



## Práctica I

# Tipología y ciclo de vida de los datos

### Componentes:

Arturo González Díez - 71454421-E

#### 1. Título del dataset. Poned un título que sea descriptivo.

IguanaSell\_GOT\_dataset

#### 2. Subtítulo del dataset. Agregad una descripción ágil de vuestro conjunto de datos por vuestro subtítulo.

Productos de Game of thrones oficiales vendidos en IguanaSell

#### 3. Imagen. Agregad una imagen que identifique vuestro dataset visualmente





#### 4. Contexto. ¿Cuál es la materia del conjunto de datos?

El dataset incluye los productos vendidos por la empresa IguanaSell referentes a la serie Juego de tronos y sus libros homónimos, en su mayoría plumas y agendas.

#### 5. Contenido. ¿Qué campos incluye? ¿Cuál es el periodo de tiempo de los datos y cómo se ha recogido?

**Date:** Fecha se recogen los datos.

**Time:** Hora en la que se recogen los datos.

**Name:** Nombre del producto, el formato que sigue IguanaSell en la mayoría de los casos es 'tipo marca nombre tipo\_de\_laca edicion\_limitada(opcional) referencia(opcional)'.

**Price:** Precio del producto en la tienda correspondiente.

**PriceCurrency:** Moneda de curso legal en la que se vende el producto

**Seller:** Distribuidor encargado de hacer las gestiones. IguanaSell actúa como distribuidora propia además de tienda pero también trabaja con otras distribuidoras como Aurora o MontBlanc.

**Availability:** Disponibilidad del producto en el distribuidor indicado. Existen dos valores: 'in\_stock' para los productos que se encuentran actualmente en el almacén y 'no\_stock' para los que no se tiene disponibilidad actualmente.

**ItemCondition:** Especifica información adicional a día de hoy, o al menos en los datos almacenados, únicamente guarda el valor 'New' no obstante, como es recomendable, se está almacenando en vista a que en el futuro pueda aportar información relevante.

Se adjuntan los datos desde el momento de creación el dataset el día 14/4/2018.

Para la generación de este dataset se ha empleado el lenguaje de programación Python ya que es uno de los más potentes para la realización de scraping. Se ha empleado el módulo de BeautifulSoup que, aunque su tiempo de ejecución es más lento que otras opciones, es de fácil uso e instalación.

La primera vez que se ejecuta, obtendrá los datos de una única página web creando objetos. Una vez analizada la url, se procederá a la creación del fichero que en caso de no existir se crearía (creando el encabezado correspondiente), si existe, se añade a continuación las siguientes líneas.

#### 6. Agradecimientos. ¿Quién es propietario del conjunto de datos? Includ citas de investigación o análisis anteriores.



La empresa poseedora de los datos es IguanaSell, empresa ubicada en Madrid focalizada en la venta de artículos de lujo para caballeros, destacando estilográficas y relojes, con precios entre los 50 y los 60.000€. Para un público muy selecto, IguanaSell se encuentra entre las más prestigiosas del país, así como la mejor posicionada a nivel SEO. Actualmente se encuentra expandiendo sus artículos a otros países, entre los que cabe destacar Alemania, donde actualmente se encuentra en el top5 en venta de artículos de este tipo.

Empresa para la que trabajé durante el año 2017 como analista programador y e-commerce e información que me ha corroborado mi compañero Fermín Fernández, product manager de IguanaSell.

Más información en <https://www.iguanasell.es/pages/sobre-nosotros>

### **7. Inspiración. ¿Por qué es interesante este conjunto de datos? ¿Qué preguntas le gustaría responder la comunidad?**

Al haber trabajado en esta empresa anteriormente, sé la organización interna y el mercado de este tipo de productos, es un nicho de mercado focalizado en unos productos concretos pero con grandes beneficios y ganancias. Por otra parte, estoy estudiando la viabilidad de la apertura de una tienda centrada en productos de series 'emblemáticas' de la actualidad como puede ser Perdidos y Juego de tronos y sabiendo la repercusión que tiene IguanaSell a nivel nacional e internacional se crea la necesidad de un análisis de sus productos para saber en qué precios se mueven y cómo poder hacer frente a los competidores.

También se estudia la posibilidad de hacer scraping a sus páginas de otros países como la web americana ya que en ocasiones (conocimientos internos) se producen desbarajustes en el cambio de moneda.

### **8. Licencia. Seleccionad una de estas licencias y decid porqué la habéis**

**seleccionado:**

**Released Under CC BY-SA 4.0 License:** de manera que pueda ser usada por otros usuarios, que puedan modificarla y construirla de acuerdo a sus necesidades incluso con fines comerciales siempre y cuando proporcione un enlace a la licencia e indicar si se han realizado cambios.

**9. Código: Hay que adjuntar el código con el que habéis generado el dataset, preferiblemente con R o Python, que os ha ayudado a generar el dataset**

**10. Dataset: Dataset en formato CSV**

Adjuntados

