

Web Scraping – Business Intelligence Project



Contenido

Introducción.....	2
Web Scraping	2
Power BI	5
Observaciones y Conclusión	7
Reference List.....	10

Introducción

Este proyecto tiene como finalidad extraer información valiosa de la página web de Lluesma Muebles (<https://www.muebleslluesma.com/>), una empresa dedicada a la venta de muebles de diseño, calidad e innovación. El propósito es conocer el mercado y la demanda de los productos que ofrece, y en particular comparar la marca Doos Design con otras marcas competidoras.

El objetivo es recolectar y analizar los datos sobre los productos y precios para poder evaluar el desempeño de Doos Design frente a sus competidores e identificar las oportunidades y amenazas que se presentan en el sector.

Mediante el uso de la biblioteca BeautifulSoup de Python, se aplicarán técnicas de web scraping para extraer los datos de forma automática y estructurada. Entre los datos que se buscarán se encuentran el nombre, la descripción, el precio y la categoría de cada producto. Estos datos nos permitirán analizar y comparar los productos de forma eficiente y precisa.

Una vez realizada la extracción de datos mediante web scraping, se utilizará la librería Pandas de Python para procesar y organizar los datos obtenidos. Se crearán tablas que contengan las columnas y los valores relevantes para el análisis, y se eliminarán los datos innecesarios o duplicados. También se generarán columnas de identificación única para cada fila, que servirán como clave primaria en Power BI.

En este proyecto, se utilizará Power BI para crear visualizaciones que muestren los resultados del análisis realizado. Estas visualizaciones permitirán a Doos Design obtener una visión más completa de su oferta y la de sus competidores. Así, podrá identificar sus fortalezas, debilidades, oportunidades y amenazas en el mercado.

Web Scraping

El web scraping es una técnica que nos permite obtener datos de cualquier sitio web. En este ejemplo, se ha utilizado el web scraping para extraer la información relevante de la página web Lluesma, una tienda online de muebles y decoración. Con estos datos, se pretende realizar un análisis de mercado que compare Doos Design con otras marcas competidoras.

Las marcas elegidas para hacer la comparación son:

- Andreu World. andreuworld.com
- Inclass. inclass.es
- Viccarbe. viccarbe.com
- Ondarreta. ondarreta.com

Estas marcas han sido elegidas por ser en primer lugar marcas españolas como lo es Doos Design y en segundo lugar porque son fabricantes punteros de muebles de diseño

modernos con mucha calidad y atención al detalle. Doos al ser una empresa pequeña, aspira a lograr un posicionamiento parecido a la de estas marcas en el mercado.

Elaboración de la base de datos.

En Lluesma, una web de ventas online especializada en productos de diseño, podemos acceder a un catálogo organizado por fabricantes. Esto nos facilita la búsqueda y el análisis de los datos, ya que no tenemos que filtrar por este criterio. Podemos ver la variedad y la calidad de los productos que ofrece esta web en un solo clic.

Para extraer los datos de la página web, se utilizó la librería **Beautifulsoup** de **Python** y se aplicó el filtro de búsqueda para seleccionar solo las categorías de Sillas, Sillones y Butacas. Estas categorías son las más relevantes para Doos Design, ya que se dedica al diseño de muebles de asiento. El objetivo del proyecto es analizar las características y precios de estos productos en el mercado online.

Tras inspeccionar y añadir la posición exacta HTML de cada producto, precio y descripción del código fuente de la página al código Python de BeautifulSoup, el resultado es el siguiente:

	Title	Description	Price
0	Butaca Nuez BU-2742 Andreu World	Butaca tapizada de respaldo alto con base cent...	2.059,42 €
1	Butaca Nuez madera BU-2743 Andreu World	Butaca tapizada de respaldo alto con base cent...	2.192,52 €
2	Butaca Alya BU-1538 Andreu World	Butaca respaldo alto tapizada en base patín de...	1.255,98 €
3	Reposapiés Beat Andreu World	Reposapiés tapizado doble función disponible e...	568,70 €
4	Butaca Alya Executive BU-1489 Andreu World	Butaca respaldo alto con base central giratori...	1.783,54 €
5	Butaca Alya Executive BU-1490 Andreu World	Butaca respaldo bajo con base central giratori...	1.456,84 €
6	Butaca Alya BU-1523 Andreu World	Butaca respaldo bajo con base central giratori...	1.456,84 €
7	Butaca Alya BU-1524 Andreu World	Butaca respaldo bajo con base central giratori...	1.588,73 €
8	Butaca Alya Bu-1539 Andreu World	Butaca respaldo bajo con base de patín de acer...	1.032,13 €
9	Butaca Bolete BIO Andreu World	Butaca tapizada con base de termopolímero. Dim...	1.553,64 €
10	Butaca Alya BU-1521 Andreu World	Butaca respaldo alto giratoria en base de made...	1.779,91 €
11	Butaca Alya BU-1525 Andreu World	Butaca respaldo bajo tapizada con base de 4 pa...	1.206,37 €

Aunque la información obtenida es bastante interesante y reveladora, esta base de datos ha de ser manipulada para facilitar su lectura por un lado y para que sea interpretable por el programa Power BI por otro lado.

El siguiente paso es limpiar y dividir las columnas para obtener una tabla legible y fácil de interpretar. Tras manipular la table, el resultado es el siguiente:

	Description	Price	Category	Product	Company
0	Butaca tapizada de respaldo alto con base cent...	2.059,42 €	Butaca	Nuez BU-2742	Andreu World
1	Butaca tapizada de respaldo alto con base cent...	2.192,52 €	Butaca	Nuez madera BU-2743	Andreu World
2	Butaca respaldo alto tapizada en base patín de...	1.255,98 €	Butaca	Alya BU-1538	Andreu World
3	Reposapiés tapizado doble función disponible e...	568,70 €	Reposapiés	Beat	Andreu World
4	Butaca respaldo alto con base central giratori...	1.783,54 €	Butaca	Alya Executive BU-1489	Andreu World

Como se puede observar la tabla es más legible y categorizada. Este proceso se lleva a cabo en todas las tablas de cada una de las marcas.

El siguiente paso es juntar todas las bases de datos obteniendo así todos los productos, precios y fabricantes en una misma tabla para ser manipulada en Power BI.

Merging all tables to create unique IDs

```
pd.set_option('display.max_rows', None)

df_merging = pd.concat([data_andreu_butacas, data_andreu_sillas_sillones, data_doos_butacas, data_doos_sillas_sillones, data_inclass
df_merging.head()
```

[13] Python

Unnamed: 0	Description	Price	Category	Product	Company
0	Butaca tapizada de respaldo alto con base cent...	2.059,42 €	Butaca	Nuez BU-2742	Andreu World
1	Butaca tapizada de respaldo alto con base cent...	2.192,52 €	Butaca	Nuez madera BU-2743	Andreu World
2	Butaca respaldo alto tapizada en base patín de...	1.255,98 €	Butaca	Alya BU-1538	Andreu World
3	Reposapiés tapizado doble función disponible e...	568,70 €	Reposapiés	Beat	Andreu World
4	Butaca respaldo alto con base central giratori...	1.783,54 €	Butaca	Alya Executive BU-1489	Andreu World

Finalmente se crea ID (identificación) única a cada marca, categoría y producto para poder visualizar y manipular la información en Power BI.

	Price	Category	Product	Company	Product_id	Price_id	Category_id	Company_id
0	2.059,42	Butaca	Nuez BU-2742	Andreu World	1	1	3	1
1	2.192,52	Butaca	Nuez madera BU-2743	Andreu World	2	2	3	1
2	1.255,98	Butaca	Alya BU-1538	Andreu World	3	3	3	1
3	568,70	Reposapiés	Beat	Andreu World	4	4	4	1
4	1.783,54	Butaca	Alya Executive BU-1489	Andreu World	5	5	3	1

Un punto importante para tener en cuenta es que se ha creado varias tablas únicas de la marca y su ID, la categoría y su ID y el producto y su ID para unirlos en el apartado de model view de Power BI.

Creation of categorized tables and excels for use on Power BI

```
df_category_id = df_merging[['Category', 'Category_id']]
df_category_id.drop_duplicates(keep='first', ignore_index=True, inplace=True)

df_product_id = df_merging[['Product', 'Product_id']]

df_price_id = df_merging[['Price', 'Price_id']]

df_company_id = df_merging[['Company', 'Company_id']]
df_company_id.drop_duplicates(keep='first', ignore_index=True, inplace=True)

df_combined_id = df_merging[['Category_id', 'Company_id', 'Product_id', 'Price_id']]
df_combined_id.drop_duplicates(keep='first', ignore_index=True, inplace=True)

with pd.ExcelWriter('Category id.xlsx') as writer:
    df_category_id.to_excel(writer, sheet_name='Category')

with pd.ExcelWriter('Product id.xlsx') as writer:
    df_product_id.to_excel(writer, sheet_name='Product')


with pd.ExcelWriter('Price id.xlsx') as writer:
    df_price_id.to_excel(writer, sheet_name='Price')


with pd.ExcelWriter('Company id.xlsx') as writer:
    df_company_id.to_excel(writer, sheet_name='Company')


with pd.ExcelWriter('Combined id.xlsx') as writer:
    df_combined_id.to_excel(writer, sheet_name='Combined')
```


[73] Python

Como resultado final, hemos obtenido un Excel con toda la información categorizada y lista para ser analizada con facilidad posteriormente.

 all_excels_merge

 Category id

 Combined id

 Company id

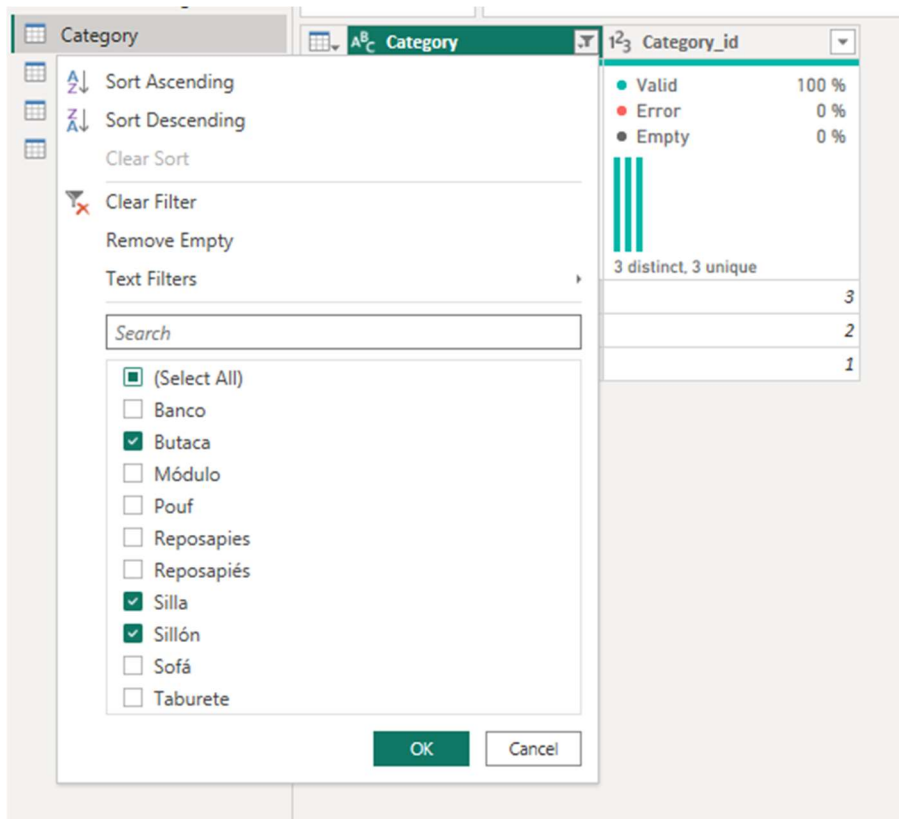
 Price id

 Product id

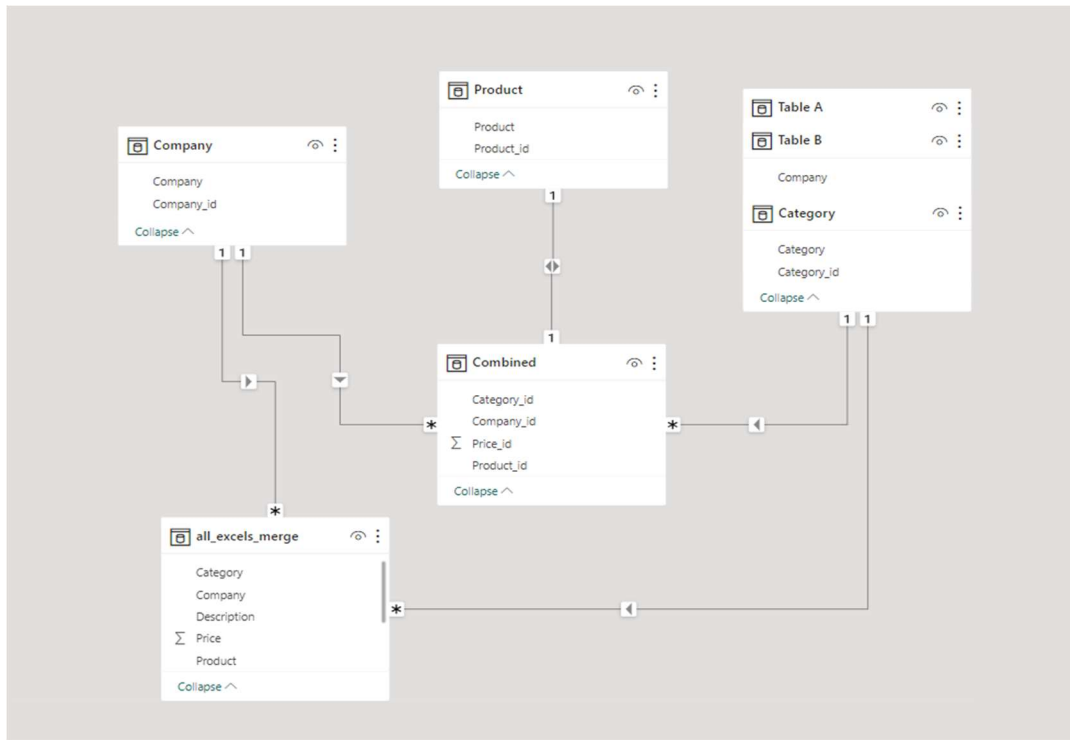
Power BI

Gracias a esta aplicación potente y versátil, podremos visualizar grandes cantidades de información de una manera fácil e intuitiva.

El primer paso es importar toda la información y manipularla en Power Query. En mi caso, la manipulación ha sido simple y rudimentaria ya que he limpiado y he preparado todo usando Python. Solamente he eliminado una columna con la enumeración de cada fila que viene predeterminada de Python y usando el mecanismo de filtrar de Power Query, he seleccionado solo las categorías de silla sillón y butaca.



Tras manipular toda la información en Power Query, llega el momento de crear los enlaces pertinentes entre las tablas para poder visualizarlas correctamente en el apartado de visualización. El resultado es el siguiente:



Tras crear las uniones entre las tablas, ya podemos empezar a visualizar nuestros datos.

Para realizar este análisis, contamos con una base de datos limitada y poco detallada, ya que no incluía información sobre las ventas totales de cada producto. Solo teníamos acceso a los productos y sus precios, por lo que el estudio se basará en una comparación simple de la cantidad y variedad de productos y el rango de precios de cada marca, información que nos permite identificar los nichos que puede atender cada marca y en particular comparar los precios de Doos y ver si son muy altos o bajos y si las otras marcas ofrecen productos similares a los de Doos.

El resultado final de la creación de visualizaciones es el siguiente:



Observaciones y Conclusión

Entre las marcas que ofrecen una gran variedad de productos, Andreu World e Inclass son las que más llaman la atención, pero también hay que destacar el rango de precios que manejan. Algunas de sus sillas tienen un precio inferior a los 150€ PVP, lo que resulta muy sorprendente si lo contrastamos con la silla más económica de Doos, que cuesta 304€ PVP. Sin embargo, al examinar con más detalle, se observa que la descripción señala que el material utilizado para fabricar las sillas de Andreu World e Inclass es el plástico, por lo que no sería justo compararlas con Doos, que no produce ese tipo de productos.

Como solución, se aplicó un filtro con Power Query para excluir los productos que mencionaban plástico, polipropileno, termo polímero o tecno polímero en la descripción. De esta manera, se obtuvo una lista de productos fabricados con materiales similares a los de Doos.

Después de realizar este cambio, notamos que los precios son más equilibrados. Considerando que todas las marcas analizadas en este proyecto tienen una gama de producto más extensa que la de Doos por lo que pueden ofrecer productos un poco más económicos que Doos, pero no son muchos esos productos.

Andreu World disponía de 145 productos

Inclass 127

Viccarbe 51

Ondarreta 44

Doos con 37

Al examinar las sillas tapizadas que es el fuerte de Doos, vemos que los precios están en el promedio y eso es algo positivo ya que es un buen indicador de que el producto puede competir con el producto de estas marcas reconocidas. Esto también se aplica a las categorías de sillones y butacas.

Marca	Butacas	Sillas	Sillones
Andreu World	1276€	552€	800€
Doos	1130€	450€	643€
Inclass	1061€	538€	639€
Ondarreta	1276€	494€	622€
Viccarbe	1288€	560€	811€

Un aspecto importante de Doos es que ofrece sillas, sillones y butacas de espuma inyectada; una tecnología innovadora que permite crear formas curvas en el diseño del producto, además de ser un material resistente que garantiza una larga duración. En esta categoría, Doos tiene un precio muy competitivo. Así, Doos brinda una calidad superior a un precio más accesible, lo que lo convierte en un factor diferencial en el mercado y en una ventaja de venta muy estratégica.

Doos ofrece otra ventaja competitiva en el mercado de los muebles, ya que sus productos tienen un precio más asequible cuando se trata de patas de madera y bases de madera con asiento tapizado. Estos elementos suelen encarecer el costo de otras marcas. Sin embargo, Doos no puede competir con las patas metálicas.

En resumen, Doos tiene una ventaja competitiva en el mercado de los muebles inyectados y tapizados con patas de madera, ya que ofrece una calidad superior a un precio asequible. Esta es una característica distintiva que lo diferencia de sus competidores, que cobran más por productos similares. Por lo tanto, Doos debería enfocar su estrategia de marketing en resaltar este atributo y atraer a los clientes que buscan muebles de diseño y confort.

Un aspecto que Doos debe tener en cuenta es que no puede competir en el mercado por el precio de sus productos, ya que los más demandados son de un material que Doos no utiliza, lo que le impide acceder a ese segmento de clientes. Estos productos requieren una producción masiva y un gran stock por modelo y por color, algo que Doos no tiene, ya sea por falta de espacio, capacidad productiva o disponibilidad de moldes y tecnología para fabricar esas sillas.

Doos ha de concentrarse en un nicho que busca calidad con materiales naturales tales como la madera como factor diferenciador a la hora de promocionarse y venderse antes el público y recomendar a su red de agentes de que su factor diferenciador ante la competencia es esta.

Doos es una empresa que se dedica a la fabricación de muebles de madera de excelente calidad, utilizando materiales naturales y ecológicos. Su propuesta de valor se basa en ofrecer productos resistentes, sofisticados y personalizados, que se adaptan a las necesidades y gustos de cada cliente. Doos ha de concentrarse en los nichos de mercado que valora la calidad por encima de la cantidad, y que busca diferenciarse del resto con un estilo propio y exclusivo.

Para aumentar las ventas de sus productos diferenciadores, Doos debe seguir una doble estrategia. Por un lado, debe capacitar a su red de agentes para que destaquen el valor añadido de estos productos frente a la competencia y los posicionen como una solución óptima para los clientes. Por otro lado, debe diseñar una campaña de marketing que resalte las características y beneficios de estos productos, utilizando medios visuales y atractivos que capten la atención del público objetivo.

Reference List

[Convert Float to String in pandas DataFrame Column in Python \(4 Examples\) | Using astype\(\) & apply\(\) \(youtube.com\)](#)

[SPLIT PANDAS COLUMN | How to split items into multiple columns in a dataframe \(Python\) \(youtube.com\)](#)

[Export Pandas DataFrames to new & existing Excel workbook - YouTube](#)

[Full Power BI Guided Project | Microsoft Power BI for Beginners \(youtube.com\)](#)