Welcome to my study

Recommendation system

Let's talk about shopping

Start →

Aurélie Lopez-Vicente

What is a product recommendation system ?

A recommendation system is a tool based on algorithms and data, designed to suggest relevant products or content to users based on their preferences, past behaviour or other specific criteria.

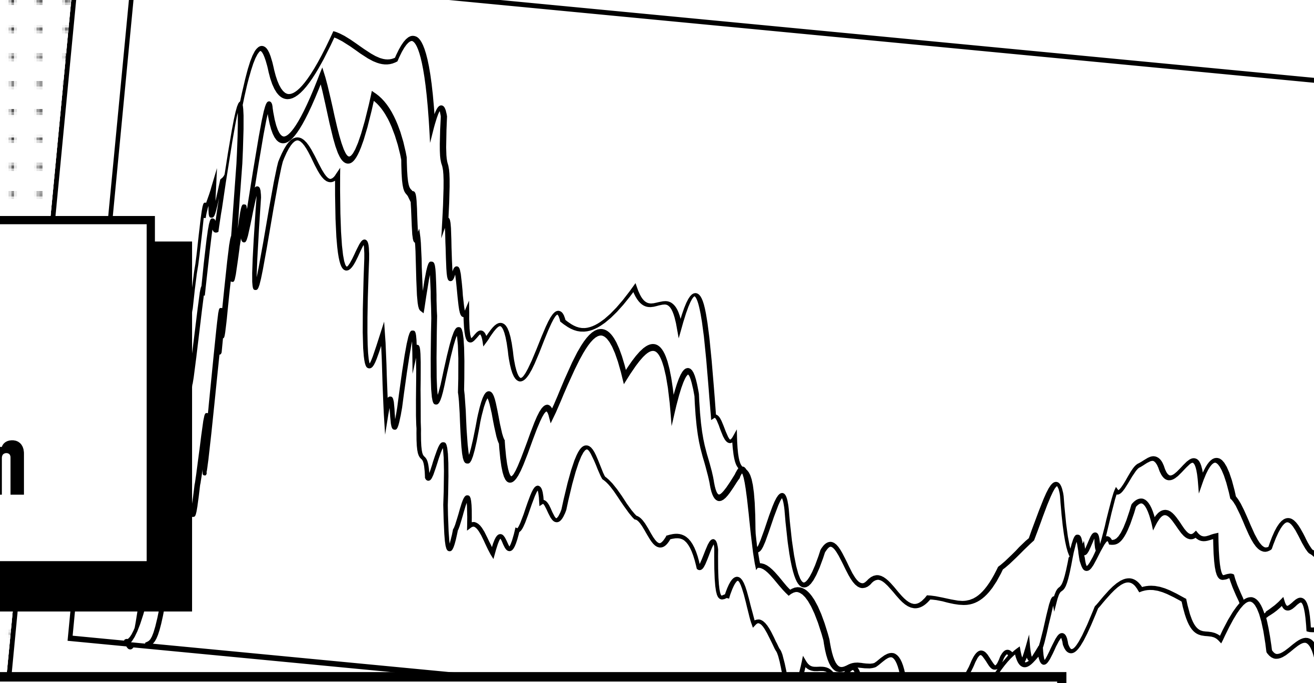
Let's go!

Business Case: Why is this important?

In a world saturated with information and options, recommendation systems enable companies to :

- Improve the user experience by making it easier to discover products,
- Increase sales and engagement through targeted suggestions,
- build customer loyalty by personalising their experience.

Giants such as Netflix, Amazon and Spotify have demonstrated the strategic impact of recommendations on revenue growth and user satisfaction.

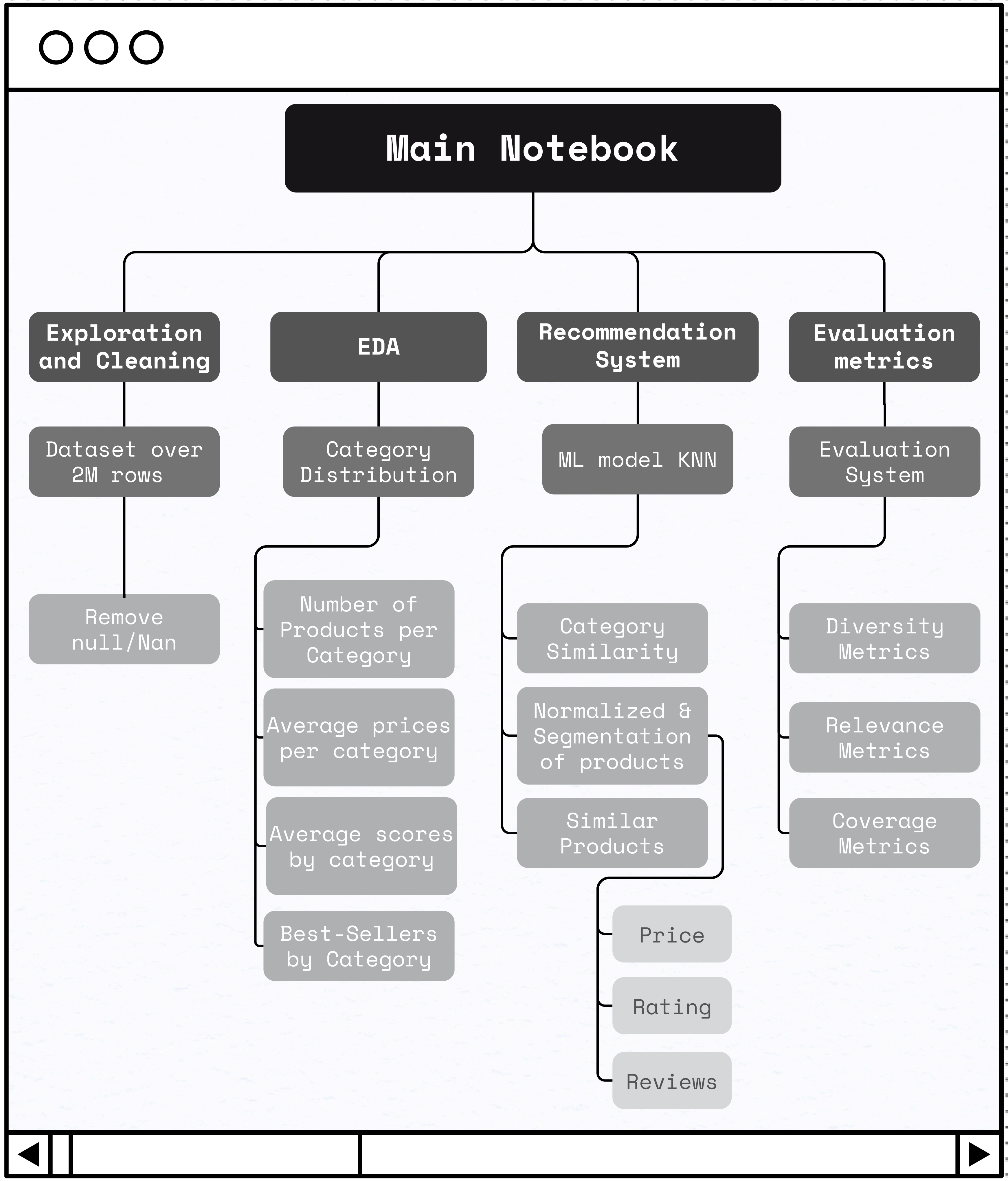


My choice : Content-Based System

Content-based filtering is a fundamental method frequently employed by modern recommendation systems.

The core concept of content-based filtering is that if a customer enjoys a particular product, they are likely to appreciate another product with similar characteristics.

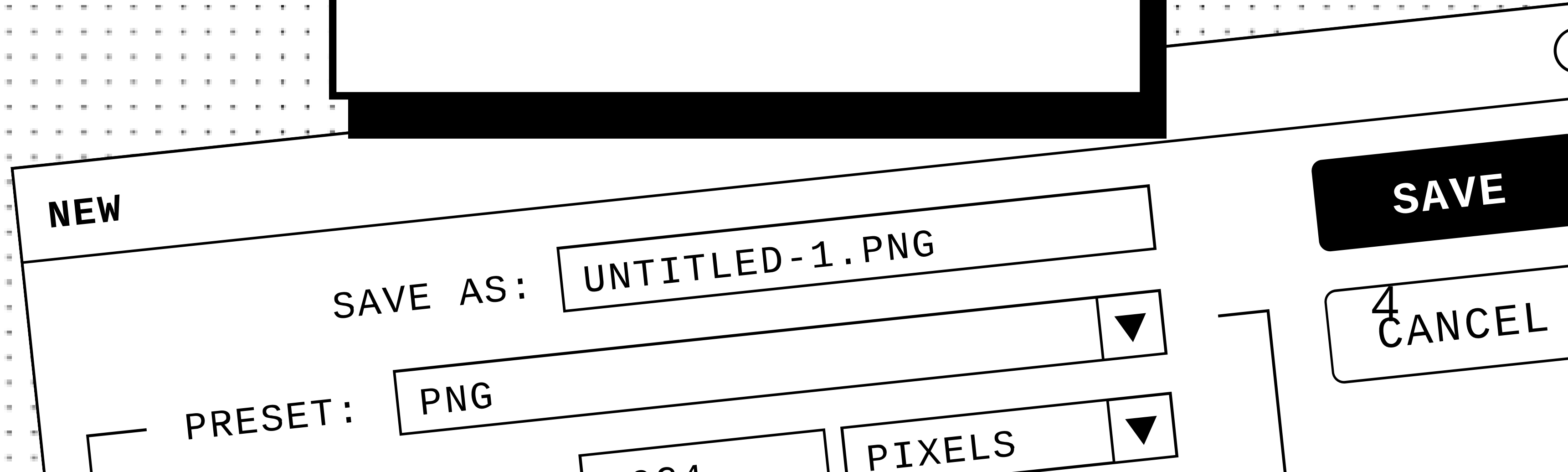
This method analyzes the attributes of products that a user has shown interest in and suggests items with comparable features. By focusing on product specifications and user preferences, content-based filtering provides tailored recommendations that align closely with individual tastes.



Conception Process

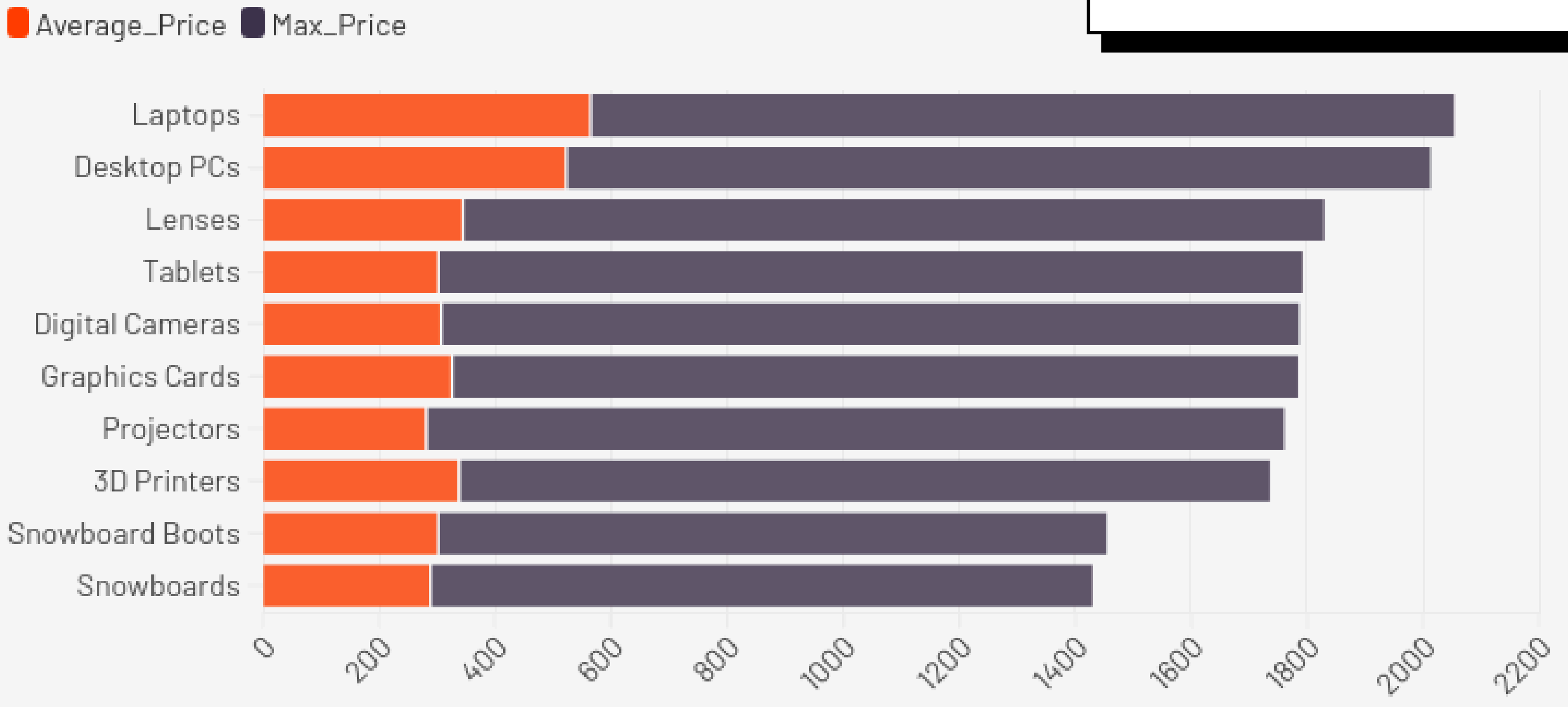
Legends :

- Main component of the structure
- The main phases of the project
- Specific tasks within each major phase
- Sub-processes

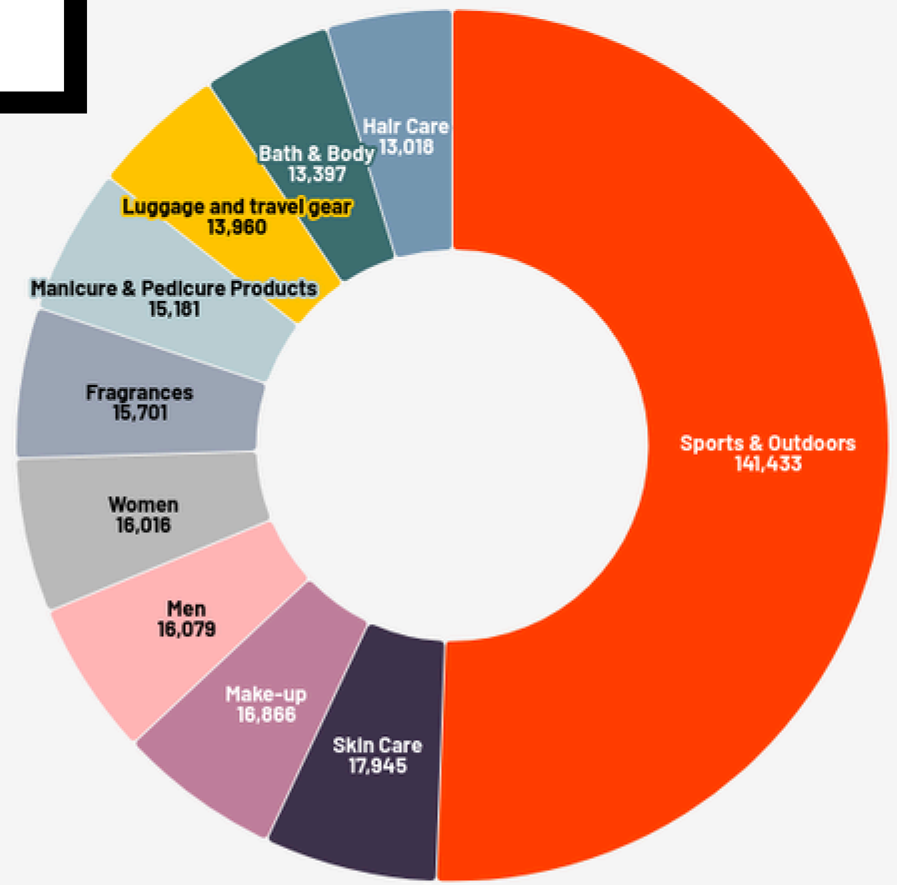


EDA

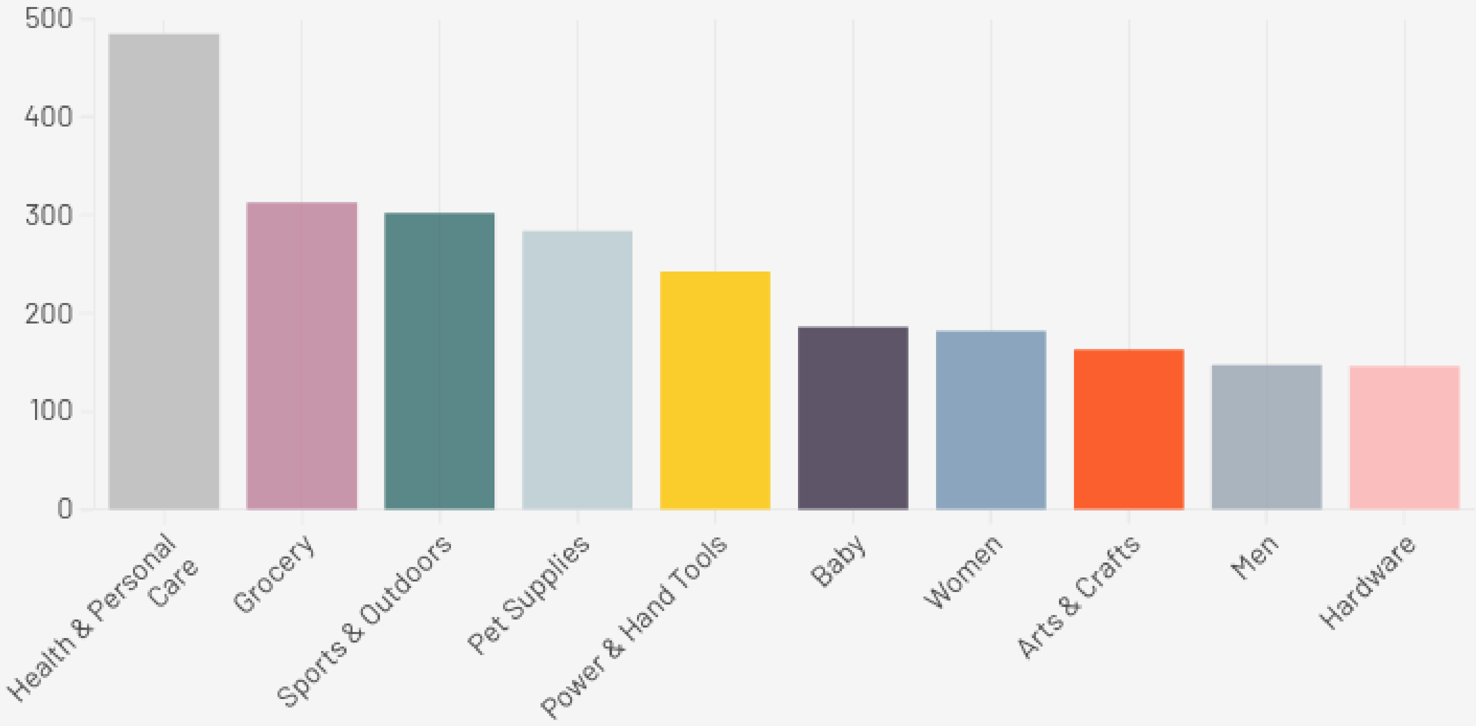
Top 10 Categories by average and max price



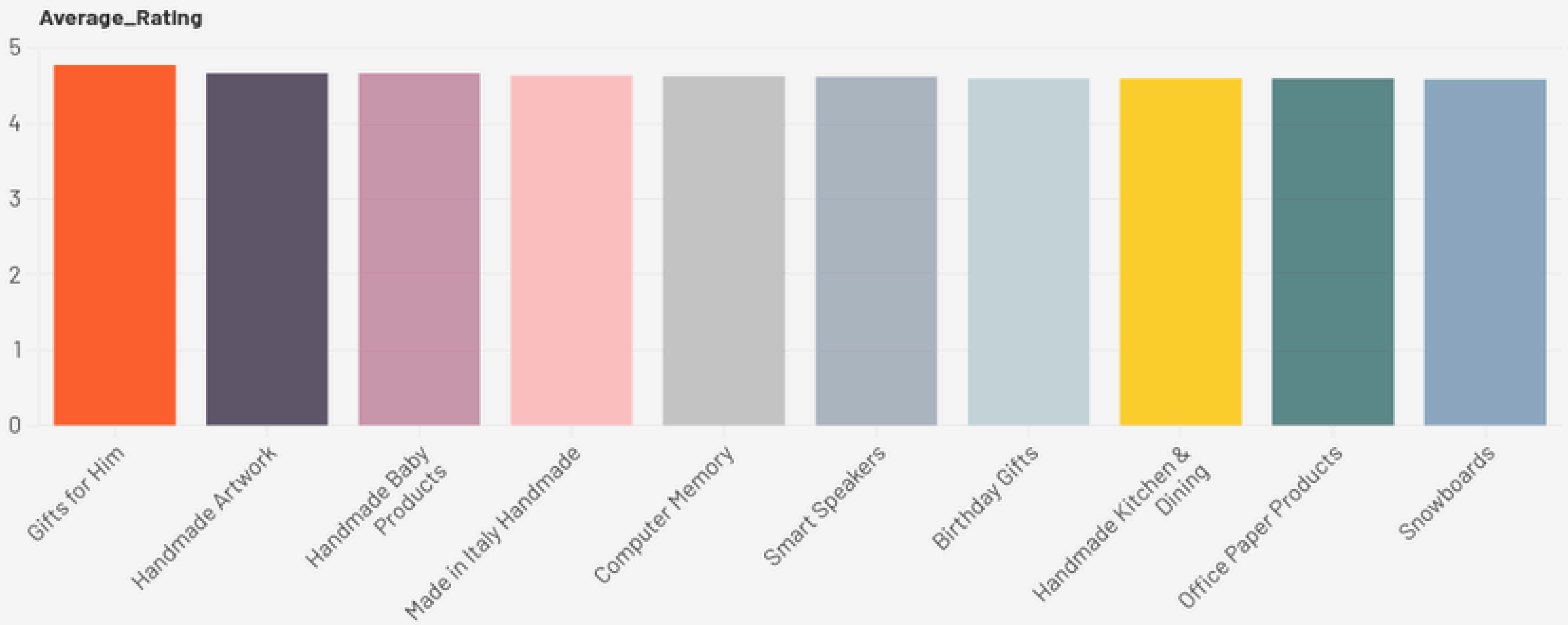
Top 10 Categories by Number of Products



Top 10 Categories by Number of Bestsellers



Top 10 Categories by Average rating



My recommendation system, why KNN ?

Choosing the KNN model for the recommendation system

The K-Nearest Neighbors (KNN) model was chosen for its key advantages:

Simplicity and interpretability:

Easy to understand and transparent in its operation.
Ideal for simple debugging and improvements.

Relevance for recommendations:

Effective for finding similar items using multiple features (price, category, ratings).
Works well with standardised feature spaces.

Technical advantages:

No training required: adapts instantly to new data.
Suitable for structured and mixed data with a metric such as cosine distance.

Adaptation to Amazon:

Efficiently manages large catalogues.
Facilitates the balance between similarity and diversity of recommendations.

Acceptable limitations:

Slowness is not a problem thanks to periodic updates and pre-calculated recommendations.

Evaluation Metrics Results

Diversity metrics :

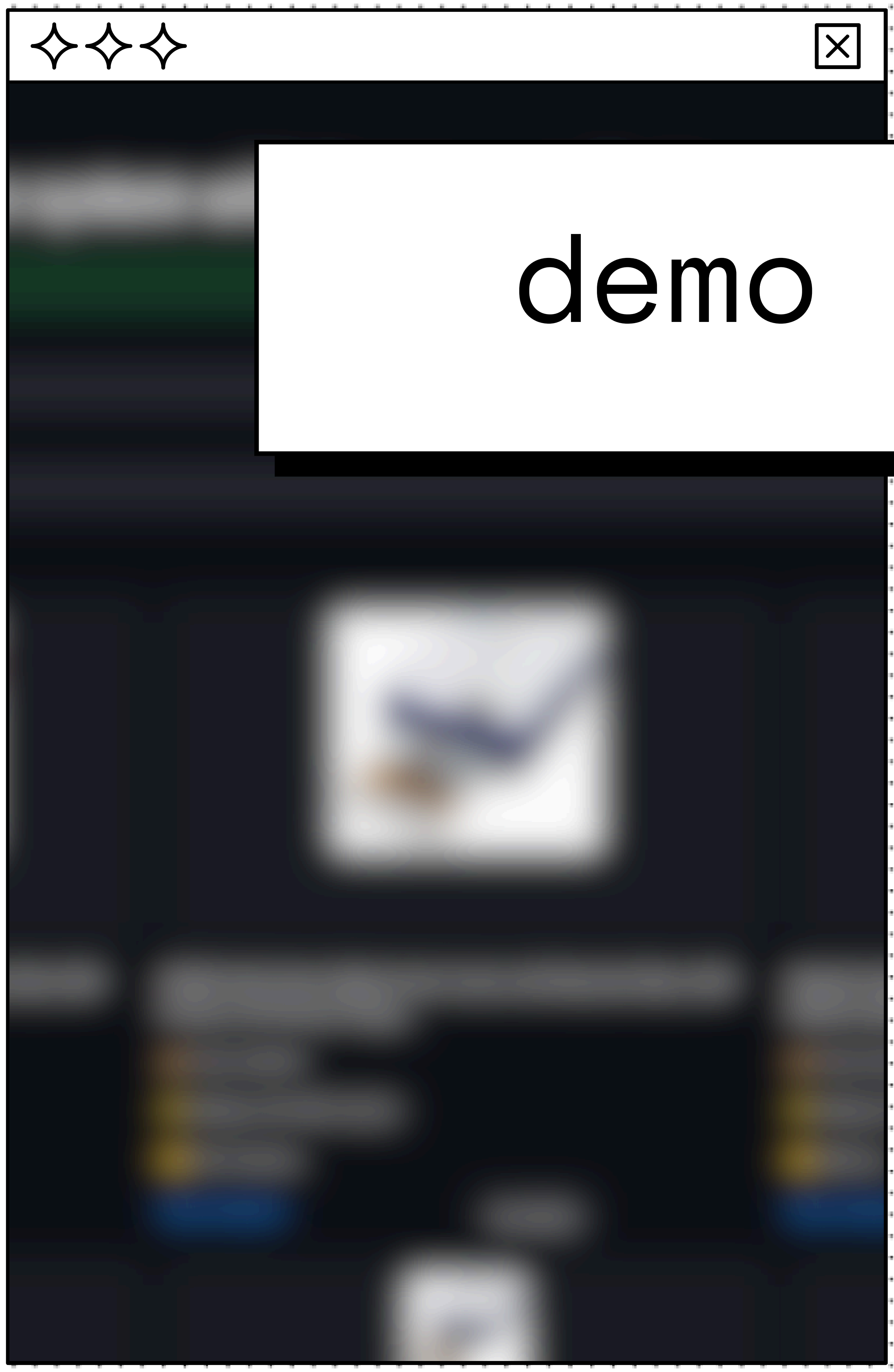
- category_diversity: ~ 0.2 → Around 20% of recommendations are in different categories, which is a bit low but consistent with our focus on similar categories
- price_range_ratio: ~ 0.8 to 1.1 → Recommendations cover around 1x the price of the original product, indicating a
- reasonable price variation
- avg_price_distance: ~ 0.4 → On average, prices differ by 44% from the original price, showing good price diversity
- rating_spread: ~ 0.5 → The difference between min and max ratings is around 0.5 stars, suggesting consistency in quality

Relevance metrics:

- avg_rating: ~ 4.4 → Excellent average score (out of 5)
- min_rating: ~ 4.1 → Even the lowest-rated products are still of good quality
- avg_reviews: ~ 1247.580 → Good average number of reviews, indicating well-established products
- weighted_rating: ~ 4.469 → The weighted score is close to the average score, confirming the reliability of the ratings

Coverage metrics :

- unique_items_ratio: ~ 1 → No duplications in recommendations
- success_rate: ~ 1 → System successfully generates recommendations for all products tested



demo

Learnings and Challenges

- Simplifying the architecture of features with .py files and executing them with main notebook saves time
- Streamlit ✨
- You can add CSS on streamlit ✨
- Debug functions
- KISS (Keep It Simple, Stupid)
- Python Class ✨

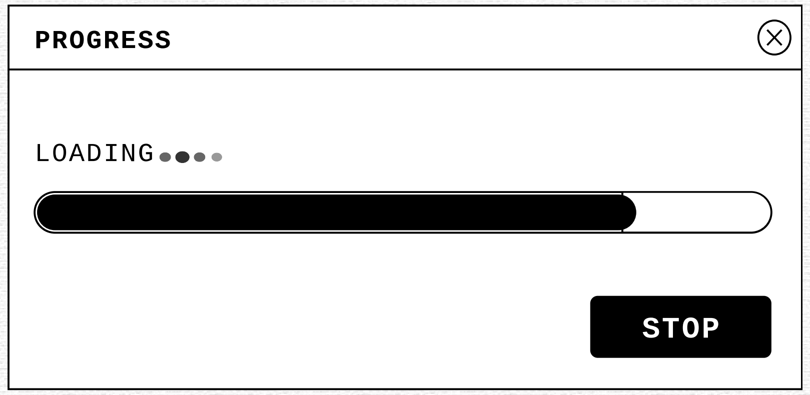
- Huuuuuuuuge dataset
- Very long debugging session for the recommender sys
- Search bar = hell
- Random bugs

OOPS!

YOUR FILE HAS STOPPED RESPONDING.

CLOSE CRY

Features to add in the future



- Better search bar
- Better filter system
- More user friendly interface

- New segmentation : Create major categories with product sub-categories for greater diversity in recommendations

Thanks y'all