

Section I: Evaluate data quality

- A. The data is a category level summary, it shows each product subcategory: how many listings exist, what the average price is, and how much revenue the “push up” promotion feature generated.

This data set provides a few interesting features:

1. WOMENS products dominate “push up” revenue.
 2. 15 subcategories have no “push up” revenue.
 3. “Push up” features are more popular with higher priced, niche products.
- B. To perform testing on this data for its quality, I would run a short python script, using panda’s library and I would check if there were any missing data, are there any duplicates and if every row has values where they are expected to be.
- C. Each row in category_2 represents parent category, and each category_3 represents a subcategory of its parent category. So, unit observation is a category_2 and category_3 combination. That means the dataset contains 238 observations and 5 variables.
- D. Suggested descriptions for the columns in the dataset:

category_2	Parent product category
category_3	Products subcategory inside category_2
number_of_listings	Total number of active listings in a subcategory
avg_listing_price_eur	Average asking price, across all listings
revenue_from_push_ups	Revenue generated by “push up” feature in that subcategory

Included:

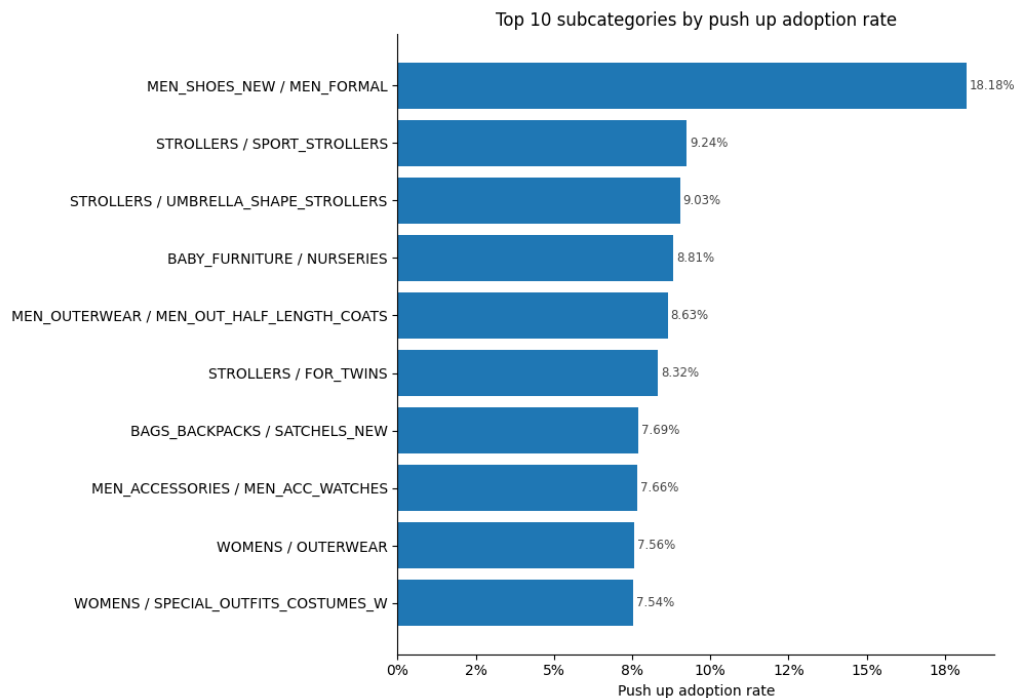
1. Parent and child relationship between two categories.
2. “asking price” instead of “price”, since those transactions are not completed.

Omitted:

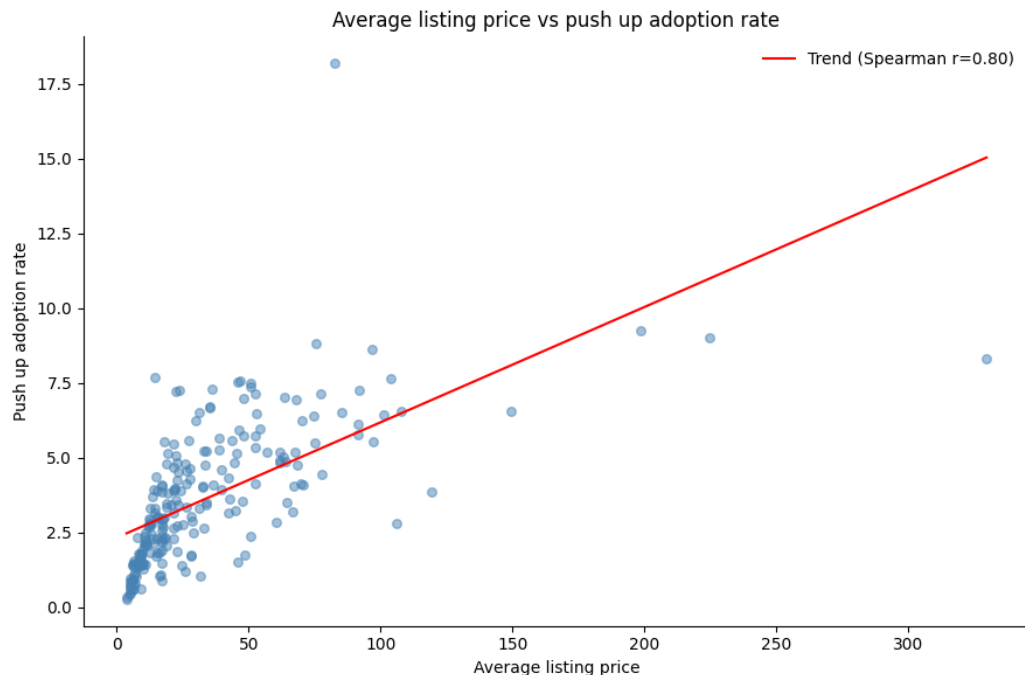
1. Any mentions of missing values or data quality issues.
2. Any guesses of what drives the values.

Section II: Evaluate current situation

- A. Based on given data, the best metric to show user interest in this feature is “push ups” purchased per items listed. $(\frac{revenue}{2€}) \div listings$. Because “push ups” are a paid feature and users buying it makes a direct signal of their interest. Dividing by number of listed items normalizes our data and shows how many users are willing to spend on “push ups” relative to how many of them there are.
- B. When calculated by the revenue per item listed metric, we can see that MEN_SHOES_NEW are leading with 18.18% “push up” adoption rate.



- C. Another metric that would be relevant to the “push ups” would be average listing price, since users who are selling higher priced items are more willing to purchase “push ups” because of the potential for a larger return from a sale.



The scatter plot confirms a strong relationship (Spearman $r=0.8$). Users are more willing to pay for “push ups” when their listing price is higher.

- D. GIRLS_CLOTHING / FOR_BABIES category is still gaining revenue even if the price of the “push up” feature is more than half of the average listing price, because of a few reasons:

1. Underpricing: The average listing price is 3.73 Eur, but the individual listings vary, there is a possibility that some people place more expensive baby clothing, but the average masks its spread.
2. Speed: Babies outgrow their clothes in weeks. Selling them for less profit empties wardrobe quicker and makes more room for bigger clothing.
3. Profile traffic: Some people may use “push up” features to increase the visibility of their profile rather than just a single listing. For example, users with 10 other active item listings could use “push up” on one item to attract more visitors to their profile, potentially increasing sales for all their other listings.

Section III: Analyse alternative strategies

A. Price change analysis:

1. Price increase:

A

2. Price decrease:

A

B.