, A., & Rocha, M. (2017). Valoración de inmuebles comerciales mediante la aplicación del método de capitalización directa y el método de flujos de caja descontados. *INVESTIGACIÓN & DESARROLLO*, *16*(2). https://doi.org/10.23881/IDUPBO.016.2-6E