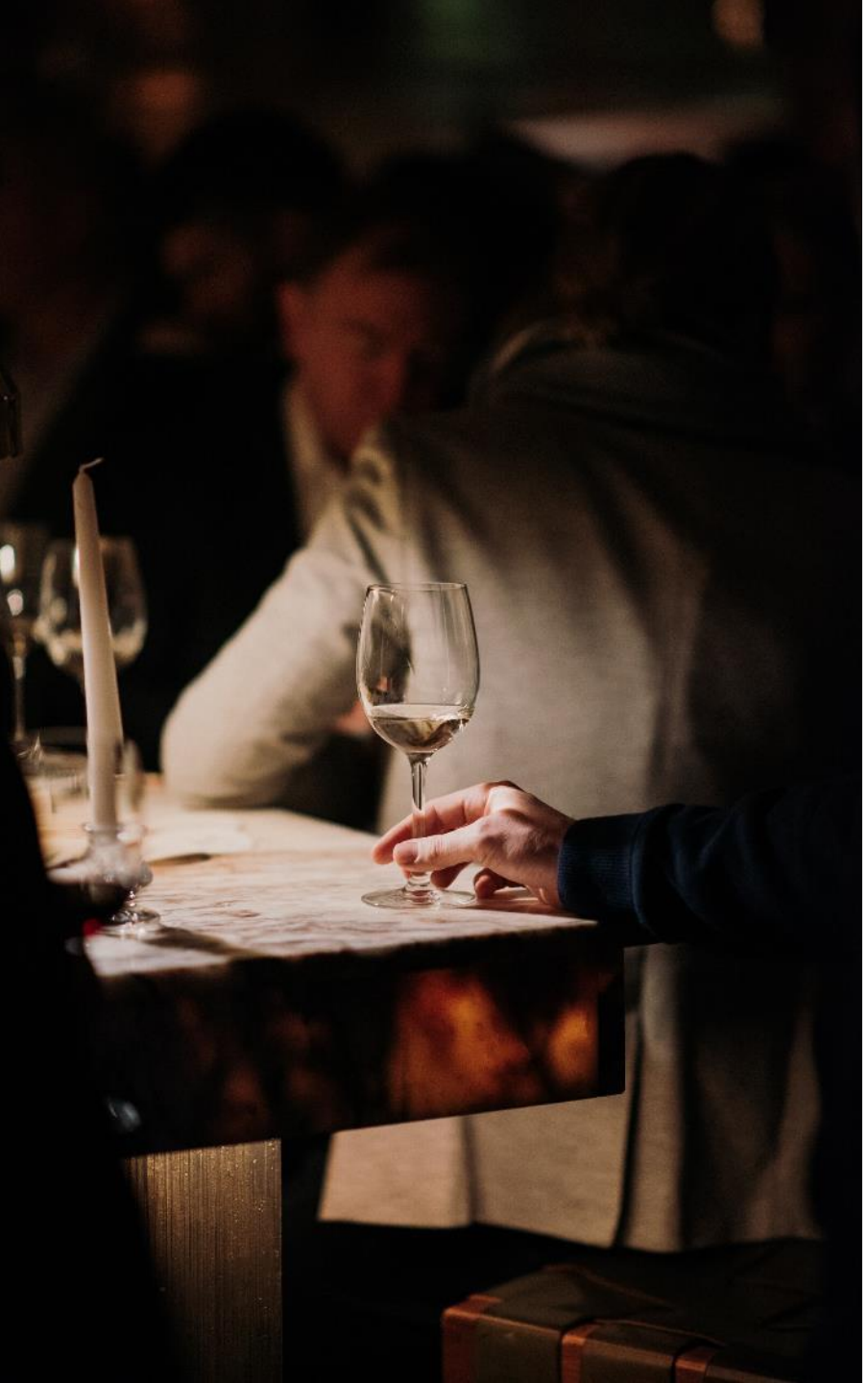
OPTIMISATION OF DATA MANAGEMENT

bottleneck
fine wine spirit



Context

- The ERP is not linked to our online sales site
- We need to improve our visibility in terms of online sales analyses
- While waiting for a more centralised solution, I created a connection between our 2 databases (ERP and CMS).

Methodology

3 Files

- ERP – product references, product prices, and stock status
- Web – information about products sold online
- Link – with this we can link the product ID in the ERP file to the SKU in the web file.

Methodology

- Clean data - types, outliers, duplicates, nulls, inconsistencies
- Determine primary key
- Merge DataFrames
- Check joins
- Undertake the analysis

File : Liaison

My approach

- Check the shape and the head of the dataframe
- Check data types – unexpected values but not true outliers
- Sum of zero values
  - 91 valeurs nulles dans la colonne id_web.
- As we want to see the turnover for online sales, I deleted all the lines with nan values in id_web.
- Check for duplicates – none
- Check for outliers in the product_id list - none.
- Change column name ‘id_web’ to ‘sku’.
- Determine primary key – 734 unique values for product id and shape = (734, 2).

File : Web

My approach

- Check the shape and the head of the dataframe
- Data types - total sales and post author
- Sum of zero values
  - 4 columns with all values missing
  - 2 rows where there is no SKU, but other columns have values
  - 83 rows with only null values
- Search for duplicates
  - There are duplicate row for SKU because each product has a line for post_types ‘attachment’ and ‘product’.
- Identify and remove columns with all NaN or zero values
- Remove columns that will not be used for analysis
- Determine primary key – 714 unique values for sku and shape = (714, 3)

File : ERP

My approach

- Check the shape and the head of the dataframe
- Data types
- Sum of zero values
- Find duplicates
- Outlier detection – negative values for price and stock quantity.
  - 2 columns with negative price values
  - 2 columns with negative stock_quantity values
- Detecting values with inconsistencies
- 1 line where stock quantity is 0, but stock_status is in stock.
- 5 lines where stock quantity is >0, but stock_status is out of stock
- Determine primary key – 823 unique values for product_id and shape = (823, 5)

Merging the Dataframes

My approach

- Merge the 'erp' and 'liaison' files on Product Id with a left join.
- Merge the erp-liaison and web files on SKU with an ‘inner’ join.
- Create the turnover column.

Inconsistencies

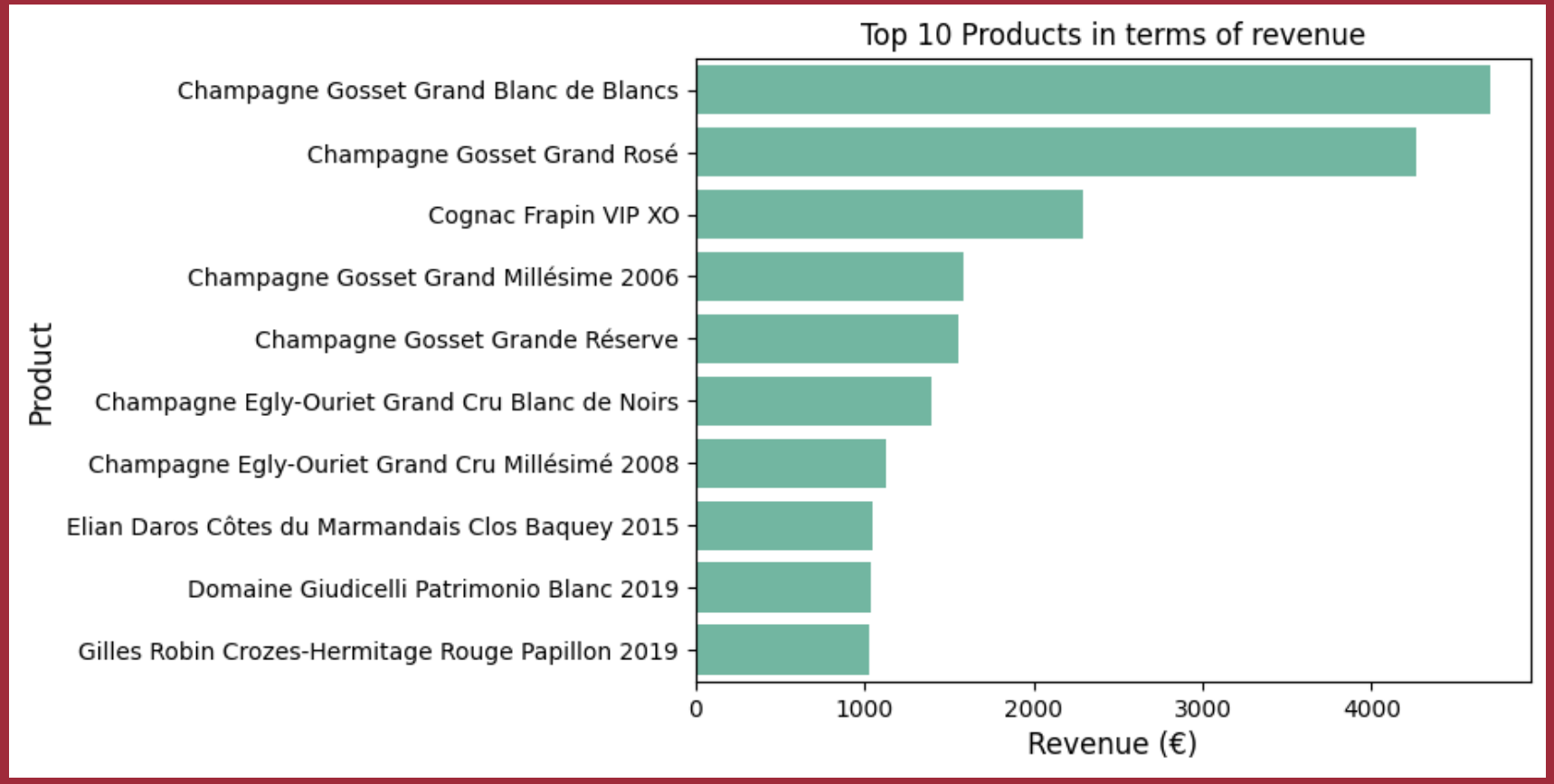
Web

- 2 rows with NaN values for SKU, but data in other columns

ERP

- The price column has negative values
- The stock_quantity column has negative values
- For 1 row, the ‘stock status’ column indicates instock when the ‘stock quantity’ = 0.
- For 5 row, the 'stock status' column indicates out of stock when the 'stock quantity' was > 0.

Top 10 Products in terms of revenue



70568.60 €

Total du chiffre d'affaires réalisé en ligne

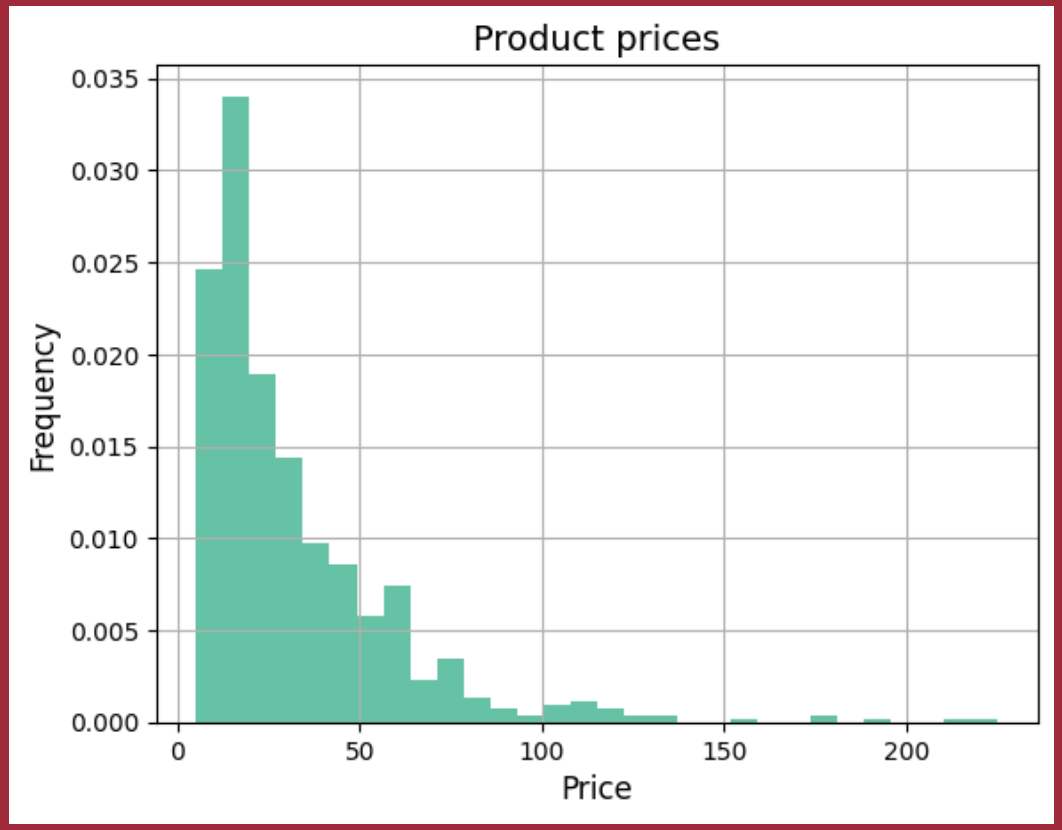
15.95 %

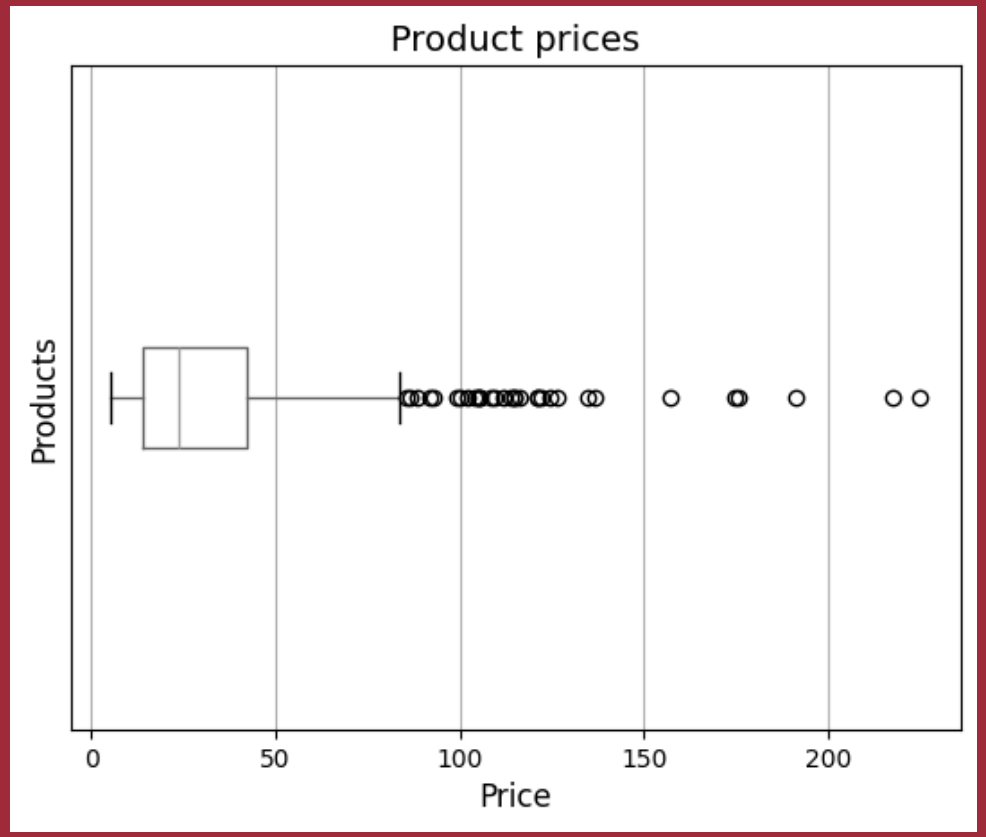
Pourcentage de notre chiffre d'affaires constitué par le top 3 produits

Price Analysis

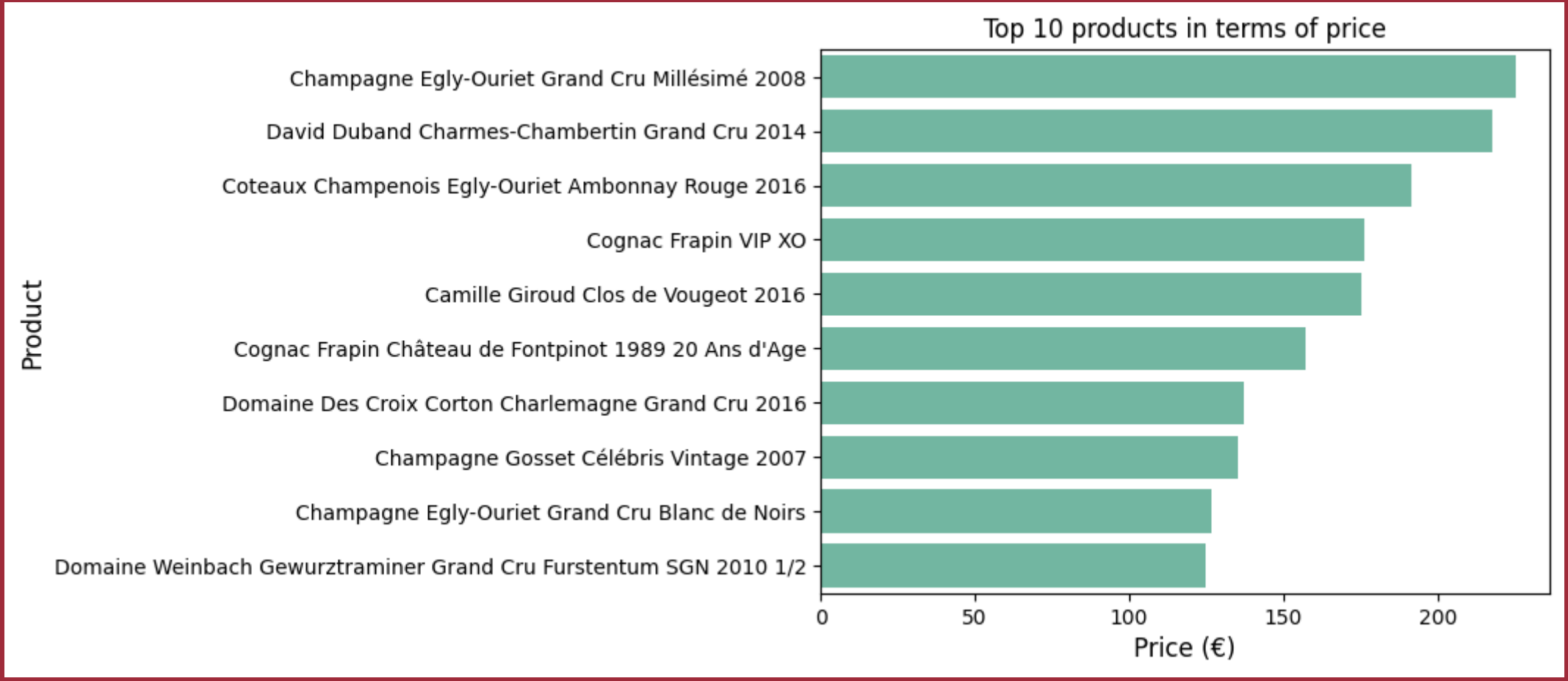
32.49 €	772.34		
Mean product price	Variance		
€ 5.20		27.79	28.08
Minimum price		Standard deviation	IQR
	23.55 €		
	Median product price		
€ 225		2.58	56
Maximum price		Skewness	Number of outliers
	19 €	10.09	
	Mode product price	Kurtosis	

Price Analysis: Graphs





Top 10 products in terms of price



Conclusion

1

No outliers – atypical values

Excluding negative values identified during data cleaning, I did not find any outliers. Instead, the values are simply atypical. Wine prices can vary widely, and the most expensive products are mostly cognacs, champagnes and grands crus, which you would expect to be expensive.

2

A lack of data

We have a number of NaN values in our data, particularly in the 'web' file. We should examine this to see if we can improve data collection.

3

Top products and turnover

Our total online turnover is €70,568.60. Our top 3 products contribute to 15.95% of this total. It could be interesting to study how to market our less popular products to further increase turnover.