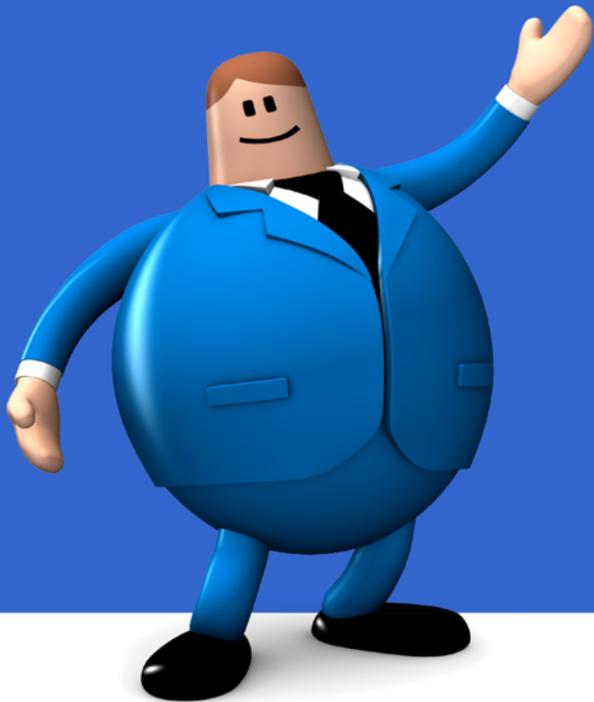
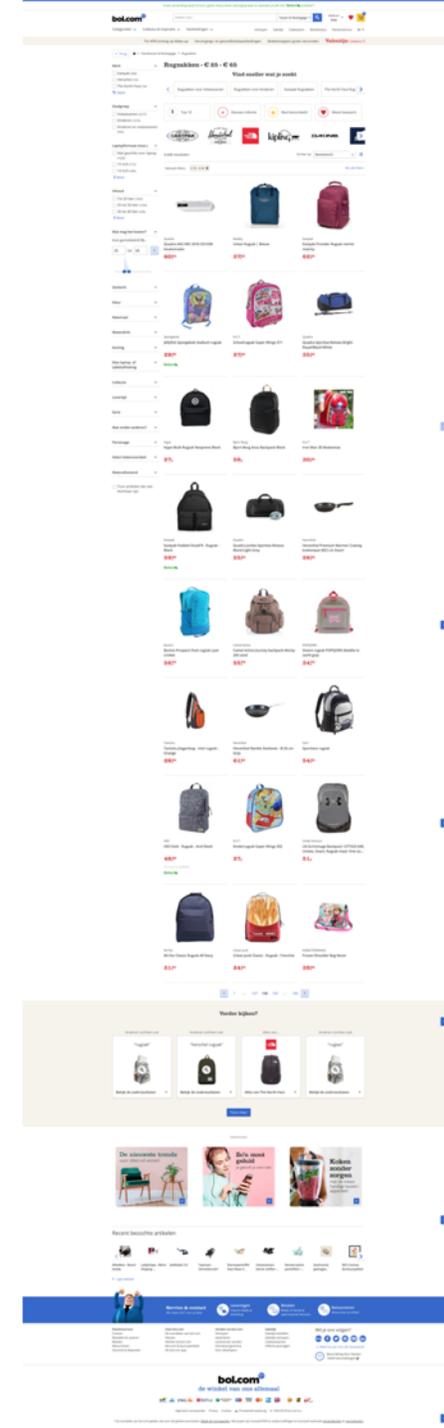


Cleaning up the Bol.com product catalogue with CRaPS



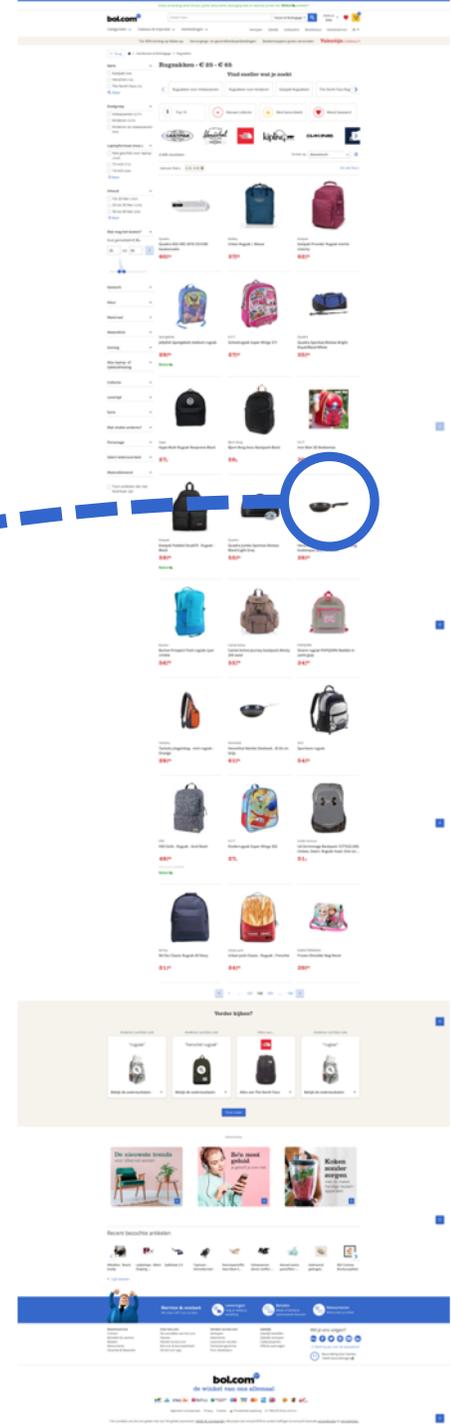
Let's look for a backpack



Let's look for a backpack



herenthal
Herenthal Premium Marmer Coating
koekenpan Ø22 cm Zwart
28,90



Let's look for a backpack



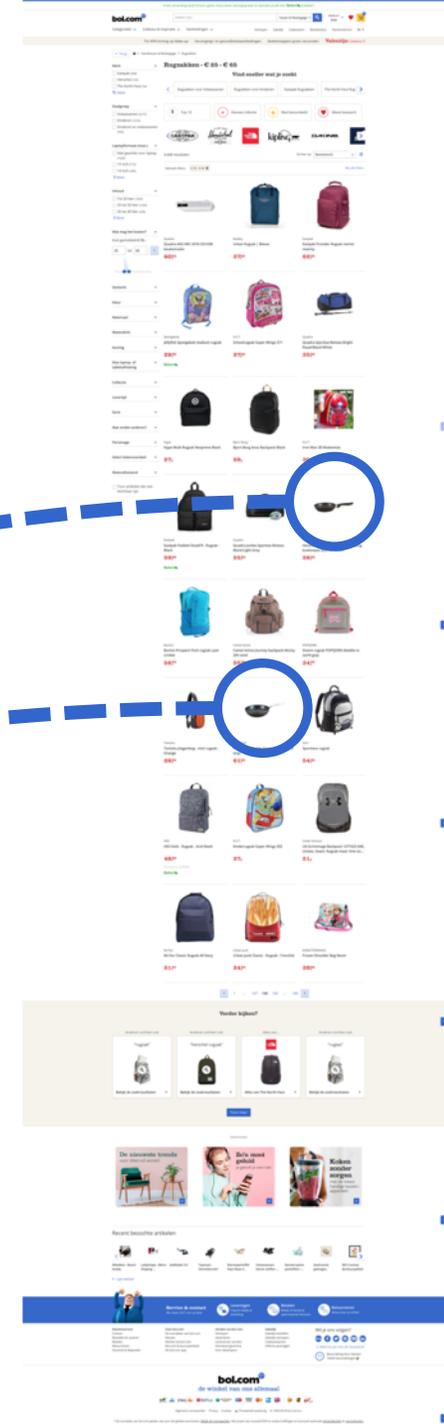
herenthal
Herenthal Marble Steelwok - Ø 26 cm
Grijs

61,95



herenthal
Herenthal Premium Marmer Coating
koekenpan Ø22 cm Zwart

28,90



Let's look for a backpack



Quadra
Quadra AEG KRC 4376 CD/USB
keukenradio

60⁹⁴



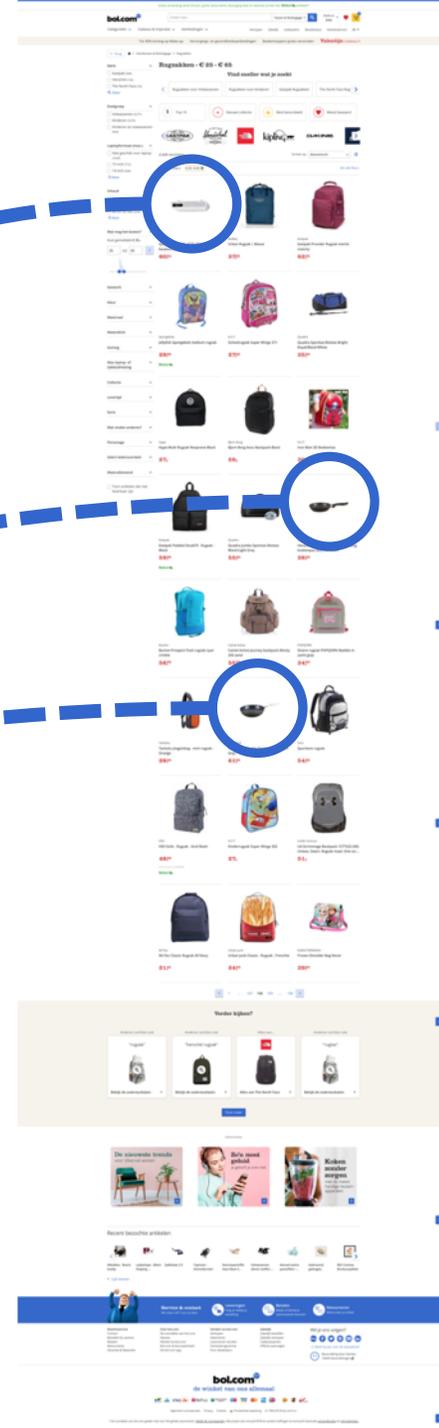
herenthal
Herenthal Marble Steelwok - Ø 26 cm
Grijs

61⁹⁵

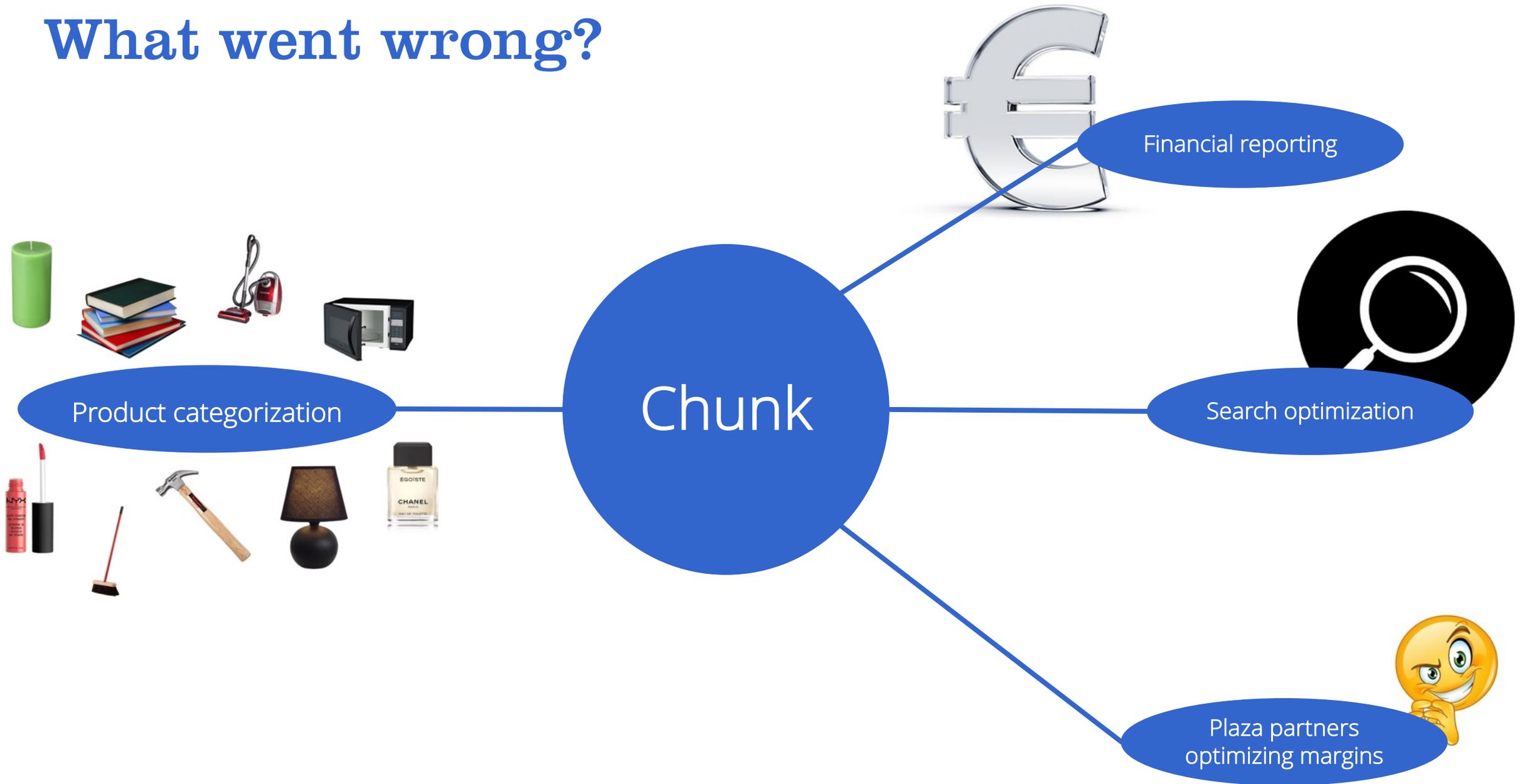


herenthal
Herenthal Premium Marmer Coating
koekenpan Ø22 cm Zwart

28⁹⁰

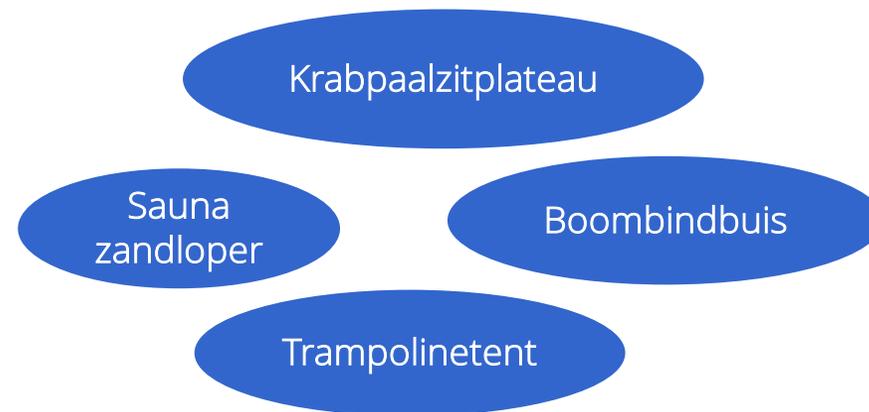
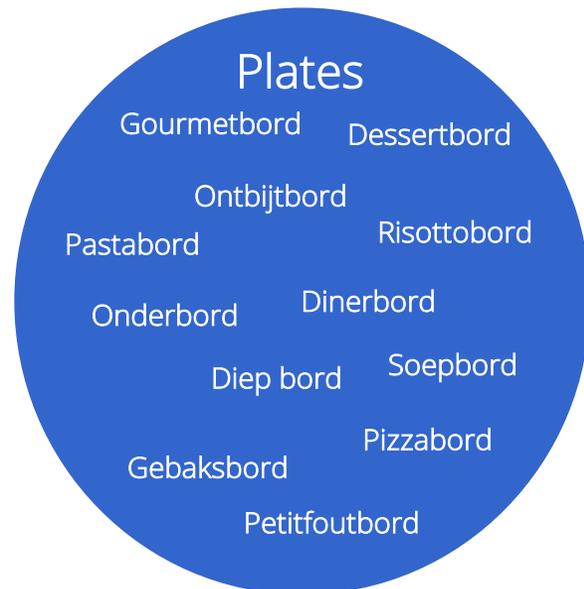


What went wrong?



You get a chunk! And you get a chunk!

- >6000 chunks
- Much ambiguity
- 50% has less than 50 products with an offer
- Partners have to choose a chunk themselves



You get a chunk! And you get a chunk!

- >6000
- Much
- 50% h
- an offe
- Partne
- thems





500 products /
hour

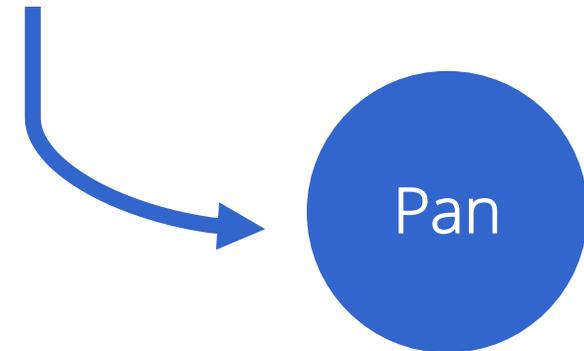
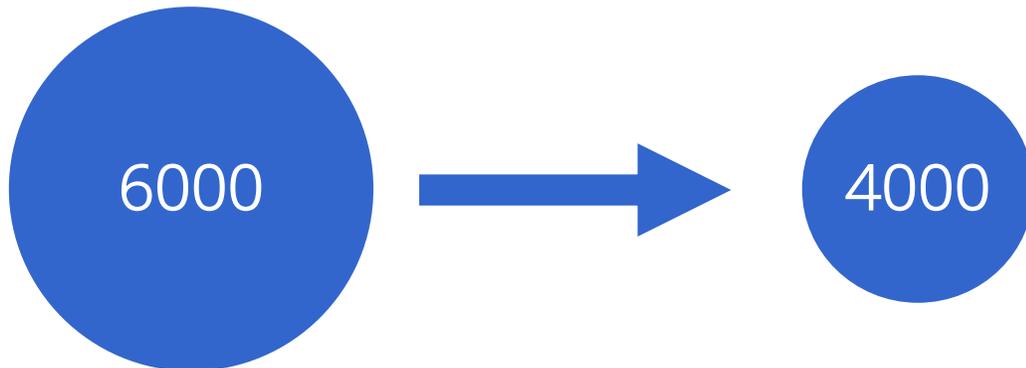
Number of
products in the
wrong chunk:
Unknown

1 FTE for ~20
Year to scroll our
catalogue *once*

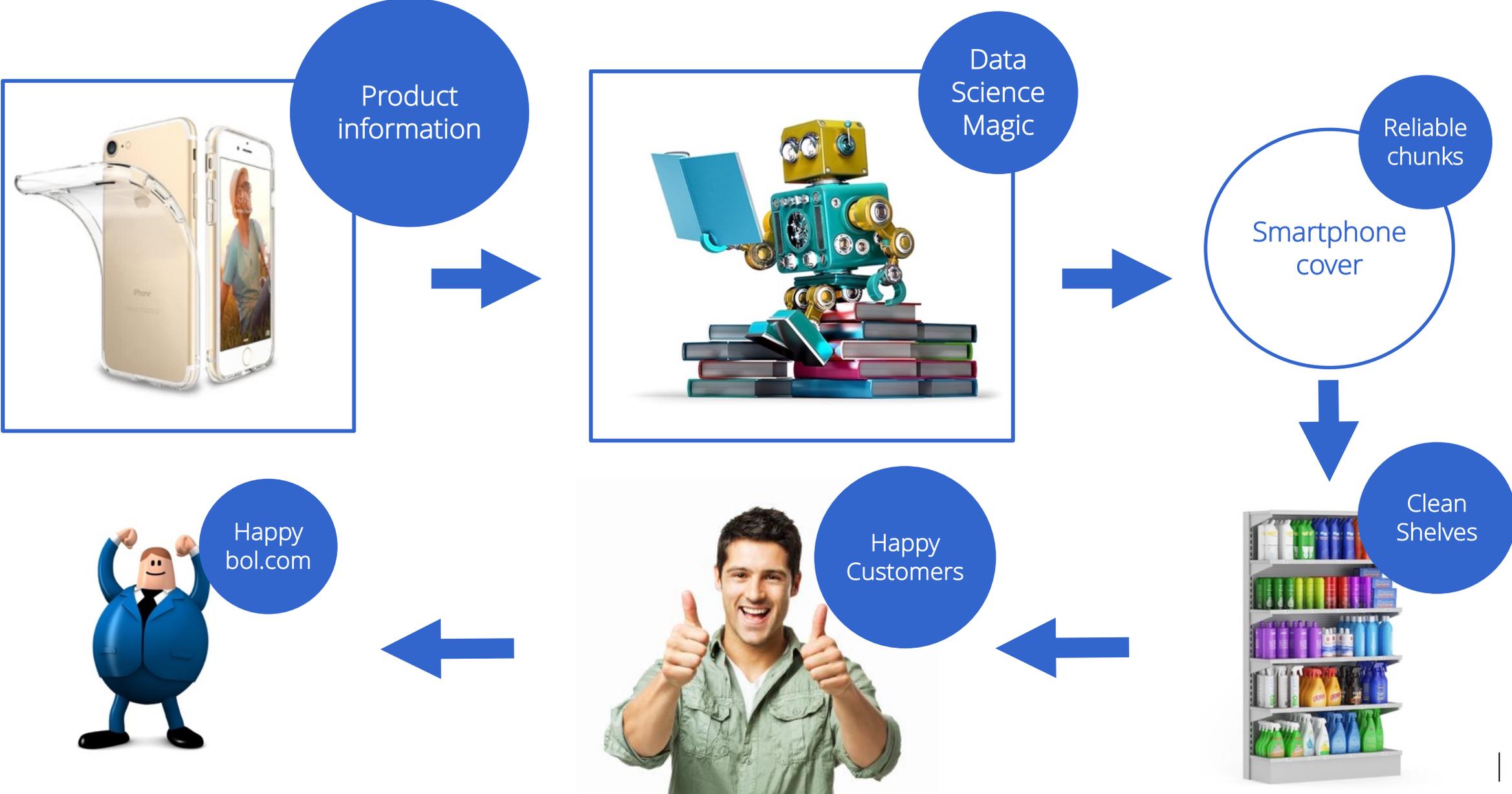
How to fix this mess?



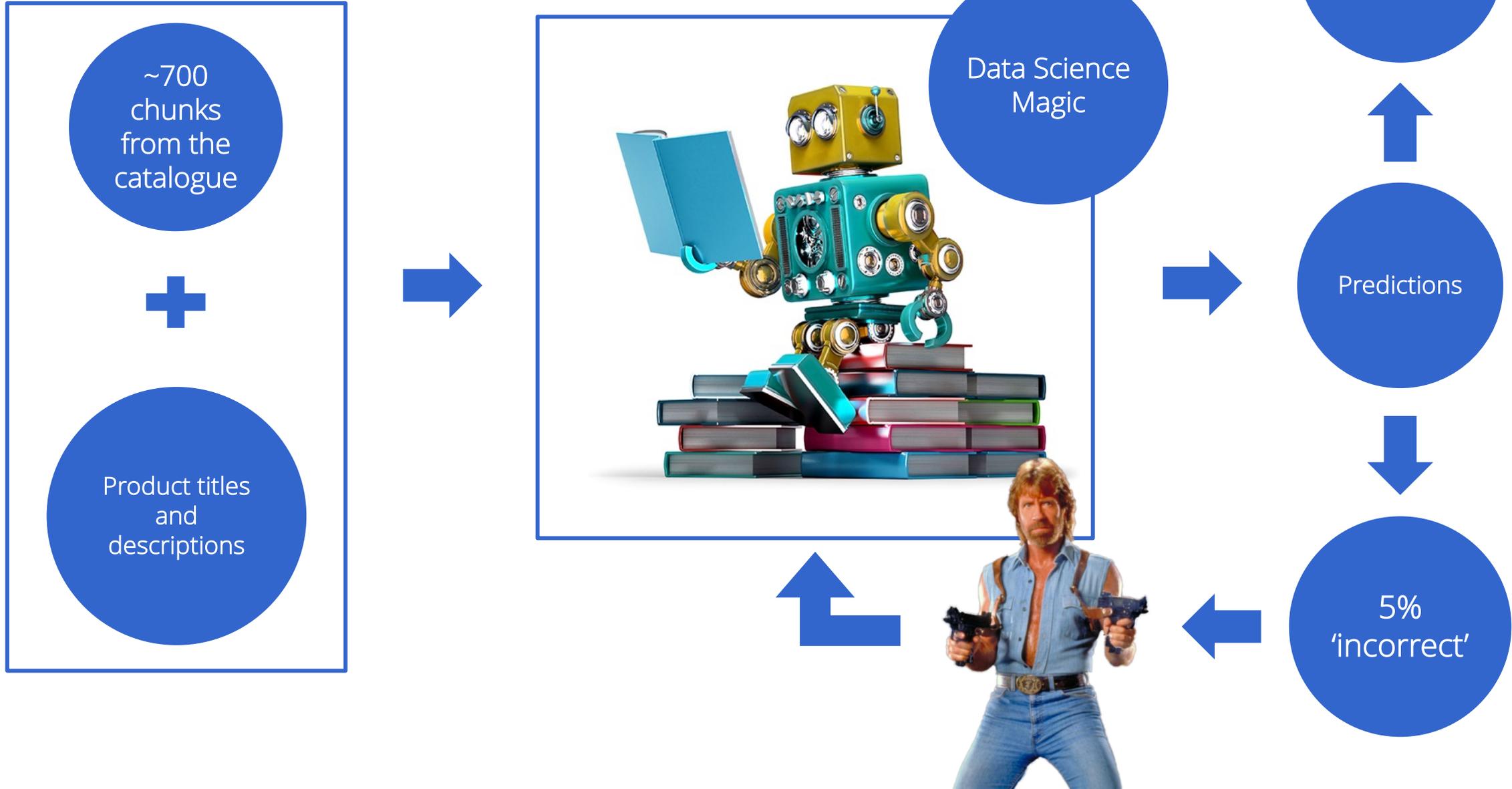
V.l.n.r. (een soort van): Aspergepan, Au-bain-mariepan, Braadpan, Cocotte, Flensjespan, Fonduepan, Grillpan, Hapjespan, Koekenpan, Kookpan, Mosselpan, Paellapan, Pannenset, Pastapan, Poffertjespan, Sauteerpan, Slowcookerpan, Snelkookpan, Soeppan, Steelpan, Stoompan, Vispan, Wokpan



Never a product on the wrong shelf with Data Science!



The machine: CRaPS



Validation: Introducing Chunk Norris



Does Valueline VLCP60220B20 USB-kabel belong on *Kabel*?

Product



Deze kabel is geschikt voor het verbinden van een apparaat met een computer

Global ID	9200000036146382
Predicted chunk	Kabel
Current chunk	Usb-kabel
Content or Image	Flag for review
Link	Check

Does this product belong on *Kabel*?

Accept

Decline

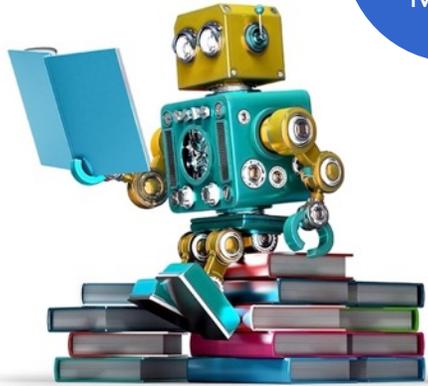
No clue



Polish Night: Data Science in action!



Data Science Magic



CHUNK norris

Does Valueline VLCP60220B20 USB-kabel belong on *Kabel*?

Product	Global ID 9200000036146382
 Deze kabel is geschikt voor het verbinden van een apparaat met een computer	Predicted chunk Kabel
	Current chunk Usb-kabel
	Content or Image Flag for review
	Link Check

Does this product belong on Kabel?

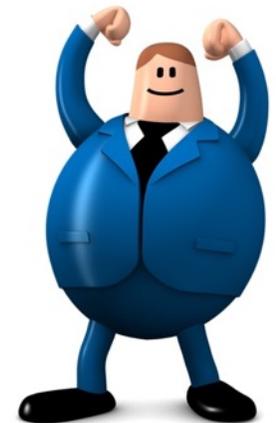
Awesome UI



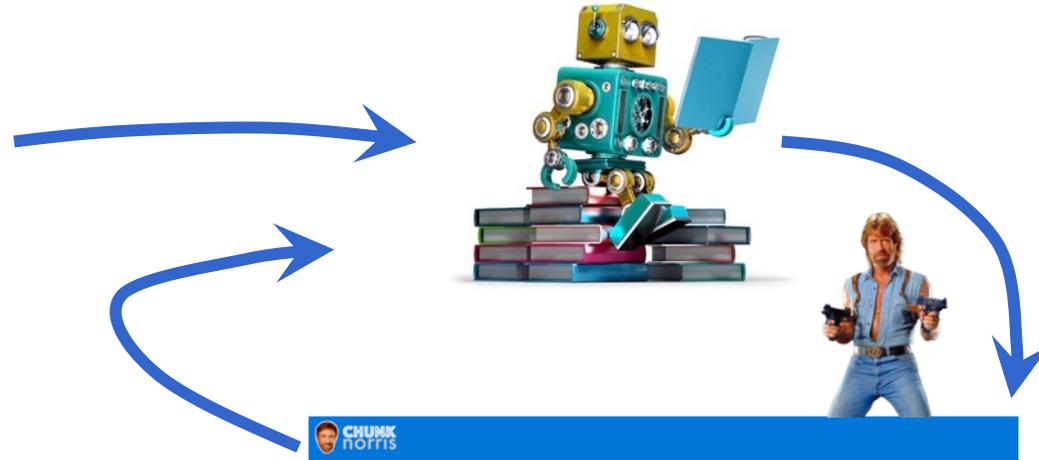
Thousands of products moved to the right chunk!



Awesome People



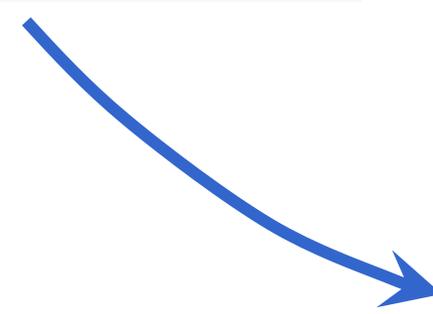
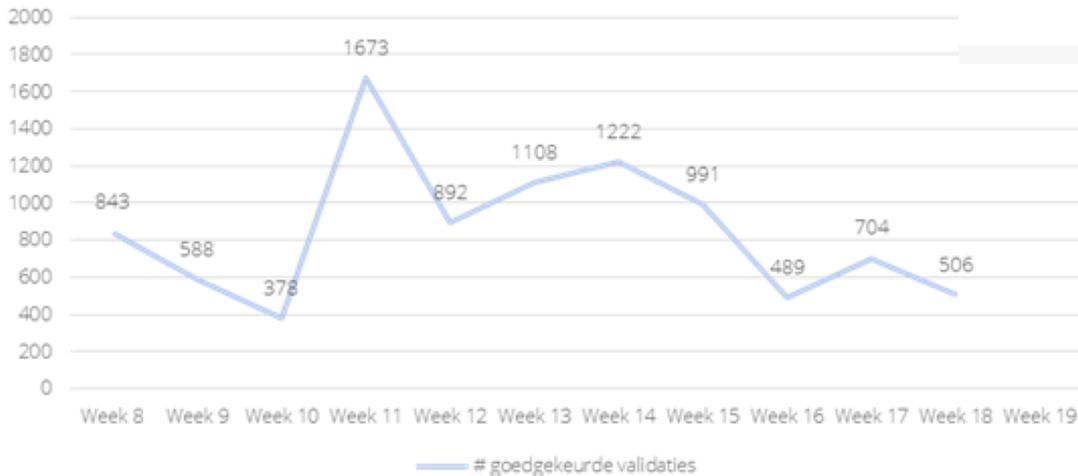
And the clean-up is ongoing!



Does Valuline VLCP60220B20 USB-kabel belong on Kabel?

Product	Global ID	920000036146382
 Deze kabel is geschikt voor het verbinden van een apparaat met een computer	Predicted chunk	Kabel
	Current chunk	Usb-kabel
	Content or image	Flag for review
	Link	Check

Does this product belong on Kabel?



Summary

Data Science offers a **scalable solution** for cleaning up the Bol.com product catalogue

~700 chunks

~70% of the catalogue

~70% of revenue

Already thousands of products have been put in the right chunk

Roughly 5% of products need to be checked

