



**ISPC**

INSTITUTO SUPERIOR  
POLITÉCNICO CÓRDOBA

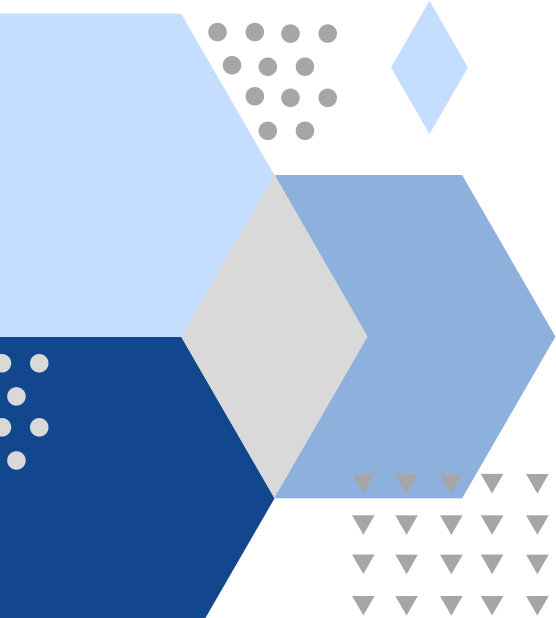
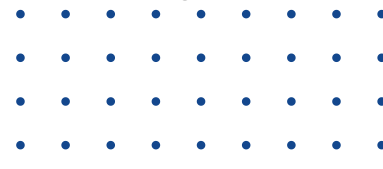
COHORTE 2023 -  
TECNICATURA SUPERIOR EN  
CIENCIA DE DATOS E INTELIGENCIA  
ARTIFICIAL

# ALQUILERES EN CORDOBA

PRACTICA PROFESIONAL

GRUPO SKYNET

INTEGRANTES:

- GUTIERREZ MATEO
  - PAGANO DANTE JAVIER
  - MICELI HECTOR EMMANUEL
  - LOPEZ ERICK
  - CORTEZ GONZALO
  - PAULINA MARÍA ITATÍ
  - BONZON GABRIEL IGNACIO
  - GRANDE FERNANDO
- 
- 

## **Tema: Seguimiento de precios de viviendas en Córdoba**

Informe estadístico y comparativo de los alquileres en la ciudad de Córdoba tomando de referencia los principales barrios de la ciudad de cordoba utilizando diferentes lenguajes de programación, aplicaciones y plataformas para la visuliazación de los datos obtenidos. Para este proyecto, nos centraremos en los principales barrios la ciudad de Córdoba Capital. Esto permitirá satisfacer las necesidades de los usuarios interesados en alquilar - comprar viviendas en diferentes partes de la ciudad.

### **Objetivos**

#### **Objetivo general:**

Buscar diferentes páginas de inmobiliarias, clasificados y de particulares precio de alquileres y cartacterísticas de los departamentos en arrendamiento y elegir una de ellas para realizar un análisis estadístico y comparativo.

#### **Objetivos específicos**

- Utilizar el método web scraping para extraer los datos: ubicación del alquiler, precio, tipo de moneda , mts2. de la página elegida, en este caso “Clasificados de la Voz del Interior”:
- Utilizar librerías de Python como **requests**, **BeautifulSoup** para realizar la extracción de datos
- Hacer la conexión y cargar la base de datos en MongoDB con los datos elegidos
- Comparar los datos obtenidos y graficarlos comparativamente con tableau
- Realizar informe

### **Metodología**

**Web Scraping:** Se utilizarán técnicas de web scraping con Python, utilizando para ello la librería Beautiful Soup. Se extraerán datos de las páginas web mencionadas. Incluyendo información sobre precios, ubicaciones, tamaño de la propiedad, cantidad de habitaciones, etc.

Se utilizará el protocolo HTTP (Protocolo de Transferencia de Hipertexto). El mismo rige la estructura y el lenguaje de las peticiones y respuestas que tienen lugar entre clientes y servidores. Los clientes suelen ser los navegadores web, pero pueden presentarse de muchas formas, como los robots de los motores de búsqueda. Sobre los elementos encontrados se extraerá la información utilizando **find** para buscar un elemento dentro de otro elemento y el método **get\_text** para obtener el texto que contiene. **GET** es el método de petición HTTP más utilizado, ya que el mismo solicita al servidor una información o recurso concreto. Cuando uno conecta a un sitio web, el navegador suele enviar varias peticiones GET para recibir los datos que necesita para cargar la página.

Luego, los datos extraídos se guardarán en una lista vacía, usando el método **append**.

- Almacenamiento de Datos: Utilizaremos una base de datos realizada en MongoDB, ya que la consideramos fácilmente consultable para guardar toda la información extraída. Esto nos permitirá una gestión eficiente de los datos a medida que se recopilan.
- Análisis de Datos y Visualización: Realizaremos un análisis comparativos y estadístico sobre los precios de alquiler en diferentes barrios de Córdoba. Incluyendo e identificando precios promedio, precios por metro cuadrado, análisis de tendencias temporales y demás. Crearemos visualizaciones que permitirán a los usuarios explorar la información de forma intuitiva. La plataforma elegida será My SQL, Tableau, Phytion.
- Para graficar los resultados, se utilizará el **software Tableau**, ya que nos permitirá realizar un análisis completo de los datos obtenidos para que los usuarios puedan llegar a una toma de decisiones efectiva.
- Informe Final y Presentación: Con motivo de presentar los resultados del análisis de datos, crearemos un informe final. Este incluirá información valiosa sobre el mercado de alquiler. Además, grabaremos una presentación que resumirá los hallazgos clave del proyecto.

### **Usuarios Beneficiarios:**

- Personas que buscan alquilar - comprar en Córdoba: Podrán tomar decisiones al encontrar propiedades que se ajusten a sus necesidades y presupuesto
- Inversionistas del mercado de alquileres: Pueden identificar oportunidades de inversión en el mercado inmobiliario.
- Profesionales del sector inmobiliario: Pueden usar los datos para mejorar sus servicios y estrategias.

**Beneficios:**

El proyecto proporciona una solución para quienes buscan alquilar viviendas en la ciudad, ya que brindará información actualizada y herramientas de búsqueda efectivas. Además de tratarse de una plataforma útil para la investigación de mercado en el sector inmobiliario de Córdoba Capital

**Fuente de Datos:**

- Clasificados la Voz del Interior (2023) <https://clasificados.lavoz.com.ar/inmuebles>
- Argenprop (2023) <https://www.argenprop.com>
- Kinsta (19/12/2022) ¿ Qué es una petición HTTP? <https://kinsta.com/es/base-de-conocimiento/que-es-una-peticion-http>

## Proceso de Web Scraping en Clasificados La Voz

### **Introducción:**

El presente informe detalla el proceso llevado a cabo por el equipo *Skaynet* del Instituto Superior Politécnico de Córdoba para realizar un Web Scraping en la página web “Clasificados La Voz”. El objetivo principal fue extraer información relevante sobre anuncios de inmuebles, incluyendo ubicación, precio, tipo de moneda, tamaño, cantidad de habitaciones y comodidades.

### **Métodos Utilizados:**

Para llevar a cabo el proceso de scrapping se utilizaron varias librerías como requests, lxml y BeautifulSoup.

Luego de definida la página fuente de los datos a ser levantados, se identificaron las etiquetas que nos serían útiles. En esta primera instancia la página principal de La voz ya llevaba aplicados filtros, el primero limitado a la ciudad de Córdoba y el segundo a las operaciones inmobiliarias de tipo alquiler.

Se observaron un máximo de 24 resultados con inmuebles que cumplían con los filtros principales. Para poder obtener datos relevantes se observó que se debía ingresar a cada link que redireccionaba a una página con los detalles de la publicación. Además, se identificó el máximo de páginas disponibles (paginación).

Ya ubicadas las etiquetas que poseían los datos de interés (links de redirección y máximo de páginas), se realizó un primer scrapping para extraer estos datos, lo catalogamos como “metadatos” que, si no nos apegamos a la definición estricta, serían los datos sobre los datos.

Con esos “metadatos” realizamos la consulta a cada página disponible concatenando el número de página en la url.

Así, situados en esta página entramos a cada una de estas publicaciones y de allí se extrajeron los datos para realizar el análisis, los datos obtenidos como precio, tipo de inmueble, tipo de operación, barrio, entre otros fueron guardados después de hacer una pequeña limpieza y normalización de forma automática. Por lo que en principio los datos ya contaban con una estructura y un cierto nivel de integridad.

Este proceso se repitió por cuantas publicaciones de inmuebles había en la página en cuestión. Por lo que, si hacemos un cálculo por encima, si el resultado de la paginación nos entregaba 100 páginas se realizó el scrapping de 24 x 100 páginas web. En un considerable número de veces que se consultó el servidor en un corto tiempo por lo que el trabajo de lectura de los datos tomó un tiempo considerable. También se separaron los tipos de moneda utilizados.

Una vez obtenidos los datos fueron almacenados por medio de una base de datos no relacional, en mongoDB, en un archivo tipo json.

Para llevar a cabo este proceso, se utilizaron las siguientes herramientas y técnicas:

## **Librerías de Python:**

Se emplearon librerías como requests y BeautifulSoup para realizar peticiones HTTP y analizar el código HTML de la página web.

## **Pasos del Proceso:**

### **a. Obtención de la URL:**

Se seleccionó Clasificados La Voz como la fuente de datos para el Web Scraping.

### **b. Petición HTTP:**

Se utilizó la librería requests para obtener el código HTML de la página web a través del método get.

### **c. Análisis del Código HTML:**

Se empleó BeautifulSoup para identificar y extraer los elementos relevantes que contienen los anuncios de inmuebles.

### **d. Extracción de Información:**

Se iteró sobre los elementos encontrados y se utilizó el método find para buscar y get\_text para obtener el texto deseado.

### **e. Almacenamiento de Datos:**

La información extraída se guardó en una lista para su posterior procesamiento.

## **Procesamiento Adicional:**

Javier Pagano y Gonzalo Cortez llevaron a cabo la clasificación de la moneda (pesos argentinos o dólares) y separaron los distintos tipos de inmuebles, la cantidad de habitaciones, la zona y ubicación, la fecha de actualización y la descripción de cada anuncio. Base de Datos:

Se decidió utilizar MongoDB para almacenar los datos extraídos, estableciendo así una conexión para su posterior análisis y consulta.

## **Análisis en tableau:**

Posteriormente, se procedió a utilizar tableau para el análisis de los datos extraídos:

## **Creación de Gráficos:**

Se generaron gráficos para visualizar tendencias de precios, distribución de ubicaciones, etc.

### **Tendencia de Precios:**

Se crearon gráficos que muestran la evolución de los precios a lo largo del tiempo, permitiendo identificar posibles patrones y tendencias.

### **Distribución de Ubicaciones:**

Se elaboraron visualizaciones que muestran la distribución geográfica de los inmuebles disponibles, proporcionando una visión clara de las zonas más demandadas.

## **Conclusión**

El proceso de Web Scraping en la página “**Clasificados La Voz**” se llevó a cabo de manera exitosa. La extracción de información relevante sobre anuncios de inmuebles y la posterior carga de datos en MongoDB facilitó el análisis y consulta de la información recopilada, y el posterior análisis en Tableau proporcionaron valiosa información sobre los anuncios de inmuebles que se pudieron ver reflejados en los gráficos realizados. Lo que estamos en condiciones de afirmar que cumplimos con nuestro objetivo ya que tal análisis permitirá tomar decisiones asertivas en el ámbito inmobiliario.