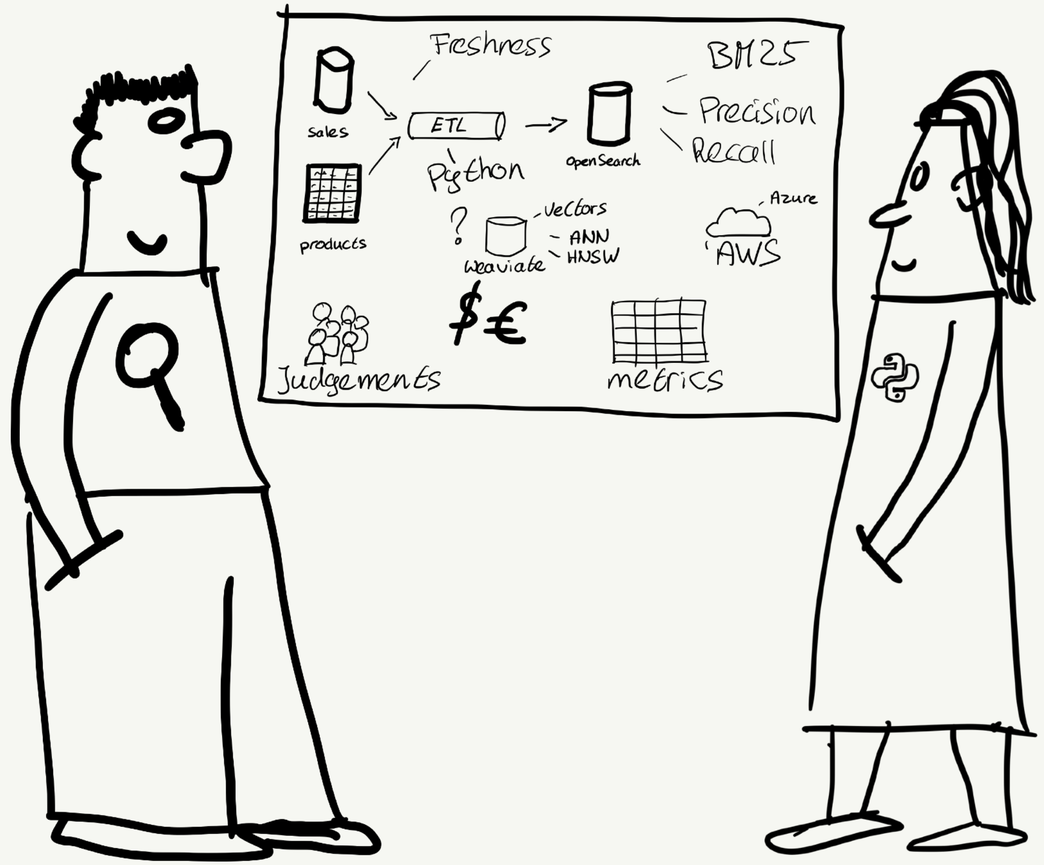
HAYSTACK 2023

Top 8 search topics to teach your team members

Jettro Coenradie

luminis.



A story about Search, by Frank and Amy

This story is fictional, any resemblance with real people you know are purely coincidental and definitely not on purpose.

Jettro Coenradie

Search Relevance guy @ Luminis

Live in the Netherlands

Father of 2 Kids (or Adults these days)

Love to be outside, ride my bike, walk, cycle

Like to learn and share knowledge

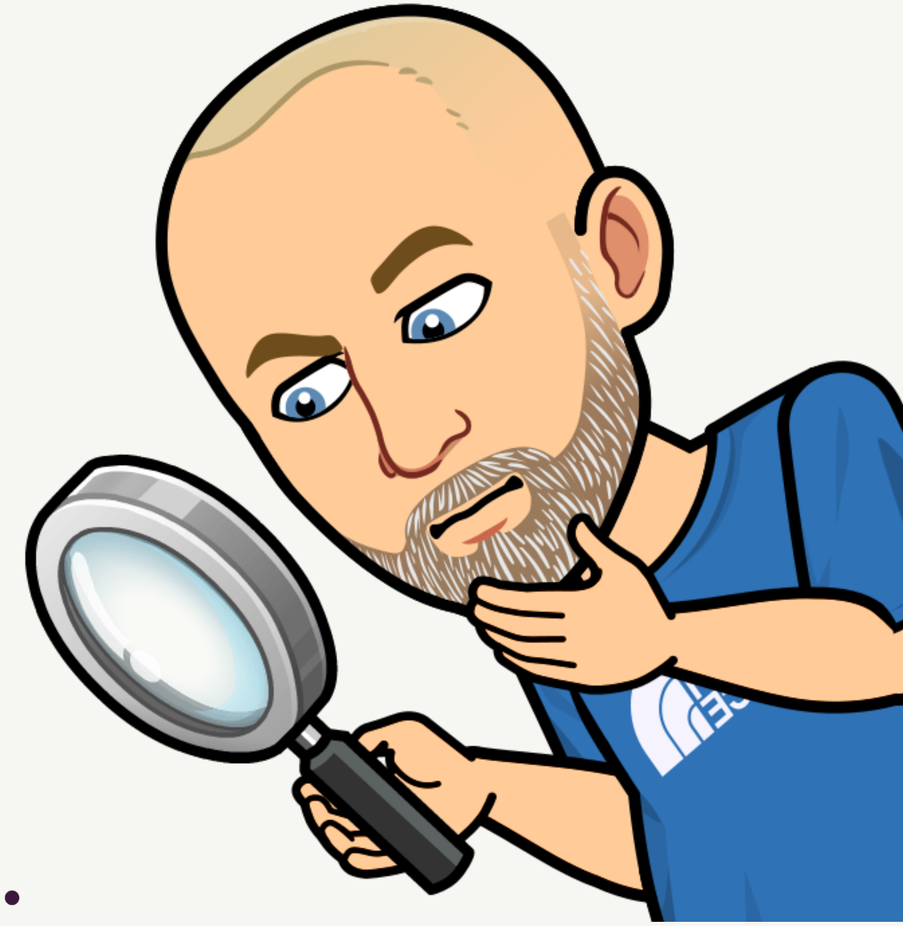
Talk to me if you want to know more about me

https://www.linkedin.com/in/jettro/

https://www.luminis.eu/expert/jettro-coenradie/

https://jettro.dev

luminis.



https://www.bitmoji.com/

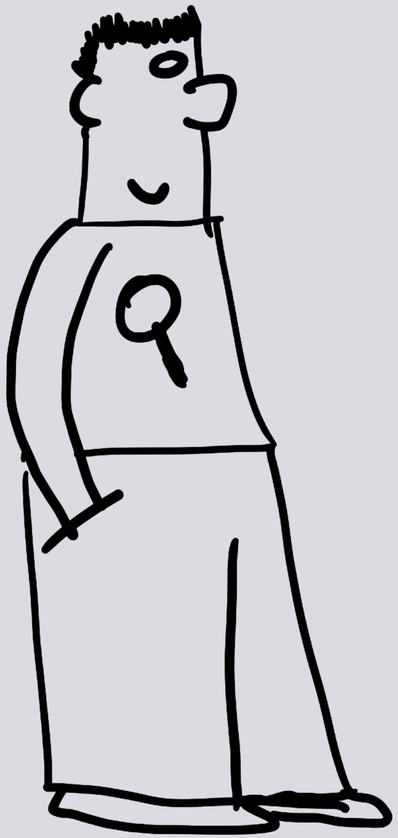
Amy

Senior Software Engineer / Architect

Has been working for the company for years.

Always on the lookout for new projects to improve her skills.

Likes to program in Python.



Frank

Seasoned Search Relevance Engineer

Started his career as a software engineer and grew into the search domain.

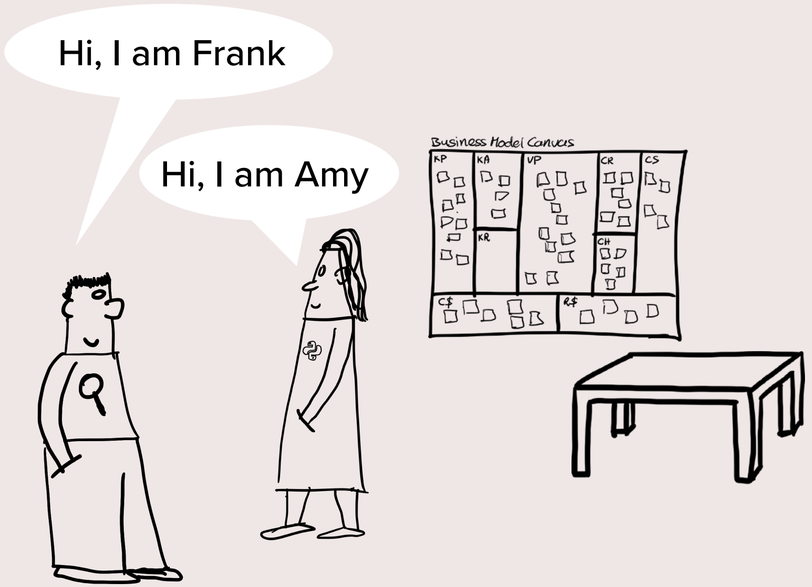
Joined the company only recently

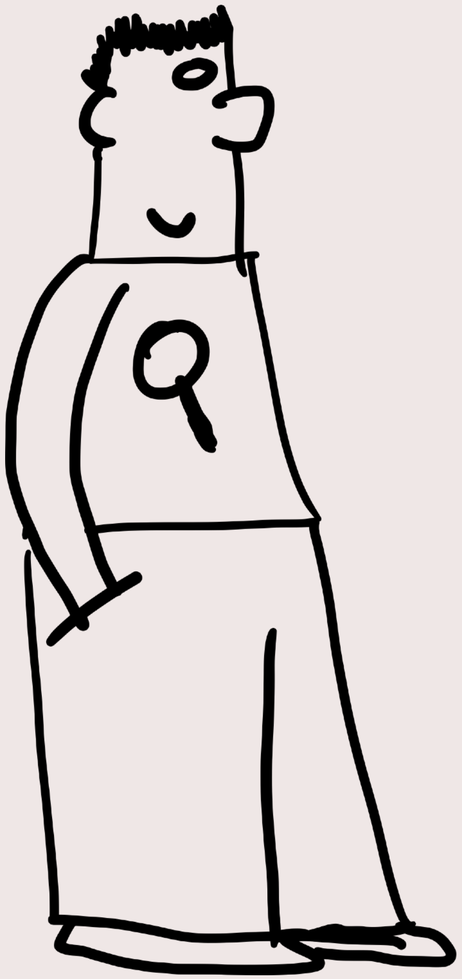
Feels that all users deserve the best search experience everywhere.

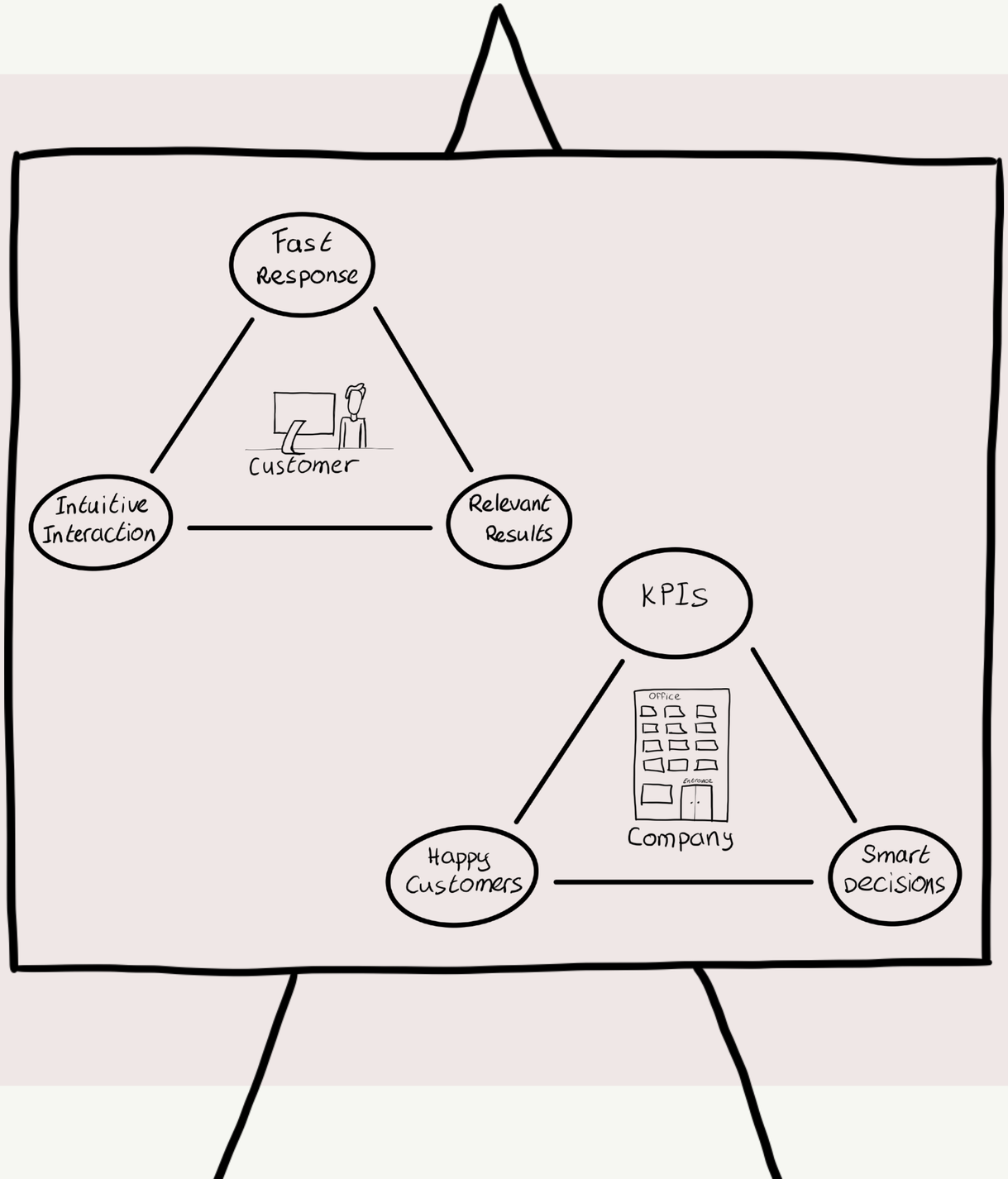
Frank walks to the office for the first time

Amy walks to the office, excited to start the new project.

Frank and Amy meet for the first time.






Fast Response
Customer
Intuitive Interaction
Relevant Results
KPIs
Office
Entrance
Company
Happy Customers
Smart decisions

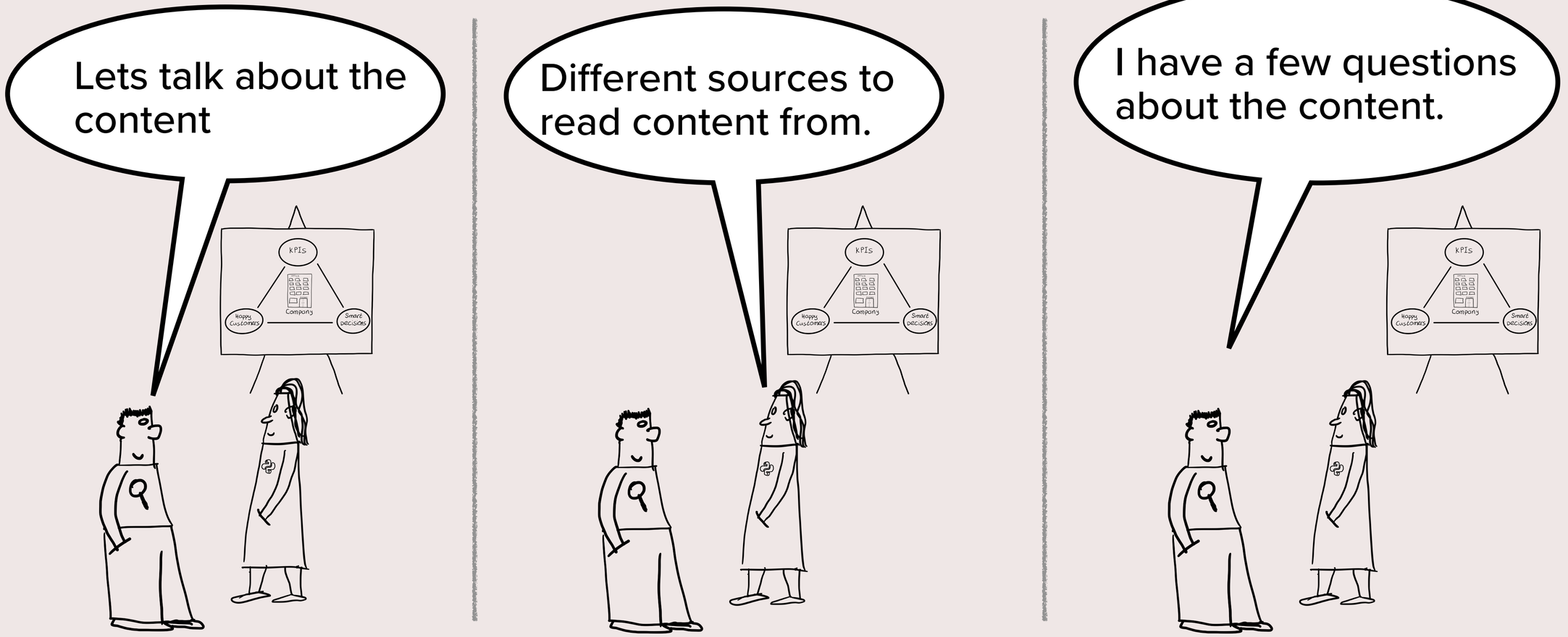
We need a bigger team to do all this

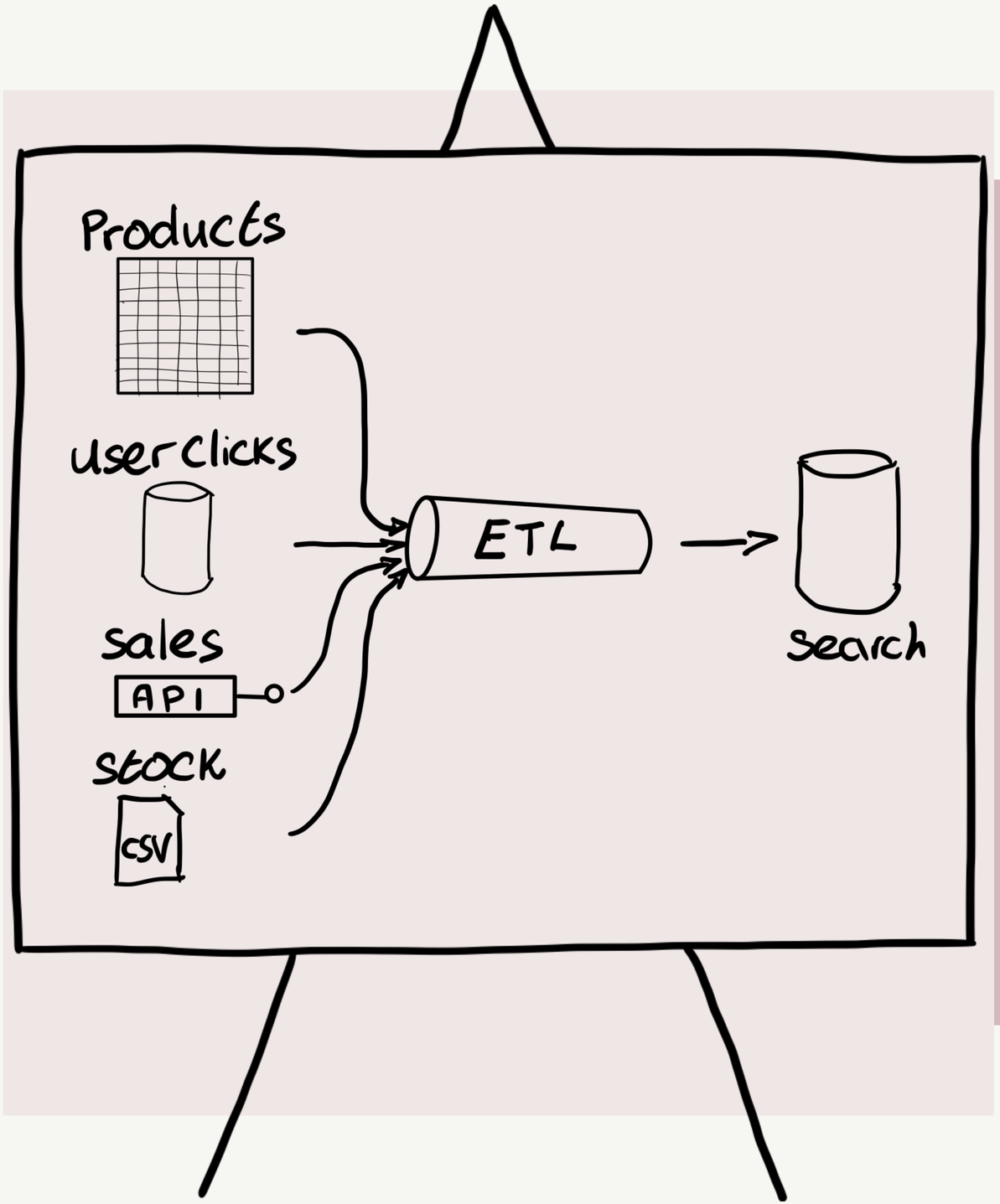
- Content
- User Experience
- Search Technology
- Analytics
- Business

- We have different stakeholders
- We need multiple disciplines in the team

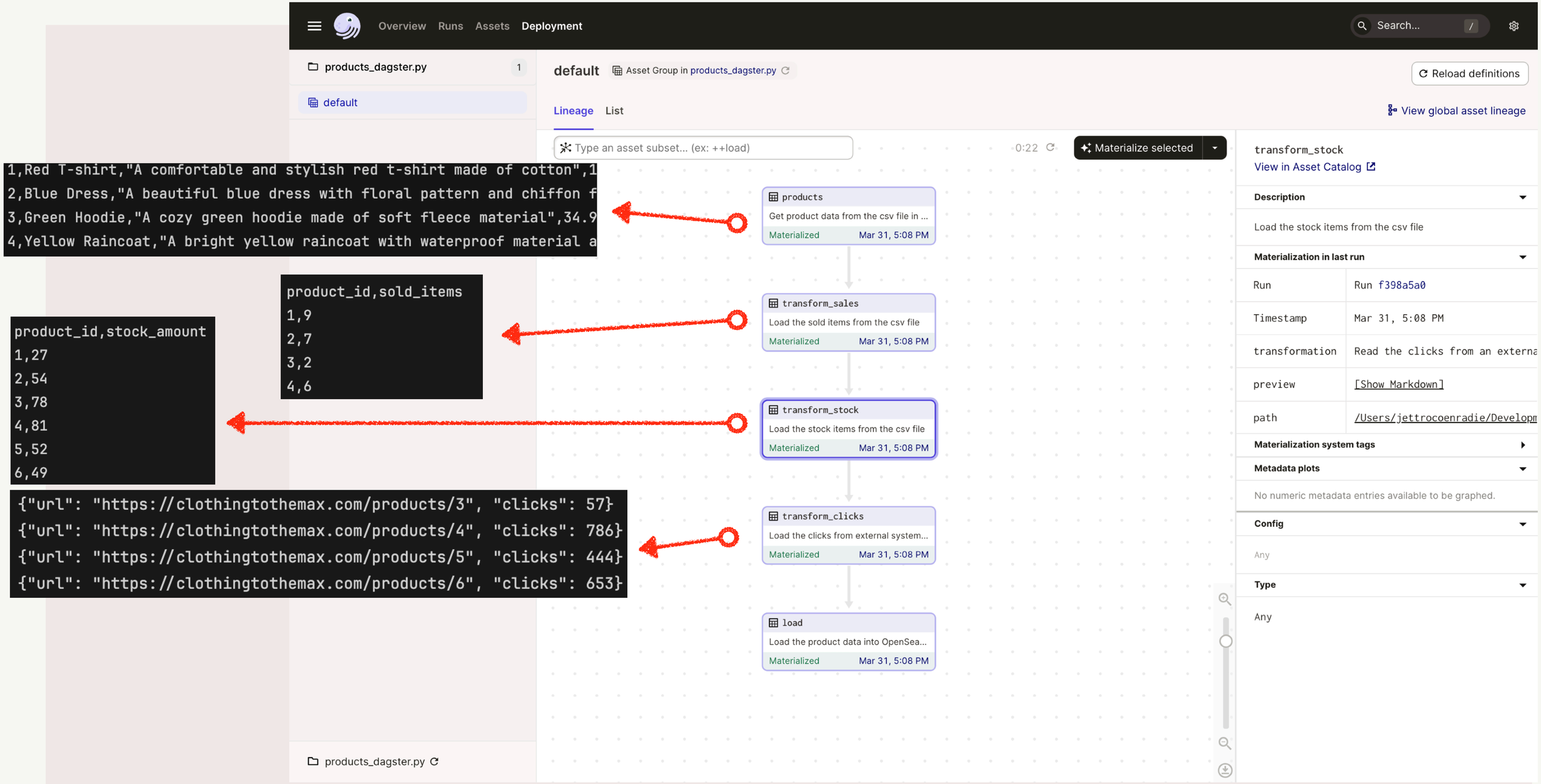
Search is more than Technology

Topic 1

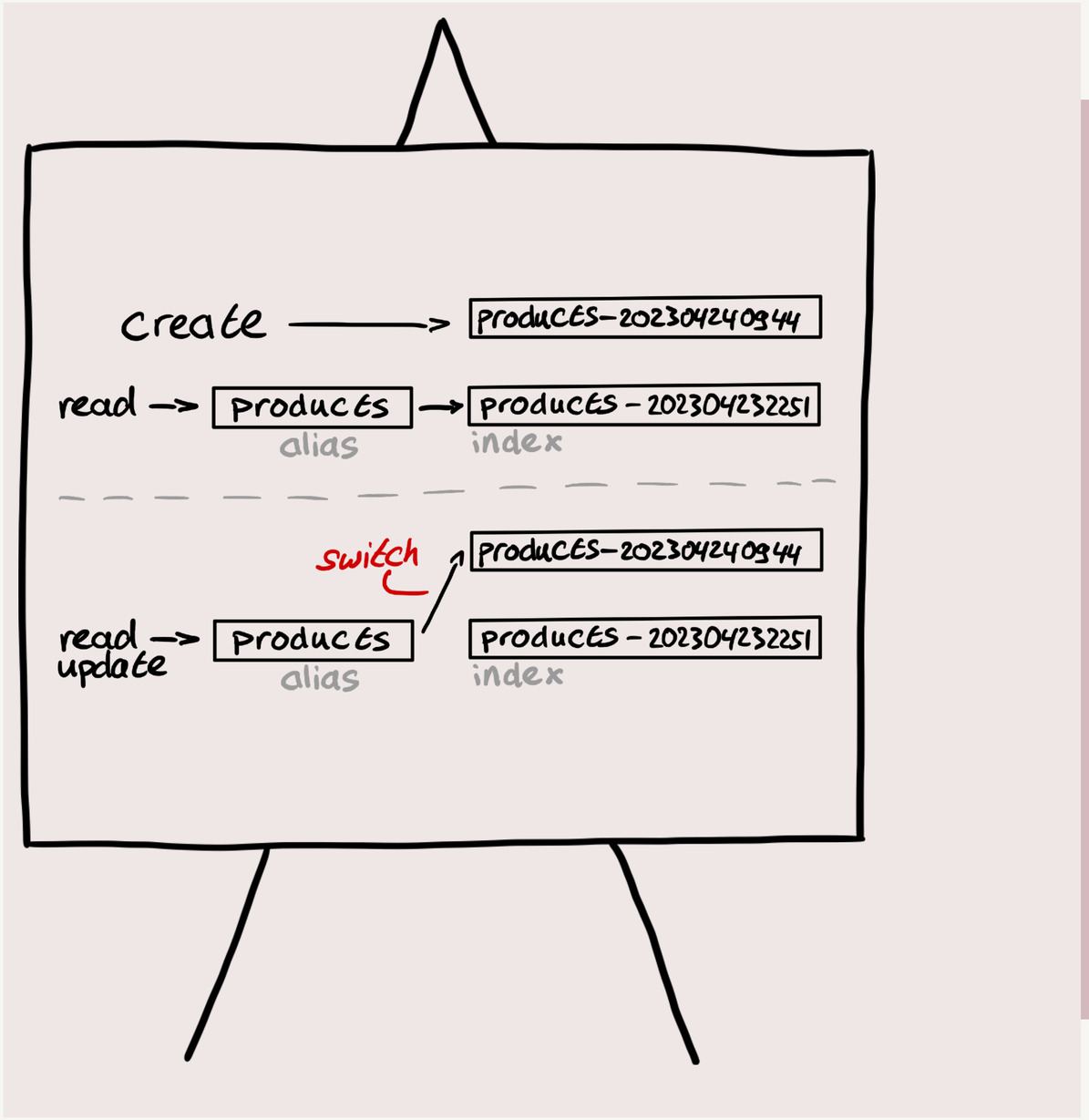

Lets talk about the content
KPIs
Happy customers
Company
Smart decisions
Different sources to read content from.
KPIs
Happy customers
Company
Smart decisions
I have a few questions about the content.
KPIs
Happy customers
Company
Smart decisions



- What does the content look like?
- Update Frequency of content?
- Full Import or updates?
- Quality of content?
- Who controls the content?
- Extract info from raw content?


Overview
Runs
Assets
Deployment
Search...
products_dagster.py
1
default
default
Asset Group in products_dagster.py
Reload definitions
Lineage
List
View global asset lineage
Type an asset subset... (ex: ++load)
0:22
Materialize selected
1,Red T-shirt,"A comfortable and stylish red t-shirt made of cotton",1
2,Blue Dress,"A beautiful blue dress with floral pattern and chiffon f
3,Green Hoodie,"A cozy green hoodie made of soft fleece material",34.9
4,Yellow Raincoat,"A bright yellow raincoat with waterproof material a
products
Get product data from the csv file in ...
Materialized
Mar 31, 5:08 PM
product_id,sold_items
1,9
2,7
3,2
4,6
transform_sales
Load the sold items from the csv file
Materialized
Mar 31, 5:08 PM
product_id,stock_amount
1,27
2,54
3,78
4,81
5,52
6,49
transform_stock
Load the stock items from the csv file
Materialized
Mar 31, 5:08 PM
{"url": "https://clothingtothemax.com/products/3", "clicks": 57}
{"url": "https://clothingtothemax.com/products/4", "clicks": 786}
{"url": "https://clothingtothemax.com/products/5", "clicks": 444}
{"url": "https://clothingtothemax.com/products/6", "clicks": 653}
transform_clicks
Load the clicks from external system...
Materialized
Mar 31, 5:08 PM
load
Load the product data into OpenSea...
Materialized
Mar 31, 5:08 PM
transform_stock
View in Asset Catalog
Description
Load the stock items from the csv file
Materialization in last run
Run
Run f398a5a0
Timestamp
Mar 31, 5:08 PM
transformation
Read the clicks from an externa
preview
[Show Markdown]
path
/Users/jettrocoenradie/Developm
Materialization system tags
Metadata plots
No numeric metadata entries available to be graphed.
Config
Any
Type
Any
products_dagster.py

```
{
  "id": 1,
  "name": "Red T-shirt",
  "description": "A comfortable and stylish red t-shirt made of cotton",
  "price": 15.99,
  "category": "T-shirts",
  "color": "Red",
  "create_date": "2022-01-01",
  "sold_items": 9,
  "stock_amount": 27,
  "num_clicks": 325
}
```

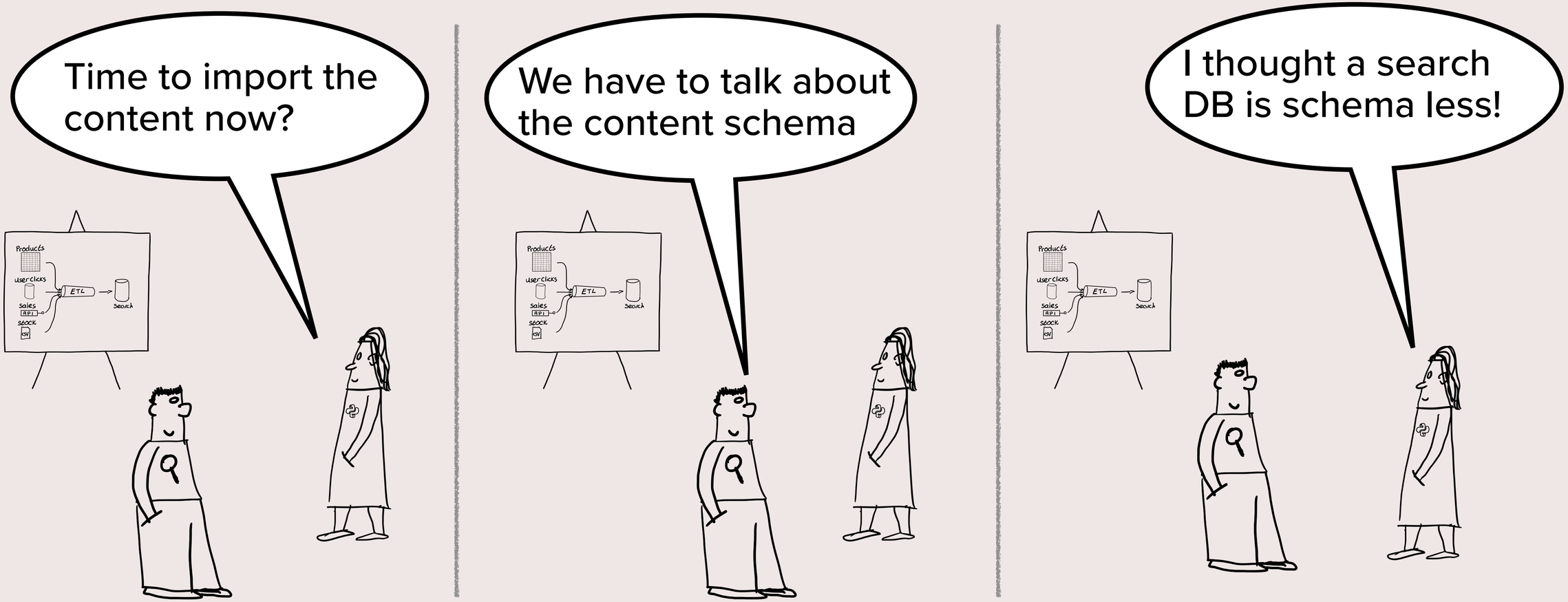

create
products-202304240344
read
products
alias
products-202304232251
index
switch
products-202304240344
read
update
products
alias
products-202304232251
index


Create a new index
without downtime.

- What your content looks like
- Frequency of updates and full import
- Sources and how to combine them

Understand you content

Topic 2


Time to import the content now?
Products
user clicks
sales
API
stock
SQL
ETL
Search
We have to talk about the content schema
Products
user clicks
sales
API
stock
SQL
ETL
Search
I thought a search DB is schema less!
Products
user clicks
sales
API
stock
SQL
ETL
Search

Match, Filter, Aggregation

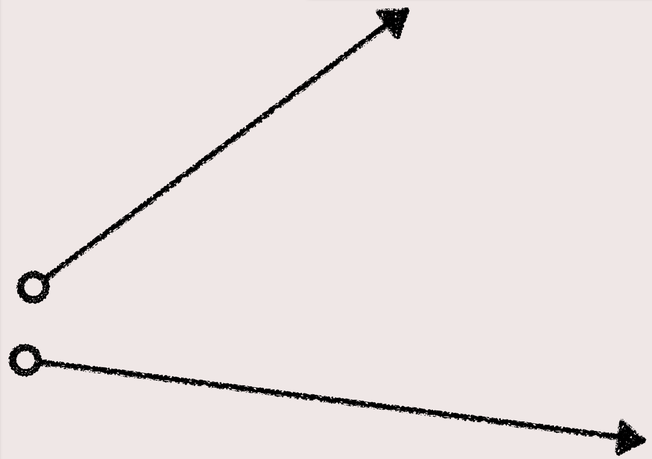
Show the content

Dynamic Mapping

Explicit field mapping

Multi-purpose field mapping

```
{
  "index_patterns": ["products-*"],
  "priority": 10,
  "template": {
    "mappings": {
      "properties": {
        "category": {
          "type": "keyword"
        },
        "description": {
          "type": "text"
        }
      }
    }
  },
  "composed_of": [
    "jc_settings",
    "jc_dynamic_template"
  ],
  "version": 2
}
```



```
{
  "template": {
    "settings": {
      "number_of_shards": 1,
      "number_of_replicas": 0
    }
  },
  "version": 1
}
```

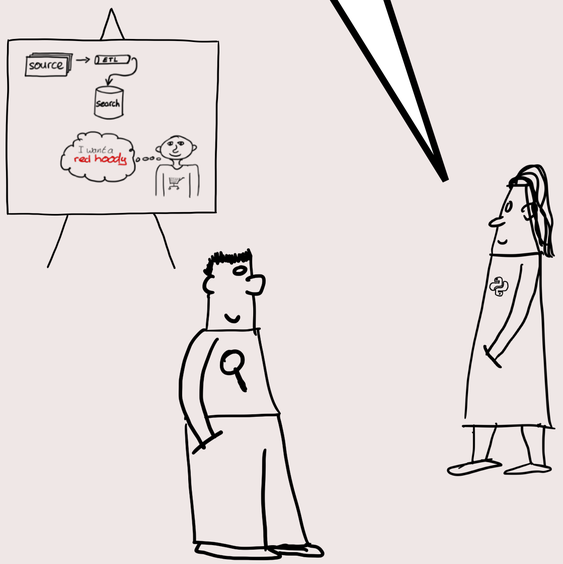
```
{
  "template": {
    "mappings": {
      "dynamic_templates": [
        {
          "strings": {
            "match_mapping_type": "string",
            "mapping": {
              "type": "text",
              "fields": {
                "keyword": {
                  "type": "keyword",
                  "ignore_above": 256
                }
              }
            }
          }
        }
      ]
    }
  },
  "version": 1
}
```

- Add a mapping for fields based on usage patterns
- Use dynamic, explicit and multi-purpose fields

Use a schema for your content

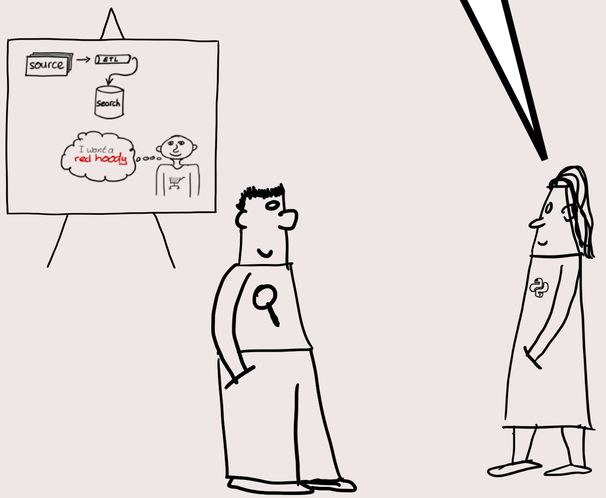
Topic 3

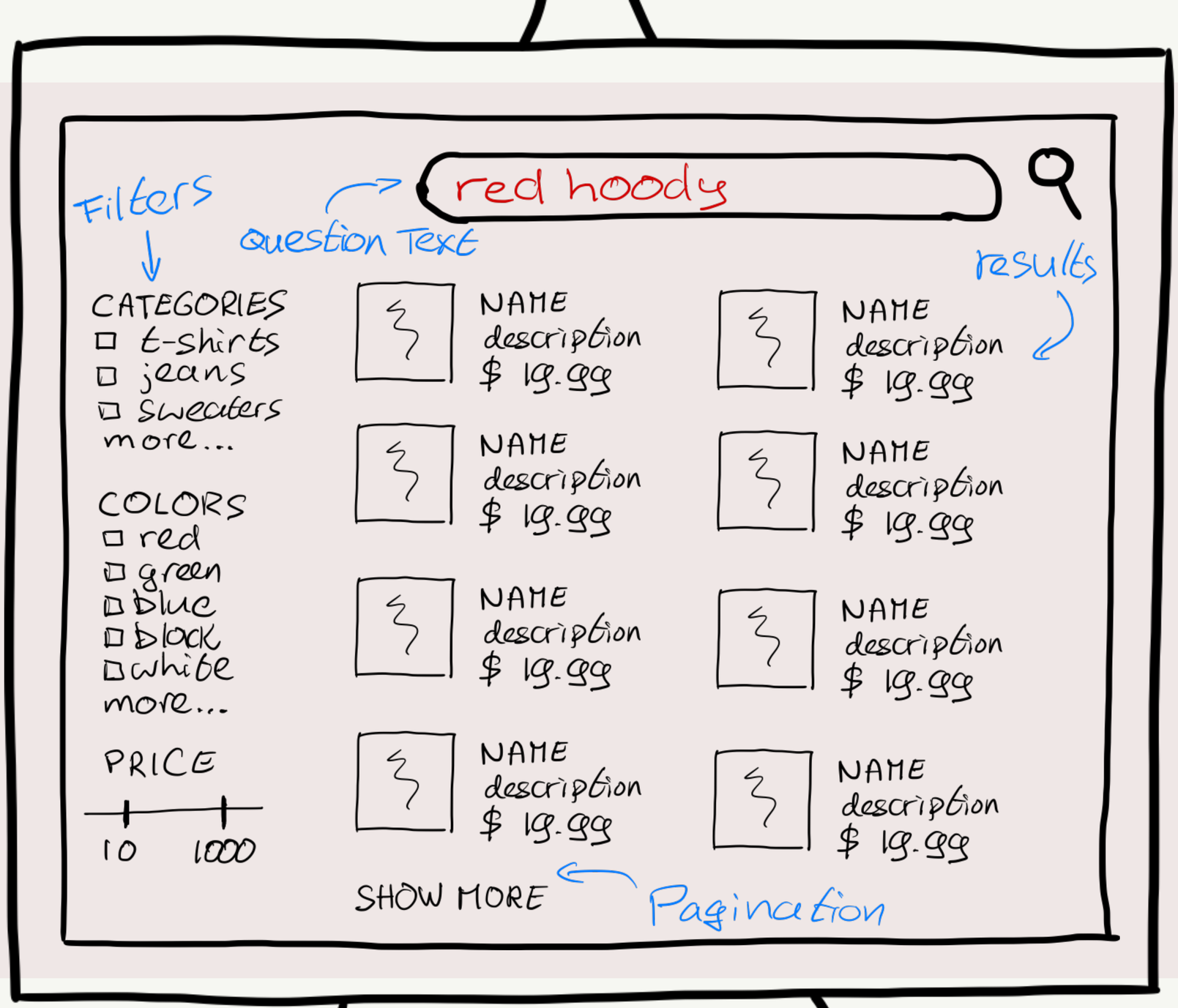
How can customers find products?



We transform a question into a query

Search, query, question, what do you mean?



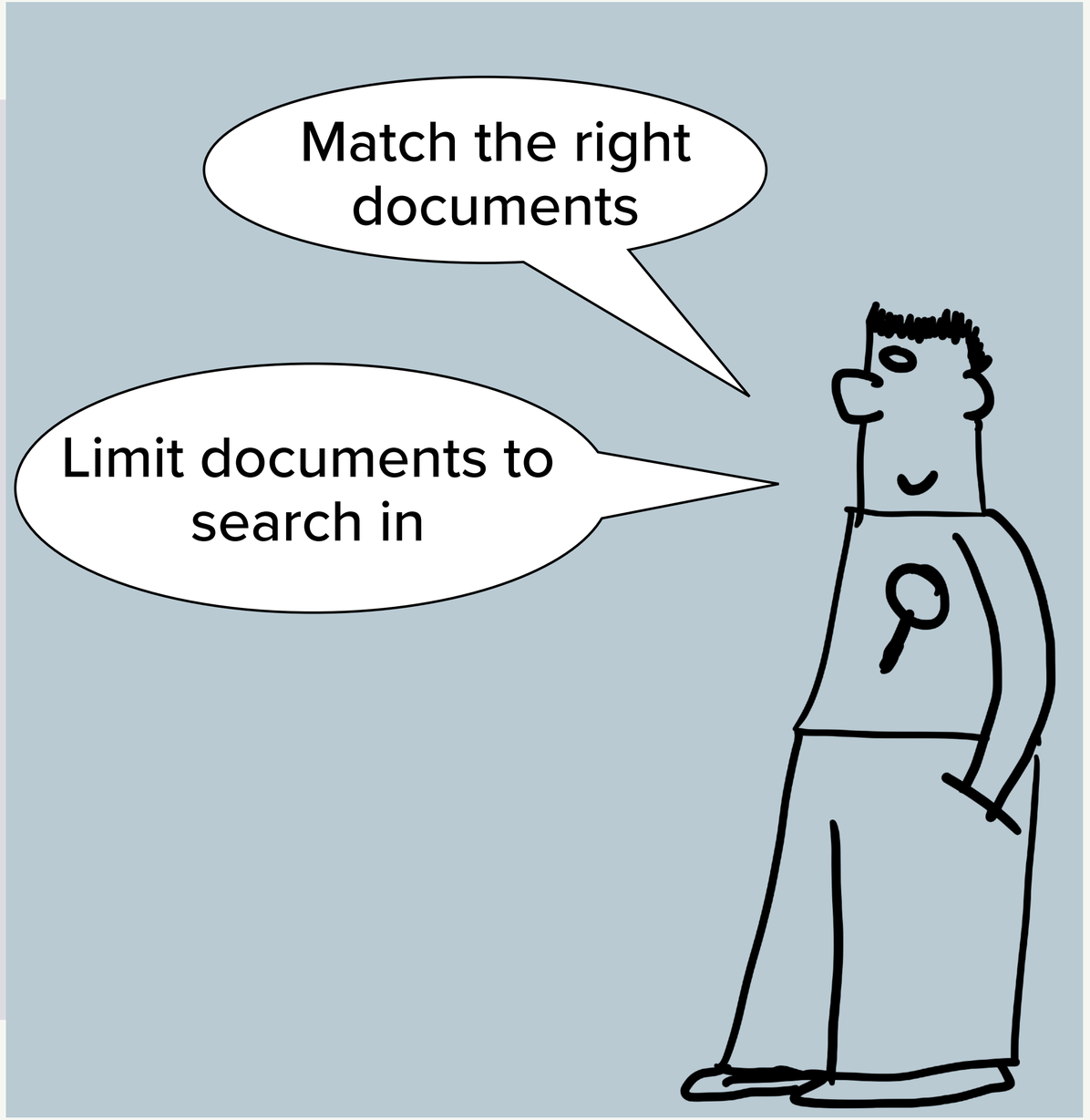

red hoody
Filters
question Text
results
CATEGORIES
t-shirts
jeans
sweaters
more...
COLORS
red
green
blue
black
white
more...
PRICE
10
1000
NAME
description
$ 19.99
SHOW MORE
Pagination

Match, Phrase, Multi-match

Bool, must, filter

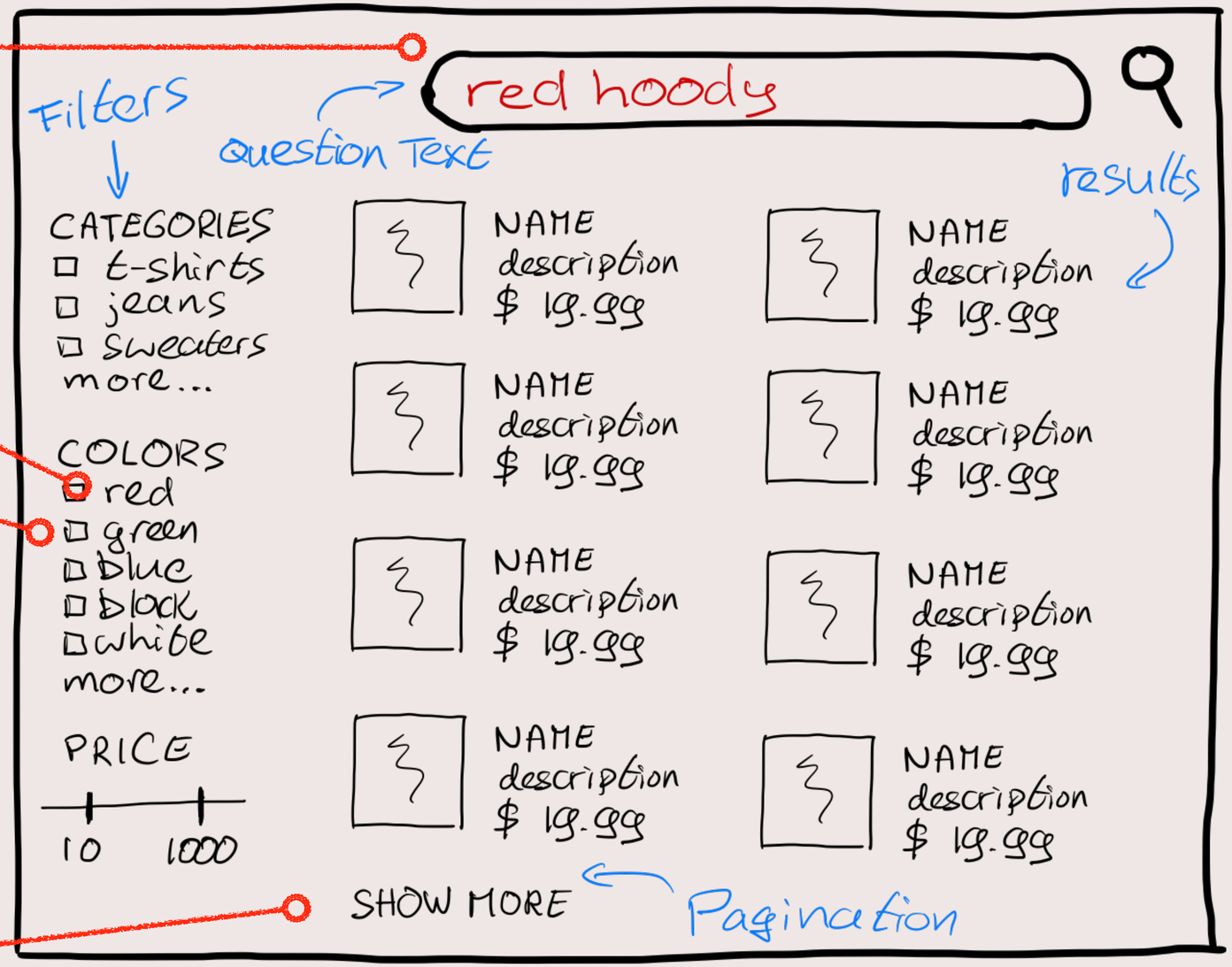
Term, Range

Terms, Range aggregations



```
GET products/_search
{
  "query": {
    "bool": {
      "must": [
        {
          "multi_match": {
            "query": "red hoody",
            "fields": ["name", "description"],
            "type": "cross_fields",
            "operator": "or"
          }
        }
      ],
      "filter": [
        {
          "term": {
            "color.keyword": "red"
          }
        }
        ]
    }
  },
  "aggs": {
    "by_category": {
      "terms": {
        "field": "category",
        "size": 10
      }
    },
    "by_color": {
      "terms": {
        "field": "color.keyword",
        "size": 10
      }
    },
    "by_price": {
      "range": {
        "field": "price",
        "ranges": [
          {
            "from": 0,
            "to": 50
          },
          {
            "from": 50,
            "to": 100
          },
          {
            "from": 100
          }
        ]
      }
    }

  },
  "from": 0,
  "size": 10
}
```



- A users question contains text and filters
- Response contains results plus help to make question more specific

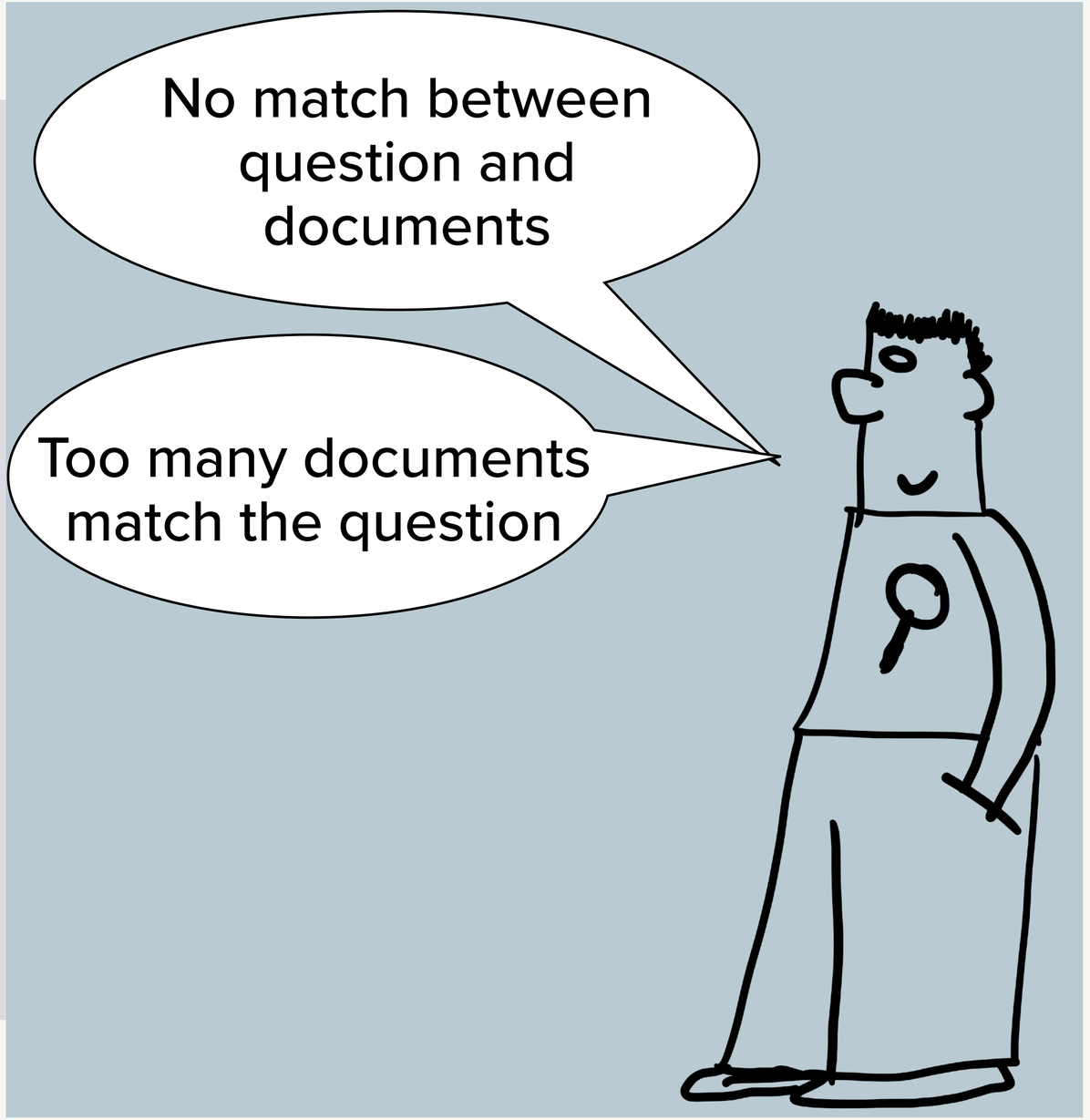
---

Transform a question into a query

Topic 4

- What if user makes a typo?
- What if user uses other words?
- What if we have to many results?
- What if one category is more important?
- What if some products are more popular?

- Synonyms, stemming
- Fuzzy, ngrams
- LLM, Vector Search
- Boosting / LTR



- What if a question does not match results?
- What if a question has to many results?

Relax matching and use boosting

Topic 5

- Abbreviations
- Opening times
- Return policies
- Narrow vs Wide

jeans

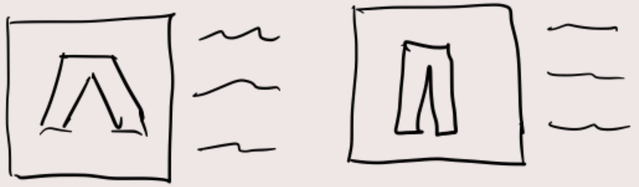
CATEGORIES
- straight
- slim
- relaxed
more...

COLORS
- black
- blue
- dark wash
- light wash
- grey
more...

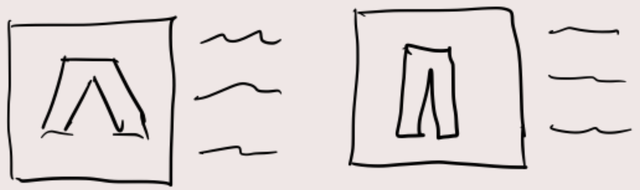
PRICE

10 200

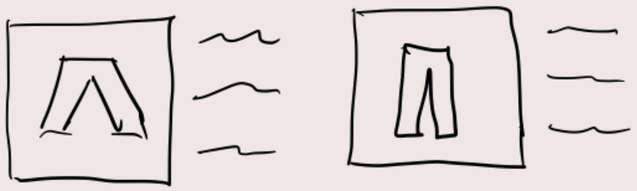
NEW



POPULAR



SALE



SIZE

WAIST

23 26 27 28
29 30 31 32
33 34 35 36
38 40 42 44
46 48 50 52

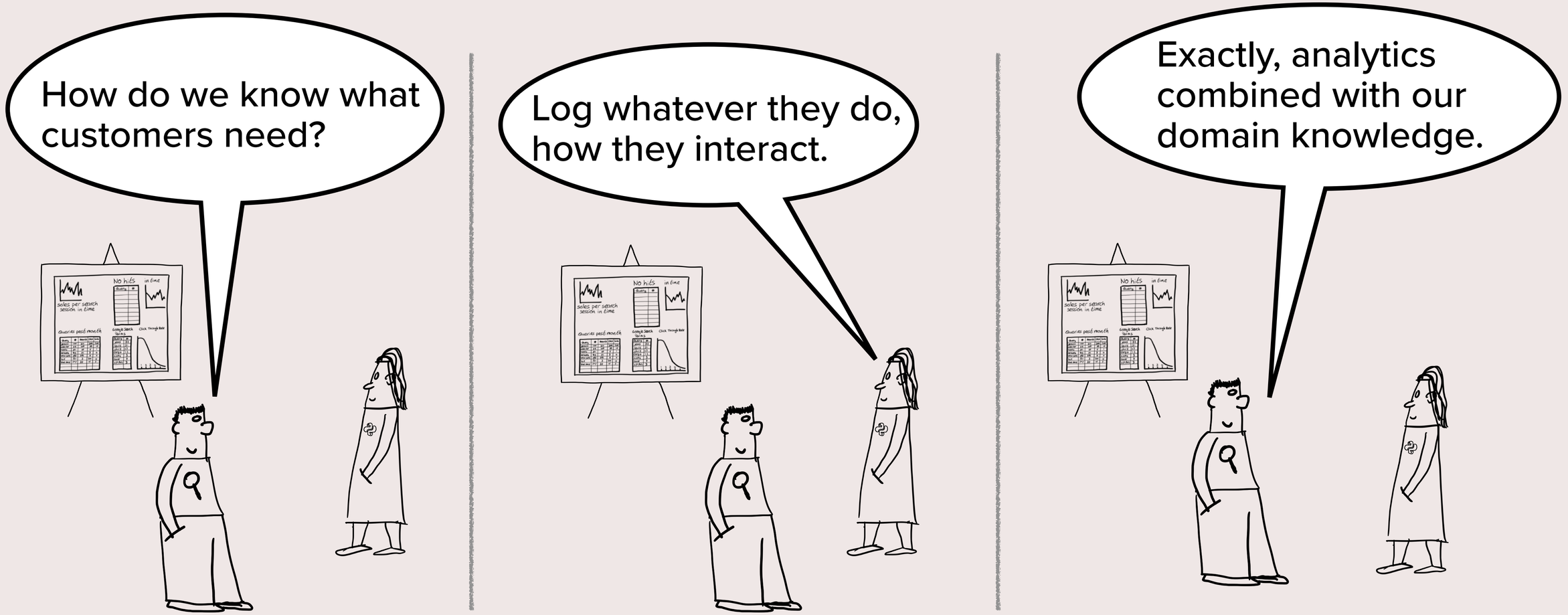
LENGTH

28 29 30 32
34 36 38 40

- **Not all questions point to a single or set of results (products in our case)**

When search is not the answer

Topic 6


How do we know what customers need?
Log whatever they do, how they interact.
Exactly, analytics combined with our domain knowledge.

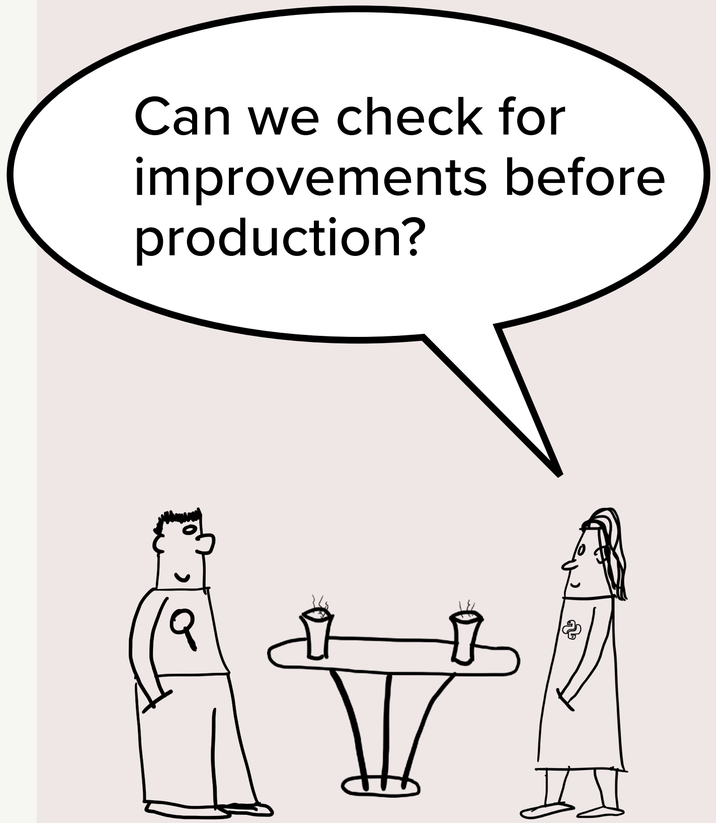
- Business Effect of Search
- Search terms used
- Clicks
- Link with Google Analytics
- Click Through Rate

- Create a search session id
- User search terms with results
- Clicks with position in result for search terms
- Sales after a search

- Think about analytics when building your search solution
- Have metrics to support KPI’s in a dashboard

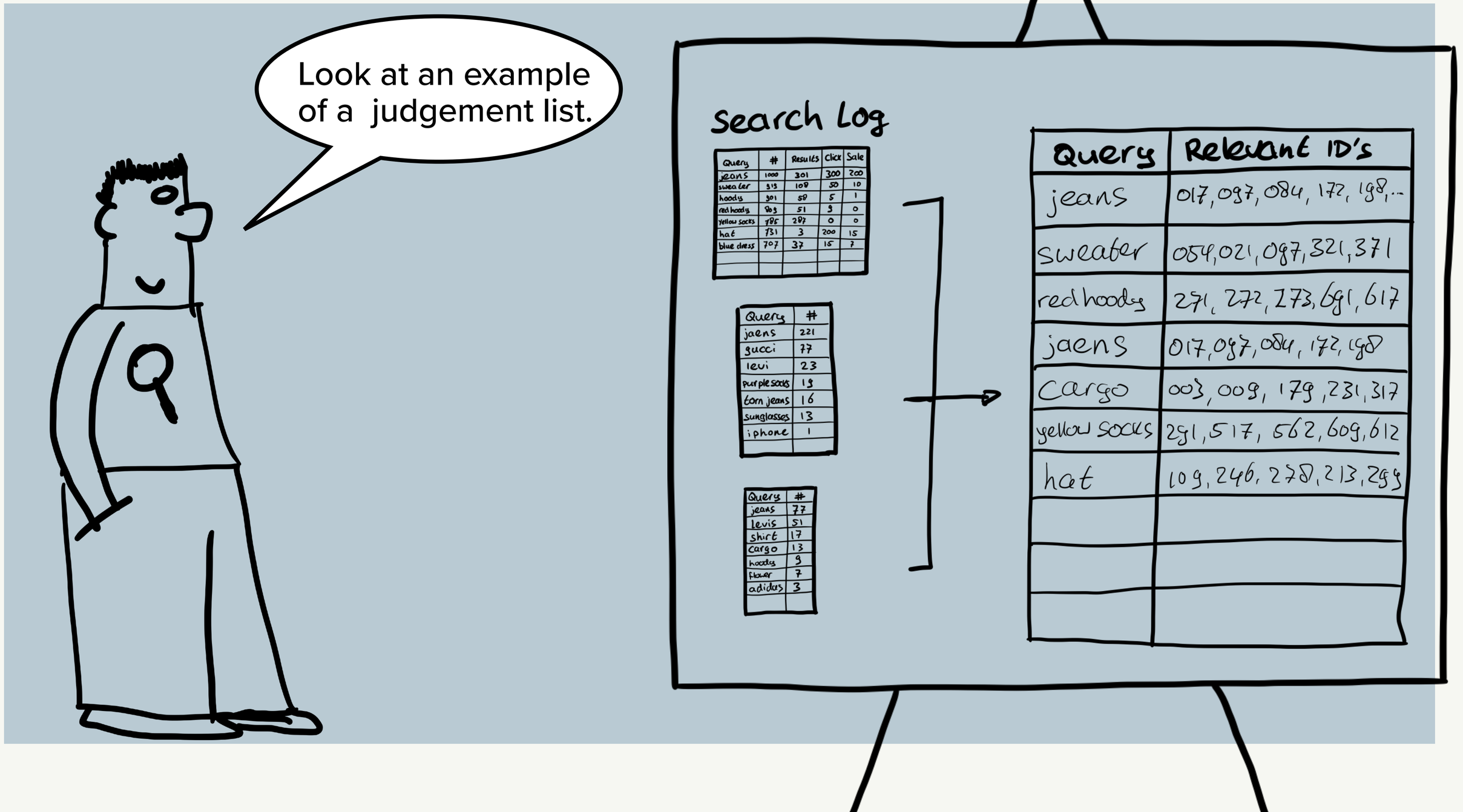
Log visitor activities and learn from them

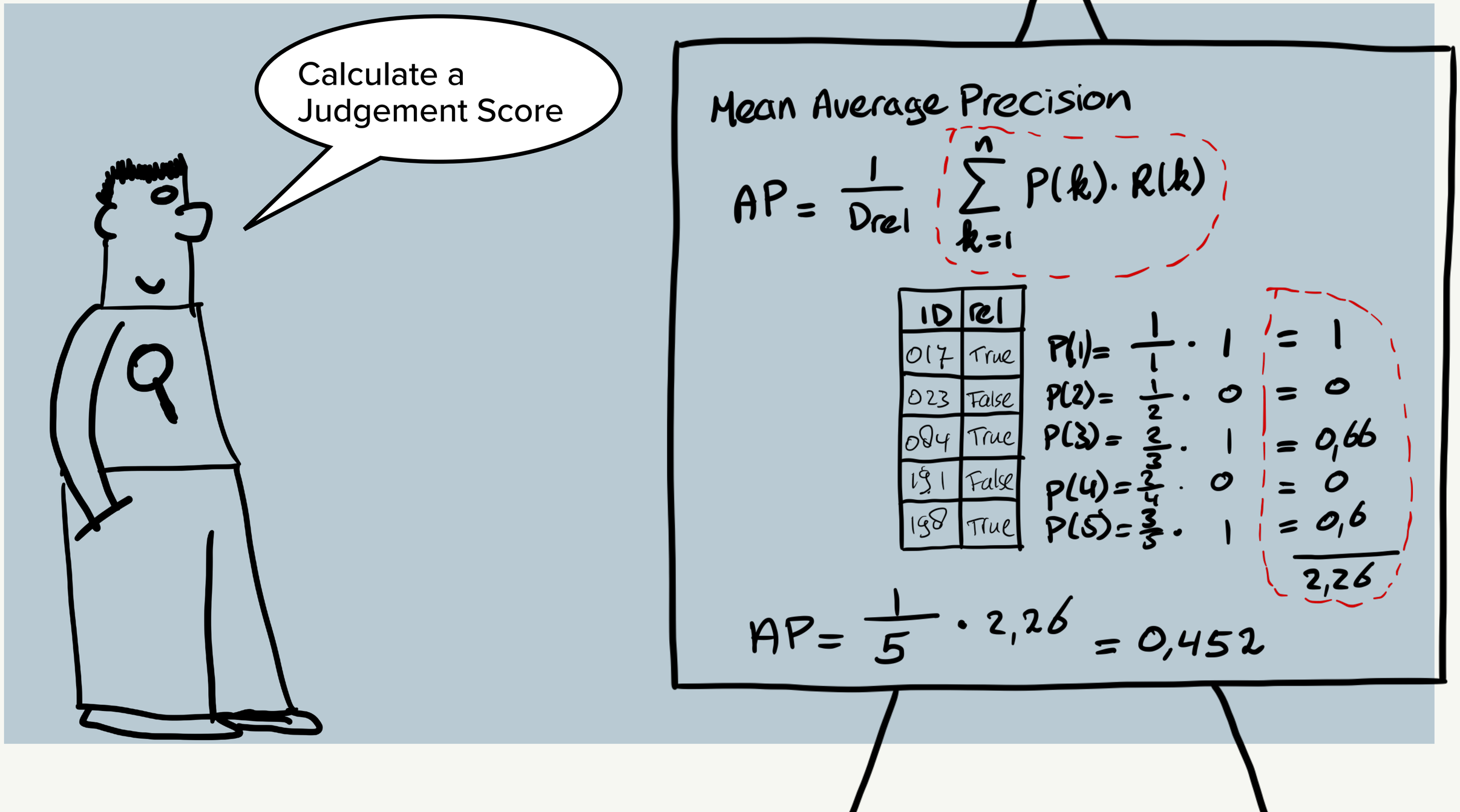
Topic 7

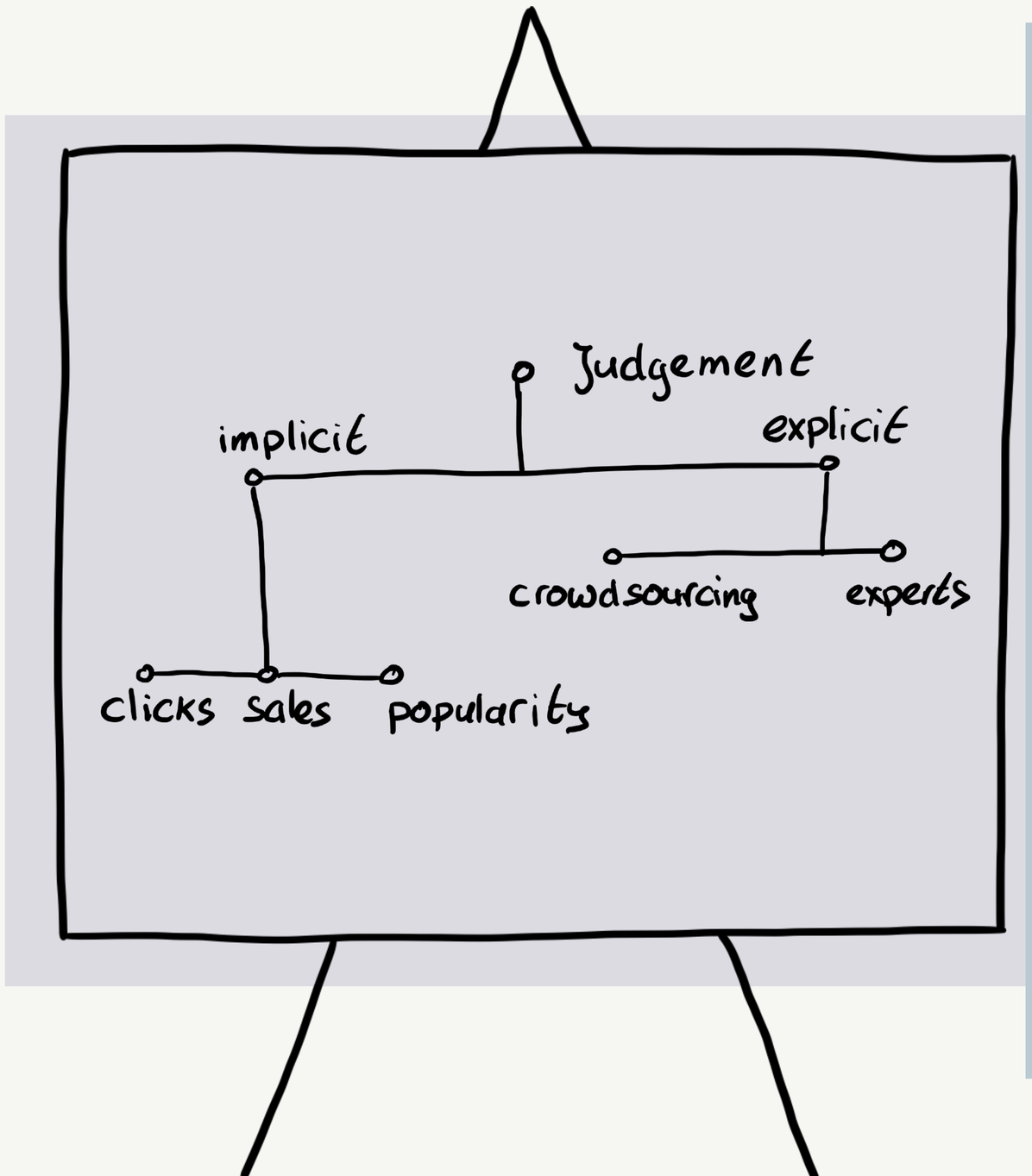

Can we check for improvements before production?


I like your thinking, yes we can do that.


We need a judgement list and verify our results against the list.
?


Look at an example of a judgement list.
Search Log
Query | # | Results | Click | Sale
jeans | 1000 | 301 | 300 | 200
sweater | 313 | 108 | 50 | 10
hoody | 301 | 58 | 5 | 1
red hoody | 203 | 51 | 3 | 0
yellow socks | 785 | 287 | 0 | 0
hat | 731 | 3 | 200 | 15
blue dress | 707 | 37 | 15 | 7
Query | #
jaens | 221
gucci | 77
levi | 23
purple socks | 19
torn jeans | 16
sunglasses | 13
iphone | 1
Query | #
jeans | 77
levis | 51
shirt | 17
cargo | 13
hoody | 9
flower | 7
adidas | 3
Query | Relevant ID's
jeans | 017,097,084,172,198,...
sweater | 054,021,097,321,371
red hoody | 271,272,273,691,617
jaens | 017,097,084,172,198
cargo | 003,009,179,231,317
yellow socks | 291,517,562,609,612
hat | 109,246,278,213,299


Calculate a Judgement Score
Mean Average Precision
$AP = \frac{1}{Drel} \sum_{k=1}^{n} P(k) \cdot R(k)$
ID | rel
017 | True
023 | False
084 | True
191 | False
198 | True
$P(1) = \frac{1}{1} \cdot 1 = 1$
$P(2) = \frac{1}{2} \cdot 0 = 0$
$P(3) = \frac{2}{3} \cdot 1 = 0{,}66$
$P(4) = \frac{2}{4} \cdot 0 = 0$
$P(5) = \frac{3}{5} \cdot 1 = 0{,}6$
2,26
$AP = \frac{1}{5} \cdot 2{,}26 = 0{,}452$


Judgement
implicit
explicit
crowdsourcing
experts
clicks
sales
popularity


Where do these judgements come from?

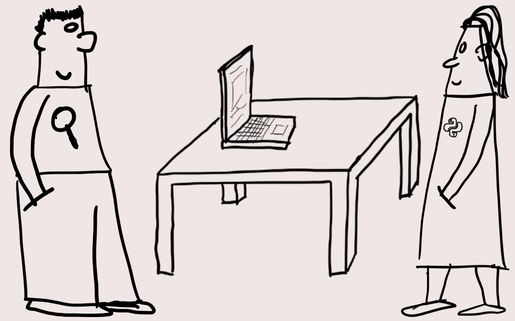
- Obtain judgements from experts or stats
- Embed calculating offline scores in your process

Use a judgement list to verify results

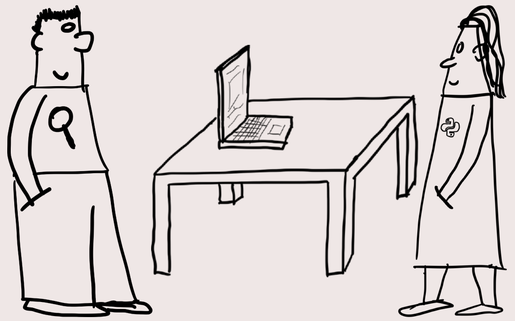
Topic 8

I implemented a search service. I can change the query and enter search terms.

Good job, good way to start testing.



What if the results are not as expected?



Data Overview - OpenSearch

localhost:5601/app/dashboards#/view/40e582a0-dc54-11ed-bfda-e3b75313765c?_g=(filters:!(),refreshInterval:(pause:!t,value:0),time:(from:now-15m,to:now))&_a=(description:'',filters:!(),fullScreenMode:!f,options:(h...

OpenSearch Dashboards

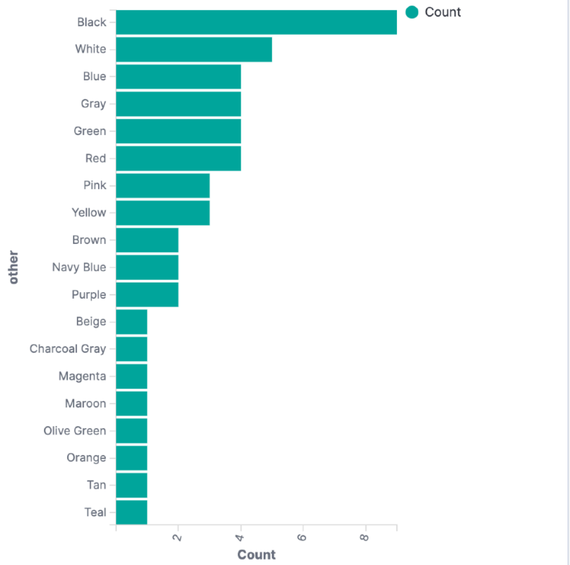
Dashboard | Data Overview

Full screen | Share | Clone | Reporting | Edit | a

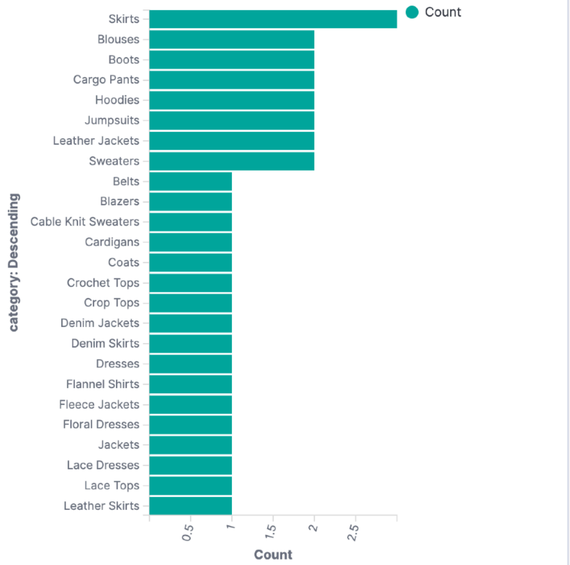
Search | DQL | Last 15 minutes | Show dates | Refresh

+ Add filter

Group by Color



Group by Category



Num Products

50

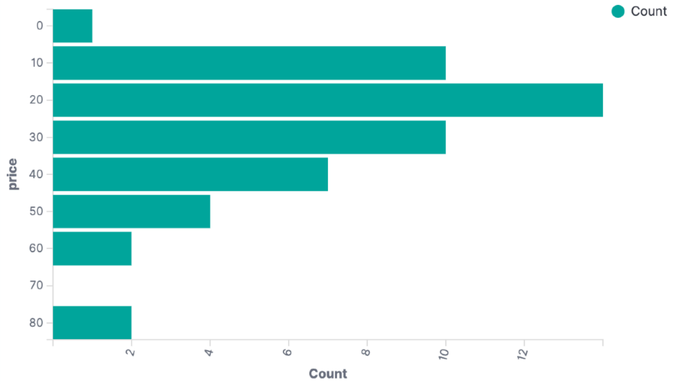
Number of Clicks

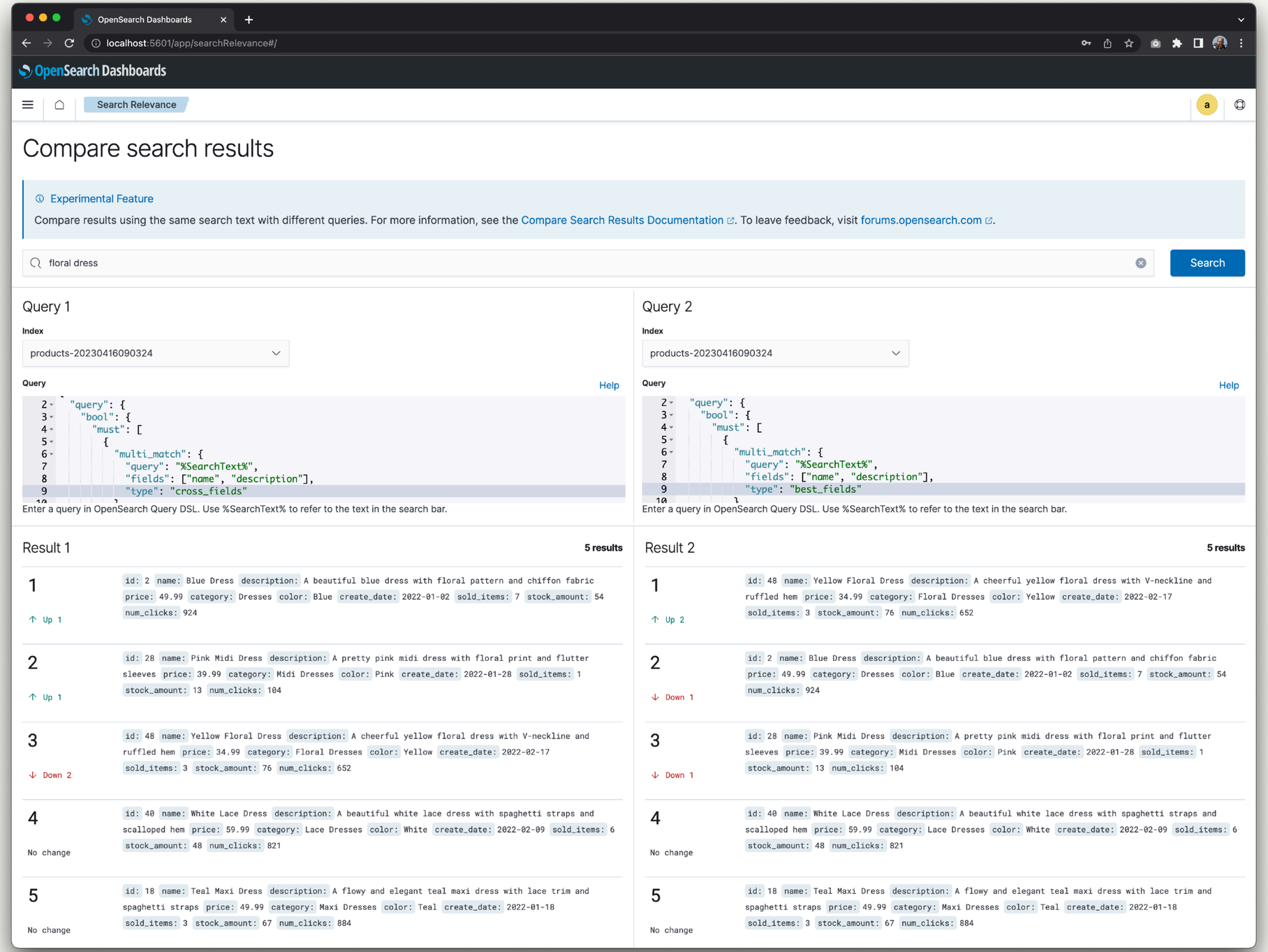
27,389

Number Sold

238

Price Ranges




OpenSearch Dashboards
localhost:5601/app/searchRelevance#/
OpenSearch Dashboards
Search Relevance
a
Compare search results
Experimental Feature
Compare results using the same search text with different queries. For more information, see the Compare Search Results Documentation. To leave feedback, visit forums.opensearch.com.
floral dress
Search
Query 1
Index
products-20230416090324
Query
Help
"query": {
"bool": {
"must": [
{
"multi_match": {
"query": "%SearchText%",
"fields": ["name", "description"],
"type": "cross_fields"
Enter a query in OpenSearch Query DSL. Use %SearchText% to refer to the text in the search bar.
Query 2
Index
products-20230416090324
Query
Help
"query": {
"bool": {
"must": [
{
"multi_match": {
"query": "%SearchText%",
"fields": ["name", "description"],
"type": "best_fields"
Enter a query in OpenSearch Query DSL. Use %SearchText% to refer to the text in the search bar.
Result 1
5 results
1
Up 1
id: 2 name: Blue Dress description: A beautiful blue dress with floral pattern and chiffon fabric price: 49.99 category: Dresses color: Blue create_date: 2022-01-02 sold_items: 7 stock_amount: 54 num_clicks: 924
2
Up 1
id: 28 name: Pink Midi Dress description: A pretty pink midi dress with floral print and flutter sleeves price: 39.99 category: Midi Dresses color: Pink create_date: 2022-01-28 sold_items: 1 stock_amount: 13 num_clicks: 104
3
Down 2
id: 48 name: Yellow Floral Dress description: A cheerful yellow floral dress with V-neckline and ruffled hem price: 34.99 category: Floral Dresses color: Yellow create_date: 2022-02-17 sold_items: 3 stock_amount: 76 num_clicks: 652
4
No change
id: 40 name: White Lace Dress description: A beautiful white lace dress with spaghetti straps and scalloped hem price: 59.99 category: Lace Dresses color: White create_date: 2022-02-09 sold_items: 6 stock_amount: 48 num_clicks: 821
5
No change
id: 18 name: Teal Maxi Dress description: A flowy and elegant teal maxi dress with lace trim and spaghetti straps price: 49.99 category: Maxi Dresses color: Teal create_date: 2022-01-18 sold_items: 3 stock_amount: 67 num_clicks: 884
Result 2
5 results
1
Up 2
id: 48 name: Yellow Floral Dress description: A cheerful yellow floral dress with V-neckline and ruffled hem price: 34.99 category: Floral Dresses color: Yellow create_date: 2022-02-17 sold_items: 3 stock_amount: 76 num_clicks: 652
2
Down 1
id: 2 name: Blue Dress description: A beautiful blue dress with floral pattern and chiffon fabric price: 49.99 category: Dresses color: Blue create_date: 2022-01-02 sold_items: 7 stock_amount: 54 num_clicks: 924
3
Down 1
id: 28 name: Pink Midi Dress description: A pretty pink midi dress with floral print and flutter sleeves price: 39.99 category: Midi Dresses color: Pink create_date: 2022-01-28 sold_items: 1 stock_amount: 13 num_clicks: 104
4
No change
id: 40 name: White Lace Dress description: A beautiful white lace dress with spaghetti straps and scalloped hem price: 59.99 category: Lace Dresses color: White create_date: 2022-02-09 sold_items: 6 stock_amount: 48 num_clicks: 821
5
No change
id: 18 name: Teal Maxi Dress description: A flowy and elegant teal maxi dress with lace trim and spaghetti straps price: 49.99 category: Maxi Dresses color: Teal create_date: 2022-01-18 sold_items: 3 stock_amount: 67 num_clicks: 884

```
{
  "_explanation": {
    "value": 4.819591,
    "description": "sum of:",
    "details": [
      {
        "value": 2.315894,
        "description": "max of:",
        "details": [
          {
            "value": 2.315894,
            "description": "weight(name:blue in 1) [PerFieldSimilarity], result of:"
            "details": [
              {
                "value": 2.315894,
                "description": "score(freq=1.0), computed as boost * idf * tf from:"
                "details": [
                  {"value": 2.2...},
                  {
                    "value": 2.0600235,
                    "description": "idf, computed as log(1 + (N - n + 0.5) / (n + 0.!
                    "details": [
                      {
                        "value": 6,
                        "description": "n, number of documents containing term",
                        "details": []
                      },
                      {
                        "value": 50,
                        "description": "N, total number of documents with field",
                        "details": []
                      }
                    ]
                  },
                  {
                    "value": 0.5110034,
                    "description": "tf, computed as freq / (freq + k1 * (1 - b + b *
                    "details": [
                      {
                        "value": 1,
                        "description": "freq, occurrences of term within document",
                        "details": []
                      },
                      {
                        "value": 1.2,
                        "description": "k1, term saturation parameter",
                        "details": []
                      },
                      {
                        "value": 0.75,
                        "description": "b, length normalization parameter",
                        "details": []
                      },
                      {
                        "value": 2,
                        "description": "dl, length of field",
                        "details": []
                      },
                      {
                        "value": 2.74,
                        "description": "avgdl, average length of field",
                        "details": []
                      }
                    ]
                  }
                ]
              }
            ]
          }
        ]
      },
      {
        "value": 2.1485512,
        "description": "weight(description:blue in
        "details": [
          {
            "value": 2.1485512,
            "description": "score(freq=1.0), comput
            "details": [
              {
                "value": 2.2,
                "description": "boost",
                "details": []
              },
              {
                "value": 2.0600235,
                "description": "idf, computed as lo
                "details": [
                  {
                    "value": 6,
                    "description": "n, number of do
                    "details": []
                  },
                  {
                    "value": 50,
                    "description": "N, total number
                    "details": []
                  }
                ]
              },
              {
                "value": 0.47407913,
                "description": "tf, computed as fre
                "details": [
                  {
                    "value": 1,
                    "description": "freq, occurrenc
                    "details": []
                  },
                  {
                    "value": 1.2,
                    "description": "k1, term satura
                    "details": []
                  },
                  {
                    "value": 0.75,
                    "description": "b, length norma
                    "details": []
                  },
                  {
                    "value": 10,
                    "description": "dl, length of f
                    "details": []
                  },
                  {
                    "value": 11.12,
                    "description": "avgdl, average len
                    "details": []
                  }
                ]
              }
            ]
          }
        ]
      }
    ]
  }
}
```

Best Fields

```
4.819591 is the max of:
   4.4713354 is the sum of:
      2.1485512 for description:blue
         BOOST: 2.2
         IDF: 2.0600235 - 6 out of 50 docs
         TF: 0.47407913 - 1.0 times in field, with 10.0 length to 11.12 average
      2.322784 for description:dress
         BOOST: 2.2
         IDF: 2.2270775 - 5 out of 50 docs
         TF: 0.47407913 - 1.0 times in field, with 10.0 length to 11.12 average
   4.819591 is the sum of:
      2.315894 for name:blue
         BOOST: 2.2
         IDF: 2.0600235 - 6 out of 50 docs
         TF: 0.5110034 - 1.0 times in field, with 2.0 length to 2.74 average
      2.5036972 for name:dress
         BOOST: 2.2
         IDF: 2.2270775 - 5 out of 50 docs
         TF: 0.5110034 - 1.0 times in field, with 2.0 length to 2.74 average
```

Cross Fields

```
4.819591 is the sum of:
   2.315894 is the max of:
      2.315894 for name:blue
         BOOST: 2.2
         IDF: 2.0600235 - 6 out of 50 docs
         TF: 0.5110034 - 1.0 times in field, with 2.0 length to 2.74 average
      2.1485512 for description:blue
         BOOST: 2.2
         IDF: 2.0600235 - 6 out of 50 docs
         TF: 0.47407913 - 1.0 times in field, with 10.0 length to 11.12 average
   2.5036972 is the max of:
      2.322784 for description:dress
         BOOST: 2.2
         IDF: 2.2270775 - 5 out of 50 docs
         TF: 0.47407913 - 1.0 times in field, with 10.0 length to 11.12 average
      2.5036972 for name:dress
         BOOST: 2.2
         IDF: 2.2270775 - 5 out of 50 docs
         TF: 0.5110034 - 1.0 times in field, with 2.0 length to 2.74 average
```

- Know the distribution of your content
- Use A/B testing against your endpoints
- Use explain to verify query results

Learn how to debug results

Bonus Topic

Topic 1: Search is more than Technology

Topic 2: Understand your content

Topic 3: Use a schema for your content

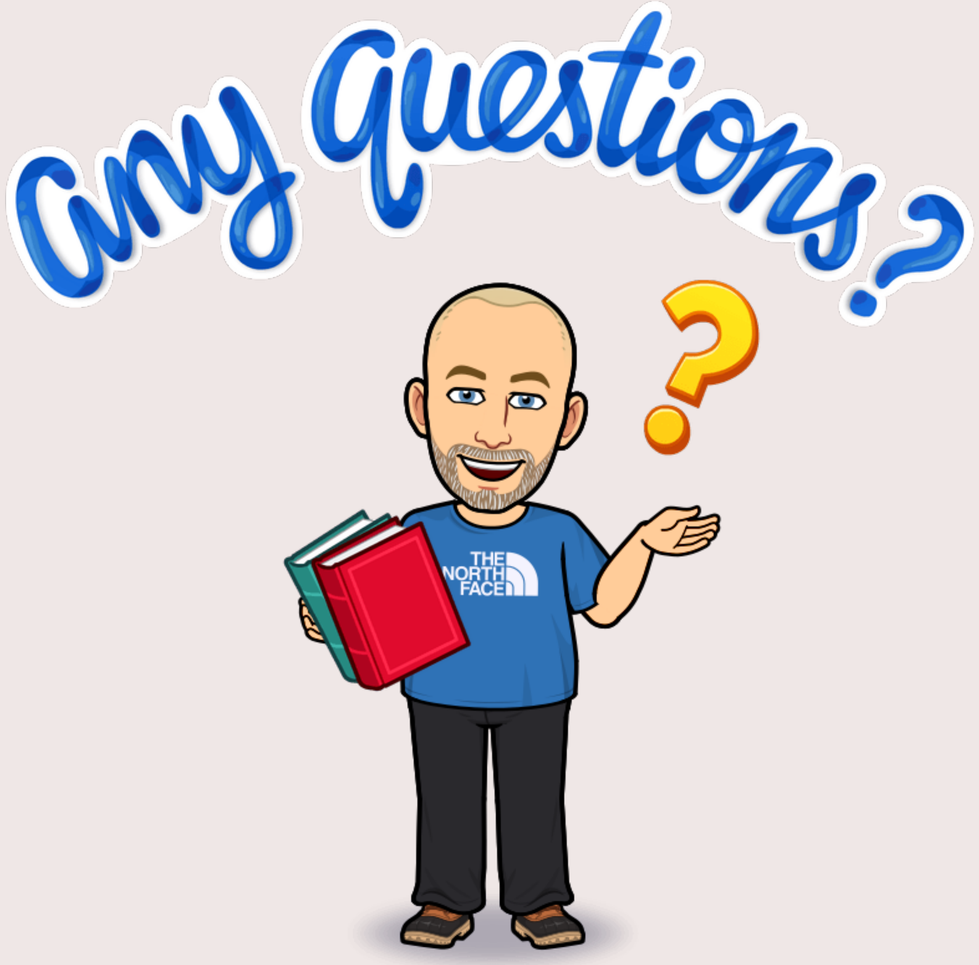
Topic 4: Transform user question into a query

Topic 5: Relax Matching and use boosting

Topic 6: When search is not the answer

Topic 7: Log customers activities and learn

Topic 8: Use a judgements list to verify results



Manning

Elasticsearch for a Search API

35% discount: coenradie35 (till the 8th of May)

http://mng.bz/d1ro

https://www.manning.com/liveprojectseries/elasticsearch-for-a-search-api-ser

luminis.

https://www.bitmoji.com/