HOW TO OPTIMIZE STRATEGIC BUSINESS DECISIONS?

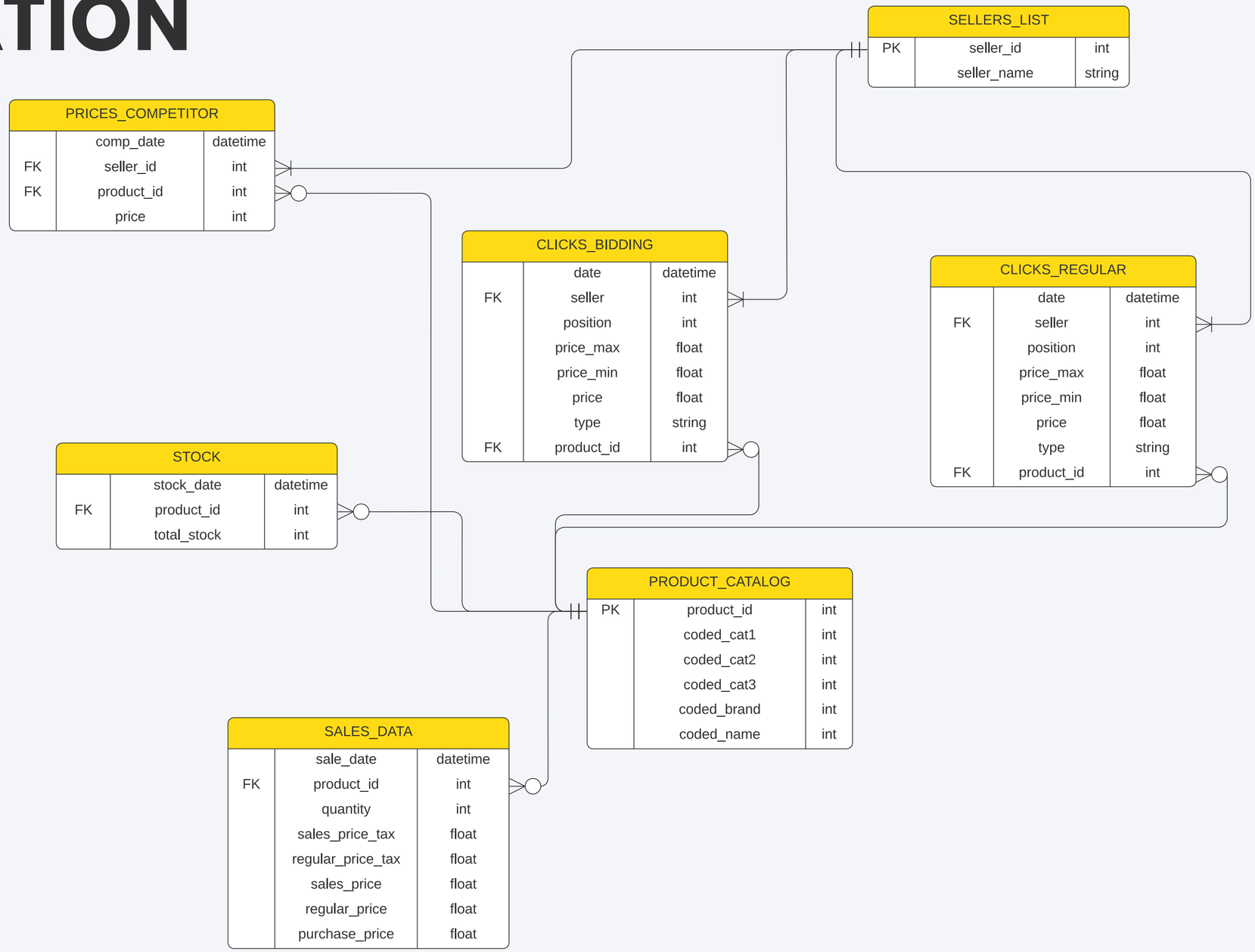
TASK 0: Getting comfortable with the data
TASK 1: Leaders and Followers & Pricing automation detection
TASK 2: Popularity Index detection

Task 0

DATA INTEGRATION

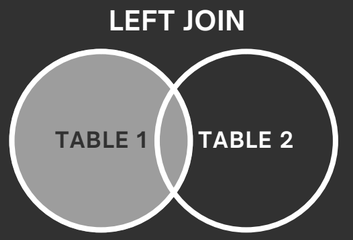
The **ER model** shows what are the entities and the relationships between them.
It was the starting point to understand how the datasets could have been merged.
For example, in *product_catalog*, product_id is the **primary key**, or the unique identifier for a product.
In *sales_data*, product_id is a **foreign key**.
What are the relationships?

- How many unique product_id's have been sold? *Zero to many*
- How many unique product_id's are in a single *sales* observation? *One and only one*



Task 0

DATA FAMILIARIZATION

Describe the data	The datasets provide data on the online sales of an e-commerce firm specializing in electronic products. Moreover, a broader perspective of this market is available through price comparison website data
What is the best way to merge the datasets with respect to my business needs?	Some data are split across multiple datasets. The best way to merge them is to use an outer join (typically a left join) used to match rows from different tables which share a key and retrieve data that are missing in one of them. For instance, in the *clicks_regular* there were 1093244 missing values for price. 87166 of them were retrieved by performing a left join with the *prices_competitor* basing on the match with date, seller, and product_id. 
How have we handled inconsistent values?	The values of the variables are mostly consistent, although there are some anomalies that we did not expect. Some examples below: For example, in *sales_data*, the minimum price of regular_price is lower than the minimum price of sales_price. It means that the minimum list price is lower than the minimum price after a promotion is applied. Moreover, the minimum value of taxation is 0, but this does not seem real because all goods are taxed. In *stock* the minimum value for the stock is negative so that the distribution appears right-skewed (slide 7).

Task 1

OUR APPROACH

Grouping the products into the categories into which they are divided would have reduced the dimensionality of the problem, but the aggregation of 7529 unique products into 13 categories, and thus the aggregation of their unique characteristics, such as price, would have neglected critical information, peculiar to each product, and sacrificed the accuracy of our predictions.

Value added: examining each product in **detail**, without aggregating valuable information.

Did not take a plot-based approach but a **rule-based** one.

Aggregation by categories could be implemented **at any time.**

Setting of abstract and general rules included in **softcoded algorithms**, whose parameters were the reference quarter and the code of the product to be analyzed.

Task 1

LEADERS AND FOLLOWERS DETECTION

GOAL: predict the price movement of the followers basing on the price movement of the leaders.

ASSUMPTION: followers will change the price with a lag of 1 day.

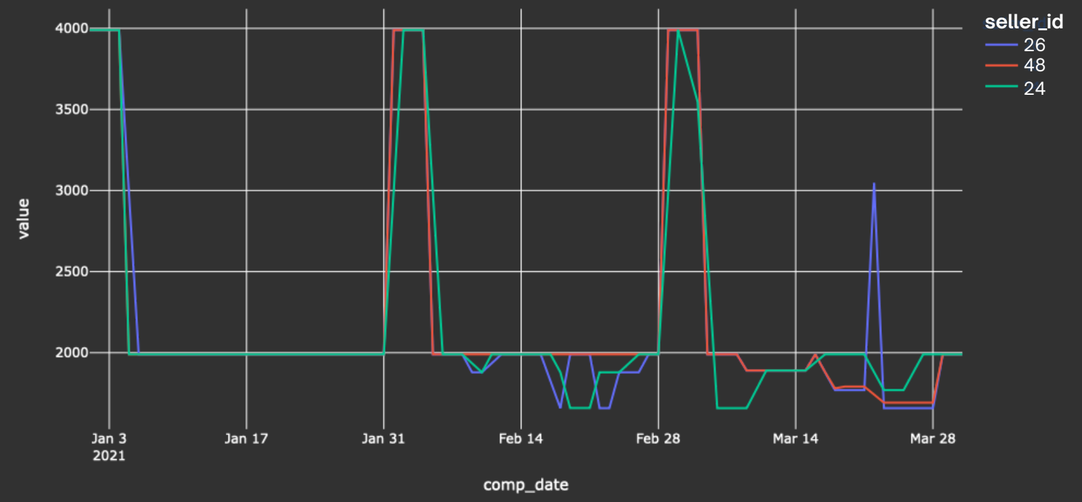
Pearson Correlation

For each seller, the correlation between every other seller is computed.
If this correlation is higher than a certain threshold (e.g. 0.8), a possible pair of leader and follower is identified.

Co-integration

For each leader-follower pair, the respective co-integration test's p-value is then computed.
Co-integration test is used to describe some underlying relationship between variables by testing the correlation between two or more non-stationary time series in the long run or for a specified period.

The leader-followers relationship is better inspected by plotting the prices they set over time and by visually analyse what is their behaviour:



Task 1

LEADERS AND FOLLOWERS DETECTION

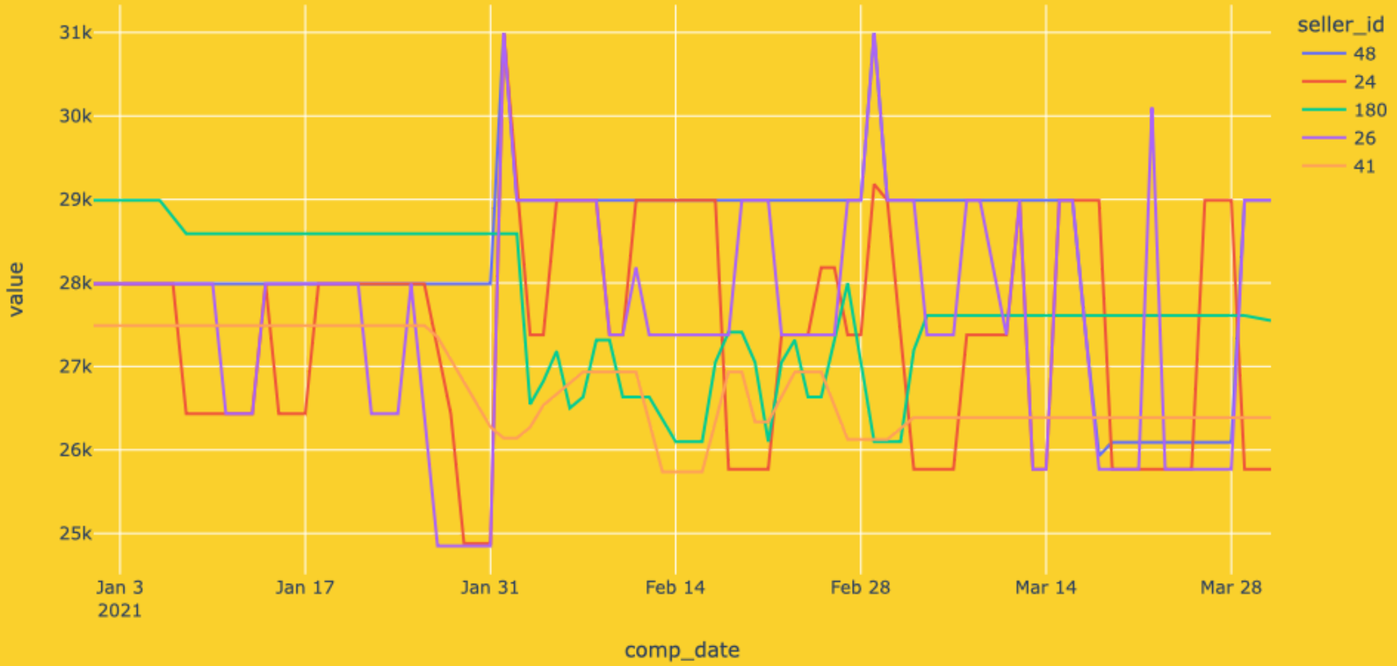
Granger causality

The Granger causality test is a statistical hypothesis test for determining whether one time series is useful in forecasting another and, therefore, in determining if a given time series and its lags is helpful in explaining the value of another series.

This test is performed for each possible pair of sellers.
The most significant ones (i.e., low p-value) will be the ones to determine who, in the analysed sellers pair, is the leader.

OUTPUT:

	Leader	Follower
0	48	26
1	24	180
2	180	26
3	180	41
4	180	48



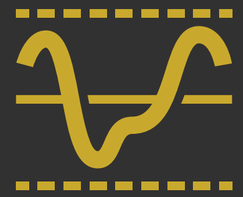
Task 1

AUTOMATED PRICING STRATEGY DETECTION

What are the rules that a pricing algorithm might follow?

It might set a price that always deviates by a **fixed percentage or value** over time (e.g., a seller sets a price that is always 10% lower than the price charged by the market leader).

It might set a new price after always the **same time lag** with respect to another seller (e.g., a seller always changes its price the day after the leader's price changes).



It might set up **maximum and minimum prices** to ensure that the company is not only gaining sales, but also gaining the highest profit margins possible. Once set up, the repricer will continuously study competitors' behaviour to automatically configure prices to allow the best results for the company.

Task 1

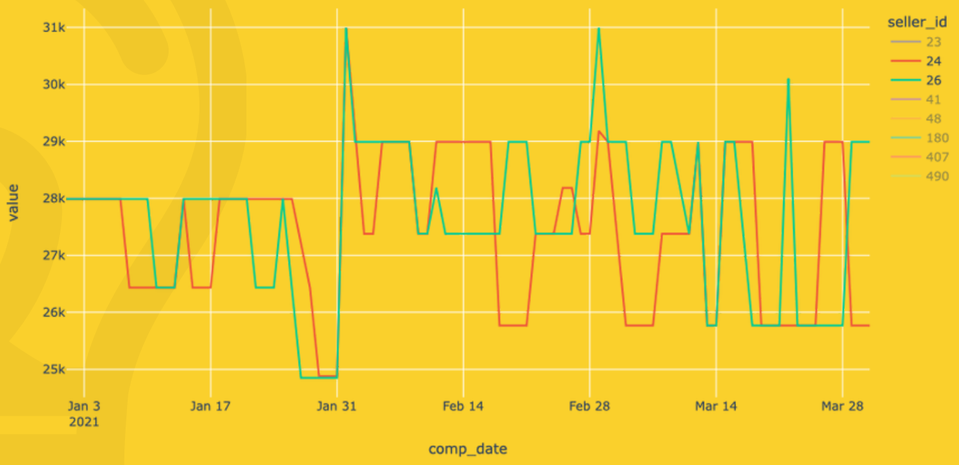
AUTOMATED PRICING STRATEGY DETECTION

Visual analysis

MANUAL PRICING AUTOMATION DETECTION

For each product id, sellers' time series are plotted in the same graph in order to visually analyse them with the aim of understanding whether sellers use pricing algorithms that follow certain rules.

Sellers that might use an automated pricing strategy:



Algorithmic approach

AUTONOMOUS PRICING AUTOMATION DETECTION

For each seller, the price difference with all other sellers is computed. If this difference is more or less constant over time, a seller always deviates by the same percentage/value from another (1st rule).
In the same way, if the number of days between when sellers change their prices is consistent over time, a seller changes its price with respect to another after always the same time lag (2nd rule).

OUTPUT:

	23	24	26	41	48	180	407	490
23	B	A	A	B	A	A	A	B
24	A	B	A	A	A	A	A	A
26	A	A	B	A	A	A	A	A
41	B	A	A	B	A	B	A	B
48	A	A	A	A	B	A	A	A
180	A	A	A	B	A	B	A	B
407	A	A	A	A	A	A	B	B
490	B	A	A	B	A	B	B	B

E.g.,
Sellers assigned the label 'B' might use an algorithm that sets prices that follows the time lag rule.

Task 1

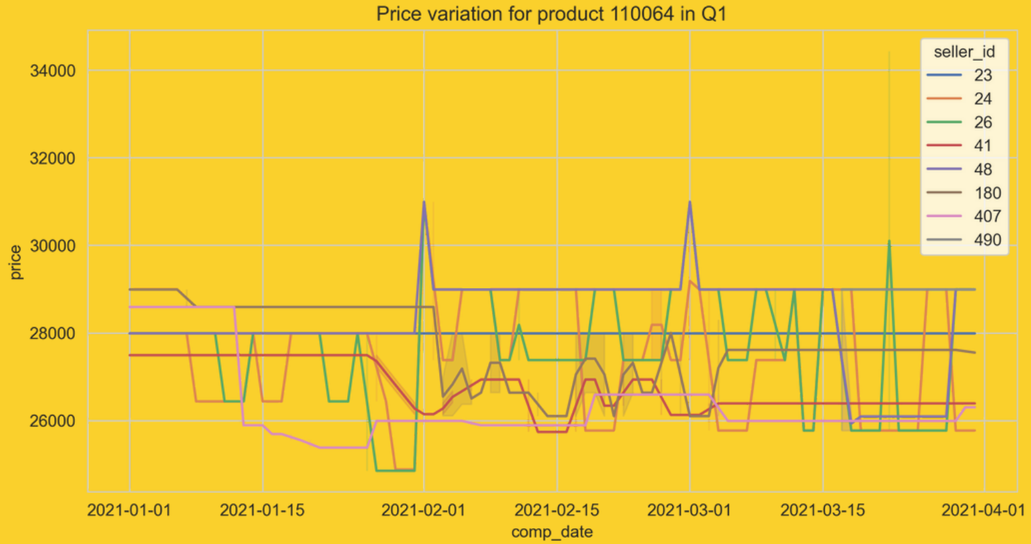
MIN/MAX PRICE STRATEGY

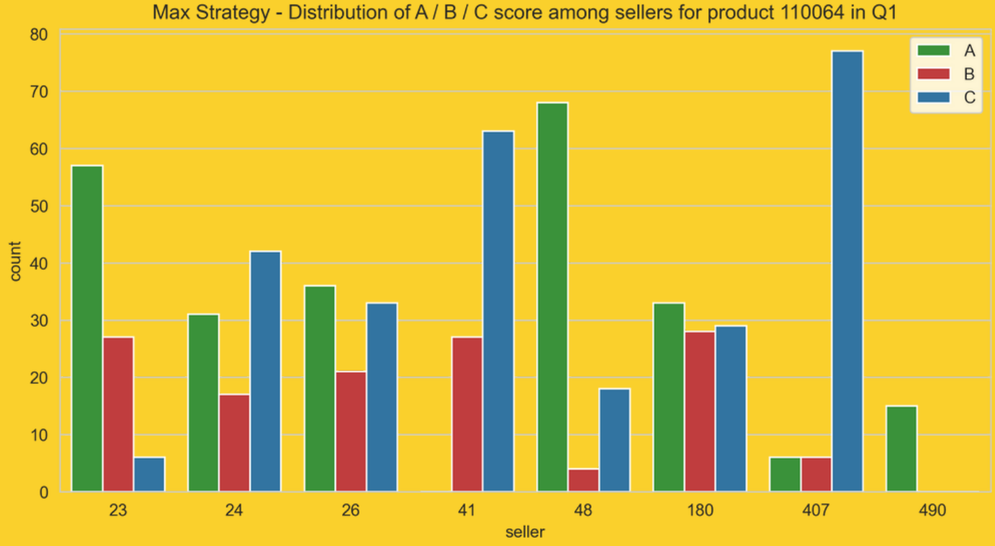
GOAL: determine the behaviour of each seller with respect to the maximum price and to the minimum price.

Take the maximum price strategy as an example:
For each day, the percentage difference between each seller's price (P of seller i at time t) and the maximum price of the other sellers is calculated.

- If the difference is positive, then the seller will charge the highest price
- If the difference is zero, then the price charged by the seller is the highest one and there is at least one other seller with the same price.
- If the difference is negative, then the seller will charge a lower price than the maximum price.

If the value of the array corresponding to seller Z belongs to the first quantile, then seller Z will score A, otherwise B or C if in the second or third. By rephrasing, if seller Z charges a higher price than the others on day 1, then it will score A.
The algorithm is iterated for each day and a countplot is rendered showing how many times A, B and C scores are assigned for each seller.





Task 2

POPULARITY INDEX

The calculation of products' popularity index is useful to help the company making future decisions including budget allocation, trend forecasting, inventory management, and more.

Sales Approach

The popularity index is obtained by dividing the quantities sold of each individual product against the total amount of products sold in that period.

- For each product, it is visually analyzed how much it contributes to the company’s profit.

- Products with similar features in terms of profitability and popularity are identified using a clustering algorithm (Kmeans ++) so as to target the same strategies to products with similar characteristics.

Task 2

POPULARITY INDEX

Strategy diversification

Products that fall in the same area are assigned the same strategies. Below is an example of division into categories with corresponding strategies.

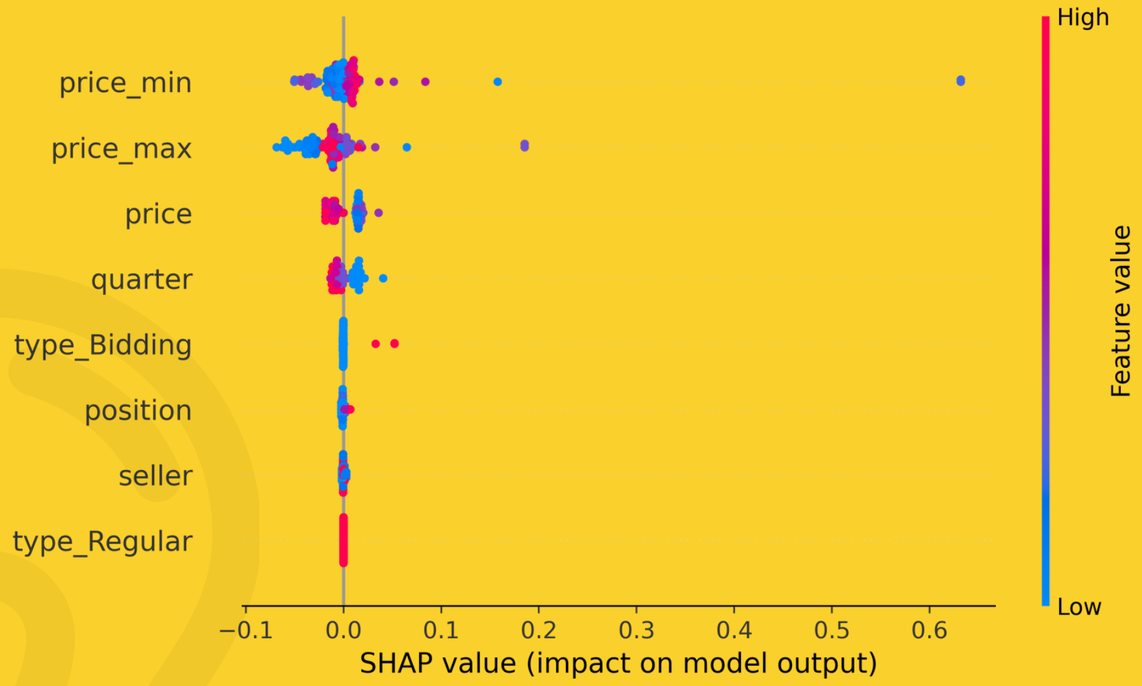
AREA	STRATEGY
Poor popularity and poor profitability	Analysis on investment suitability
Acceptable popularity and profitability	Analyse and push product to area 3 or 4. Bundeling
High profitability and poor popularity	Increase popularity by sacrificing some profit decreasing the price. (Psychological Pricing)
High popularity and low profitability	Raise the price but issue coupons in return
High profitability and popularity	Perfect situation. Keep it up!

GOAL: push all products towards area 5.

Task 2

POPULARITY INDEX

Which are the variables that most affect the products' popularity index?



Competitor prices (minimum and maximum price in the market) are the ones that most influence how much a product is clicked.

The seller does not turn out to be important in defining the popularity index for a given product.

Unexpectedly, the position within the site does not affect the popularity index.

The quarter slightly influence the target variable, showing that there are products with seasonality whose popularity changes over time.

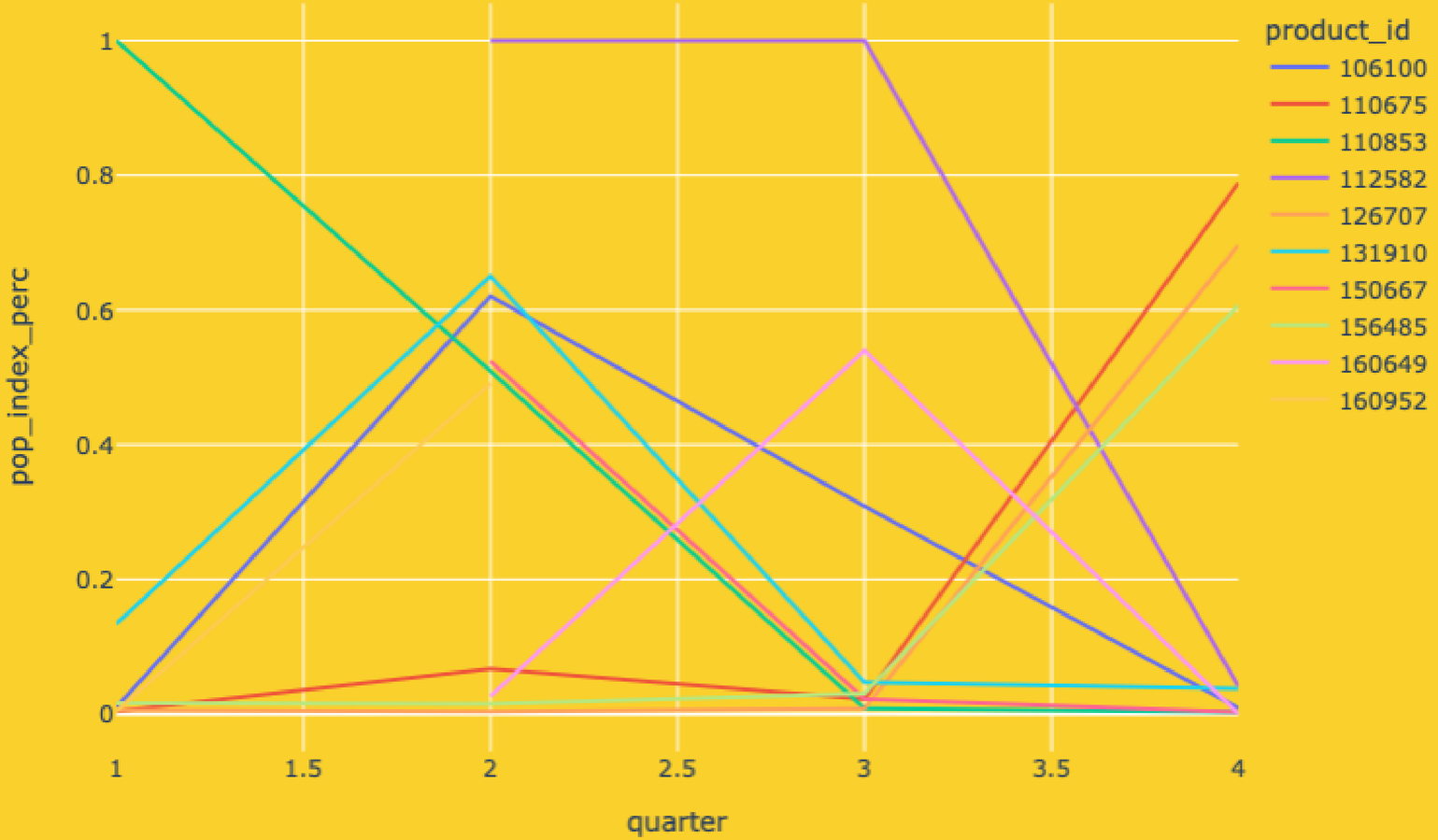
Task 2

POPULARITY INDEX

Sales Approach

The popularity index is obtained by dividing the quantities sold of each individual product against the total amount of products sold in that quarter.

- Compute the standard deviation of the popularity index of each product over quarters.
- Rank the top 10 products with the highest standard deviation
- Plot the popularity index of the top 10 products
- Analyzes the behavior of these products to detect interesting insights: could they potentially be affected by seasonality effects?



THANK YOU
FOR THE ATTENTION!

Fabiana Caccavale
Matteo Gioia
Martina Manno
Vincenzo Junior Striano