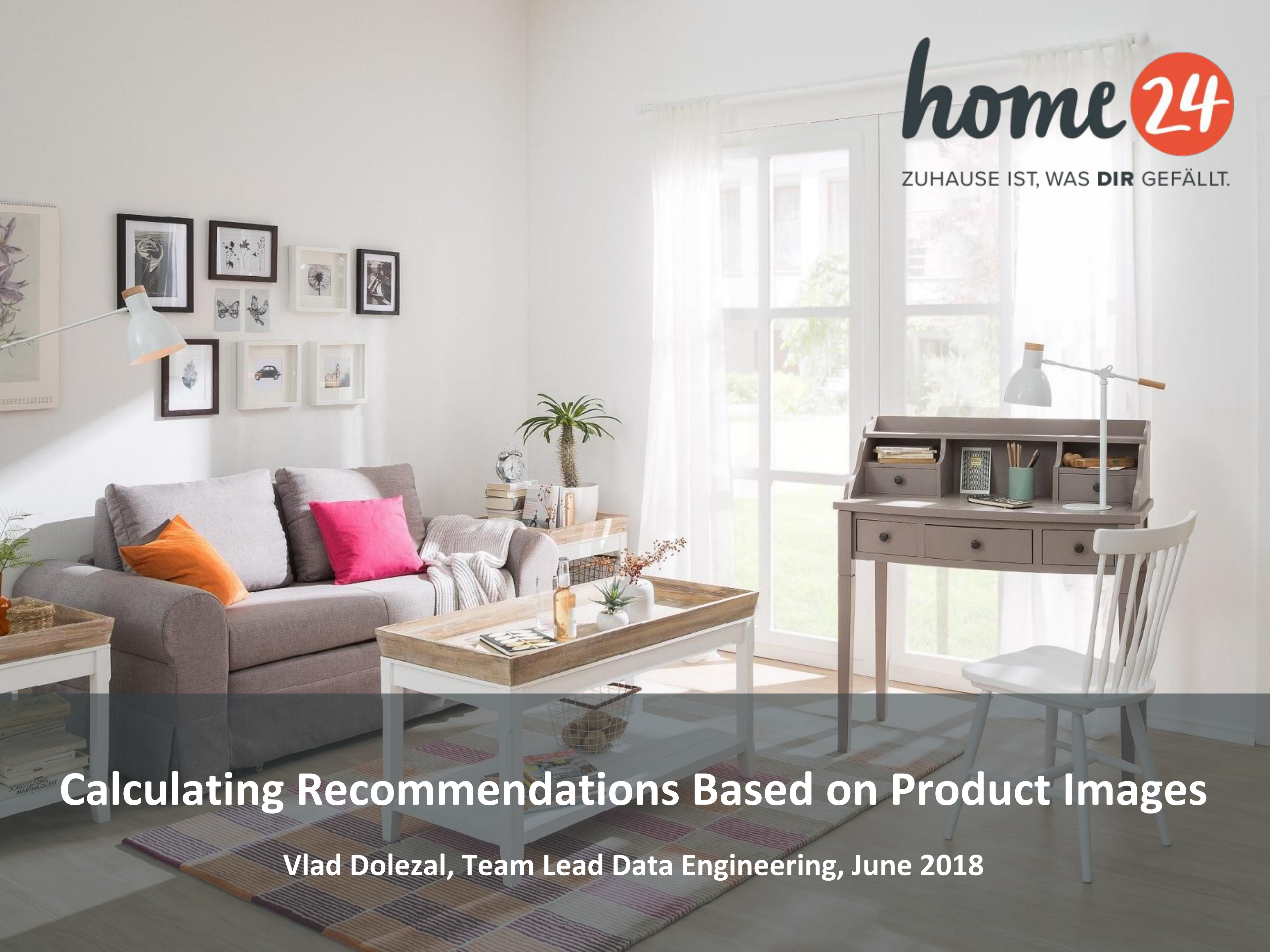


home24

ZUHAUSE IST, WAS **DIR** GEFÄLLT.

Calculating Recommendations Based on Product Images

Vlad Dolezal, Team Lead Data Engineering, June 2018

home24 "Zuhause ist, was dir gefällt"

€117b+	€276m	85%	c.€350	45%	First order	<10%
TAM[1] across 8 markets	Revenue FY2017	CAGR revenue 2010-2017	High AOV[2]	Gross margin FY2017	profitability	Low return rates[2]

1. Based on Euromonitor data for home24 markets which includes Germany, France, Italy, the Netherlands, Austria, Belgium, Switzerland and Brazil. 2. Including VAT, for Europe only, as of FY 2017.

home24 “code sweet code”

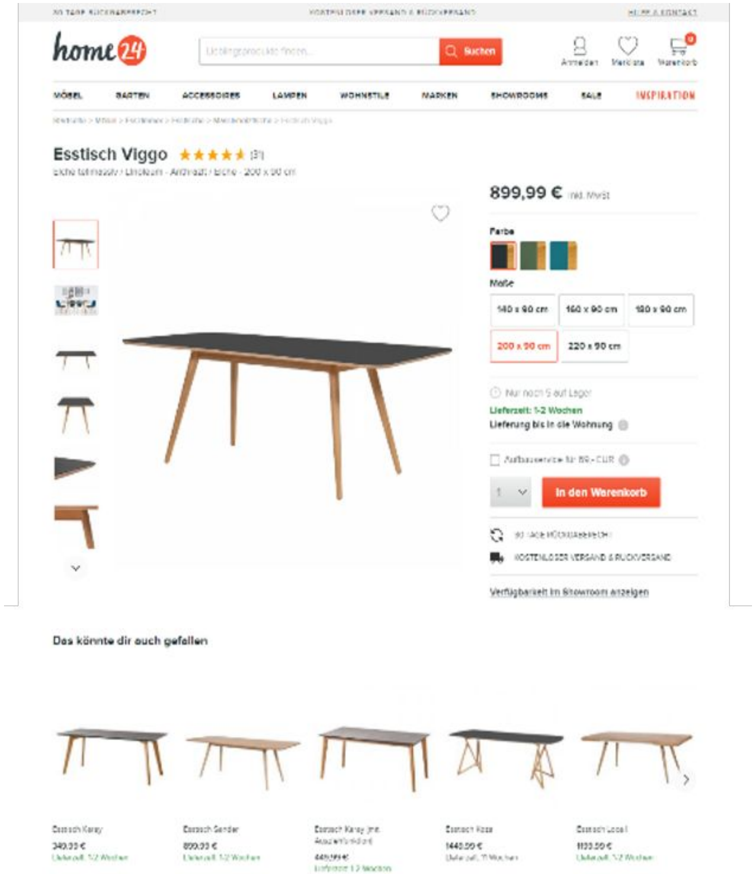
1. 100+ in tech department
2. 8 Teams
3. 16 in Data Team, 14 nationalities
4. Data Team working with Scala, Spark, R, ...

home24.tech.blog

home24.de/jobs

LinkedIn - home24 SE

How we got the idea to use article images for recommendations



Best recommender - collaborative filtering

Advantages:

- Real users shows you exactly what they prefer

Challenges:

- New products have no user behavior
- Broad assortment of products dilutes collaborative filtering signals

Improving the system - add attribute-based recommendations

Advantages:

- Covers all products

Challenges:

- Let's see...

Visualising our results

Attribute-based recommender

Click-through score on recommendations

$$\text{score}_A(B) = \frac{\#\text{clicks}_A(B) - \text{expected}_A(B)}{\text{expected}_A(B)}$$

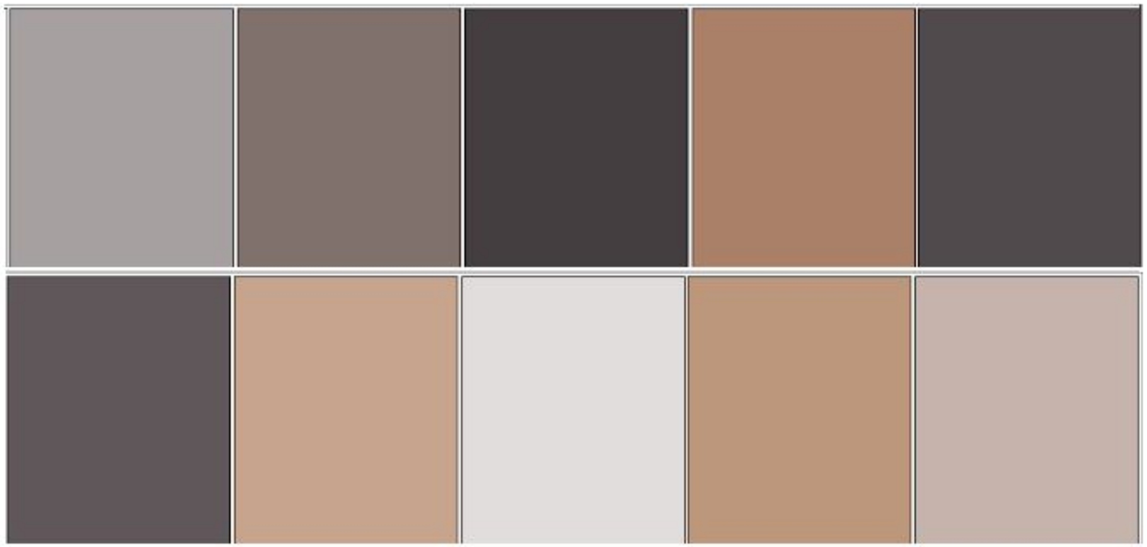
Challenge: Article attributes have limited detail

```
{
  attributes: {
    color: "brown",
    material: "solidwood",
    ...
  }
}
```

```
{
  attributes: {
    color: "brown",
    material: "solidwood",
    ...
  }
}
```

Solution: Recommend articles that have similar colors in their images

Extracted colors

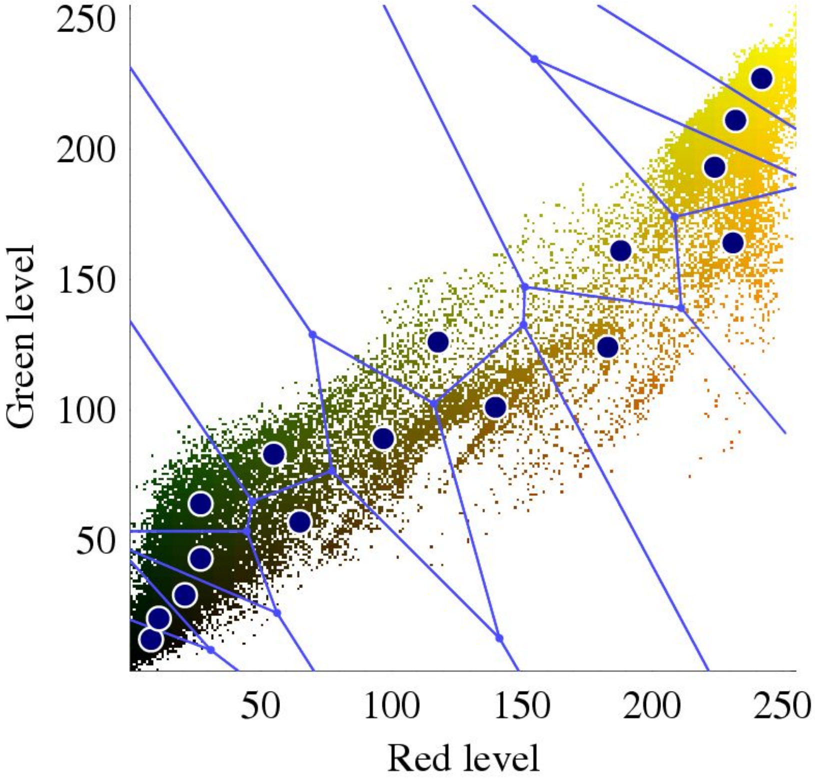


Extraction process:

- Ignore background pixels
- Cluster remaining pixels to 10 clusters using K-means clustering algorithm, to get a representative sample of colors for each image (color quantization)

Color quantization

A small photograph that has had its blue channel removed. This means all of its pixel colors lie in a two-dimensional plane in the color cube.

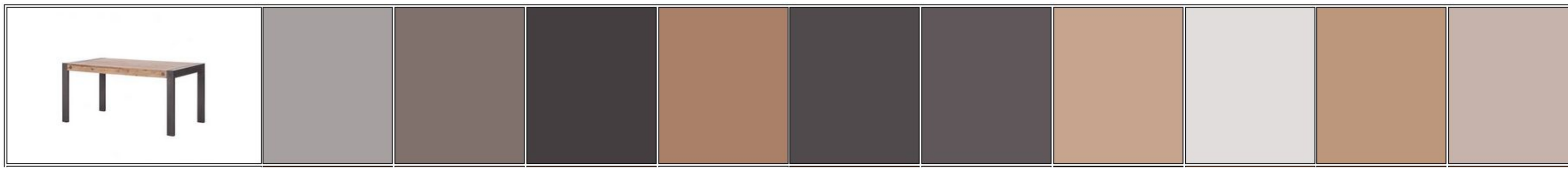


The color space of the photograph to the left, along with a 16-color optimized palette produced by Photoshop. The Voronoi regions of each palette entry are shown.

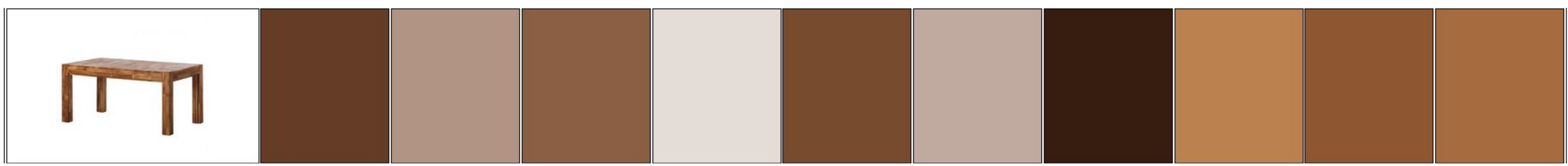
Image credit: Dcoetzee via Wikipedia

Extracted colors

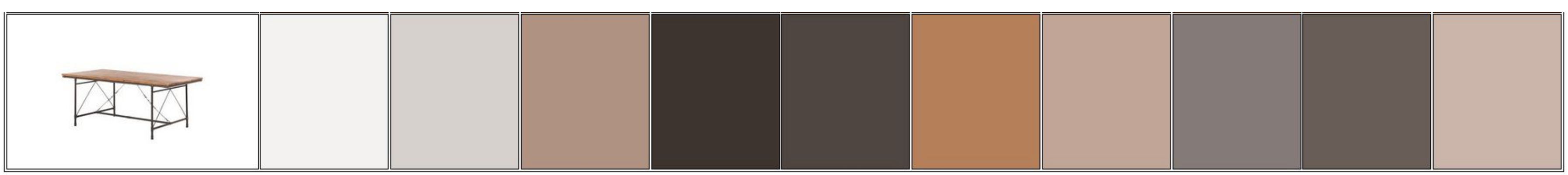
Product that we want recommendations for



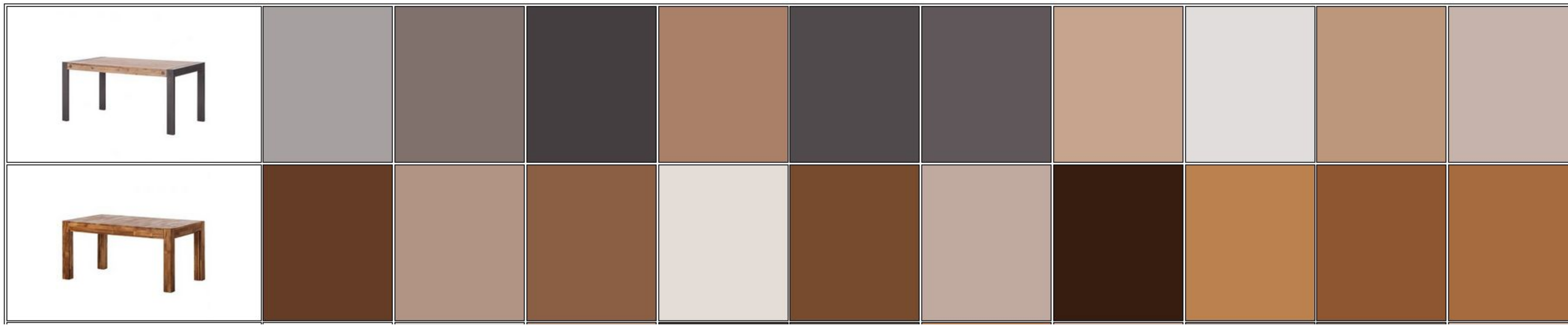
Top pick of attribute-based recommender



Most underrated recommendation (based on CTR)



Calculating the color similarity score



Let $distance(color_a, color_b)$ be the euclidean distance between the two colors in LAB color space,

and $$f(A,B) = \sum_{color_a \in A} \min_{color_b \in B} (distance(color_a, color_b))$$

Then the color similarity score for a pair of products A,B is:

$$score(A,B) = \frac{1}{1 + f(A,B) + f(B,A)}$$

Final results

Attribute-based recommender

+ Color Recommender

= Combined results

Recognizing clean vs. mood images

Solution:

Analyse image borders. Standard deviation of color components in RGB < 5 = clean image.

Dealing with broad assortment of products

Challenge:

- Calculating color similarity for ALL products of same category is expensive (can be thousands of products, complexity $O(n^2)$)

Our solution:

- Re-using the candidate filtering we have for our attribute-based recommender - locality-sensitive hashing (LSH) based on product attributes

Locality-sensitive hashing on product attributes

Basic idea: splitting a high-dimensional space to “buckets” based on lower-dimensional projections

- Products attributes are a high-dimensional space (we consider 14 different attributes)
- LSH hashes the attributes so that similar products map to the same “buckets” with high probability
- This reduces the candidate space from thousands of products to a manageable number

Results

10% uplift in Click-Through Rate

12% uplift in Click-Through Rate for products with low user behavior

4% conversion rate uplift

Key Takeaways

1. Build tools that let you visualize your results
2. Plan to iterate, because you won't get things perfect straight away
3. Think about what non-obvious data you have

```
<!doctype html>
<html lang='de-DE' class='DE lang-de'><head>
    <meta http-equiv="Content-Type" content="text/html; charset=utf-8">
<title>Couch und Beistelltische bequem online bestellen | Home24</title>
<!--
```

```
   Check out the open positions in our Berlin office at home24.de/jobs


--><meta name="description" content="Couch und Beistelltische online kaufen und von den Vorteilen auf home24 profitieren: gute Verarbeitung, große
<meta name="robots" content="index,follow" >
```

ANY QUESTIONS?

Feel free to grab any of us after the talk as well

Appendix 1 - useful links

Improving Content-Based Recommendation using Product Images - Home24 Tech Blog (includes more examples)
https://home24.tech.blog/2017/05/24/improving-content-based-recommendation-using-product-images/

Color quantization
https://en.wikipedia.org/wiki/Color_quantization
Color Reduction Using K-Means Clustering, Mikolov T., 2007 (PDF)
http://old.cescg.org/CESCG-2007/papers/Brno-Mikolov-Tomas.pdf

Locality-sensitive hashing
https://en.wikipedia.org/wiki/Locality-sensitive_hashing
LSH.9 Locality-sensitive hashing: how it works, Victor Lavrenko
https://www.youtube.com/watch?v=Arni-zkqMBA