

Exploring E-Commerce Excellence

E-Commerce: Product Range Analysis

1

Viktor Akulinin, student
victor.akulinin@gmail.com
APRIL, 2024

TripleTen



INTRO

I work as a video ad analyst at the Sterling & Draper advertising agency. Every week, new employees ask me the same questions:

- What video categories were trending last week?
- How were they distributed among various regions?
- What categories were especially popular in the United States?

I decided to create a dashboard to automate the search for answers to these questions.

TripleTen



A FEW FACTS ABOUT DATA

- Our dataset has 538,151 observations, including information on shipments, returns, and cancellations.
- 3,810 unique product codes are used, while only 3,792 unique product names are used. This is due not only to different packaging options and forms of packaging but also to other product characteristics.
- The data covers the period from 29-11-2018 to 07-12-2019, which amounts to 1 year and 8 days.

3



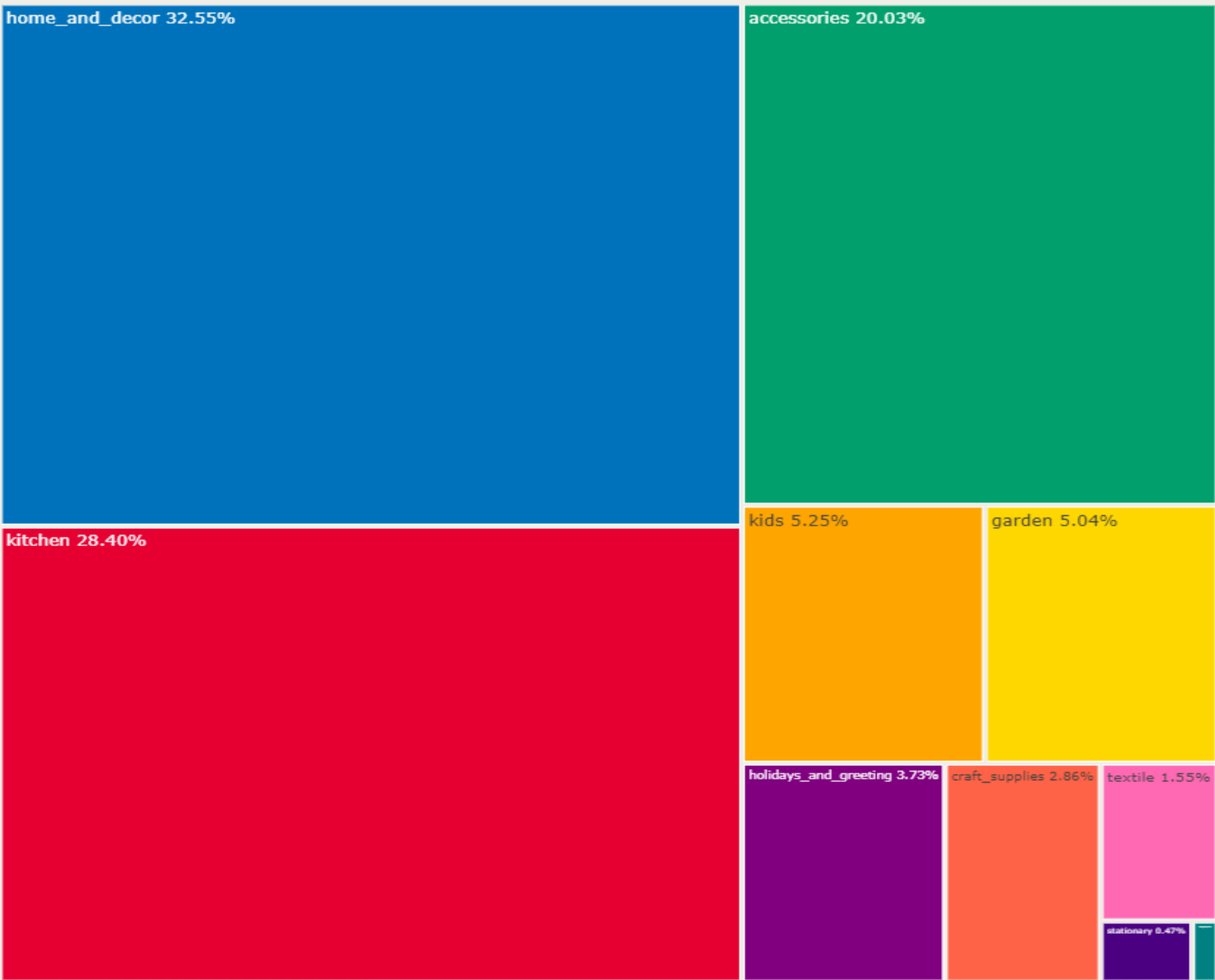
A FEW FACTS ABOUT DATA (CONTINIUS)

- During this period, at least 4,364 customers participated in at least one transaction. Of these, 4,363 customers are personalized, and one is a generalized customer, which includes data from all unregistered users or those who did not provide personal information.
- For analysis, the data was organized into 10 product categories, 4 price segments, and 7 segments⁴ based on the results of RFM analysis, which evaluates products based on criteria such as "last purchase," "purchase frequency," and "monetary value."



DISTRIBUTION ITEMS BY CATEGORY

Distribution of items by category



We have a group of leading categories contributing approximately 80% of the entire assortment: **home and decor**, **kitchen**, and **accessories**.

The second group, comprising approximately 80% of the remaining assortment, includes **kids**, **garden**, **holidays and greetings**, and **craft supplies**.

The remaining three categories are: **textile**, **stationary**, and **sports activity**) collectively account for just over 2% of the total assortment and are unlikely to yield substantial insights. While merging them with larger categories could be considered, for now, we'll maintain them as distinct categories.

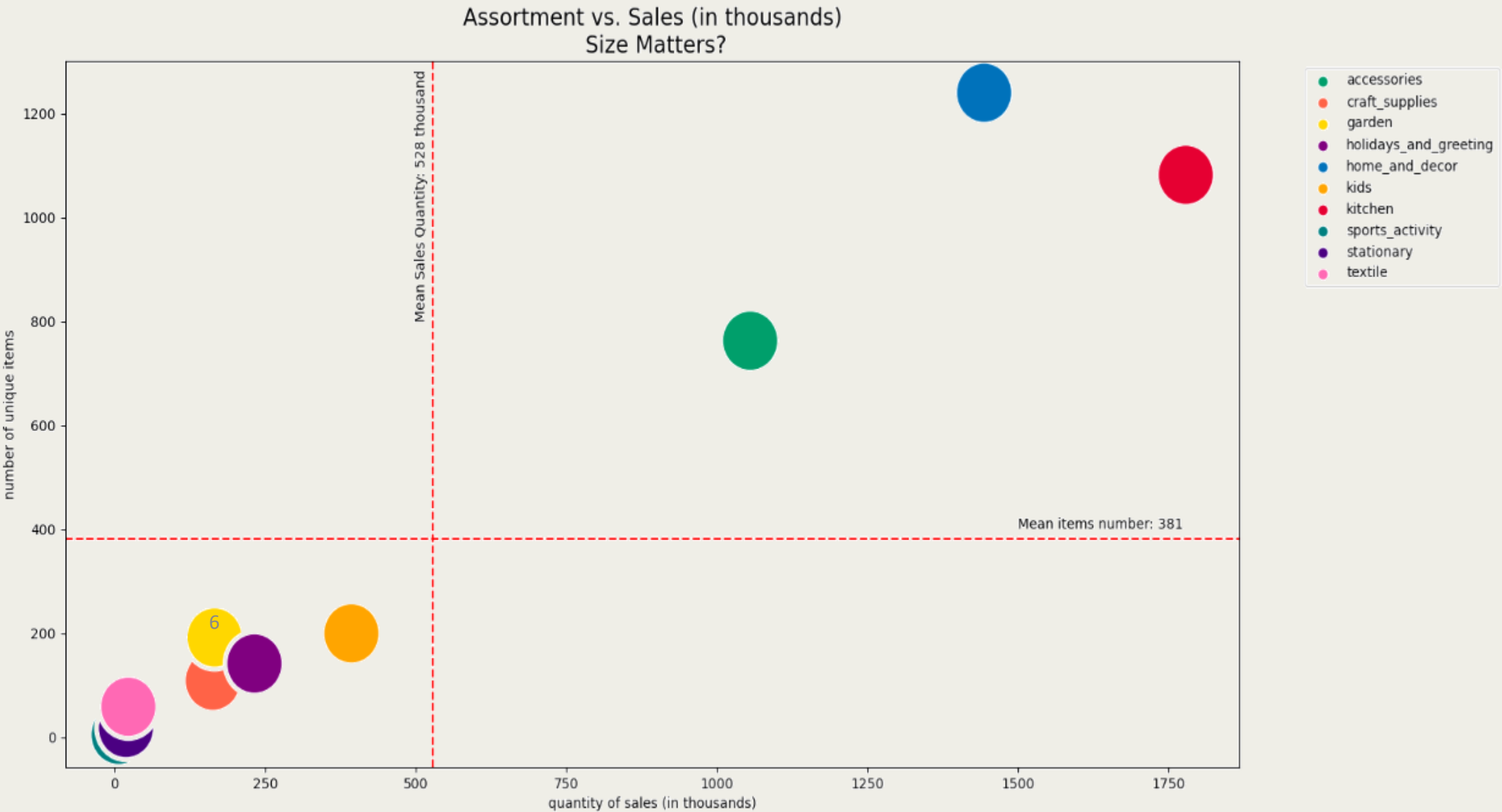
CATEGORY SIZE vs SALES

The graph illustrates a direct correlation between assortment and sales, a relationship that was anticipated from the outset.

Unlike the assortment breakdown, we observe only two distinct groups: leaders, consisting of **kitchen**, **home and decor**, and **accessories**, and all other categories, which displayed results below the overall average.

An important observation is the absence of clear "outsiders" with a vast assortment and low sales, as well as any "unicorns" characterized by high sales and modest assortment.

However, there is a noticeable asymmetry in sales. For instance, pairs such as 'kitchen and home and decor', 'kids and garden', or 'holidays and greeting and garden' demonstrate that the category with a smaller assortment tends to have higher sales.



THE SALES TO ITEMS RATIO

The absolute leader is the **kids** category, with a ratio of 1.97 (thousands of units). On the other hand, the absolute loser is the **textile** category, with a score of only 0.38 (thousands of units). Since we lack deep historical data or data on inventory levels,

The average ratio value – 1.29 (thousands of units) for evaluation.

Based on this, I see that for the "leader" group, we should take a closer look at the assortment of the home_and_decor category.

For the "other" group, attention should be paid to the assortment in the **stationary**, **garden**, and **textile** categories. It worthy to emphasize that these are not definitive conclusions about the state of the assortment, but rather indications of the need to pay attention to it.

category	sale_item	group
kids	1.966910	other
holidays_and_greeting	1.635739	other
craft_supplies	1.496560	other
sports_activity	1.313000	other
stationary	1.046167	other
garden	0.864547	other
textile	0.384322	other
kitchen	1.644429	leader
accessories	1.383717	leader
home_and_decor	1.164785	leader

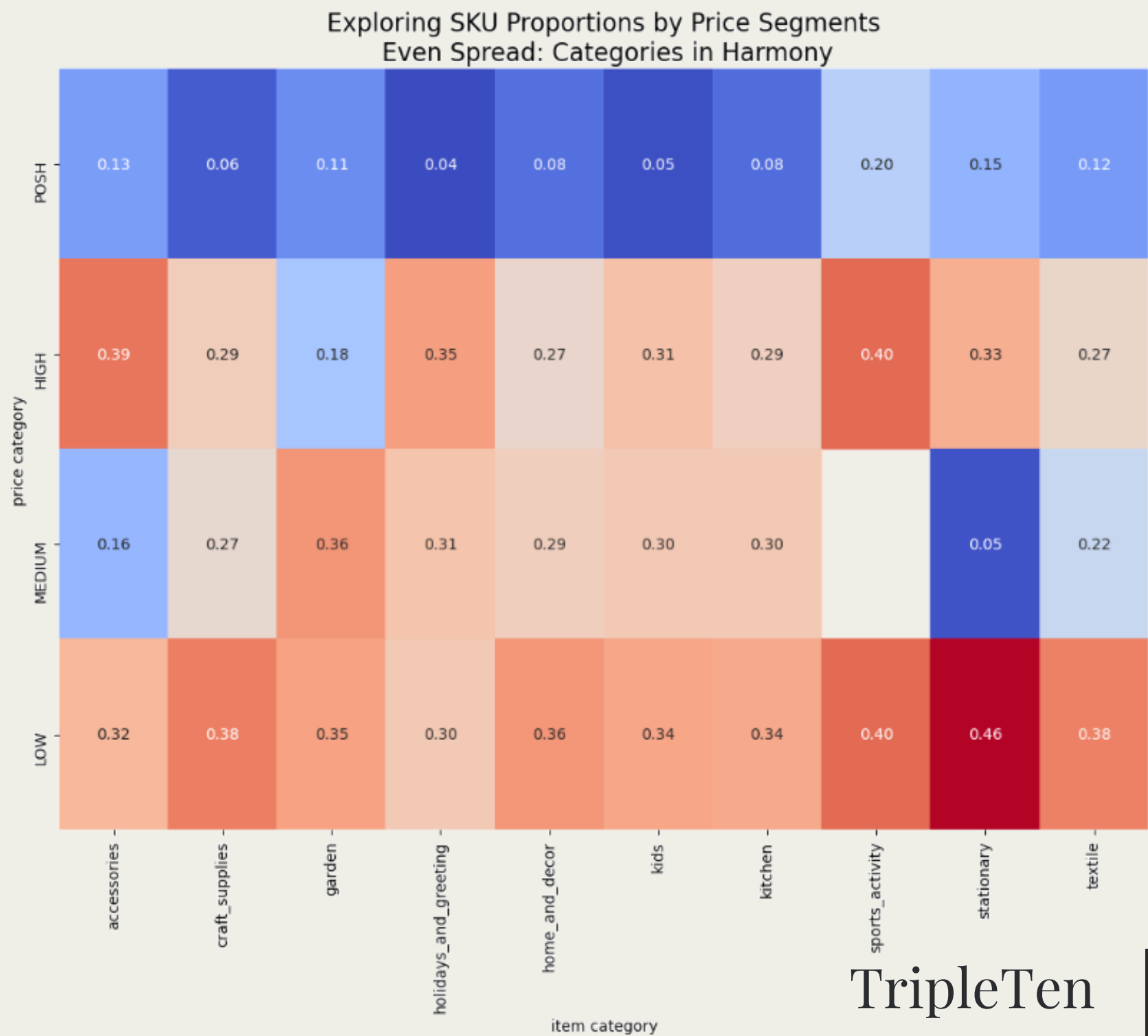
SKU PROPORTION BY PRICE SEGMENT

Overall, we have balanced price categories, with an approximate distribution of 30, 30, 30, and 10 across the categories 'LOW', 'MEDIUM', 'HIGH', and 'POSH'.

Of course, there are exceptions:

- In the **sports activity** category, there are no items with a price in the 'MEDIUM' category (there are only 5 items in this category, so it's not surprising).
- In the **stationary** category, there are very few items with prices in the 'MEDIUM' category, with a bias towards the 'LOW' category
- In the **textile** and **accessories** categories, there are also a few items in the 'MEDIUM' category.

8



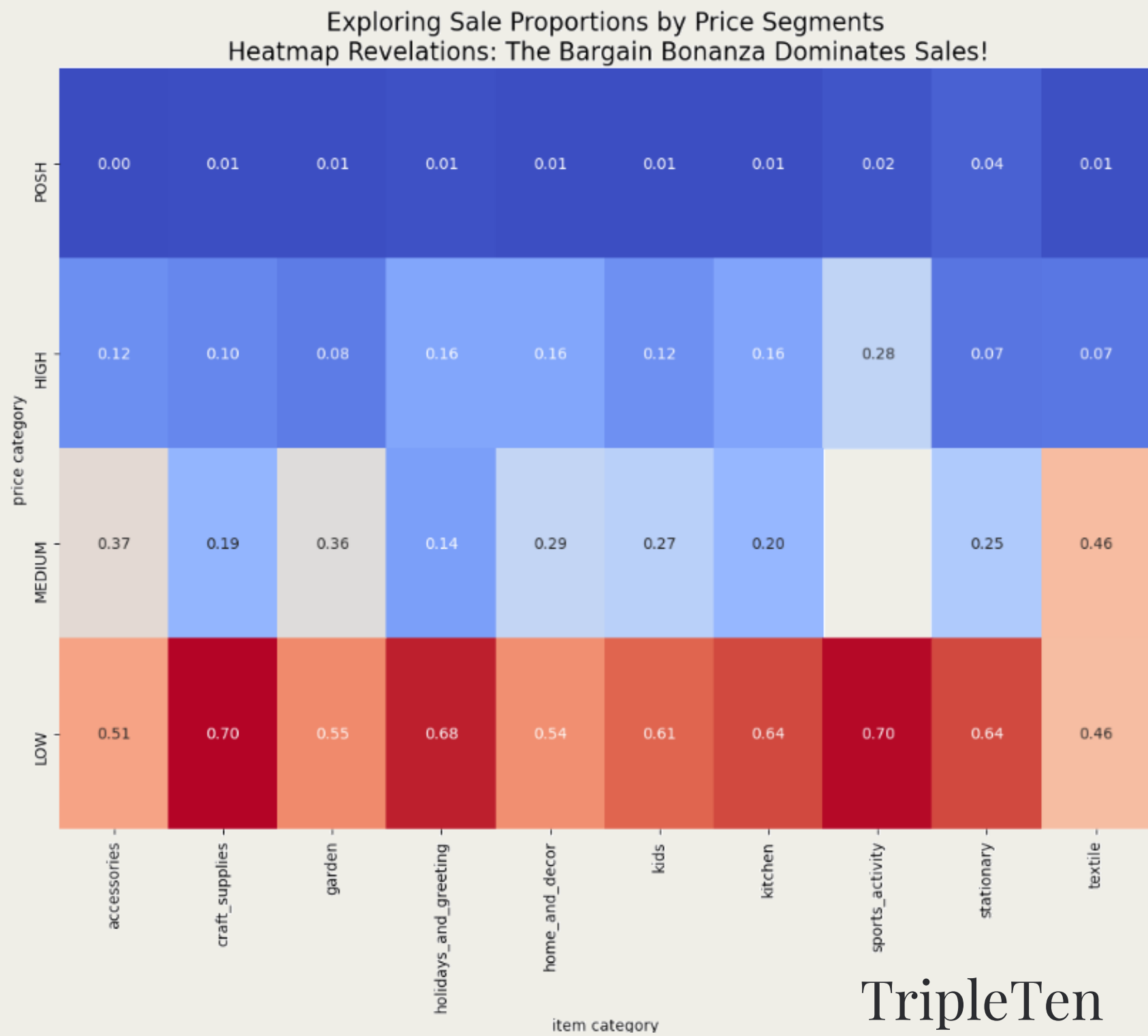
SALE PROPORTION BY PRICE SEGMENTS

The graph convincingly demonstrates the advantage of sales in the low-price segment across nearly all product categories, except for the **textile** segment.

Here, sales are evenly split - 46% each between the 'LOW' and 'MEDIUM' price segments.

However, overall, the share of the low-price segment reaches up to 70% in some categories.

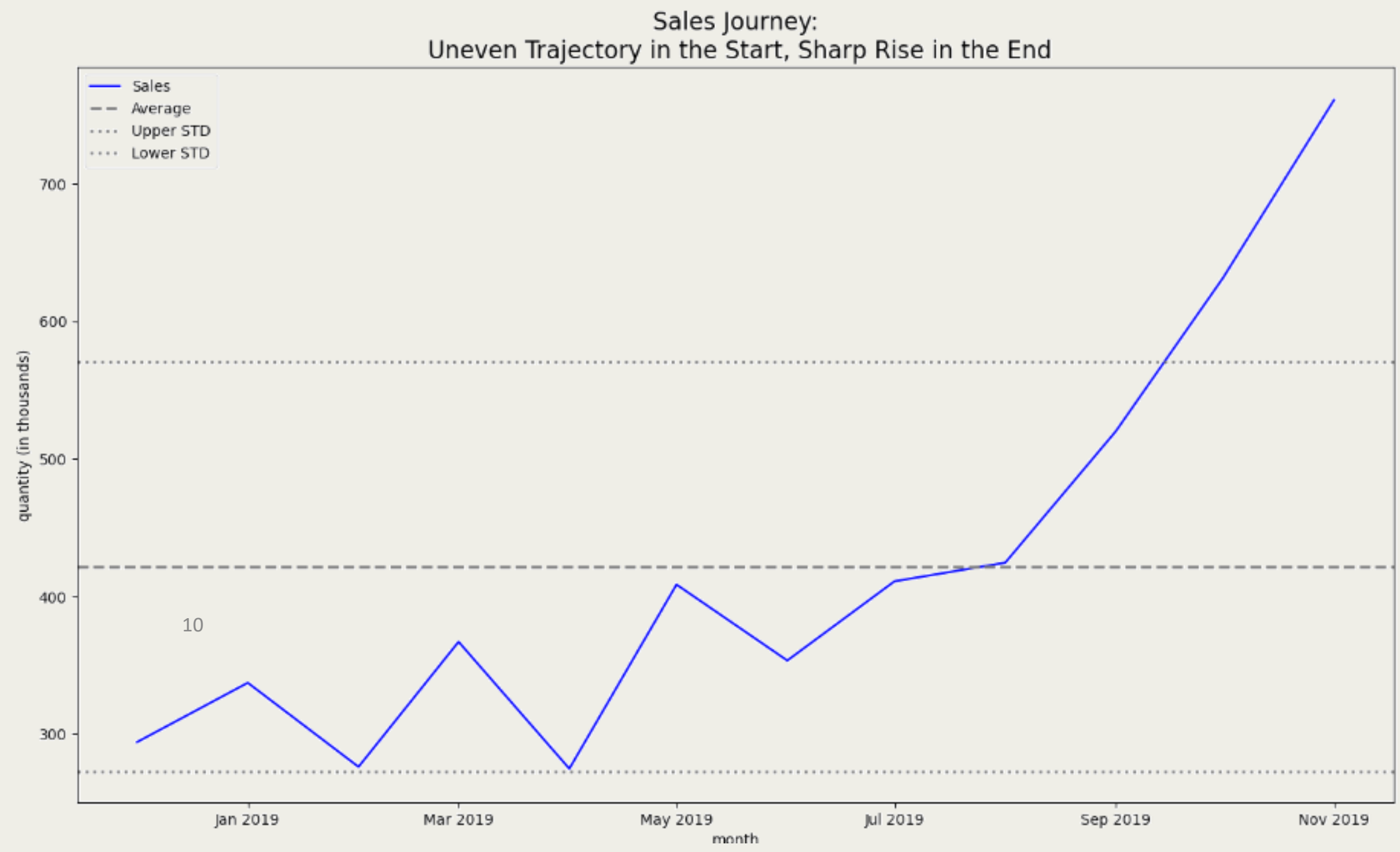
9



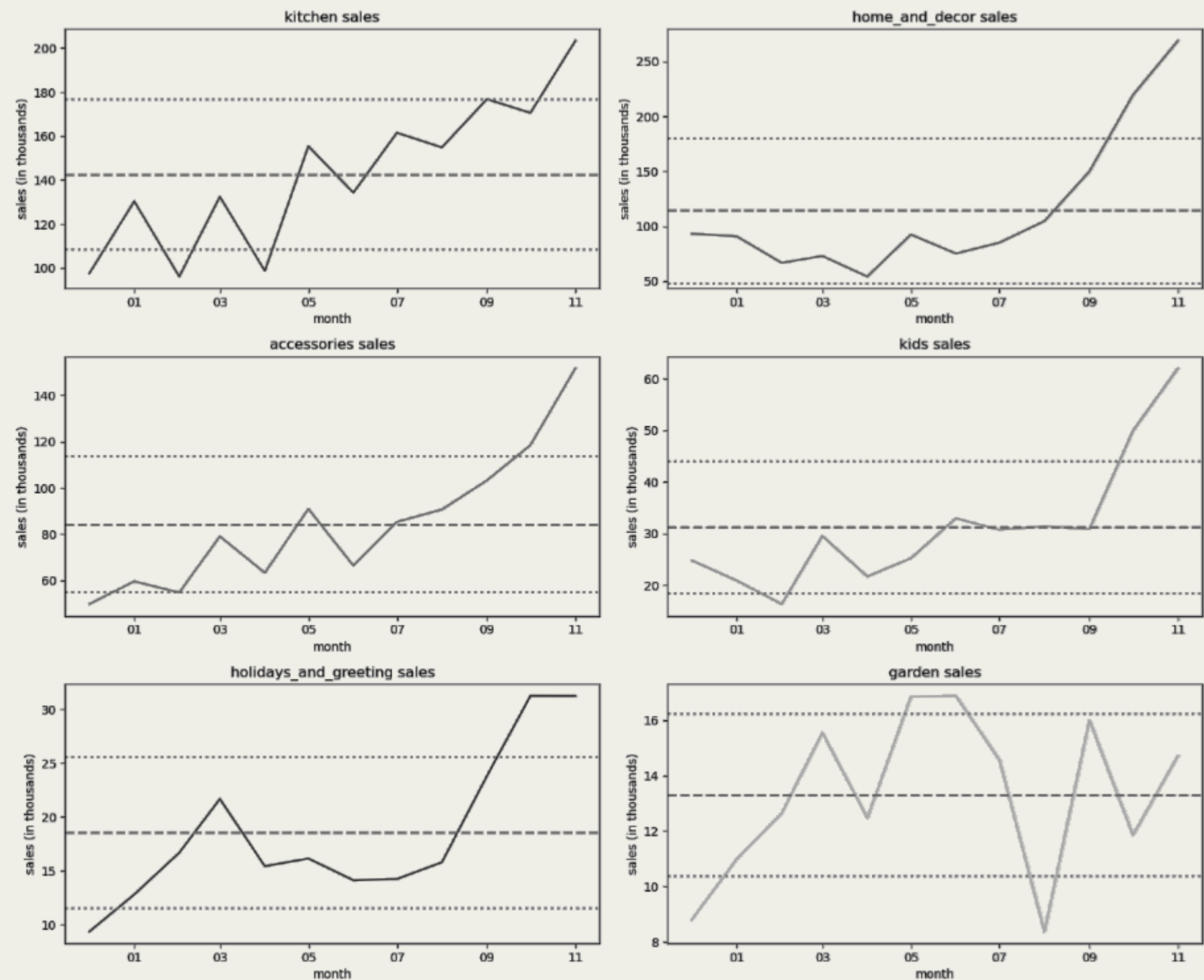
SALES JOURNEY (GENERAL)

The sharp fluctuations at the beginning of the period, but steady growth starts in June and peaks in November.

Minimum sales are recorded in **April** and maximum in **November**.



SALES JOURNEY (CATEGORIES COMPARE)



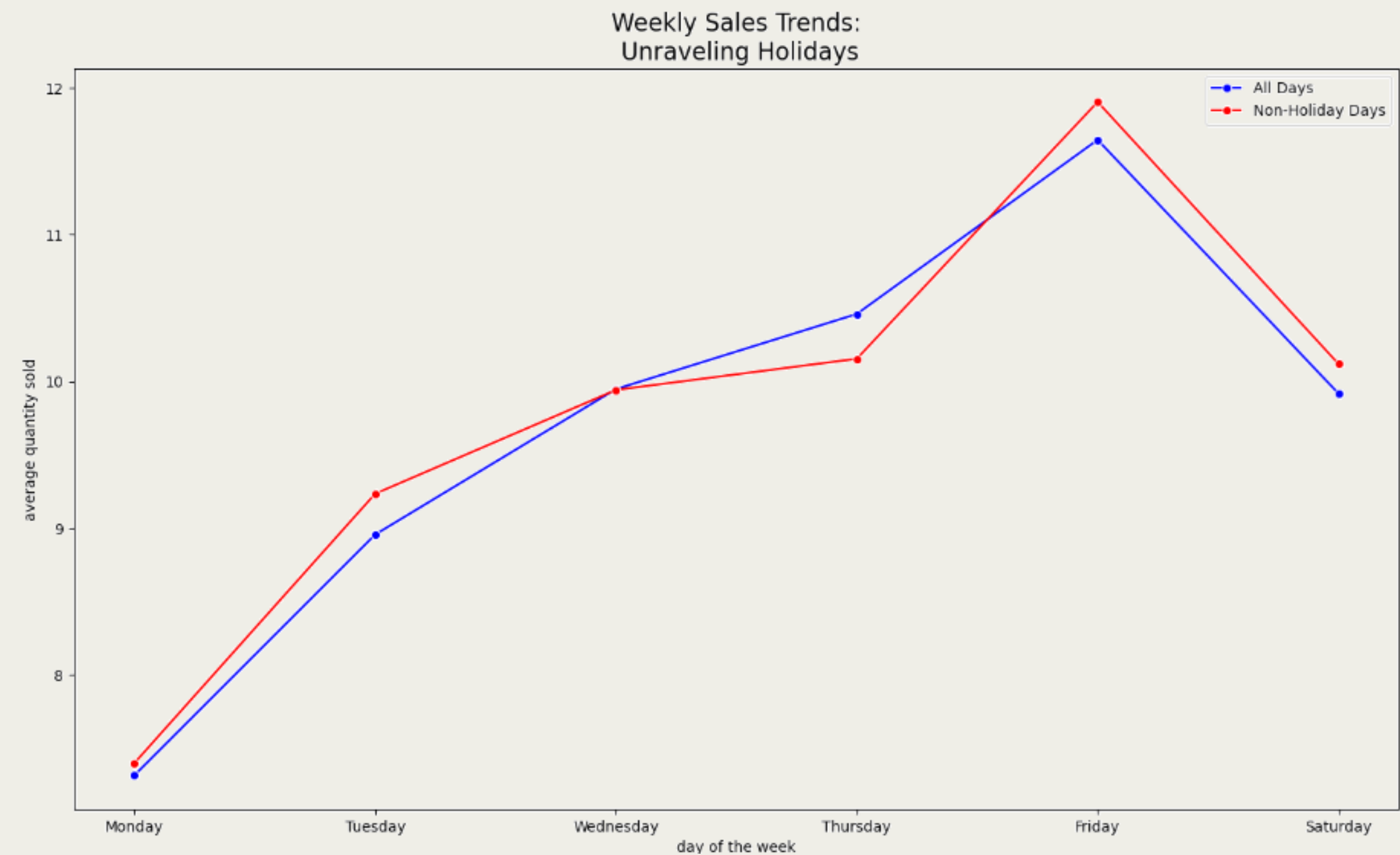
I can confirm a general decline in April. Though the extent varies, the decline is evident across all categories.

I also note a general increase towards the end of the period. The exception is the **garden** category, which diverges significantly from the overall trend.

The February drop remains uncertain, but I am inclined not to interpret it as a seasonal trend.

** The graph displays the most significant sales categories.*

WEEKLY SALES TRENDS



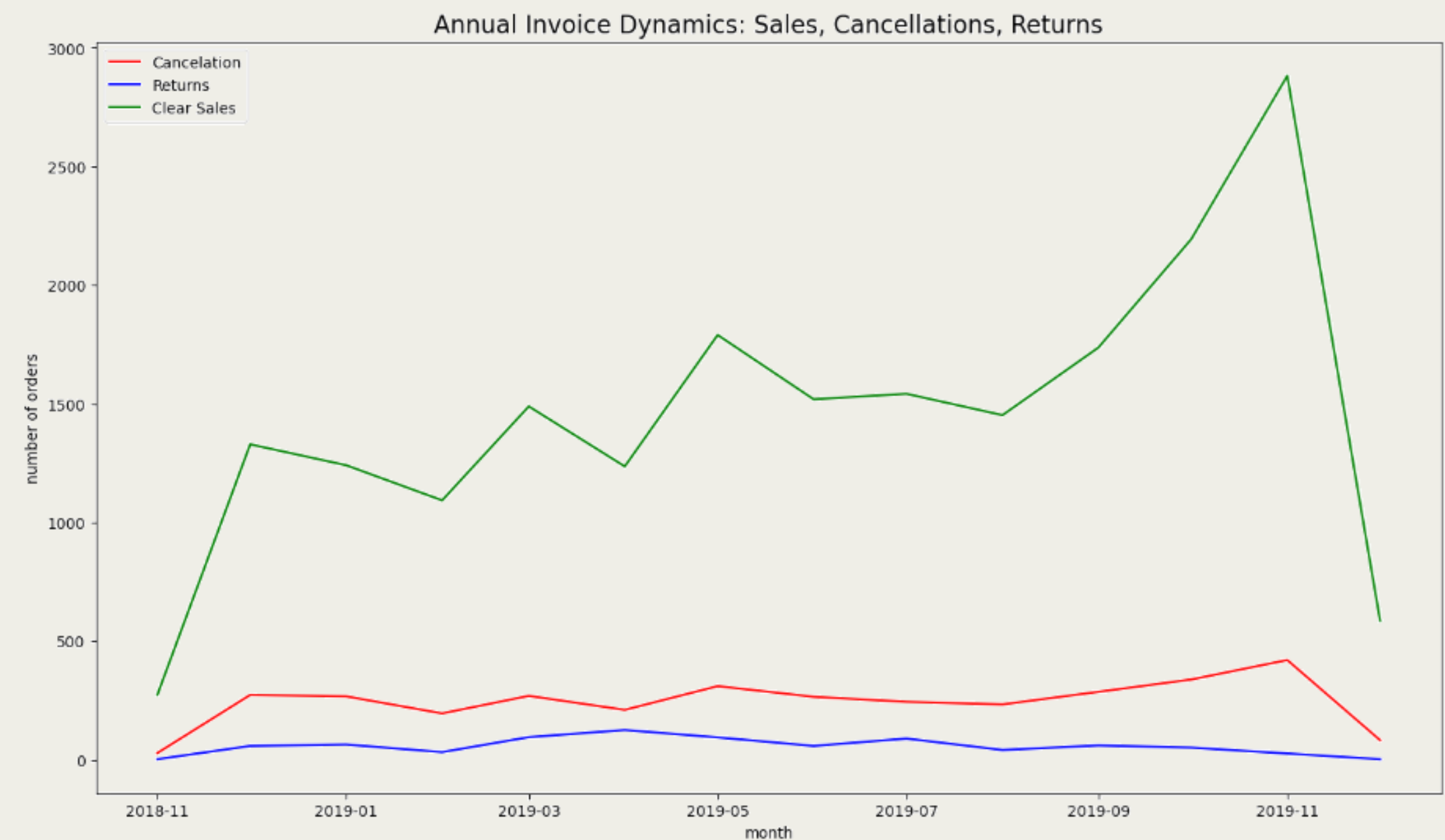
Analyzing the chart, we notice a consistent pattern in sales volumes throughout the week, with an increase from Monday to Friday and a subsequent decrease on Saturday.

It's essential to note that these trends are derived from warehouse data, primarily reflecting operational activities rather than direct customer demand. However, despite this distinction, the overall trend remains consistent.

Furthermore, when comparing sales during holidays and regular periods, we observe that, on average, sales tend to be higher during regular periods.

This suggests that while sales during holiday periods may experience an increase, it's likely not due to a surge in order volume but rather an increase in the number of orders themselves.

COMPARING SALES, CANCELLATION, AND RETURN INVOICES



The key takeaway is that cancellation and return invoice trends closely follow sales, albeit with smoother curves.

Now, let's turn our attention to the invoice count. While having one return invoice for every 26 sales may seem somewhat elevated, it's generally deemed acceptable.

However, the situation is more notable with cancellation invoices, as on average, there's one cancellation invoice for every 6 sales invoices.

RFM SEGMENTS DESCRIPTION

I categorize items into the following segments:

- 1. Jedi Knights:** These items reign supreme across all three RFM categories, attracting customers effortlessly, akin to Jedi masters harnessing the power of the Force.
- 2. Avengers:** This segment comprises items with high purchase frequency and substantial revenue, working together to offer customers top-tier products and services, just like the Avengers assembling to safeguard the world.
- 3. Sheldon's Club:** Items in this group boast high purchase frequency but low revenue, captivating attention with their unpredictability and unique characteristics, much like Sheldon from 'The Big Bang Theory.'
- 4. Force Awakens:** This category features new items with burgeoning market traction, showing promise to ascend as leaders, akin to the rising heroes in 'The Force Awakens.'
- 5. Risk Zone:** Items in this group exhibit low purchase frequency and moderate revenue, teetering on the edge of risk, necessitating attention to enhance their appeal.
- 6. Twilight Zone:** Items in this segment demonstrate moderate activity and revenue, residing in a realm between leaders and risky items, their fate uncertain, akin to the Twilight Zone.
- 7. Forgotten Treasures:** This category houses items that have long been untouched and may have faded from memory. They represent forgotten treasures awaiting rediscovery and restoration to their former glory.

TripleTen



RFM SEGMENTS

RFM Segments Distribution.
Avengers: Jedi, I'm your dady...



Nevertheless, this outcome is anticipated since **Jedi** represent a category with the highest indicators, and their numbers are naturally limited.

However, what truly surprises us is the significant presence of our **Forgotten Treasures**, which rank third and constitute nearly half of the **Avengers**. This proportion is quite substantial in my view.

When we combine the **Twilight Zone** with the **Forgotten Treasures**, their combined share accounts for almost **20%** of the total assortment. This suggests that approximately one out of every five items has not been sold for over six months.

Moreover, if these items were sold within the past quarter to half a year, they did so with low frequency and revenue. This is a matter of concern.

FINAL CONCLUSION

- The assortment from the online store is extensive, with 3810 unique items spread across 10 categories. However, 80% of the goods are concentrated in just three categories: **home and décor**, **kitchen**, and **accessories**.
- Sales trends generally follow a straight-line graph, but some smaller categories are performing surprisingly well in terms of sales.
- Pricing segments are predominantly split as 10, 30, 30, and 30, but certain categories like accessories, **sport and activity**, and **stationery** have smaller 'MEDIUM' segments. Surprisingly, over half of the sales (50% and up) are coming from the 'LOW' segment.
- Sales experience two major peaks and troughs, with November being a particularly lucrative month. Sales are typically active from January to July, peaking in November and dropping in December.
- Sales climb from Monday to Friday but dip slightly on Saturdays. During the holiday seasons, there's an increase in invoices rather than packing more goods.
- Returns are infrequent, with only every 26th invoice resulting in a returned item, and returns hardly ever exceed 1-2%. However, cancellation invoices are more common, occurring on average every 6th invoice, particularly in categories like **kitchen** and **home and decor**.
- RFM analysis reveals minimal shifts in sales leaders across different segments, except for **Force Awakening**, where accessories emerge as the top performer.
- Overall, 60% of the assortment remains active, sold within the last 30 days, while 20% sits unsold for over six months or is gathering dust, and the remaining 20% is on the verge of being forgotten but still worthy of attention.

Thank you!

E-Commerce: Product Range Analysis

Viktor Akulinin, student
victor.akulinin@gmail.com
APRIL, 2024

TripleTen

