

Web Scraping – Business Intelligence Project



Content

Introduction.....	2
Web Scraping	2
Database elaboration.....	3
Power BI	5
Observations and Conclusion	7
Reference List.....	10

Introduction

This project aims to extract valuable information from the website Lluesma Muebles (<https://www.muebleslluesma.com/>), a company that specializes in selling design, quality, and innovative furniture. The goal is to understand the market and the demand for the products it offers, and especially to compare Doos Design with other competing brands. The project will use web scraping techniques to collect and analyze data from the website, such as product categories, prices, ratings, description, etc. The project will produce a report that summarizes the findings and provides recommendations for improving the performance of Doos Design and identify the opportunities and threats that arise in the sector.

By using the **Beautifulsoup Python** library, web scraping techniques will be applied to extract the data in an automatic and structured way. This data will allow us to analyse and compare products efficiently and accurately.

Once the data extraction has been carried out through web scraping, the **Pandas Python** library will be used to process and organize the data obtained. Tables will be created that contain the relevant columns and values for analysis, and unnecessary or duplicate data will be removed. Unique ID columns will also be generated for each row, which will serve as the primary key in **Power BI**.

The goal of this project is to use Power BI to produce visualizations that display the outcomes of the analysis conducted. These visualizations will help Doos Design to gain a better understanding of its own and its competitors' products. This way, we will be able to determine your market position and the factors that affect it.

Web Scraping

Web scraping is a technique that allows us to obtain data from any website. In this example, web scraping has been used to extract the relevant information from the website Lluesma, an online furniture and decoration store. With this data, it is intended to carry out a market analysis that compares Doos Design with other competing brands.

The brands chosen to make the comparison are:

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

These brands have been chosen because they are, firstly, Spanish brands and secondly, because they are leading manufacturers of modern designer furniture with a lot of quality and attention to detail. Doos, being a small company, aspires to achieve a similar positioning to these brands in the market.

Database elaboration.

At Lluesma, an online sales website specialising in designer products, we can access a catalogue organised by manufacturers. This makes it easier for us to search and analyse the data, as we don't have to filter by this criterion. We can see the variety and quality of the products offered by this website in just one click.

To extract the data from the website, the BeautifulSoup Python library was used and the search filter was applied to select only the categories of Chairs, Armchairs and Seats. These categories are the most relevant to Doos Design as it is dedicated to the design of seating furniture. The aim of the project is to analyse the characteristics and prices of these products in the online market.

After inspecting and adding each product's exact HTML position, price, and description of the page's source code to BeautifulSoup's Python code, the result is as follows:

	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 €

Although the information obtained is quite interesting and revealing, this database must be manipulated to make it easier to read on the one hand and to be interpretable by the Power BI program on the other hand.

The next step is to clean and split the columns to get a readable and easy-to-interpret table. After manipulation, the result is as follows:

	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

As you can see, the table is more readable and categorized. This process is carried out on all the tables of each of the brands.

The next step is to bring all the databases together, thus obtaining all the products, prices, and manufacturers in the same table to be manipulated in 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

Finally, unique IDs are created for each brand, category, and product to be able to visualize and manipulate the information in 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

An important point to note is that several unique tables of the brand and its ID, the category and its ID, and the product and its ID have been created to join them together in the model view section of 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

As a result, we have obtained an Excel with all the information categorized and ready to be easily analysed later.

 all_excels_merge

 Category id

 Combined id

 Company id

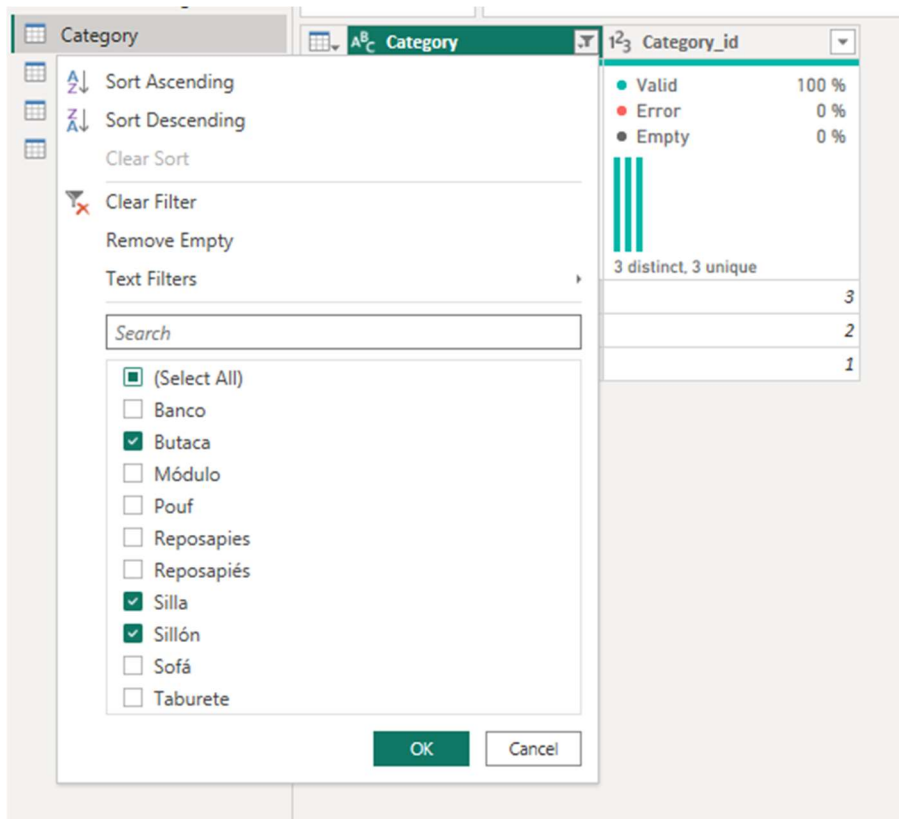
 Price id

 Product id

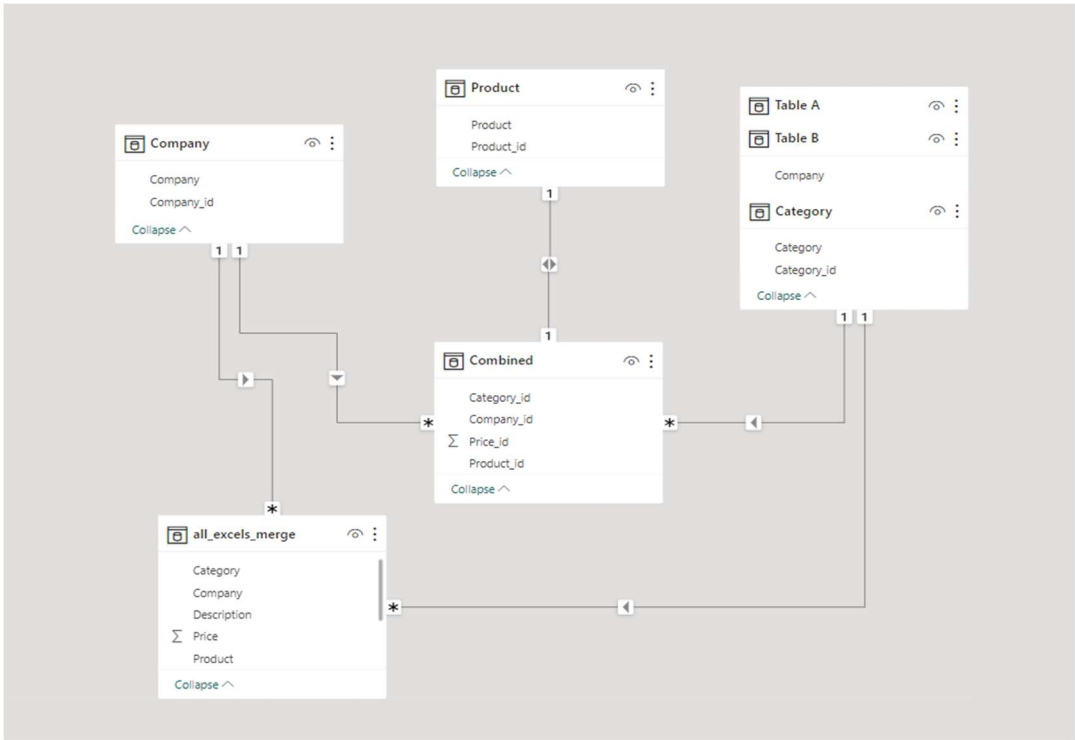
Power BI

Thanks to this powerful and versatile application, large amounts of information can be visualized in an easy and intuitive way.

The first step is to import all the information and manipulate it in Power Query. In my case, the manipulation has been simple and rudimentary as I have cleaned up and prepared everything using Python. I've only deleted one column with the enumeration of each row that comes from Python by default and using Power Query's filtering mechanism, I've selected only the categories of chair, armchair and seats.



After manipulating all the information in Power Query, it's time to create the relevant links between the tables so that they can be displayed correctly in the visualization section. The result is as follows:



After creating the joins between the tables, we can start visualizing our data.

To carry out this analysis, we had a limited and not very detailed database, as it did not include information on the total sales of each product. We only had access to the products and their prices, so the study will be based on a simple comparison of the quantity and variety of products and the price range of each brand, information that allows us to identify the niches that each brand can cater to and in particular compare Doos prices and see if they are very high or low and if the other brands offer similar products to Doos.

The result of creating visualizations is as follows:



Observations and Conclusion

Among the brands that offer a wide variety of products, Andreu World and Inclass are the ones that attract the most attention, but it is also worth highlighting the price range they handle. Some of their chairs are priced at less than €150 RRP, which is very surprising when compared to Doos' cheapest chair, which costs €304 RRP. However, on closer inspection, the description states that the material used to manufacture the Andreu World and Inclass chairs is plastic, so it would not be fair to compare them with Doos, which does not produce such products.

As a solution, a filter with Power Query was applied to exclude products that mentioned plastic, polypropylene, thermopolymer or technopolymer in the description. In this way, a list of products made with materials similar to those of Doos was obtained.

After making this change, we noticed that the prices are more balanced. Considering that all the brands analysed in this project have a wider product range than Doos, so they can offer products that are a little cheaper than Doos, but there are not many such products.

Andreu World had 145 products.

Inclass 127

Viccarbe 51

Ondarreta 44

Doos with 37

When examining the upholstered chairs that is Doos' forte, we see that the prices are in the average and that is a positive thing as it is a good indicator that the product can compete with the product of these well-known brands. This also applies to the categories of armchairs and seats.

Brand	Seats	Chairs	Armchairs
Andreu World	1276€	552€	800€
Doos	1130€	450€	643€
Inclass	1061€	538€	639€
Ondarreta	1276€	494€	622€
Viccarbe	1288€	560€	811€

An important aspect of Doos is that it offers injected foam chairs, armchairs, and seats; An innovative technology that allows the creation of curved shapes in the design of the product, as well as being a resistant material that guarantees a long life. In this category, Doos is very competitively priced. Thus, Doos provides superior quality at a more accessible price, which makes it a differentiating factor in the market and a very strategic sales advantage.

Doos offers another competitive advantage in the furniture market, as its products are more affordably priced when it comes to wooden legs and wooden bases with upholstered seat. These elements tend to make the cost of other brands more expensive. However, Doos can't compete with metal legs.

In short, Doos has a competitive advantage in the injected and upholstered furniture market with wooden legs, as it offers superior quality at an affordable price. This is a distinguishing feature that sets it apart from its competitors, who charge more for similar products. Therefore, Doos should focus its marketing strategy on highlighting this attribute and attracting customers who are looking for designer and comfort furniture.

One aspect that Doos must consider is that it cannot compete in the niche of low-price products, since the most demanded products are made of Plastic, a material that Doos does not use, which prevents it from accessing that customer segment. These products require mass production and a large stock per model and per color, something that Doos does not have, either due to lack of space, production capacity or availability of moulds and technology to manufacture those chairs.

Doos is a company that is dedicated to the manufacture of wooden furniture of excellent quality, using natural and ecological materials as a differentiating factor. Its value proposition is based on offering resistant, sophisticated, and personalized products, which adapt to the needs and tastes of each customer. Doos must concentrate on niche markets that value quality over quantity, and that seek to differentiate itself from the rest with its own exclusive style, and most importantly needs to recommend its network of agents that its differentiating factor against the competition is this.

Doos has a range of products that stand out from the competition and offer superior value to customers. To boost the sales of these products, Doos needs to implement a dual strategy. Firstly, Doos needs to train and inform its agents to emphasize the unique advantages of these products and how they can meet the customers' needs better than other alternatives. Secondly, Doos needs to launch a marketing campaign that showcases the features and benefits of these products, using attractive and interactive media that appeal to the target audience.

The products that Doos offers are innovative and high-quality, designed to solve the problems and challenges that customers face in their daily lives. These products have features that are not available in other products in the market, such as durability, convenience, efficiency, and sustainability. By training and informing its agents, Doos can ensure that they can communicate these features effectively to the customers and persuade them to choose Doos over other brands. The agents can also demonstrate how these products can fit the customers' preferences and lifestyles, and how they can provide value for money.

In addition to training and informing its agents, Doos also needs to launch a marketing campaign that showcases the features and benefits of these products, using attractive and interactive media that appeal to the target audience. The marketing campaign should highlight how these products can make the customers' lives easier, better, and more enjoyable. The marketing campaign should also use testimonials, reviews, and social proof to show how satisfied and loyal the customers are with these products. The marketing campaign should use various channels, such as social media, websites, blogs,

podcasts, videos, and newsletters, to reach the potential customers and increase their awareness and interest in these products.

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 \(youtube.com\)](#)