BURBERRY

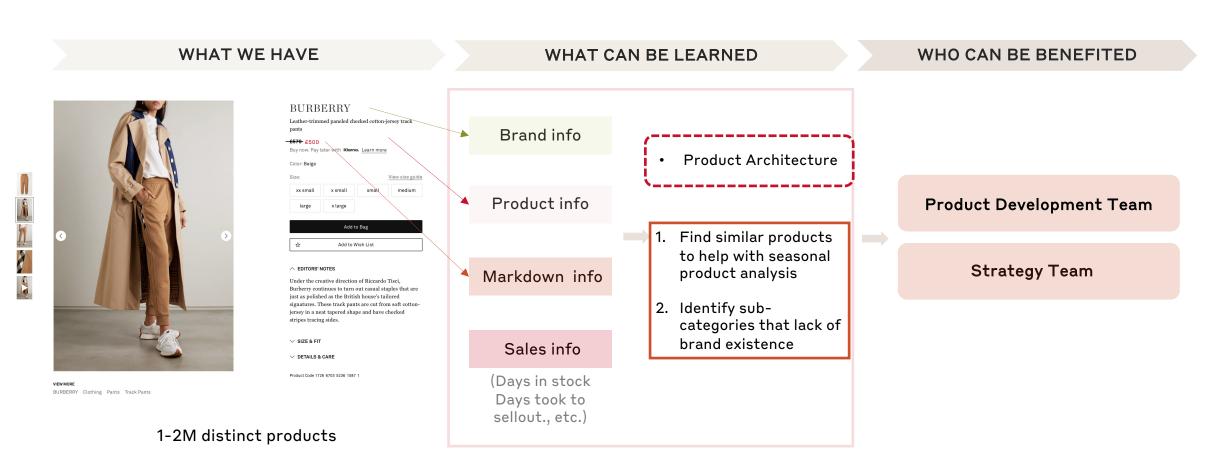
LONDON ENGLAND

TODAY'S AGENDA

1. PROJECT GOAL
2. PROBLEM IDENTIFIED AND SOLUTION PROPOSAL
3. PROJECT TIMELINE
4. DATASET

1. PROJECT GOAL

Extract market insights using web-scraped unstructured e-commerce listings historical datasets from EDITED



2. PROBLEM IDENTIFIED AND SOLUTION PROPOSAL

Product category prediction and similar image retrieval using weakly annotated e-commerce data

Problem

- Large dataset and some attributes are not partitioned correctly
- Use of unofficial product image and ID
- Inconsistent use of product ID
- Inconsistent product category across retail
- Lack of technology dedicated to e-commerce data

A Solution that

- Does not require excessive data cleaning
- · Not entirely image based
- Can incorporate all product information into the solution
- Can be used to predict product category
- Find similar product through image retrieval
- Is specialized in working with ecommerce data

Solution

Adopt a weakly supervised approach to build a visual feature extractor that was trained with fashion image and text embedding.

Then we can use the learned representation to to build a product category predicting model and image retrieval model.

We can use external Dataset

DeepFashion for category prediction
model's training and testing.

3. PROJECT TIMELINE

Project period: 03/05/2022 - 08/08/2022 (15 weeks in total)

Month 1 (03/05/2022 - 03/06/2022) - Research

```
Week 1 Until 16th May - Researching, paper reading and building datasets
```

Week 3 16th May - Start drafting the model

Week 5 3rd June - Finish first prototype using small, sampled data

Month 2 (03/06/2022 - 08/07/2022) - Implémentation

```
Week 6 6th - June - Test model on bigger dataset and a variety of architecture
```

Week 9 27th June - Thesis draft 6000-word submission date (for university review)

Week 11 11th July - Finalize Code

Month 3 (08/07/2022 - 08/08/2022) - Thesis Completion

Week 12-15 Thesis Writing

After Month 3

- Company report

4. DATASET

1 - 2 millions distinct products with 102 attributes since 2014 (681GB)

Brand info	Product info
brand	option_id
brand_slug	category
	cs_grp
	cs_subcategory
	gender
	colours
	care
	15 more

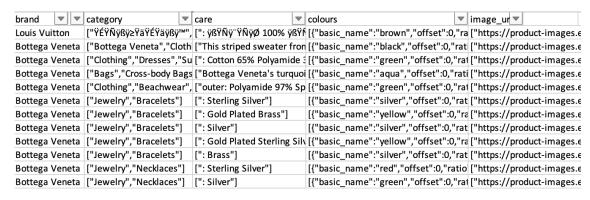
Bag of words vector for each image



Image vector

A small test set for training the feature extractor:

- 1. 50,00 data records
- 2. From Season aw18
- 3. From 5 selected brands (Burberry/ Bottega Venetta/...)
- 4. From the accessories department





```
pdf_temp_test['tokenized_text2'][5]

Out[20]: ['2ea9828a1ffa20e7e2ae8806d80e5a3ba294850f',
   'burberry',
   'burberry',
   'inspiration',
   'designers',
   'burberry',
   'accessories',
   'scarves',
   'boutiques',
   'workwear',
```