**Nombre**: Sinthia Guaigua Guanopatin

**Aula**: M2.851 - Tipología y ciclo de vida de los datos aula 1

**Fecha**: 03-04-2021

**Desarrollo de Preguntas:**

1. Contexto. Explicar en qué contexto se ha recolectado la información. Explicar por qué el sitio web elegido proporciona dicha información.

   Obtener información de Venta de casas y compra de terrenos en Ecuador. El negocio se centra en la venta de casas nuevas o usadas. Adicionalmente para construir nuevas casas se necesita revisar costos de terrenos para adquirirlos.

   El sitio web https://www.inmovision.com.ec/, provee información de ventas de casas y terrenos, el cual tiene una página principal en la cual se cargan todas las ventas disponibles en Ecuador, dando clic en cada link se puede obtener información de la propiedad de venta.

2. Título. Definir un título que sea descriptivo para el dataset.
   Titulo: Propiedades de Venta en Ecuador

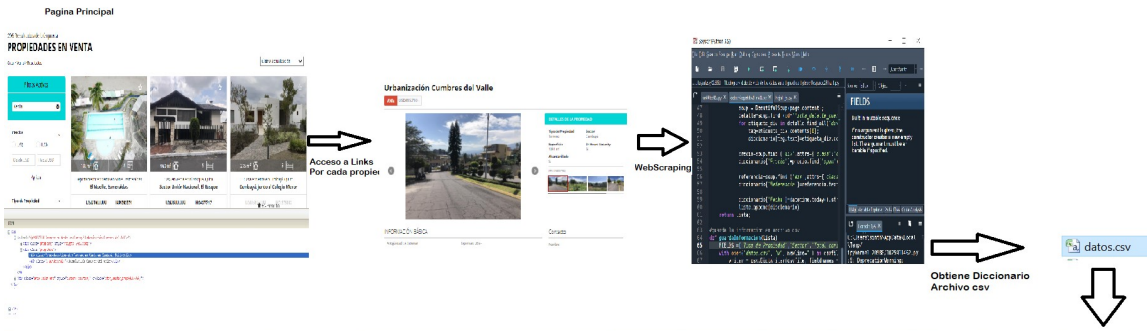
   Inicialmente se realiza Web Scraping en una página web, la idea es obtener información de otras páginas de venta de inmuebles en el Ecuador.

3. Descripción del dataset. Desarrollar una descripción breve del conjunto de datos que se ha extraído. Es necesario que esta descripción tenga sentido con el título elegido.

   Se obtiene información con respecto a propiedades de venta, para poder hacer un análisis del precio de venta de casas de la competencia y adicionalmente conocer el precio de terrenos para poder adquirirlos, entre la información relevante se tiene:

   Tipo de Inmueble, Ubicación, Referencia Ubicación, Superficie Construcción, Superficie total, Costo

4. Representación gráfica. Dibujar un esquema o diagrama que identifique el dataset visualmente y el proyecto elegido.




| Tipo de Propiedad | Sector | Total constr | Precio | Referencia | Dormitorios | Baños | Antiguedad | Superficie | Ambientes | 24 Hours Sec | Alcantarilla |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Departamento | Same | 160 m² | USD160.000 | El Muelle, Esmeraldas | 3 | 2 | 15 | | 4 | | |
| Casa | El Bosque | 960 m² | USD800.000 | Sector Unión Nacional, El Bosque | 5 | 5 | 38 | 1443 m² | 6 | | |
| Casa | Cumbayá | 275 m² | USD360.000 | Cumbayá, junto al Colegio Menor | 3 | 3 | A Estrenar | 60 m² | 4 | | |

5. Contenido. Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se han recogido.

**Campos que incluye el dataset**

Tipo de Propiedad: puede ser terreno o casas
Sector: ubicación del inmueble
Total construido: superficie de construcción
Precio: precio del inmueble
Referencia: referencia de la ubicación, puede indicar si esta dentro de una urbanización
Dormitorios: dato que estará lleno en el caso de casas
Baños: dato que estará lleno en el caso de casas
Antiguedad: dato que estará lleno en el caso de casas
Superficie: total de superficie
Ambientes: dato que estará lleno en el caso de casas
24 Hours Security: dato que estará lleno en el caso de casas
Alcantarillado:
Parqueadero fijo: dato que estará lleno en el caso de casas
Seguridad 24HsExpensas:
Fecha: fecha en la que se obtiene la información con web Scraping

**Periodo de tiempo de los datos:**

Los datos que se encuentran en la página web son de propiedades que se están vendiendo actualmente. El tiempo en que se obtenga la información dependerá de la ejecución de nuestro código fuente python

**Como se han recogido:**

1) Se carga la información de la pagina web elegida
2) Se desarrolla un función para la carga de la pagina web elegida, con scroll infinito, esto debido a que la pagina cargaba la información con mas propiedades, cada vez que se desplazaba el scroll al final de la pagina web
3) Se realiza una función para obtener los links de las propiedades que se encuentran en la página principal
4) Por cada link, se carga la página y se realiza una función para obtener la información detallada de cada propiedad y se va almacenando en un arreglo de diccionarios
5) Se realiza una función para enviar a guardar la información del arreglo de diccionarios en archivo csv, para esto se especifica las cabeceras que va a tener nuestro dataset.

6. Agradecimientos.

**Presentar al propietario del conjunto de datos. Es necesario incluir citas de análisis anteriores o, en caso de no haberlas, justificar esta búsqueda con análisis similares.**

El propietario de la información es la inmobiliaria Innovación.
Existen varios proyectos de Web Scraping a inmobiliarias, a nivel internacional y unos pocos a nivel de Ecuador, por ejemplo:

Sistema de Web Scraping orientado a portales del ámbito inmobiliario, por Álvaro Torrente Patiño
Web Scraping Masivo de Alquileres de Viviendas, por Giovanni Savio, María Paz Collinao, Bruno Lana y Rodrigo Lara (CEPAL)

**Justificar qué pasos se han seguido para actuar de acuerdo a los principios éticos y legales en el contexto del proyecto.**
No existe el archivo robots en la pagina https://www.inmovision.com.ec/, por ende no se tendría restricciones indicadas por parte del dueño de la pagina.
Se verifica el archivo robots del software Inmobiliario que construyo la pagina y en el archivo robots de dicha pagina no tiene restricciones (https://www.tokkobroker.com/robots.txt)

7. Inspiración. Explicar por qué es interesante este conjunto de datos y qué preguntas se pretenden responder. Es necesario comparar con los análisis anteriores presentados en el apartado 6.
   El negocio en el que me encuentro es construcción de casas o restauración de casas usadas, con el fin de venderlas, por lo que es muy importante conocer el precio que está ofreciendo la competencia para poder tener precios competitivos.
   Por otra parte para seguir construyendo o restaurando casas es importante conocer terrenos que se encuentran de venta y los precios, de igual forma de casas antiguas o semi terminadas que necesitan una remodelación o restauración para la venta.

8. Licencia. Seleccionar una de estas licencias para el dataset resultante y justificar el motivo de su selección: ● Released Under CC0: Public Domain License. ● Released Under CC BY-NC-SA 4.0 License. ● Released Under CC BY-SA 4.0 License. ● Database released under Open Database License, individual contents under Database Contents License. ● Other (specified above). ● Unknown License.

   Released Under CC BY-NC-SA 4.0 License, el propietario de la información es el sitio web de la inmobiliaria, a pesar que no tiene restricciones en su archivo robots, se debe citar siempre el sitio web del cual se obtiene la información, de igual forma de la persona que obtuvo el dataset. Este dataset sería para fines no comerciales.

9. Código. Adjuntar en el repositorio Git el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.

   El código fuente fue realizado en python, se encuentra en el siguiente link:

   https://github.com/sinthia2g/TipologiaCicloVidaDatos.git

10. Dataset. Publicar el dataset obtenido(*) en formato CSV en Zenodo con una breve descripción. Obtener y adjuntar el enlace del DOI.

    DOI: 10.5281/zenodo.6413272
    https://zenodo.org/record/6413272#.Yku3SSjMKK1

11. Vídeo. Se debe hacer entrega de un vídeo explicativo de la práctica en donde cada uno de los integrantes del grupo explique con sus propias palabras tanto las respuestas del proyecto como el código utilizado para llevar a cabo la extracción. El vídeo debe ser enviado a través de un enlace a Google Drive que deben proporcionar, junto con el enlace al repositorio Git, al momento de entregar la práctica

    https://drive.google.com/file/d/12jTBnpY-p7Mkxjuz_m_6Kh3FdLvjMQoA/view?usp=sharing

| Contribuciones | Firma, Integrante 1 |
|---|---|
| Investigación previa | Sinthia Guaigua  Firmado electrónicamente por: **SINTHIA ELIZABETH GUAIGUA GUANOPATIN** |
| Redacción de las respuestas | Sinthia Guaigua  Firmado electrónicamente por: **SINTHIA ELIZABETH GUAIGUA GUANOPATIN** |
| Desarrollo del código | Sinthia Guaigua  Firmado electrónicamente por: **SINTHIA ELIZABETH GUAIGUA GUANOPATIN** |
| | |