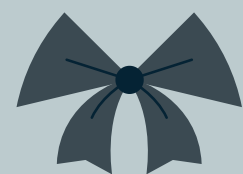


DATA DRIVEN DECISIONS



LUISS



Group



Lorenzo Antolini, Enrico Grandi,
Simone Lu, Enrico Romano



INDEX

1

OVERVIEW OF THE PROJECT

2

Exploration

- EDA
- Data Familiarization
- Data Cleaning
- Data Visualization

3

MARKET INSIGHTS

- Hidden patterns and trend in the data

4

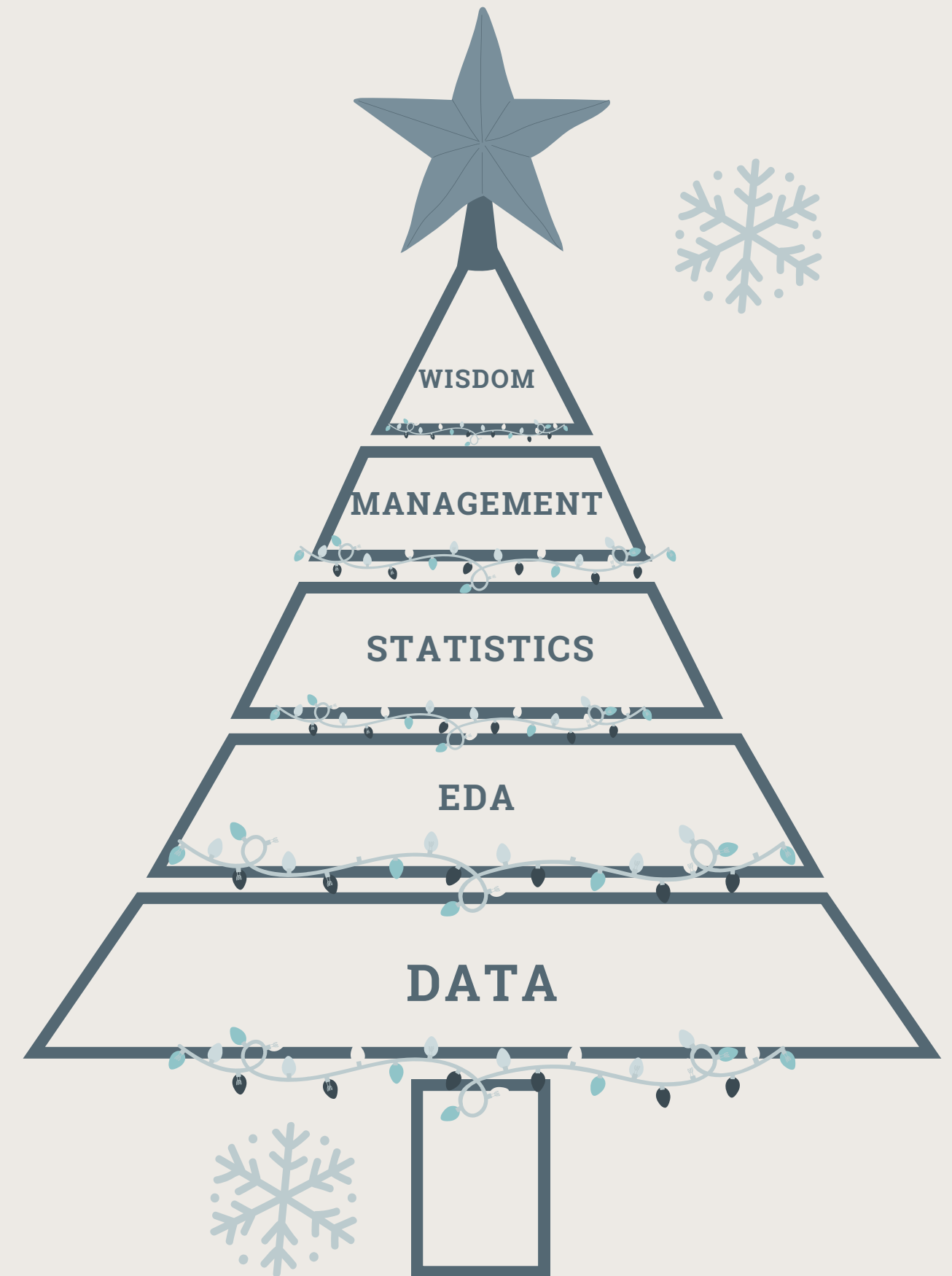
PRODUCTS SEGMENTATION

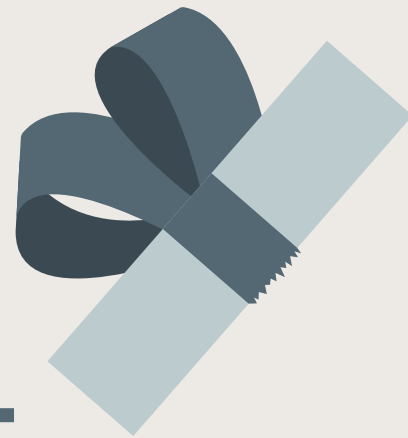
- Popularity index



ABOUT THE PROJECT

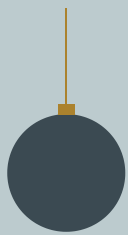
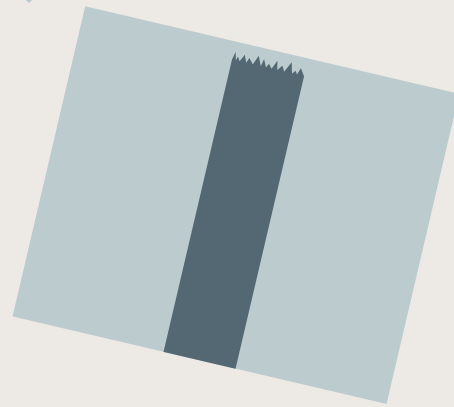
For the fall semester, our university project was carried out in partnership with the company x . The project was part of the course on Big Data and Smart Data Analytics, and it focused on using advanced data analysis, statistical and management theory and techniques to gain insights from large datasets.



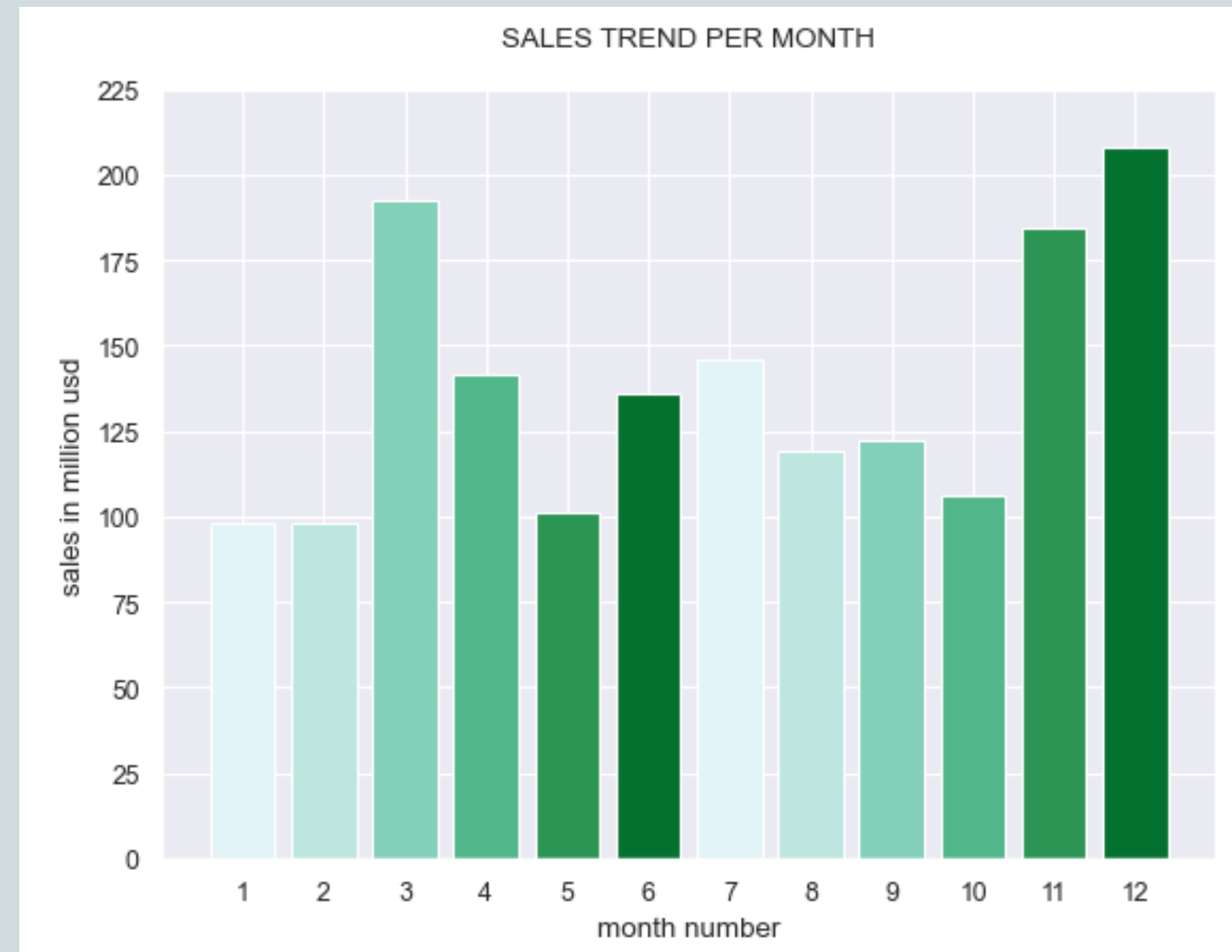


Exploration

- EDA
- Data Familiarization
- Data Cleaning
- Data Visualization

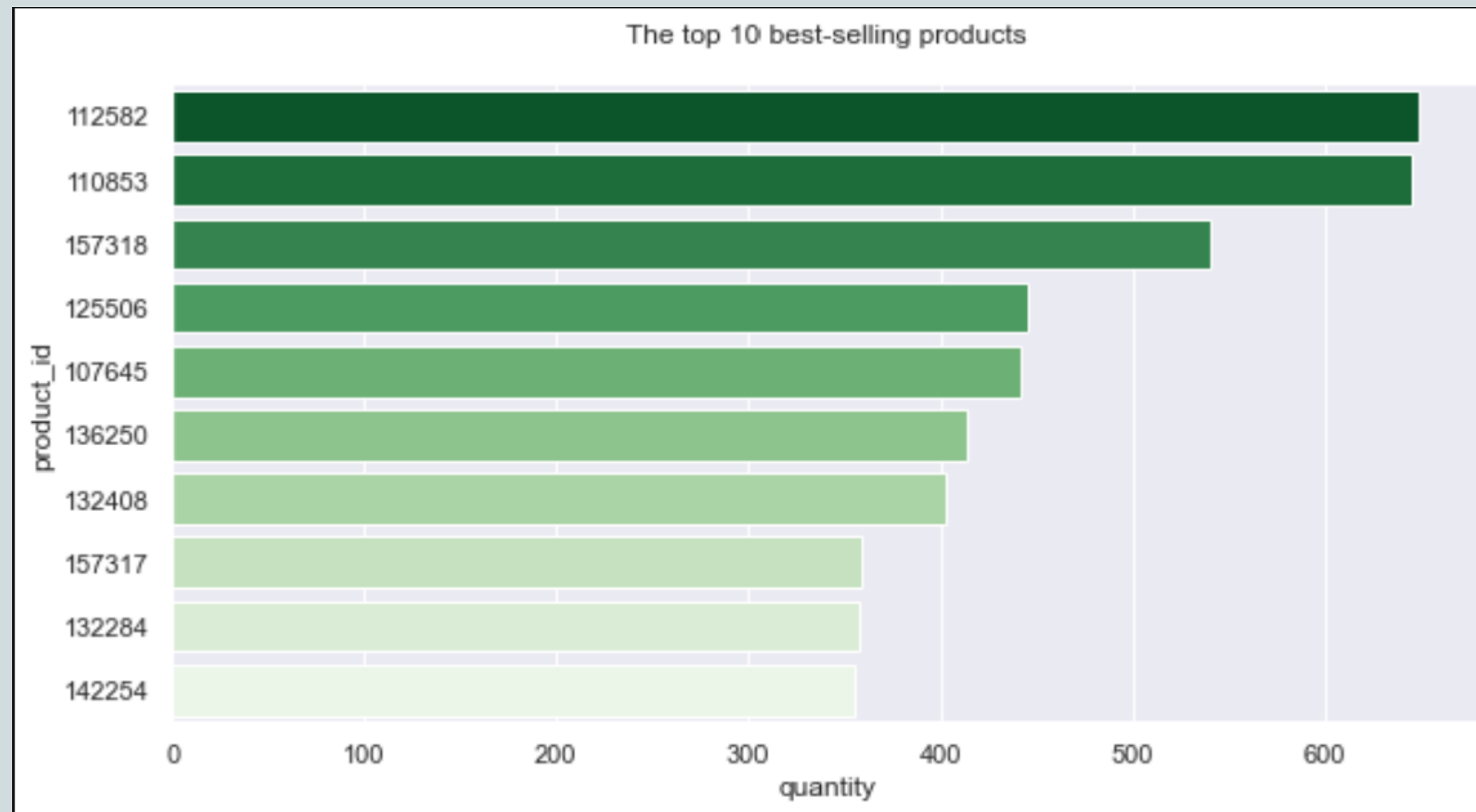


SALES TREND OF SELLER 24



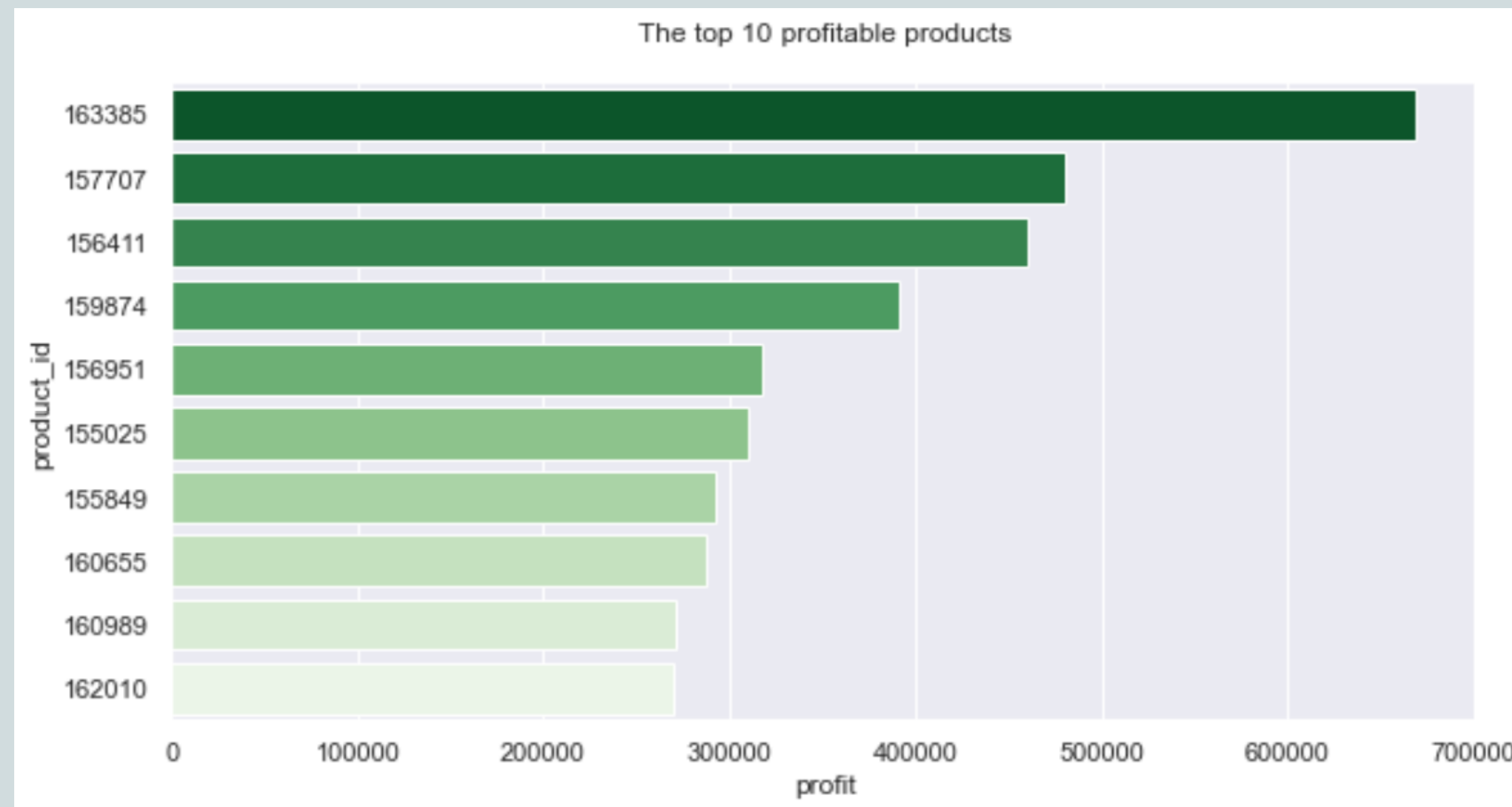
Best sales month: December, March and November.

TOP-10 BEST-SELLING PRODUCTS



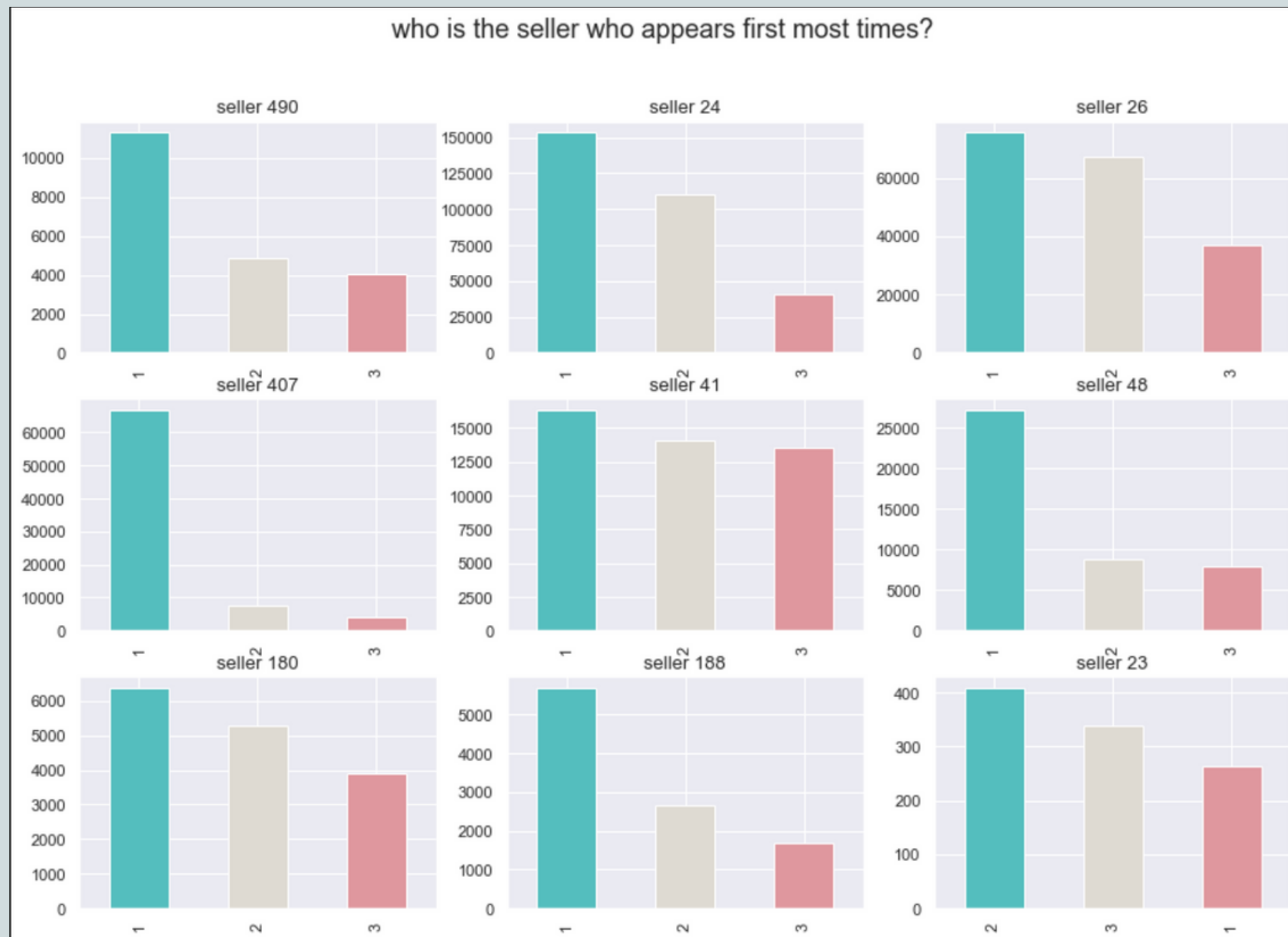
Best selling products are: 112582, 110853, 157318

TOP-10 PROFITABLE PRODUCTS



Best profitable products are: 163385, 157707, 156411

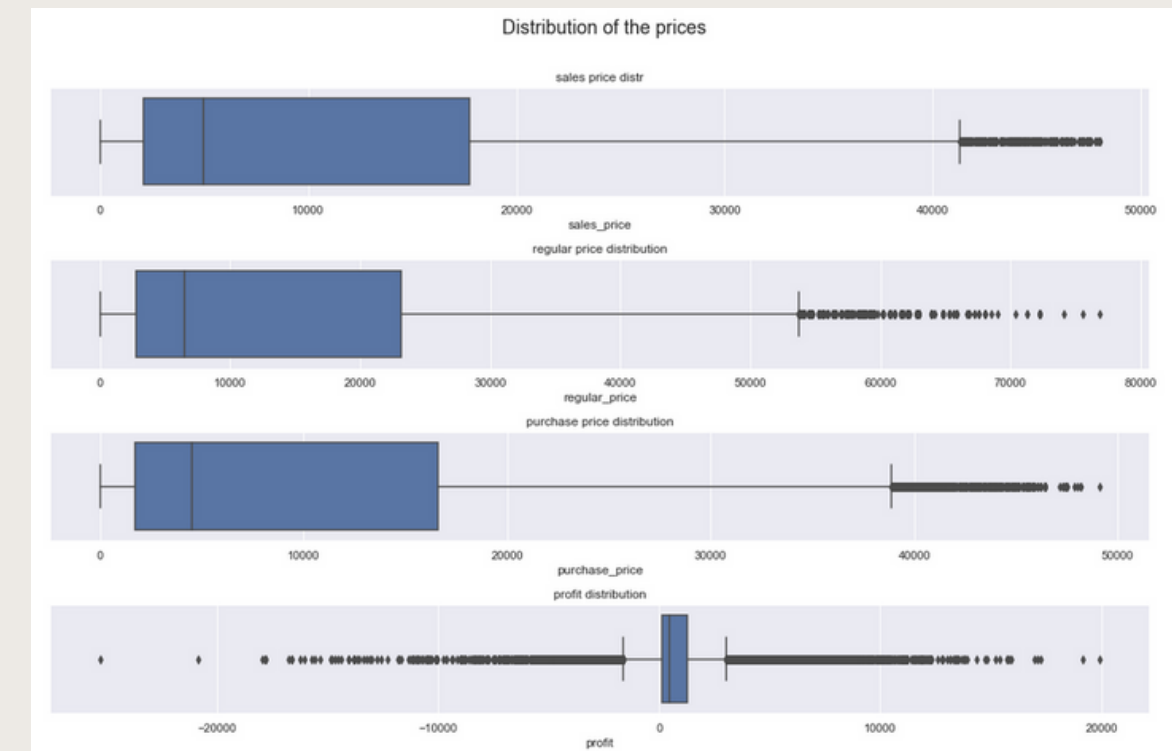
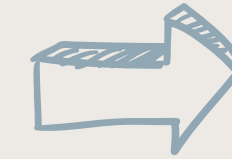
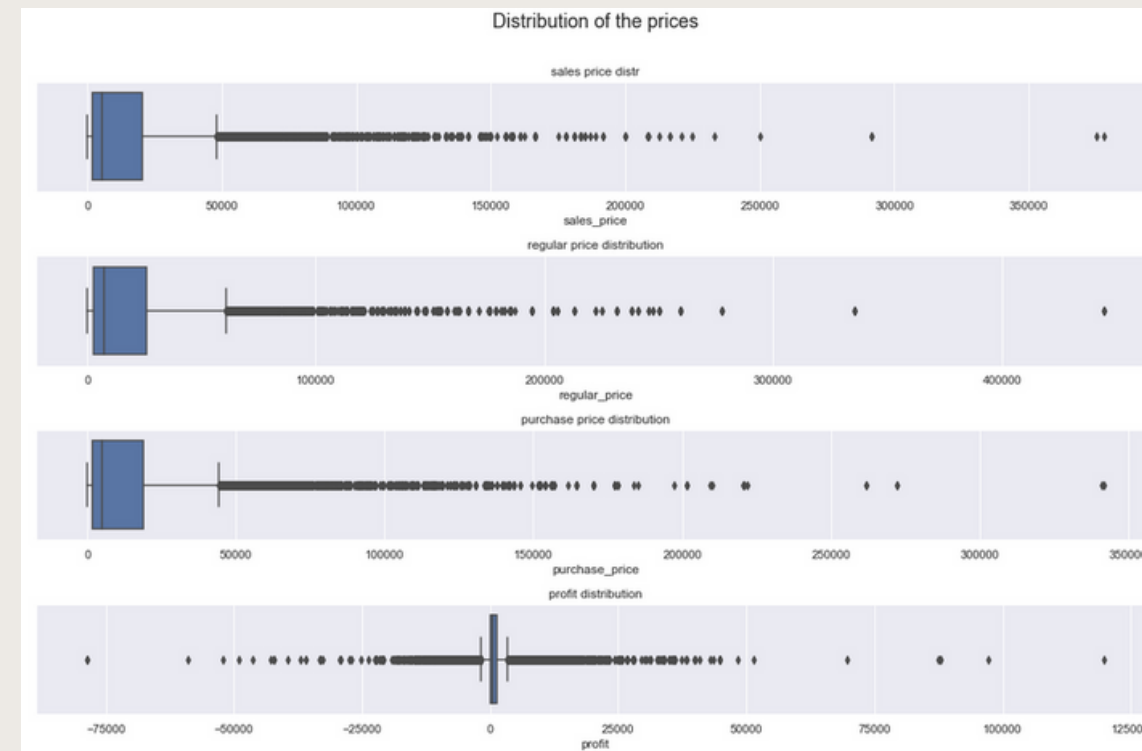
SELLERS RANK IN THE SEARCH ENGINE



Sellers that appear most frequently in first position (paying) are:

- **our client (24);**
- **407;**
- **48.**

OUTLIER DETECTION



Outlier detection is a key aspect of data exploration, and the Tukey method is one approach that can be used to identify outliers in a dataset.

1

Calculate interquartile
range IQR

2

Identify the observations that falls
outside the range as potential
outliers


3

Remove **outliers**



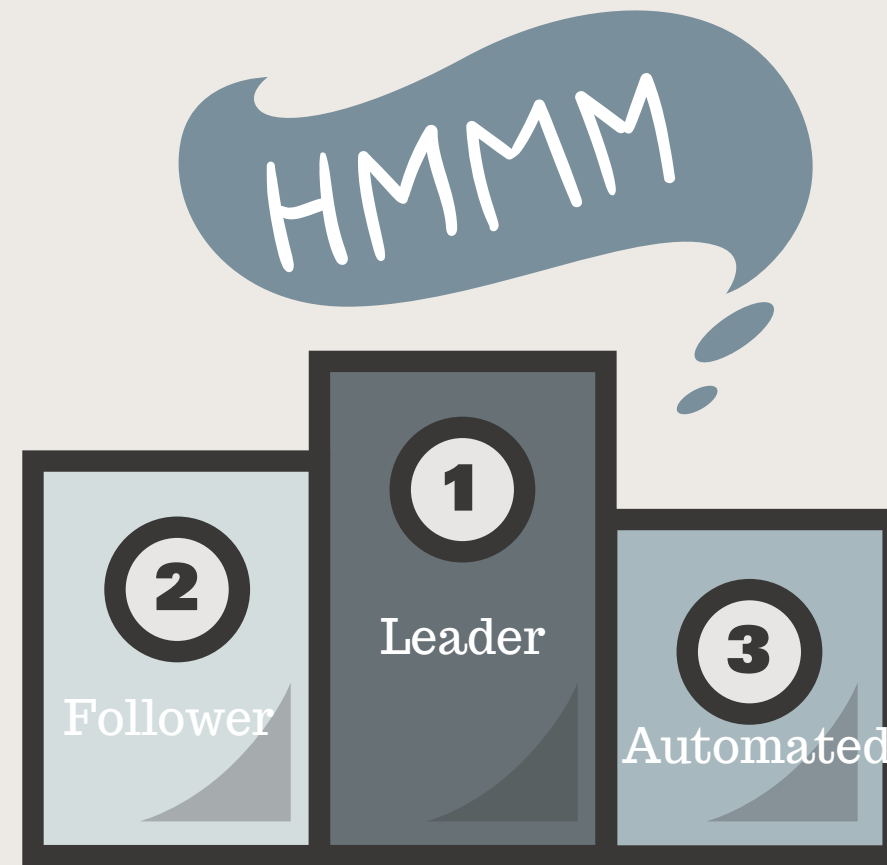
MARKET INSIGHTS

- Hidden patterns and trend in the data



Market insights are information and insights gained from analyzing data related to a particular market. This data can include things like consumer behaviour, market trends, and economic indicators. By analyzing this data, businesses can gain a better understanding of the market and make more informed decisions about their products, services, and strategies.

MARKET INSIGHTS



Price leaders are often considered to be influential in shaping the overall direction of prices in a market, while **followers** may be more reactive to changes in prices. By analyzing the behavior of these different types of players, businesses and investors can gain insights into market dynamics and make informed decisions.



Intervals

We split 2021 into **five** main intervals to conduct the analysis :

- First quarter 2021;
- Second quarter 2021;
- Third quarter 2021;
- Fourth quarter 2021;
- Black Friday : November 2021.



Select Sample to conduct analysis

Moreover, we perform an analysis in order to identify **top 10 product sold**. The aim of this, is to identify leader and followers for this specific products in each time interval.

Who is the price leader?

OUR STRATEGY



Our approach to identify the leaders and followers:

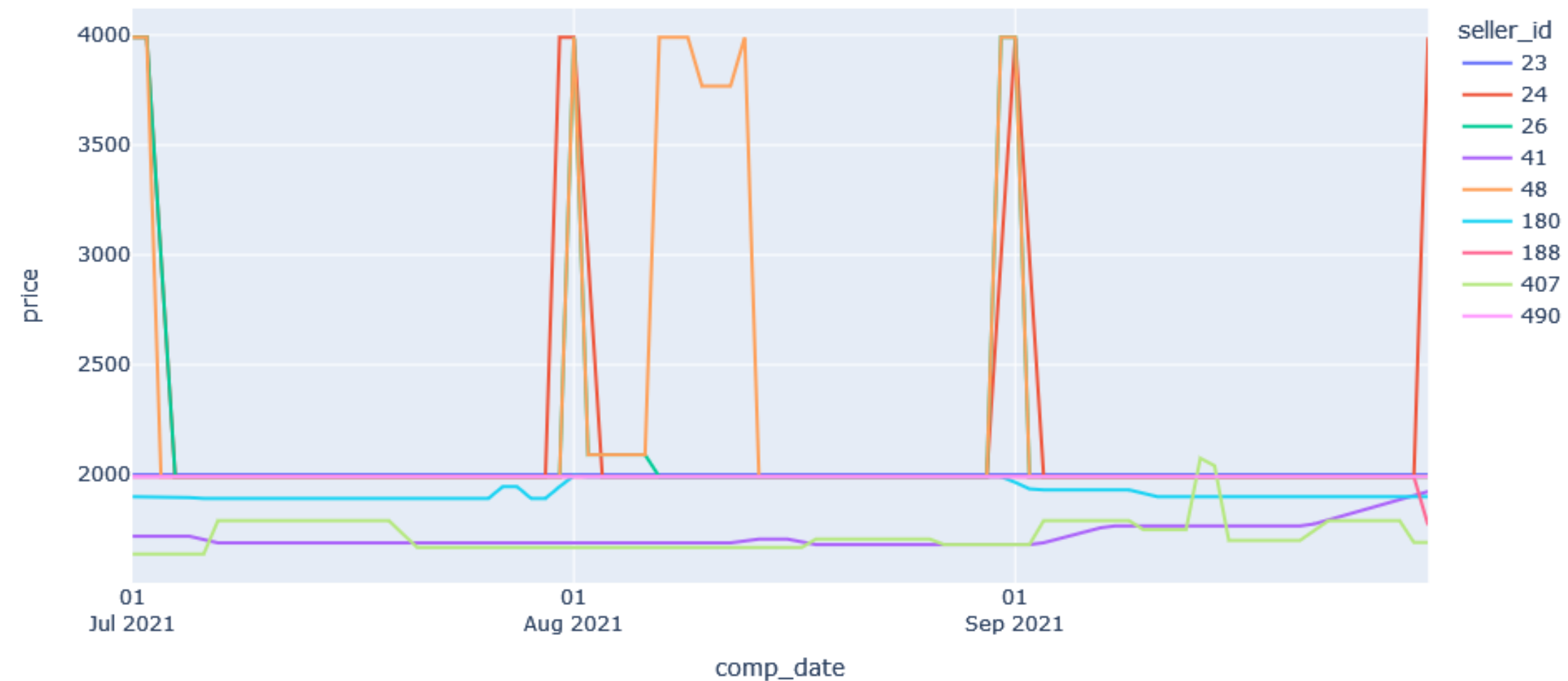
- By analyzing their trend in different time windows.
- Numerical support: such as **correlation** and **cointegration**.



GRANGER CAUSALITY TEST

To test our hypotheses on leaders and followers (developed through observation of price trend graphs), we used Granger's econometric test.

Time series for product 107645 in quarter [3]



EXAMPLE

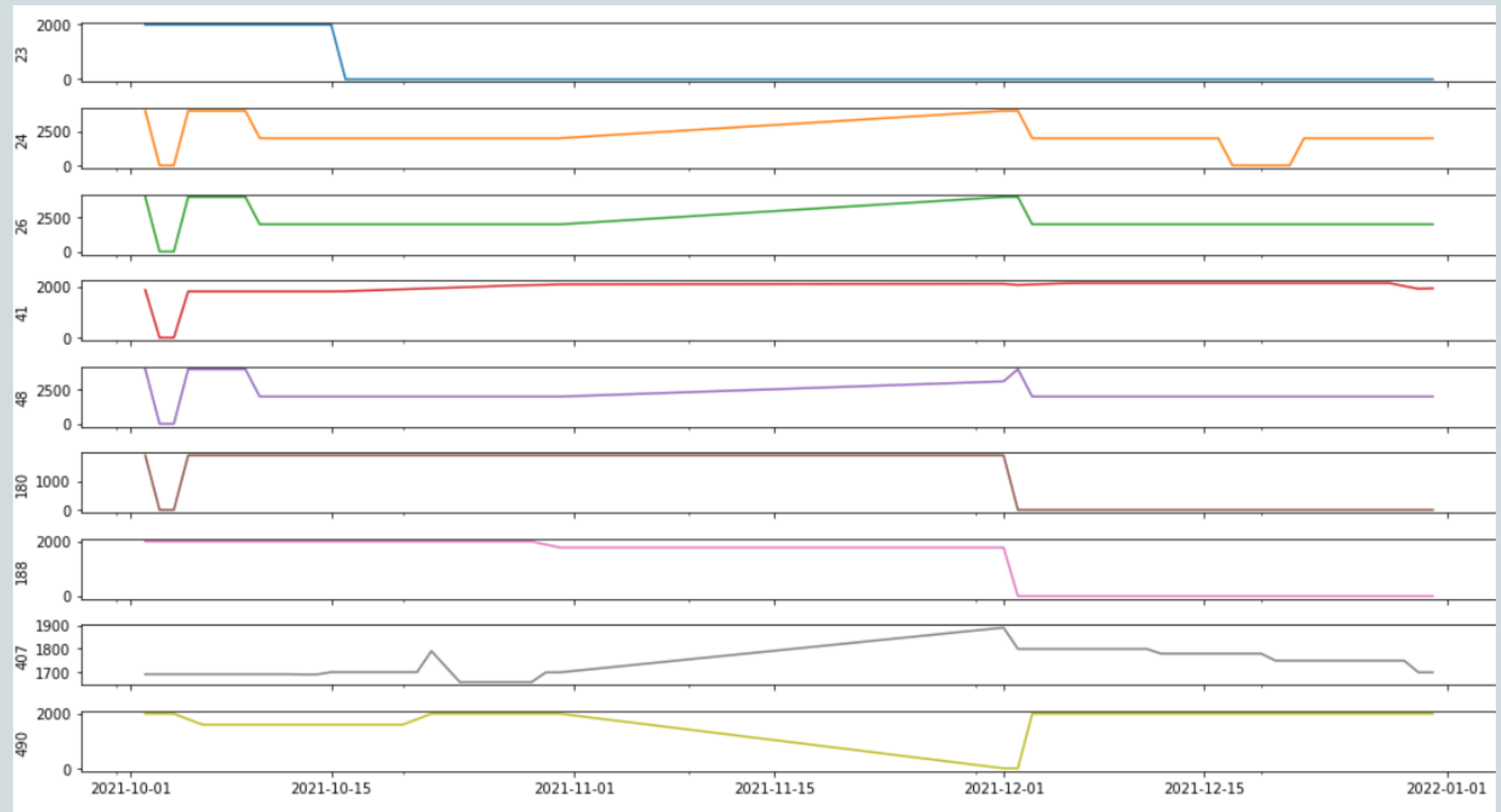
For example, we first formulated a hypothesis (H_0) from observing the graph like this one above on trends with various sellers in order to see who might be leaders, followers, and "automated followers." After that, in pairs, we went to test them, in order to confirm or reject our hypothesis, via the p-value produced by the Granger test, which indicates how likely is that the null hypothesis is true.

Who is the price leader?

WHO USES AUTOMATED PRICING SYSTEM ?



1. Analyzed correlation between sellers to determine potential use of automated pricing systems.
2. Graphically represented sellers with similar price variations through network.
3. Analyzed price trend for individual sellers to identify potential leaders and followers.
4. Used Granger test to confirm or reject hypotheses about use of automated pricing systems.
5. Results vary depending on product and time period analyzed.



Who is the price leader?

PRODUCT 107645




Results for most sold product 107645:

- 1

Roles can change depending on the product
- 2

Roles can change depending on the time window
- 3

Sellers that keep price fixed or do not follow market trends

	Leader 	Follower 	Automation 
Q1	48	26/24	24
Q2	26	48/24	48/24
Q3	24	26	48
Q4	24	26/48/41	26/48
Black Friday November	48	24/26	26

Who is the price leader?

RESULT OF THE ANALYSIS

These results only refer to 10 best-selling products.

24

26, 48

26, 48, 41

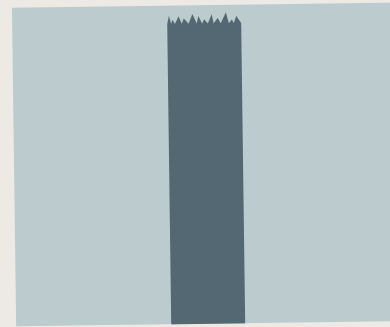
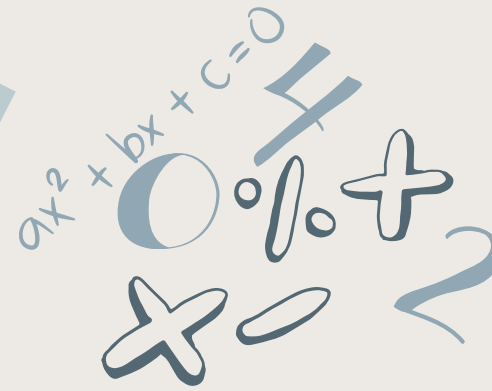
Leaders

Market followers

**Competitors that applies
Aut. pricing system**

Also, sellers 26 and 48 are market leader, in certain time windows

In some cases, also our client (24) became a follower



PRODUCTS SEGMENTATION

- Popularity index



Product segmentation involves using data to identify and understand the different segments of a product or product line.

Furthermore, by segmenting your product catalogue according to variables such as profit, quantity sold and stock, you can develop a popularity index for each product.



ABC ANALYSIS

ABC analysis is a method used in inventory management to categorize inventory based on its importance. It's based on the Pareto principle—the idea that 20% of your top-performing products account for 80% of your store's total revenue.



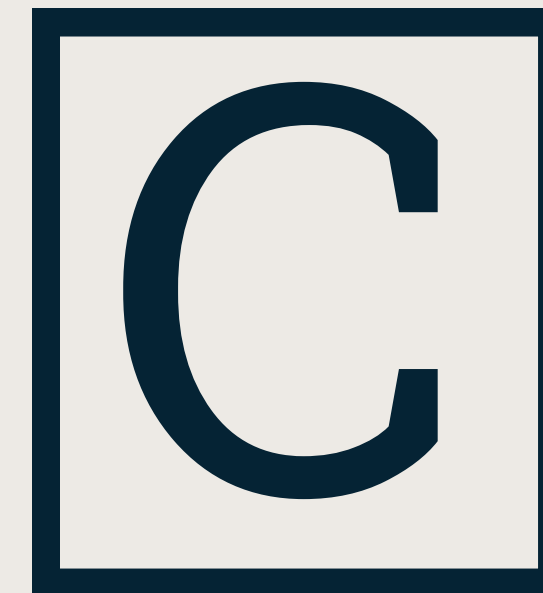
Most important

All products where the cumulative percentage of contribution is up to 80%



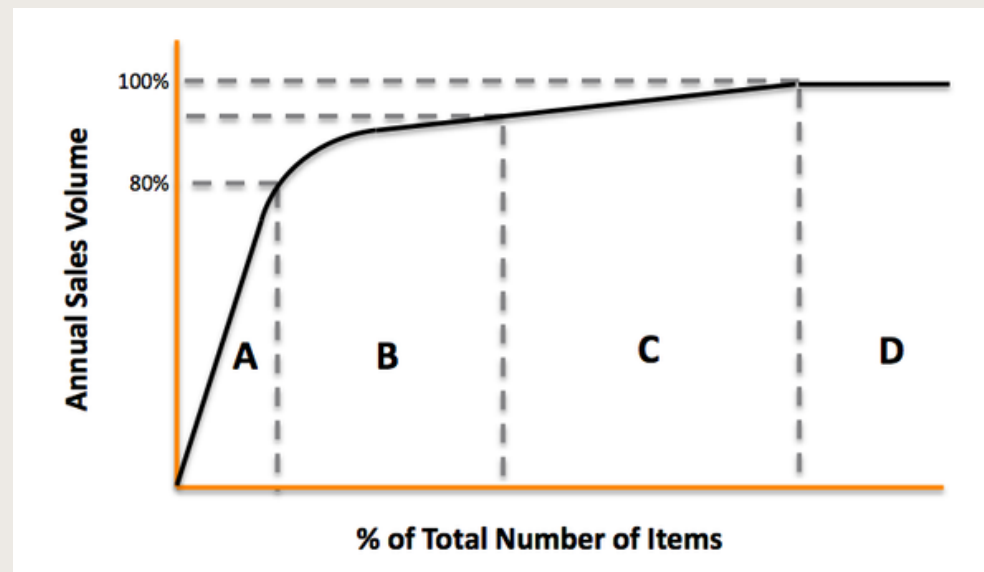
Moderately important.

Those where the cumulative percentage contribution is between 80% and 90%



Least important.

All the remaining products that have small contribution. They require attention.



OUR APPROACH

ABC

AAA,BCA ETC.



SEGMENTATION

Metrics used to segment:

revenue: quantity * sales price

profit: sales prices - Purchase price

quantity: counting of the products sold

CLASSIFICATION AND ANALYSIS

Attribute a Class to each product for their profit, revenue and quantity performance.

Combine those class

Class_Profit+
Class_Rev+
Class_Quant.

Popularity index computation

1 Step: Preparation

Profit,Revenue,Quantity = Grouped by prod_id and Sum

2. Step: Math transformation

$$\text{Profit} = (\text{Profit}^{-1}) * 100$$

$$\text{Revenue} = (\text{Revenue}^{-1}) * 100$$

$$\text{Quantity} = (\text{Quantity}^{-1}) * 100$$

3. step: Calculation

$$\text{Popularity index}_i =$$

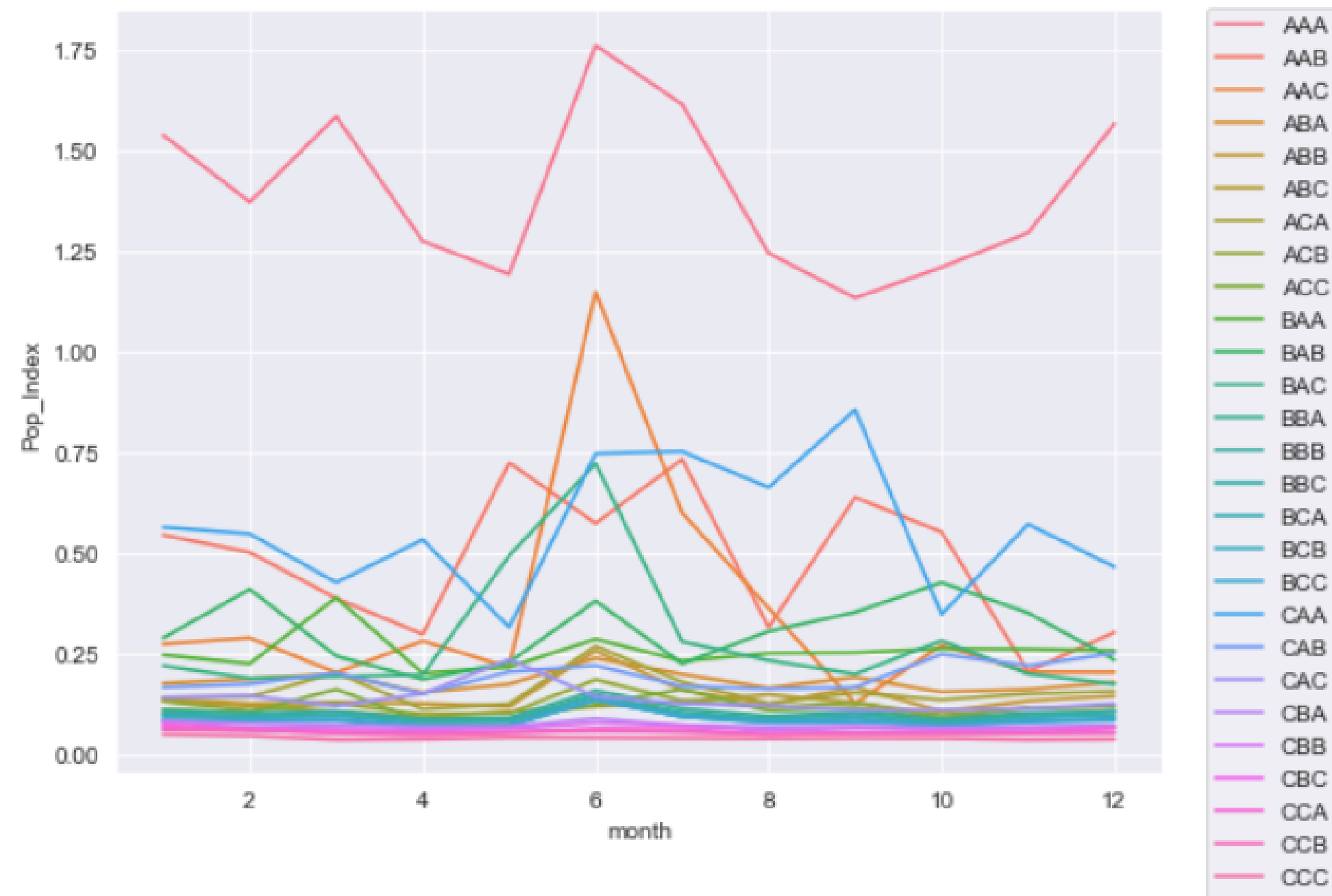
$$0.35 * \text{Profit normlized score}$$

$$+0.05 * \text{Revenue norm. score}$$

$$+0.60 * \text{Quantity norm. score}$$

Popularity Index

PRODUCT CLASS AND POPULARITY INDEX



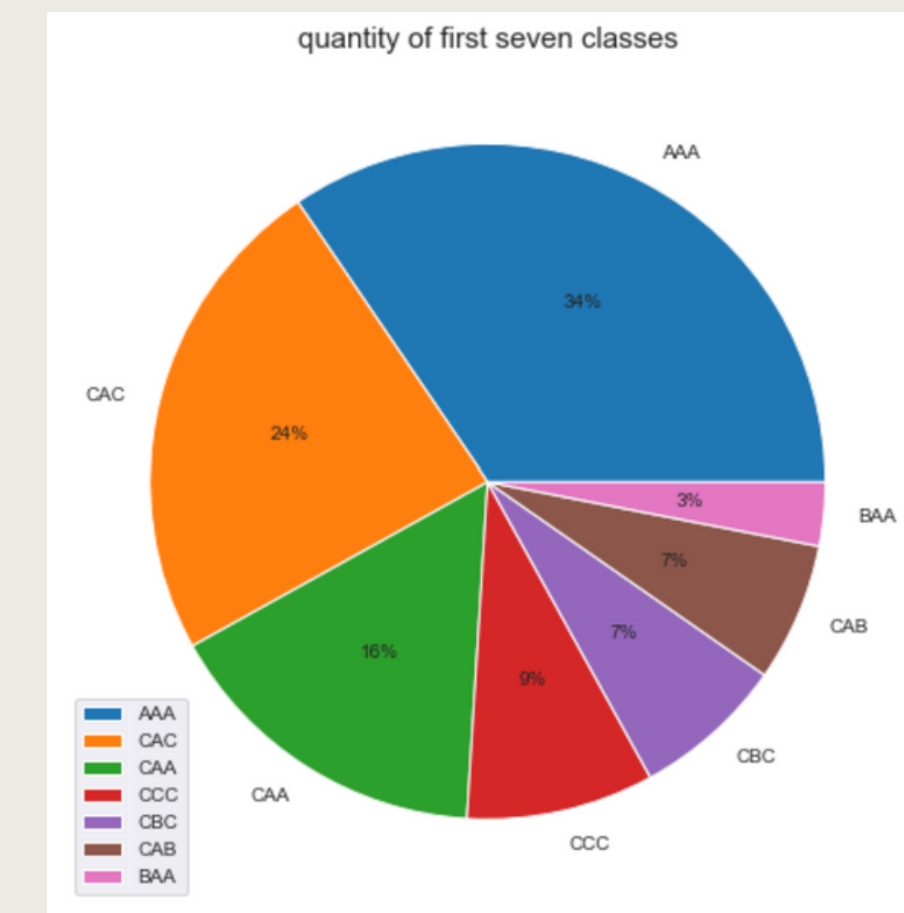
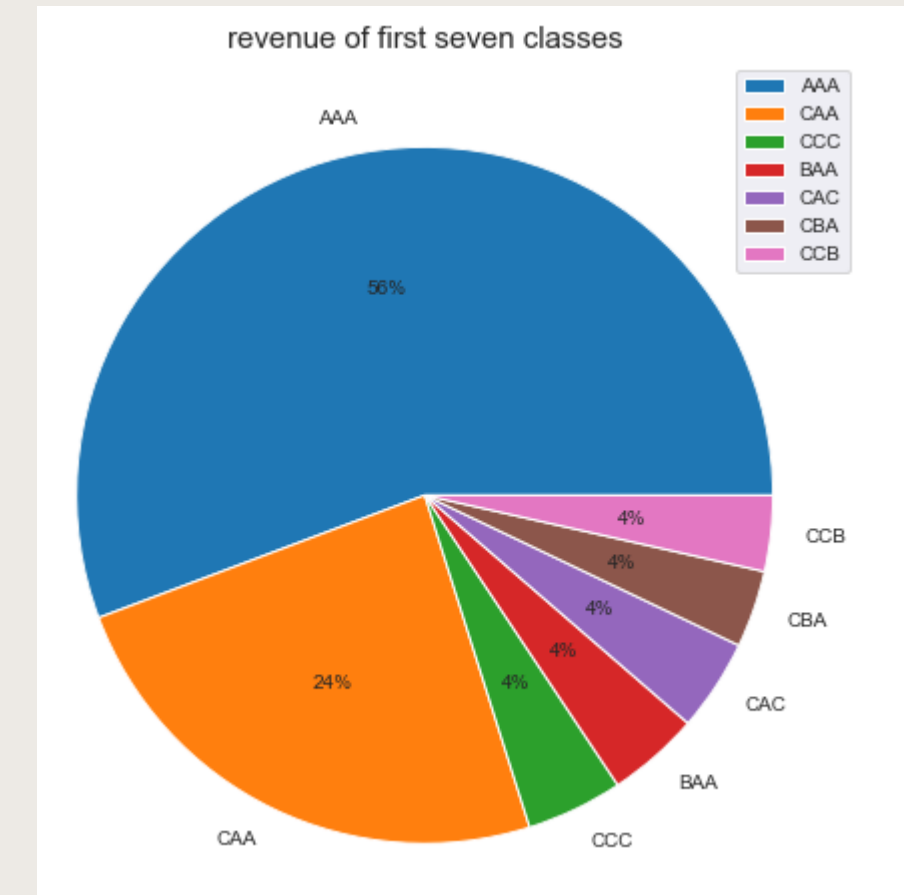
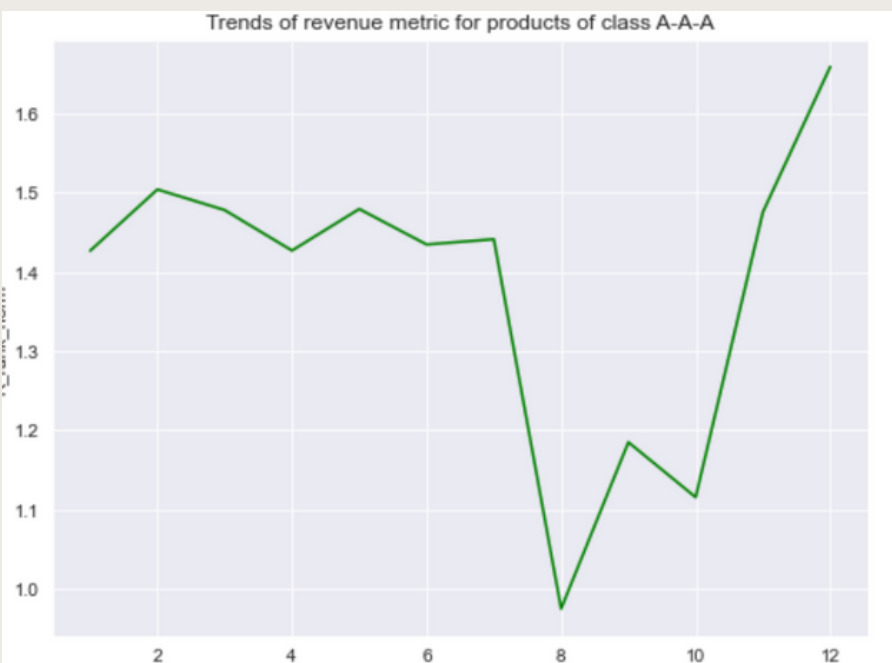
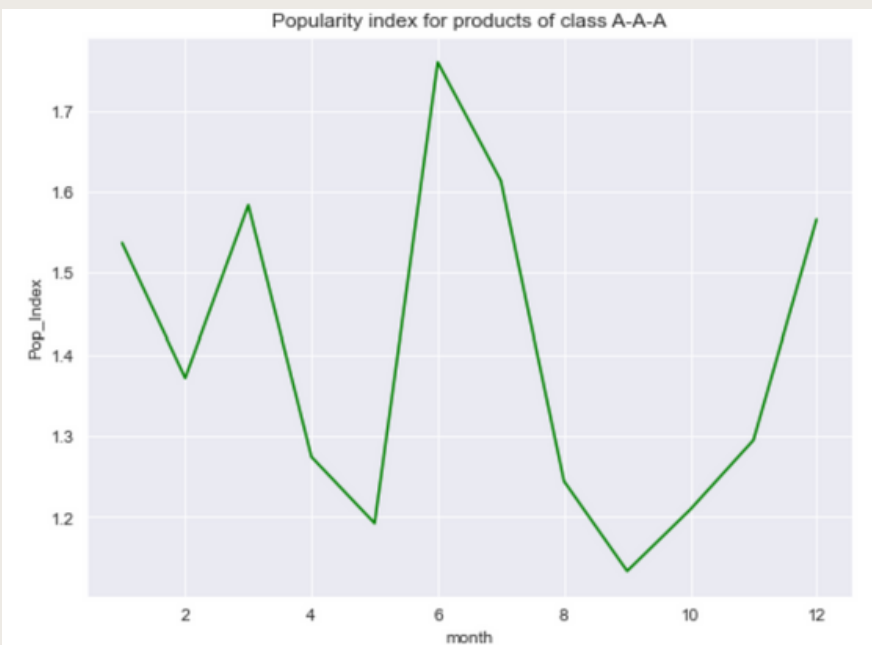
A graph depicting all the combination of Product's Class and their performance in ABC score

Popularity Index

ANALYSIS

Line plots represent the change in the popularity index over various time intervals. Each of the three line-plots represents a different metric: revenue, profit and demand

The Pie Chart describes the Revenue or the quantity contribution of each Class to the seller total sales.



Popularity Index

RESULT OF THE ANALYSIS

	CLASS	month	revenue	class_rank_revenue	quantity	class_rank_quantity	P_rank_norm	R_rank_norm	Q_rank_norm	Pop_Index
product_id										
110853	AAA	1	482125.0	16	71238.5000	1	25.000000	6.250000	100.000000	69.062500
156786	AAA	1	1598586.0	1	389481.8431	4	100.000000	100.000000	25.000000	55.000000
155431	AAA	1	167172.0	107	29267.9132	2	2.500000	0.934579	50.000000	30.921729
154218	AAA	1	982334.0	3	78260.3095	8	33.333333	33.333333	12.500000	20.833333
157315	AAA	1	137471.0	149	20947.2800	3	1.351351	0.671141	33.333333	20.506530
...
144991	CCC	1	408.0	2547	48.5976	2505	0.039216	0.039262	0.039920	0.039641
130769	CCC	1	575.0	2515	116.0000	2589	0.041203	0.039761	0.038625	0.039584
100043	CCC	1	382.0	2554	112.0000	2598	0.040984	0.039154	0.038491	0.039397
147823	CCC	1	231.0	2588	85.0500	2576	0.040241	0.038640	0.038820	0.039308
131898	CCC	1	660.0	2486	3.0000	2577	0.038506	0.040225	0.038805	0.038771
2285	CCC	1	10.0	2599	1.0000	2599	0.038506	0.040225	0.038805	0.038771



Popularity Index

RESULT OF THE ANALYSIS

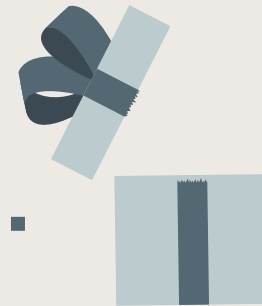
product_id	CLASS	month	revenue	class_rank_revenue	quantity	class_rank_quantity	P_rank_norm	R_rank_norm	Q_rank_norm	Pop_Index
110853	AAA	1	482125.0	16	71238.500	1	25.000000	6.250000	100.000000	69.062500
110853	AAA	2	402171.0	13	39457.400	1	16.666667	7.692308	100.000000	66.217949
110853	AAA	3	695583.5	13	46666.200	1	7.142857	7.692308	100.000000	62.884615
110853	AAA	4	229425.0	89	32390.800	7	2.439024	1.123596	14.285714	9.481267
110853	AAA	5	216125.0	53	31699.980	2	4.000000	1.886792	50.000000	31.494340
110853	CAC	6	12931.0	1601	776.404	847	0.051813	0.062461	0.118064	0.092096
110853	CBC	8	6096.0	2088	428.600	1906	0.040833	0.047893	0.052466	0.048166
110853	CAC	9	5975.0	2133	307.600	1361	0.037092	0.046882	0.073475	0.059412
110853	CCC	12	6650.0	2354	982.600	2262	0.045893	0.042481	0.044209	0.044712



KEY TAKEAWAYS



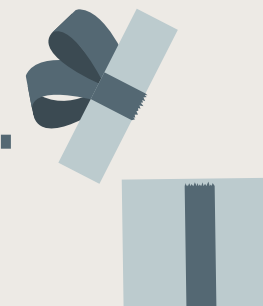
Most sold products generates money, but they are not necessarily the most profitable!



Leader and follower are not always the same, their roles may change depending on products and time windows.



Data Science is not just Python! It involves statistics, CS skills, Economics and Management knowledge! So, multidisciplinary team is crucial!



Having hard skills is not the key to success! but combining it with soft skills, imagination and creativity is!



THANKS FOR YOUR ATTENTION!

MERRY
XMAS



Lorenzo Antolini



[linkedin.com/in/lorenzo-antolini-9a4b90160/](https://www.linkedin.com/in/lorenzo-antolini-9a4b90160/)



Enrico Grandi



[linkedin.com/in/enrico-grandi/](https://www.linkedin.com/in/enrico-grandi/)



Simone Lu



[linkedin.com/in/simone-lu/](https://www.linkedin.com/in/simone-lu/)



Enrico Romano



[linkedin.com/in/enrico-romano-912171232/](https://www.linkedin.com/in/enrico-romano-912171232/)