Vancouver Meetup:
Beginner's Crash Course to Elastic Stack Series

Part 2: Understanding the relevance of your search with Elasticsearch & Kibana

Lisa Jung
Developer Advocate @Elastic

Connect with the Elastic Community

Find your local User Group:

- https://community.elastic.co/

Virtual User Group:

https://community.elastic.co/amer-virtual/

Connect with the Elastic Community

Community Slack Workspace:

https://join.slack.com/t/elasticstack/shared_invite/zt-an1h0etg-04FI2hA9vvASBkYPe~QZmw

Vancouver Meetup:
Beginner's Crash Course to Elastic Stack Series

Part 2: Understanding the relevance of your search with Elasticsearch & Kibana

Lisa Jung
Developer Advocate @Elastic

Beginner’s crash course to Elastic Stack Series

- **Part 1: Intro to Elasticsearch and Kibana**
  - use case of Elasticsearch and Kibana
  - the basic architecture of Elasticsearch
  - perform CRUD(Create, Read, Update, and Delete) operations with Elasticsearch and Kibana

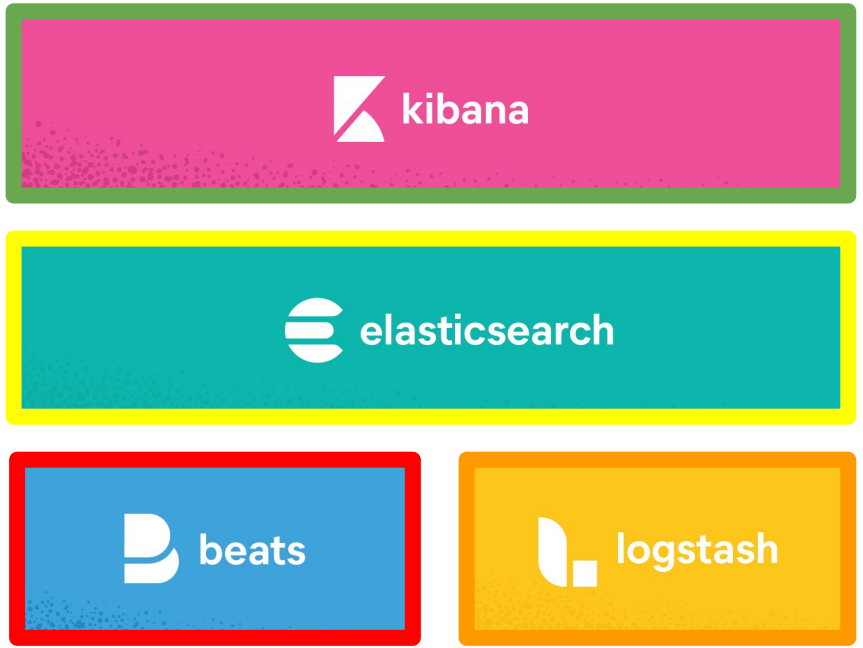
elastic

Missed the first workshop? No worries!

- **Part 1: Intro to Elasticsearch and Kibana**
  - **Repo: https://ela.st/vancouver-workshop-1**

The Elastic Stack

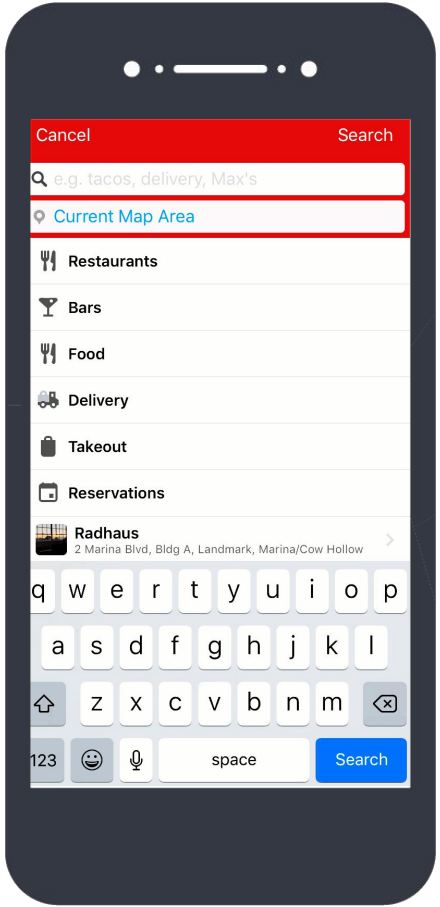
Reliably and securely take data from any source, in any format, then search, analyze, and visualize it in real time.



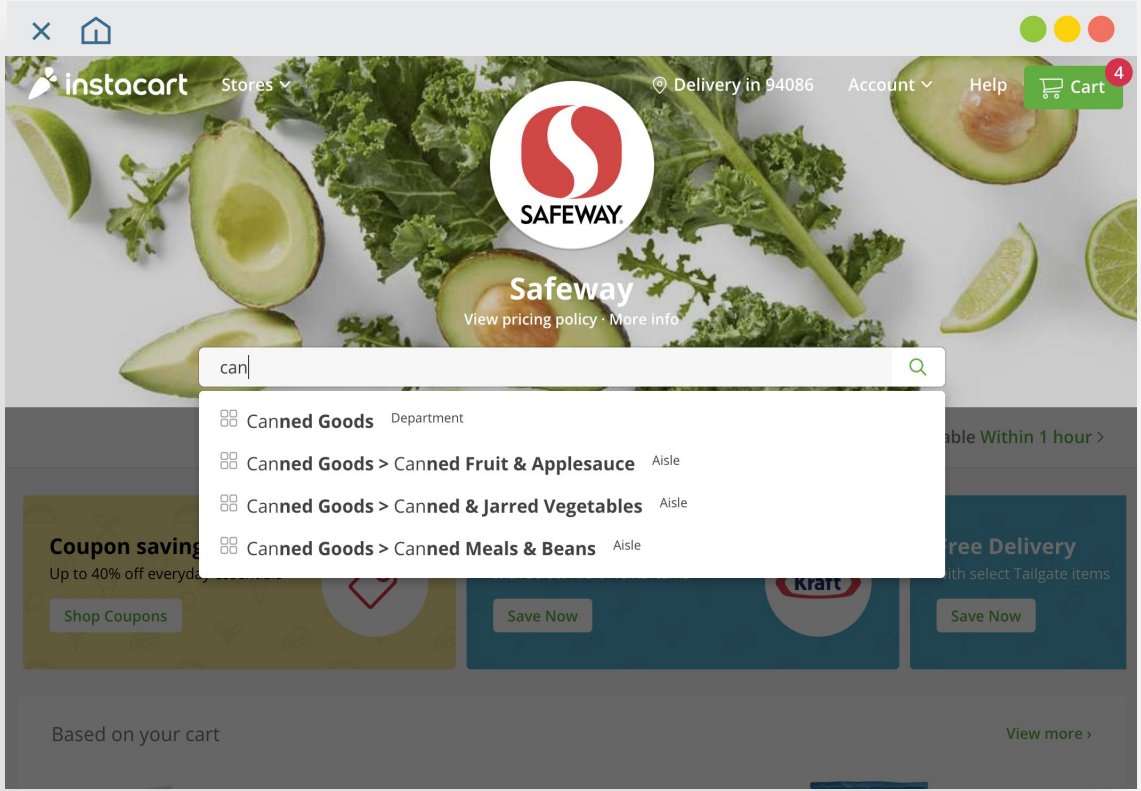
Elasticsearch

Store | Search | Analyze

yelp



instacart



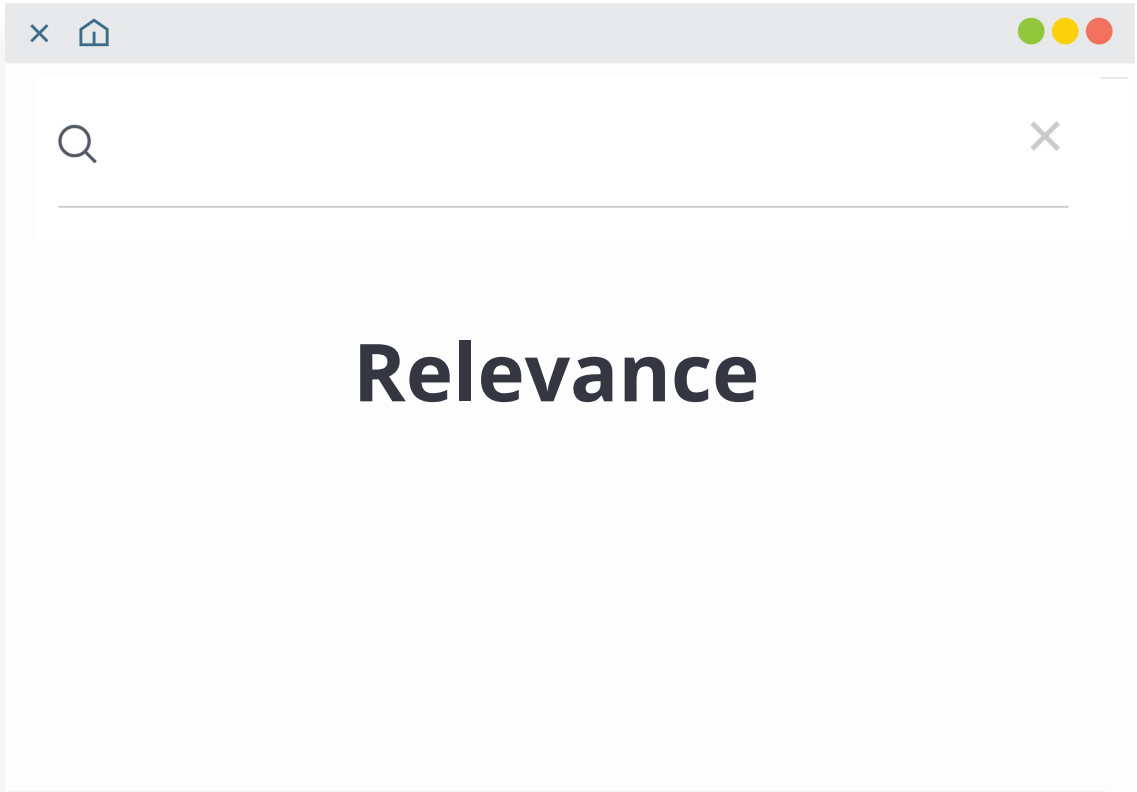
Speed, Scale,

Relevance

Elastic is a search company.

We focus on value to users by producing fast results that operate at scale and are relevant. This is our DNA. We believe search is an experience. It is what defines us, and makes us unique.

elastic



elastic

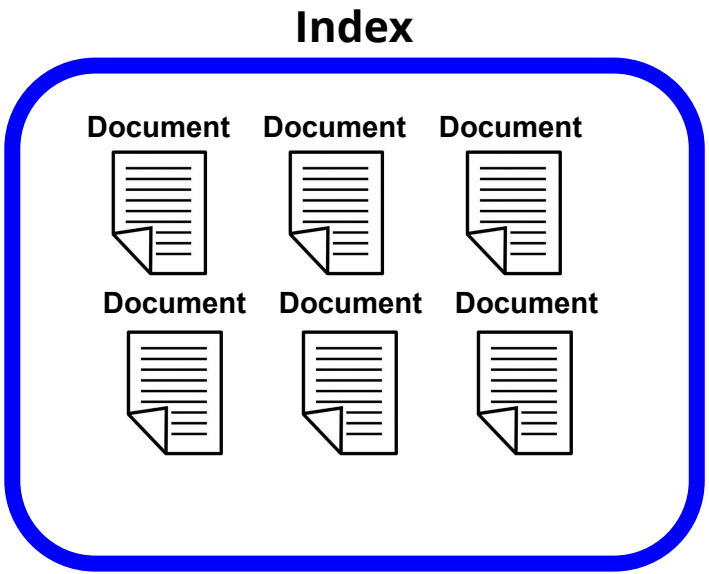
How do we measure relevance?

- **Precision**
- **Recall**

I store data as documents!

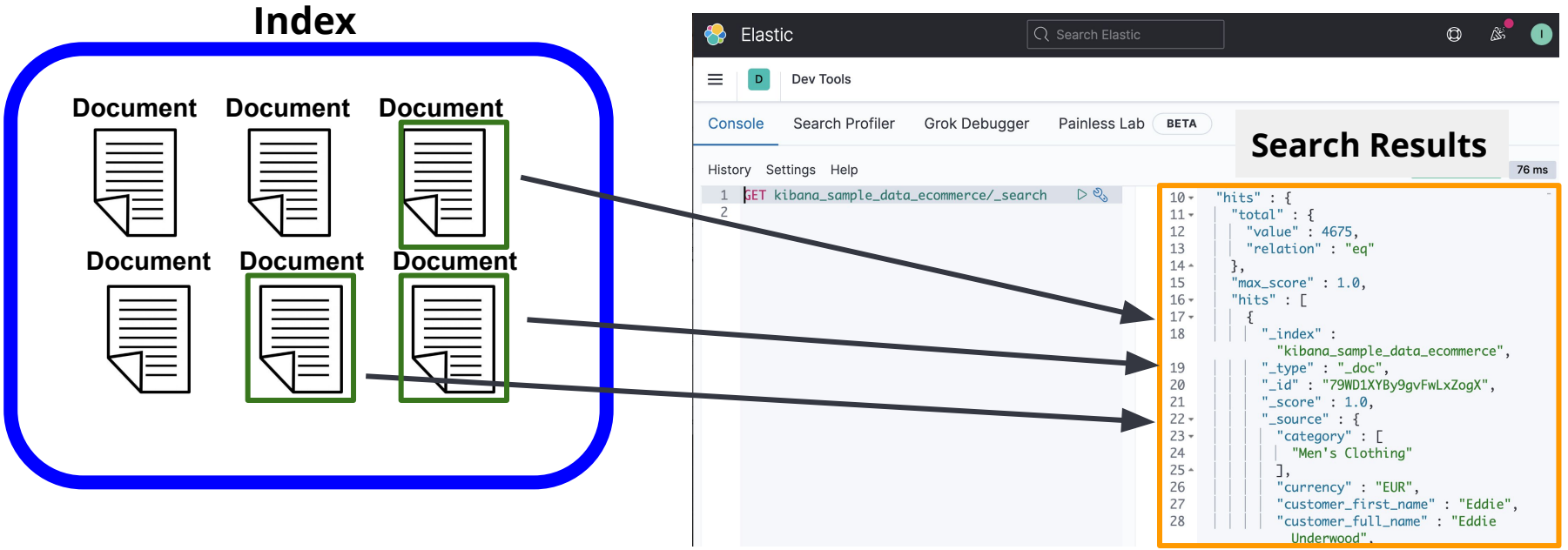
Elasticsearch

Store | Search | Analyze



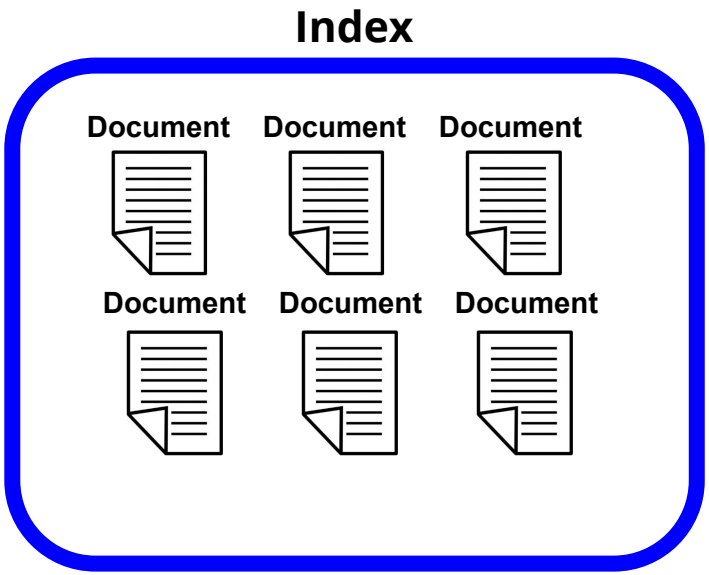
Documents with similar traits are grouped into an index!

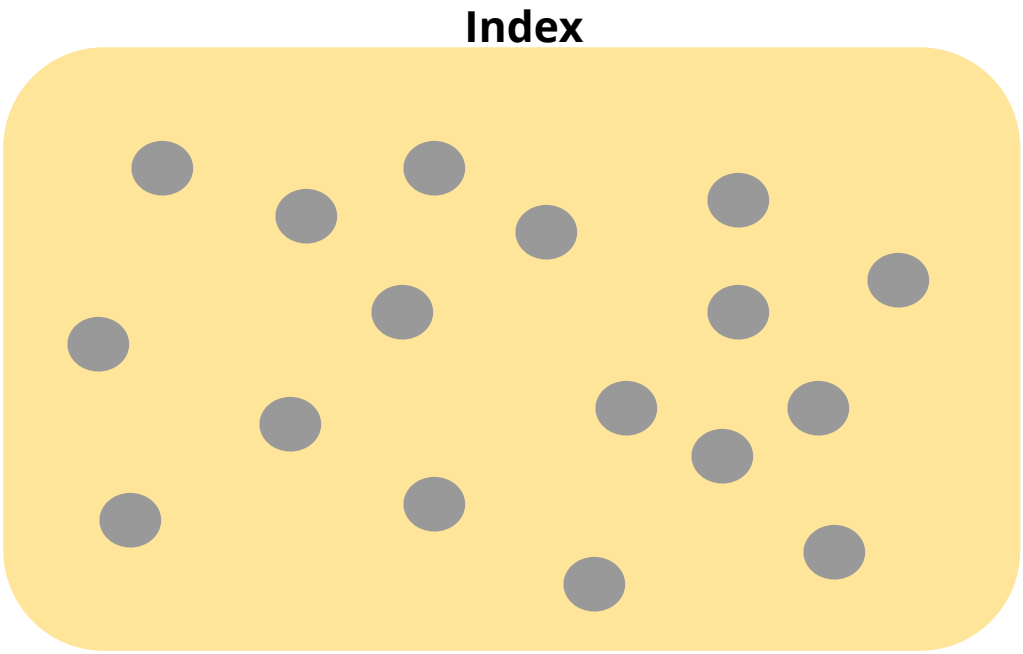
When search query is sent, Elasticsearch retrieves relevant documents and presents the documents as search results.



elastic

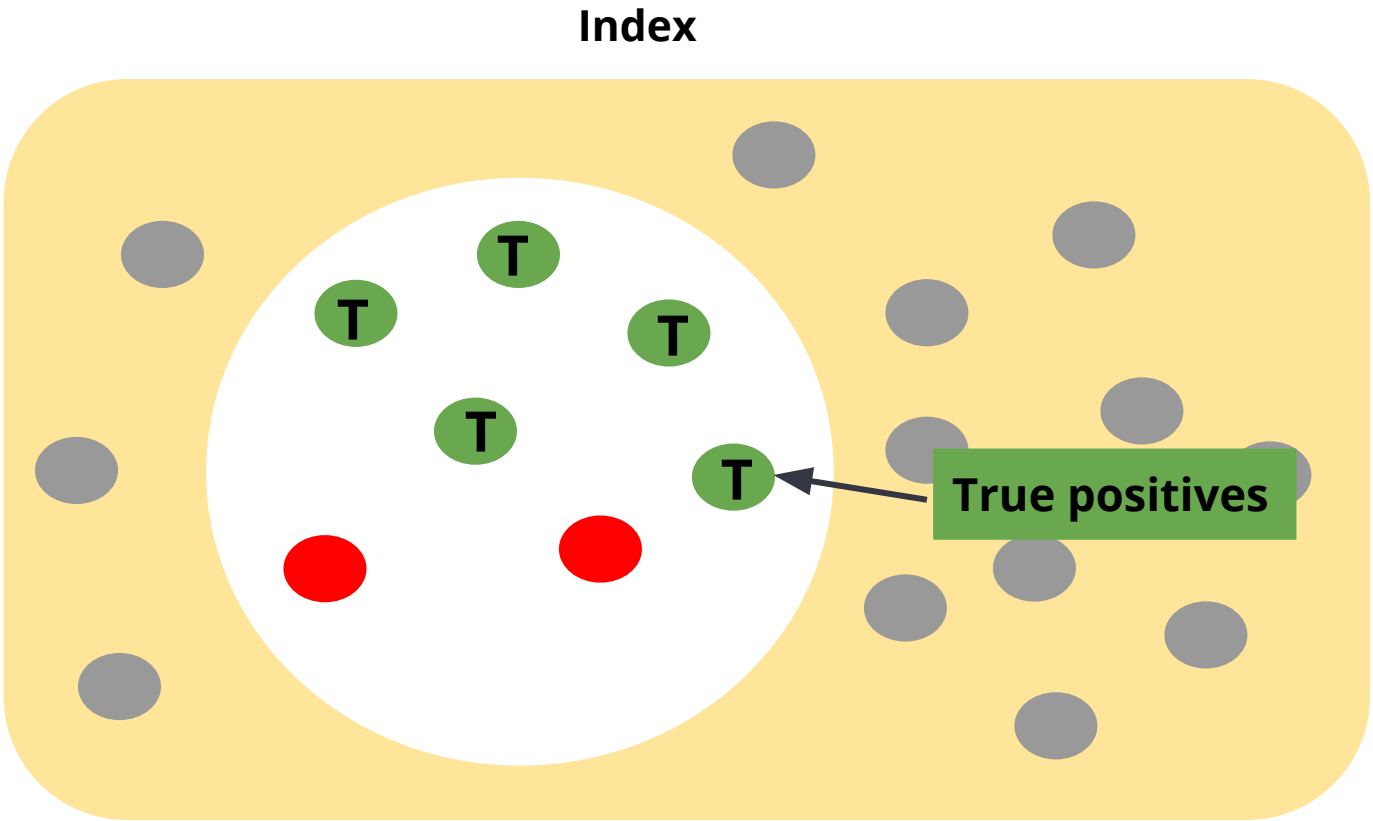
These two diagrams depict the same thing!





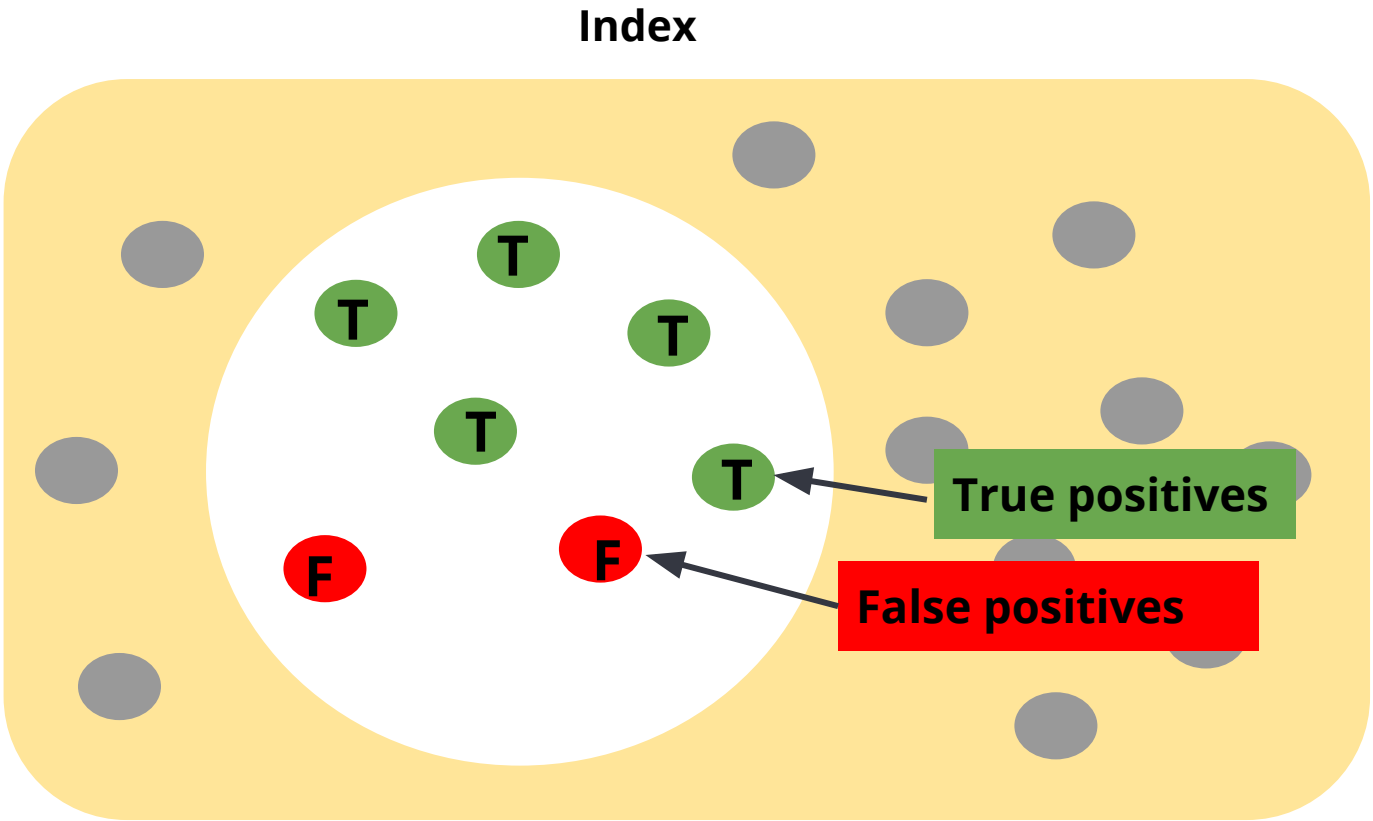
elastic

True positives are relevant documents that are returned to the user.

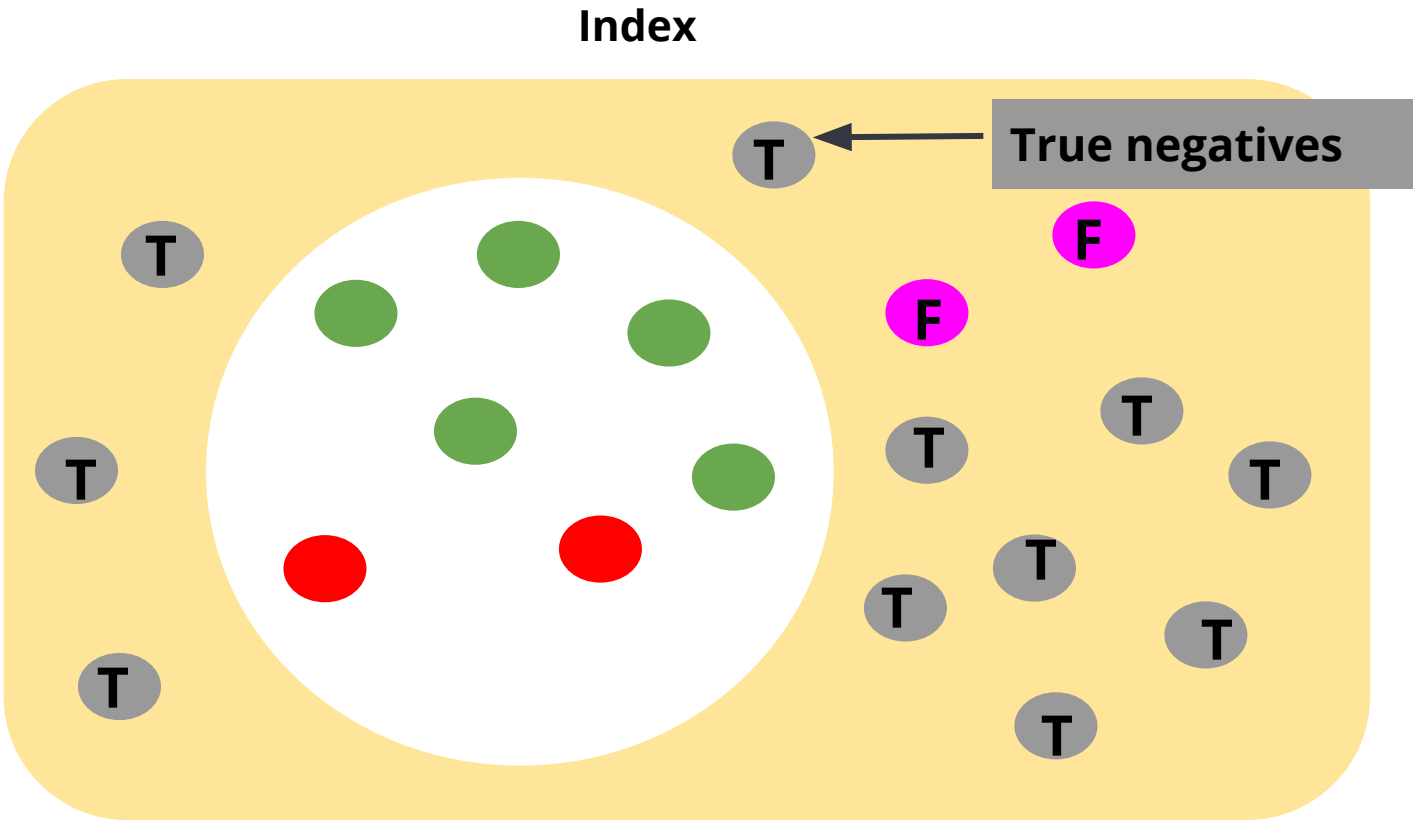


elastic

False positives are irrelevant documents that are returned to the user.

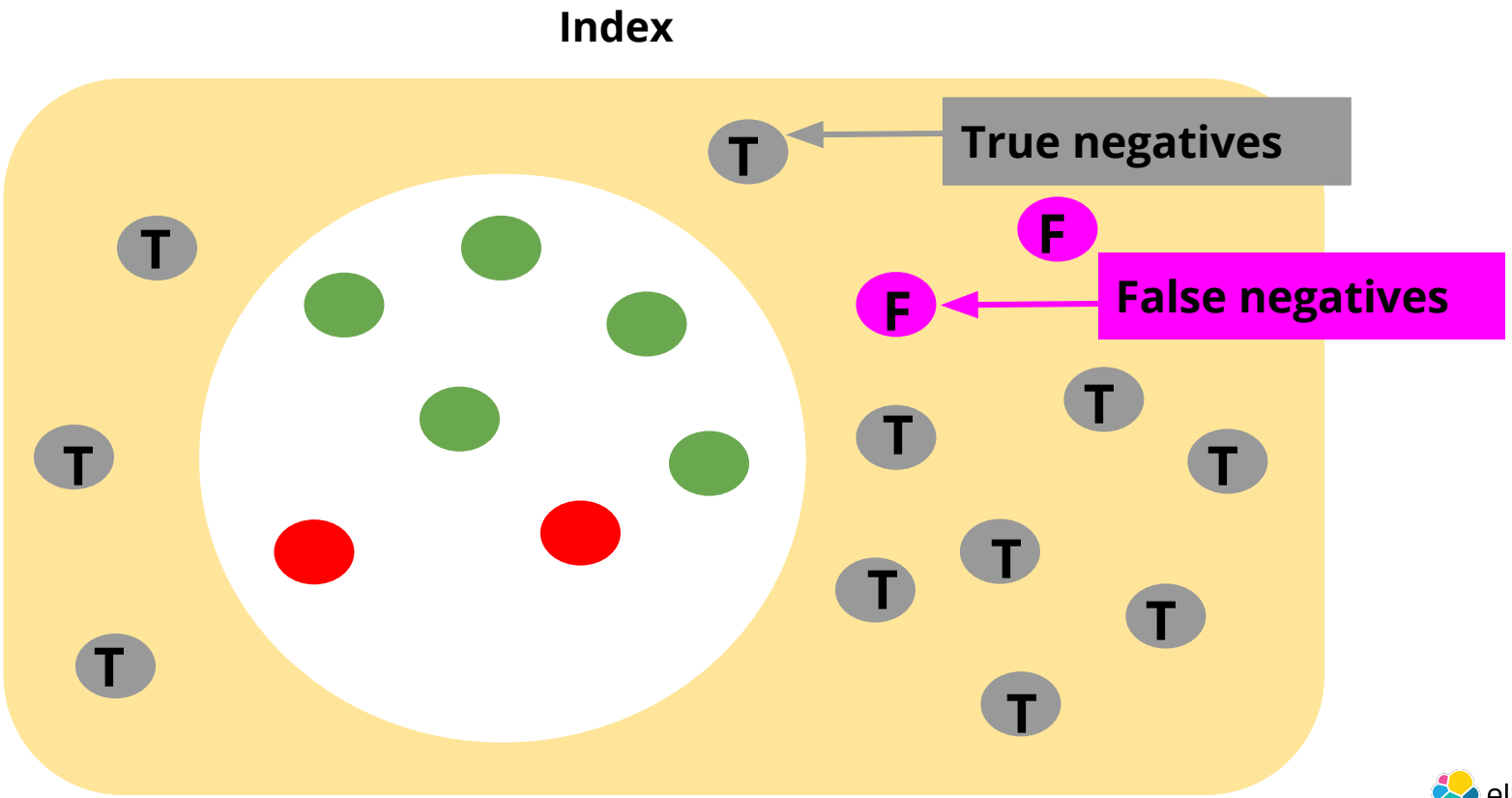


True negatives are irrelevant documents that are not returned to the user.



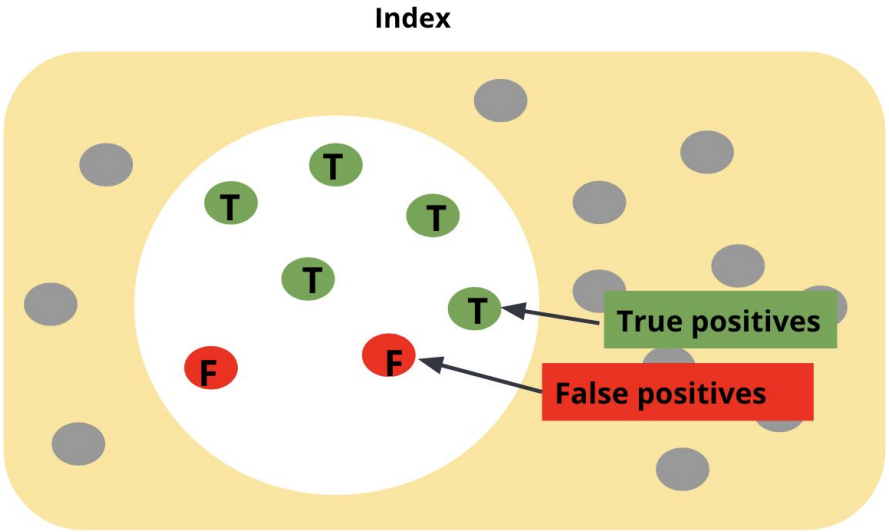
elastic

False negatives are relevant documents that were not returned to the user.



elastic

What is precision?

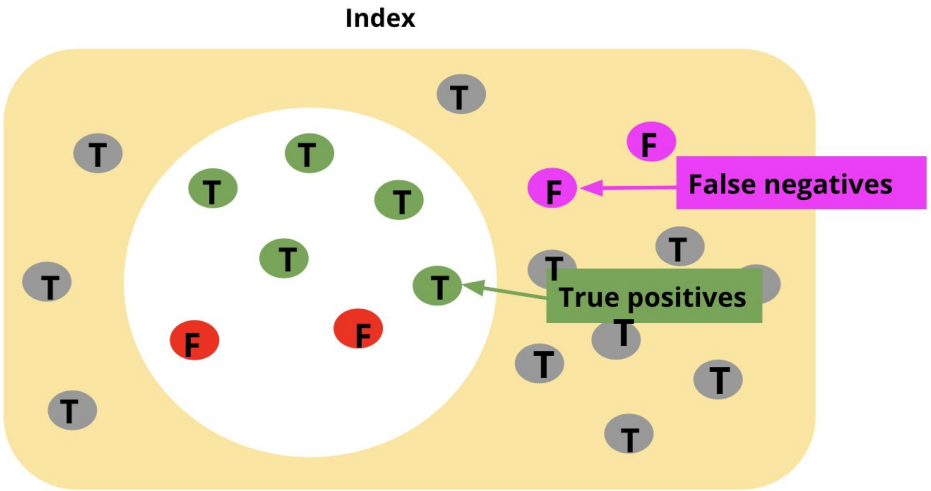


$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$$

What portion of the retrieved data is actually relevant to the search query?

elastic

What is recall?

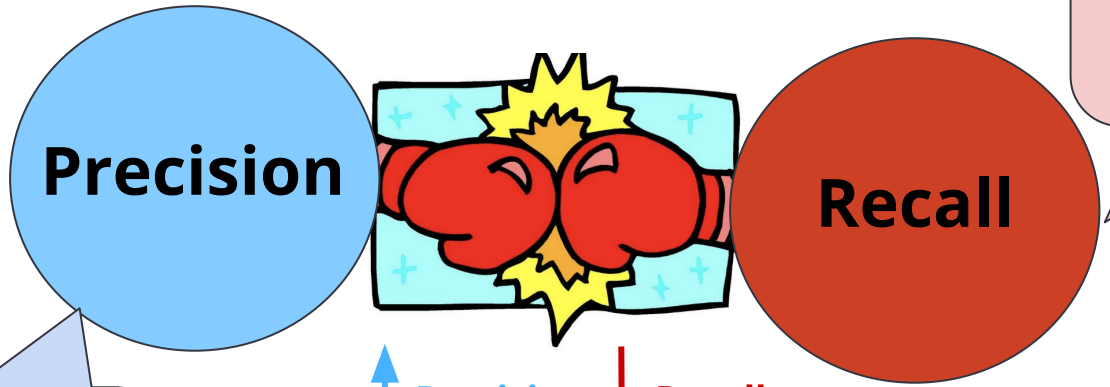


$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$$

What portion of relevant data is being returned as search results?

elastic

Precision and Recall are inversely related



I want to retrieve more results even if documents may not be a perfect match to the query.

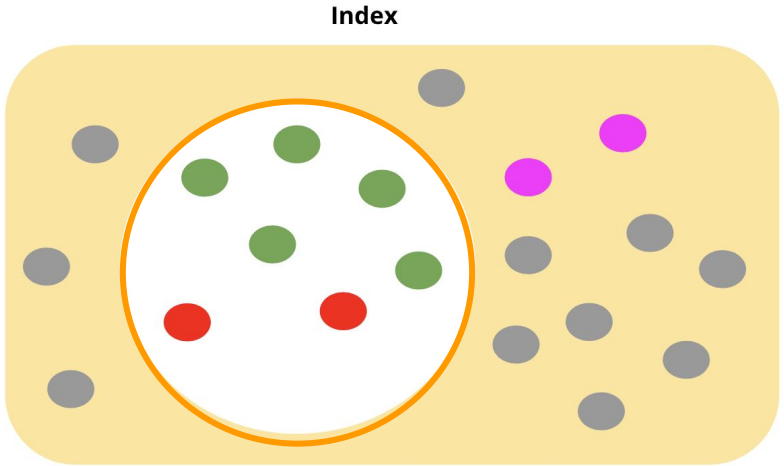
I want all the retrieved results to be a perfect match to the query, even if it means returning less documents.

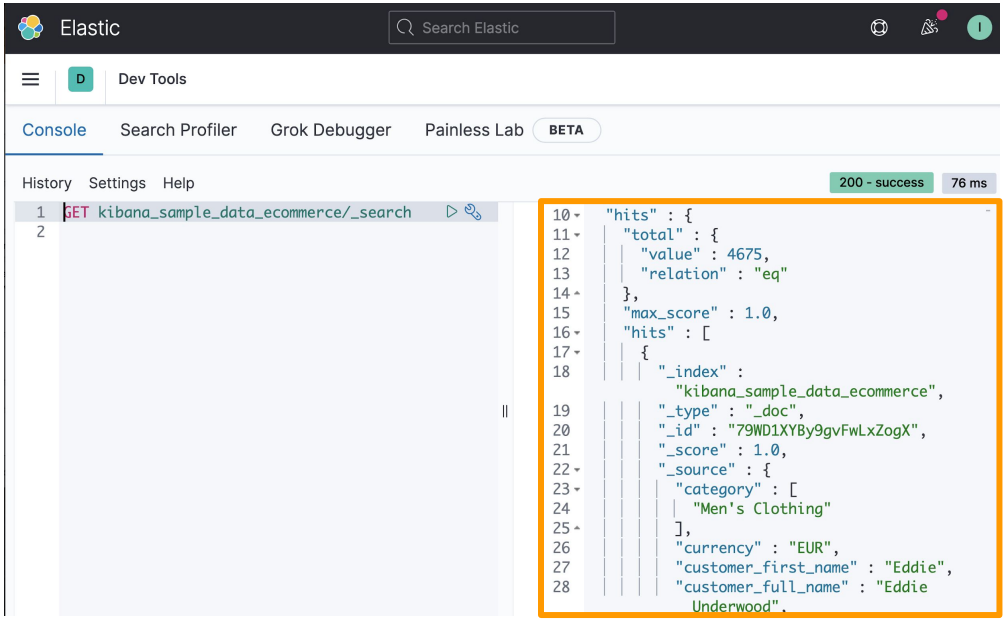
↑ Precision ↓ Recall

↓ Precision ↑ Recall

elastic

Precision and recall determine which documents are included in the search results.

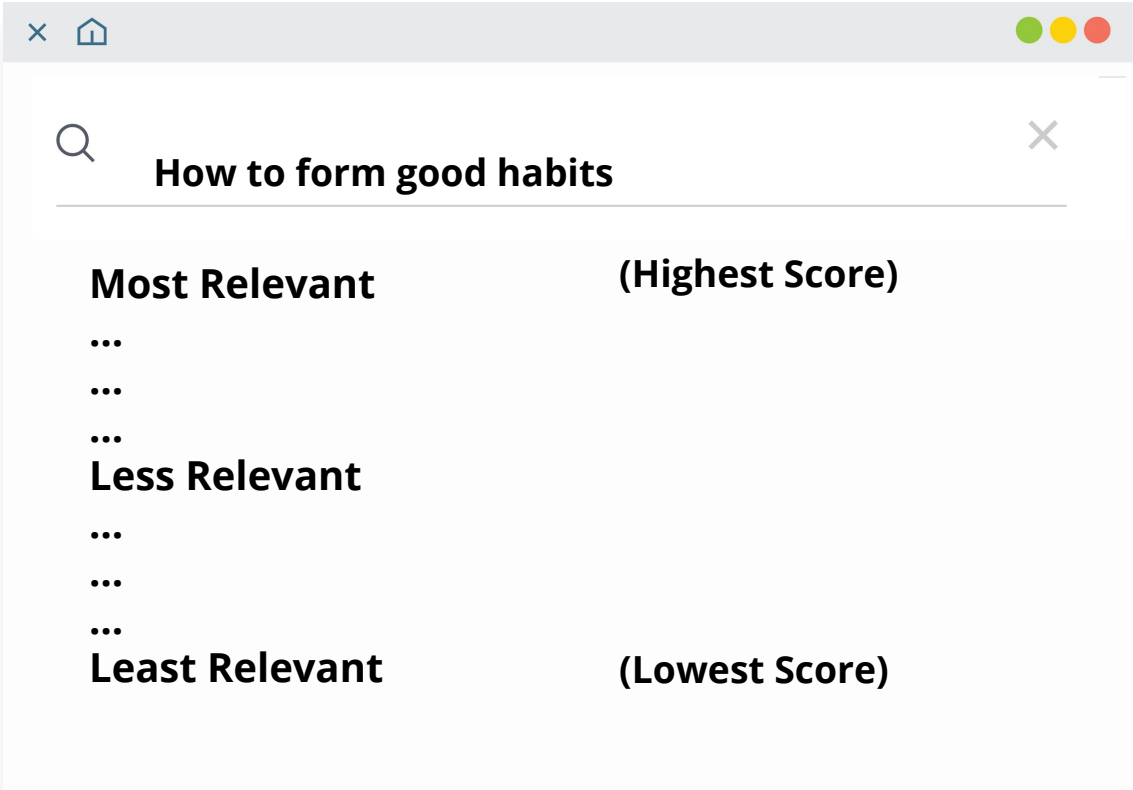




Precision and recall do not determine which of the returned documents are more relevant compared to the other!

elastic

Ranking refers to ordering of the results (from most relevant results at the top, to least relevant at the bottom).



elastic

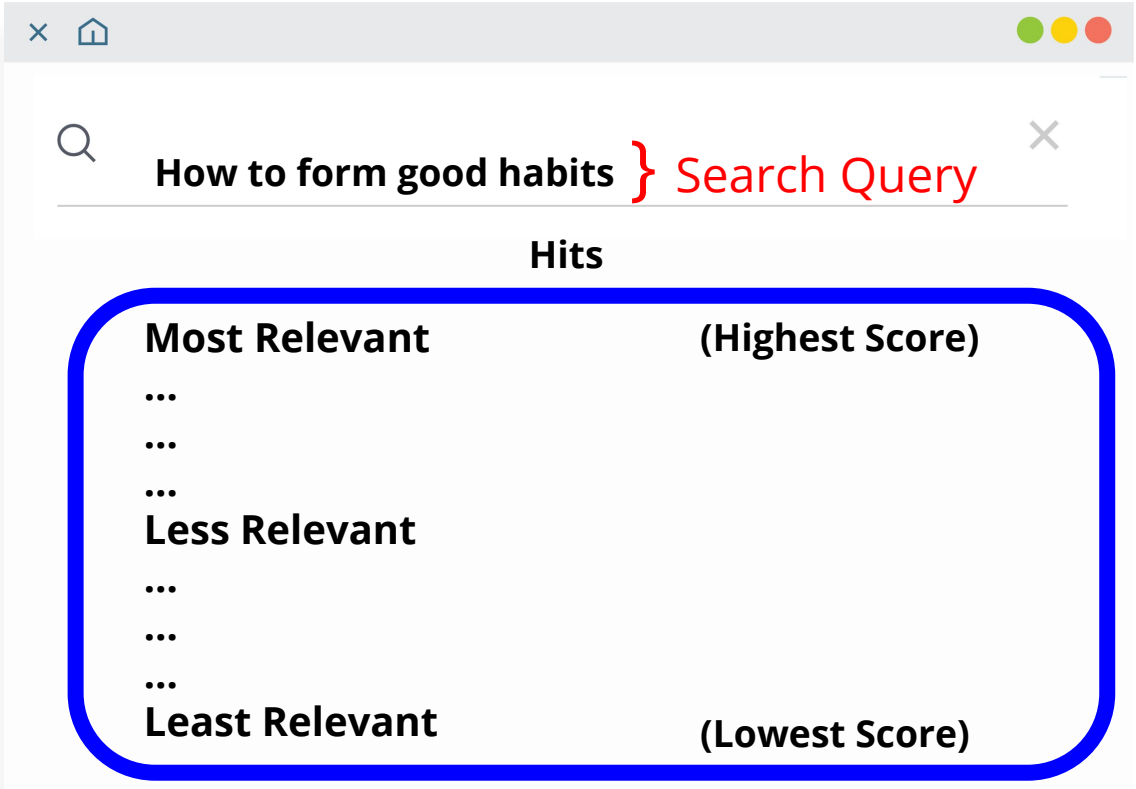
What is score?

- The score is a value that represents how relevant a document is to that specific query
- A score is computed for each document that is a hit

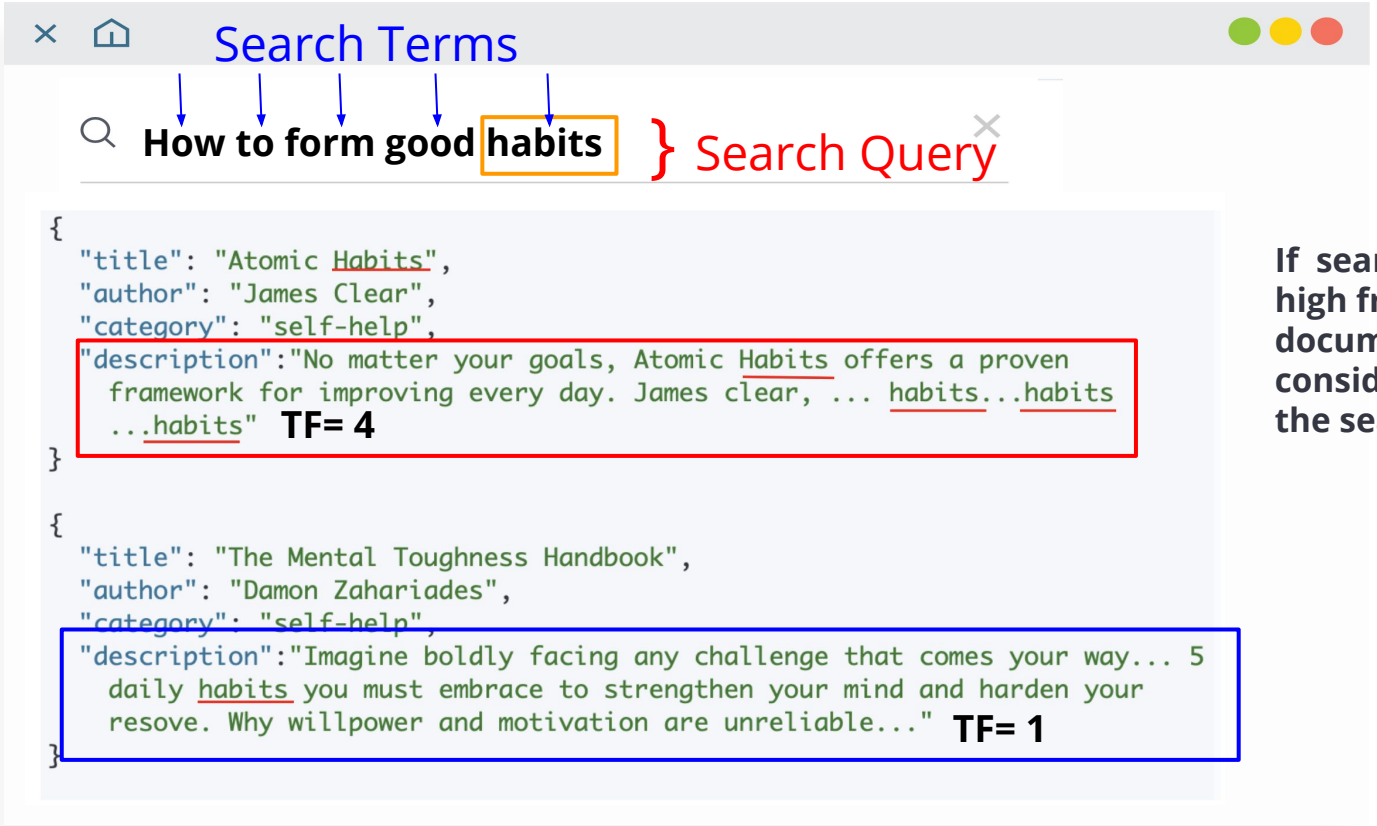
What is score?

- Term Frequency(TF)
- Inverse Document Frequency(IDF)

What is score?



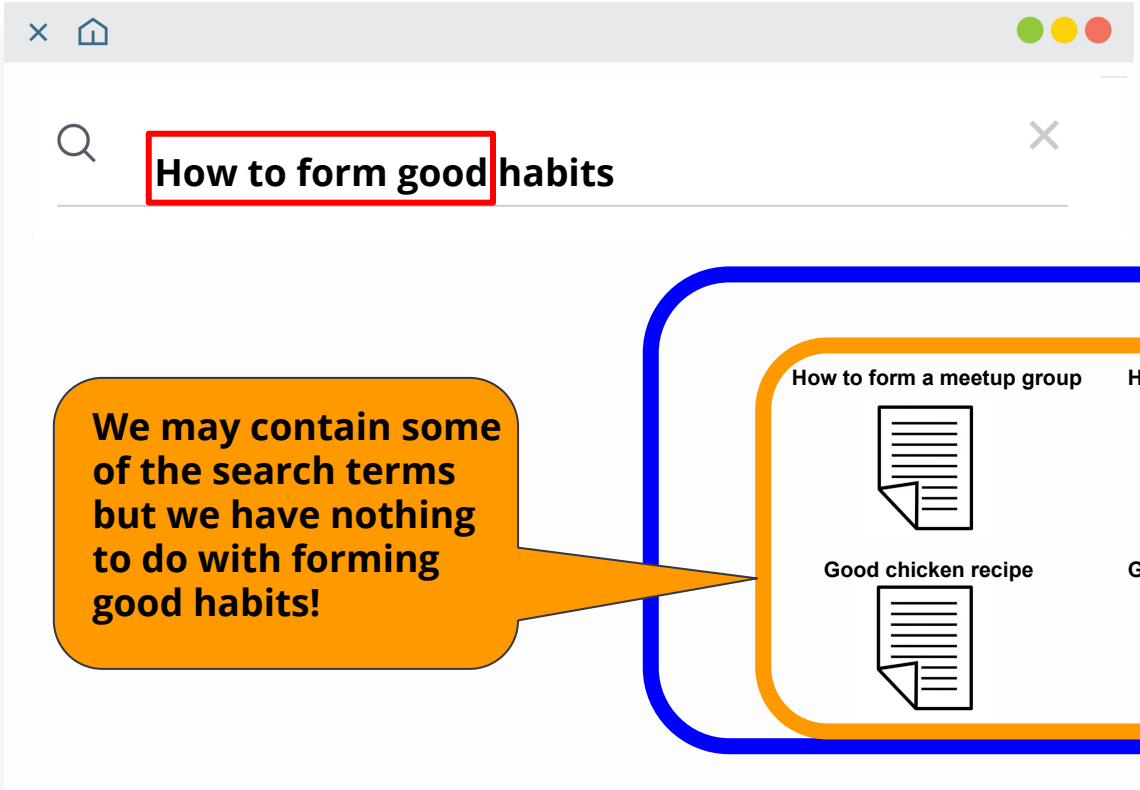
Term Frequency(TF) determines how many times each search term appears in a document.



If search terms are found in high frequency in a document, the document is considered more relevant to the search query.

elastic

What is Inverse Document Frequency(IDF)?



IDF diminishes the weight of terms that occur very frequently in the document set and increases the weight of terms that occur rarely!

elastic

Fine tuning precision or recall using Elasticsearch and Kibana

Click on the link to the workshop repo.

https://ela.st/vancouver-workshop-2

elastic

Scroll down to the Resources section & click on Instructions for downloading Elasticsearch & Kibana

Beginner's Crash Course to the Elastic Stack Series

[Vancouver] Part 2: Understanding the relevance of your search with Elasticsearch and Kibana

Welcome to the Beginner's Crash Course to the Elastic Stack!

This repo contains all resources shared during Part 2: Understanding the relevance of your search with Elasticsearch and Kibana.

Resources

Free Elastic Cloud Trial

Instructions for downloading Elasticsearch and Kibana

Presentation

Dataset from Kaggle used for tutorial

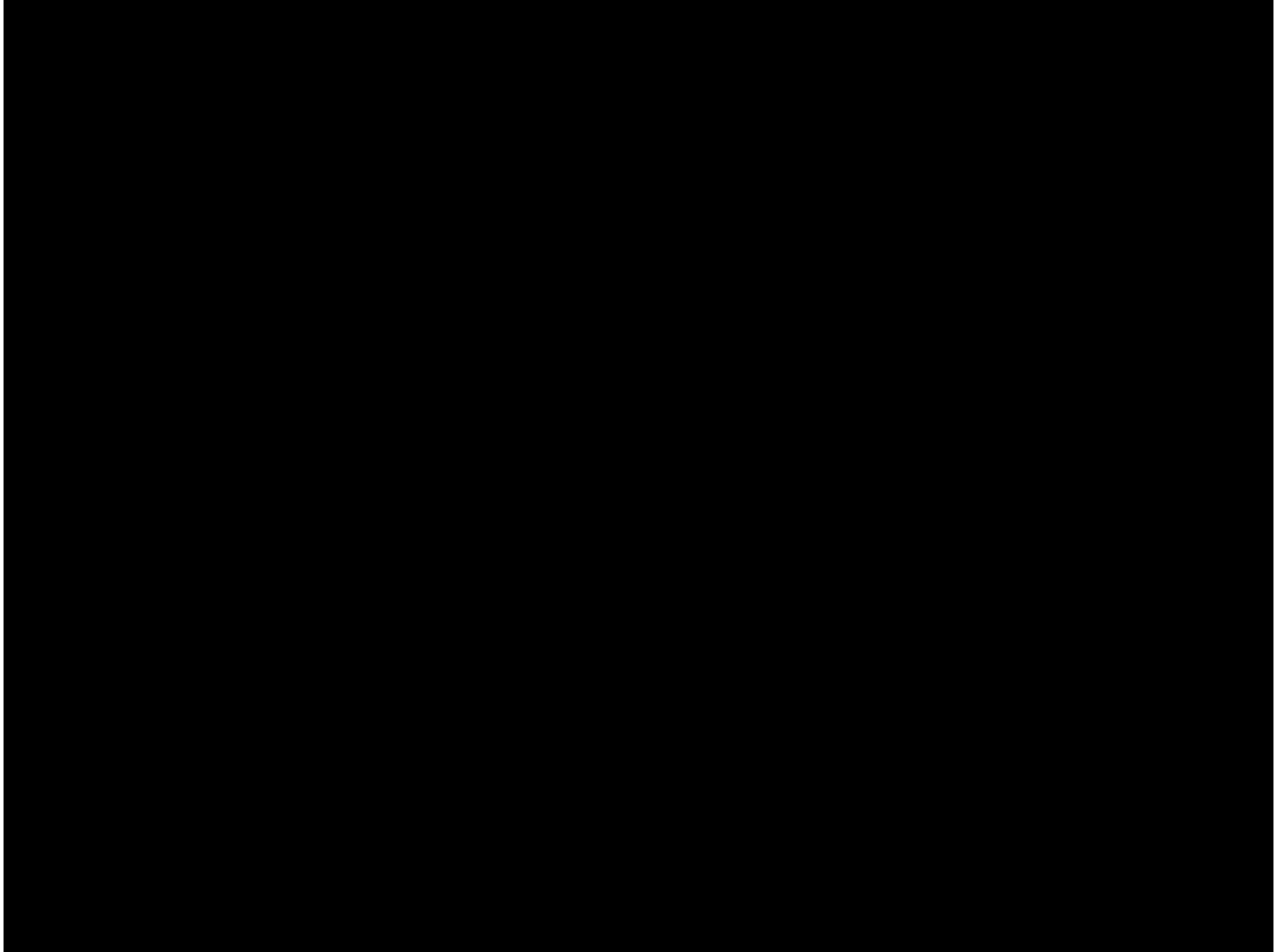
Elastic Vancouver Chapter Want to attend live workshops? Join the Elastic Vancouver Chapter to get the deets!

Search for information

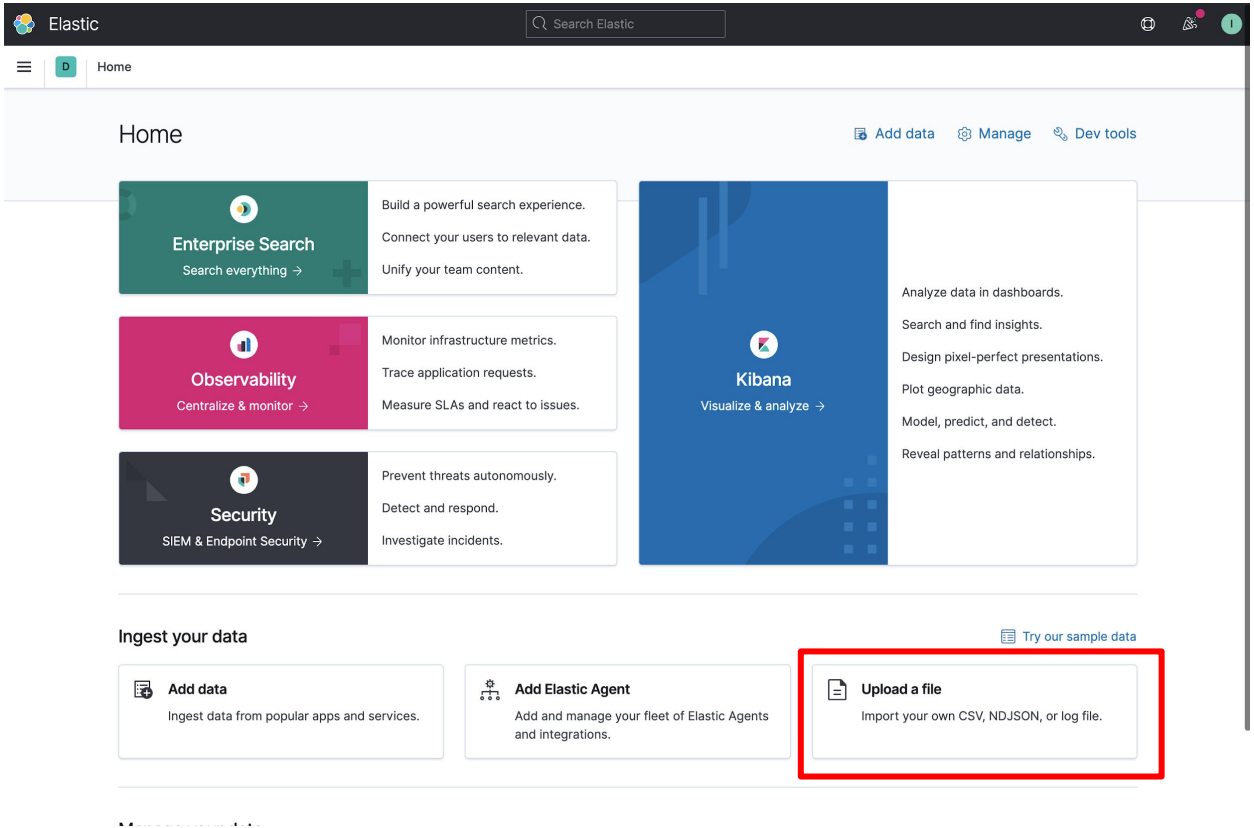
There are two main ways to search in Elasticsearch:

1. Queries
2. Aggregations

elastic

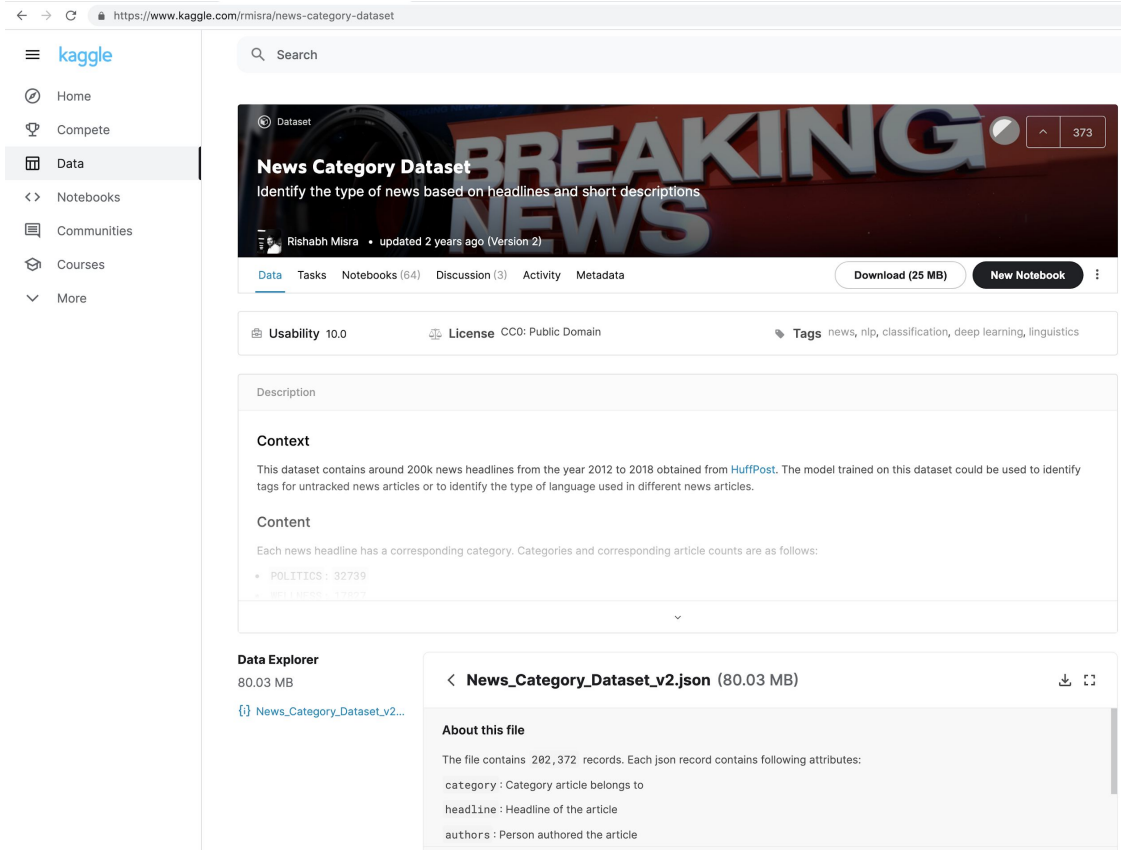


Click on Upload a file option



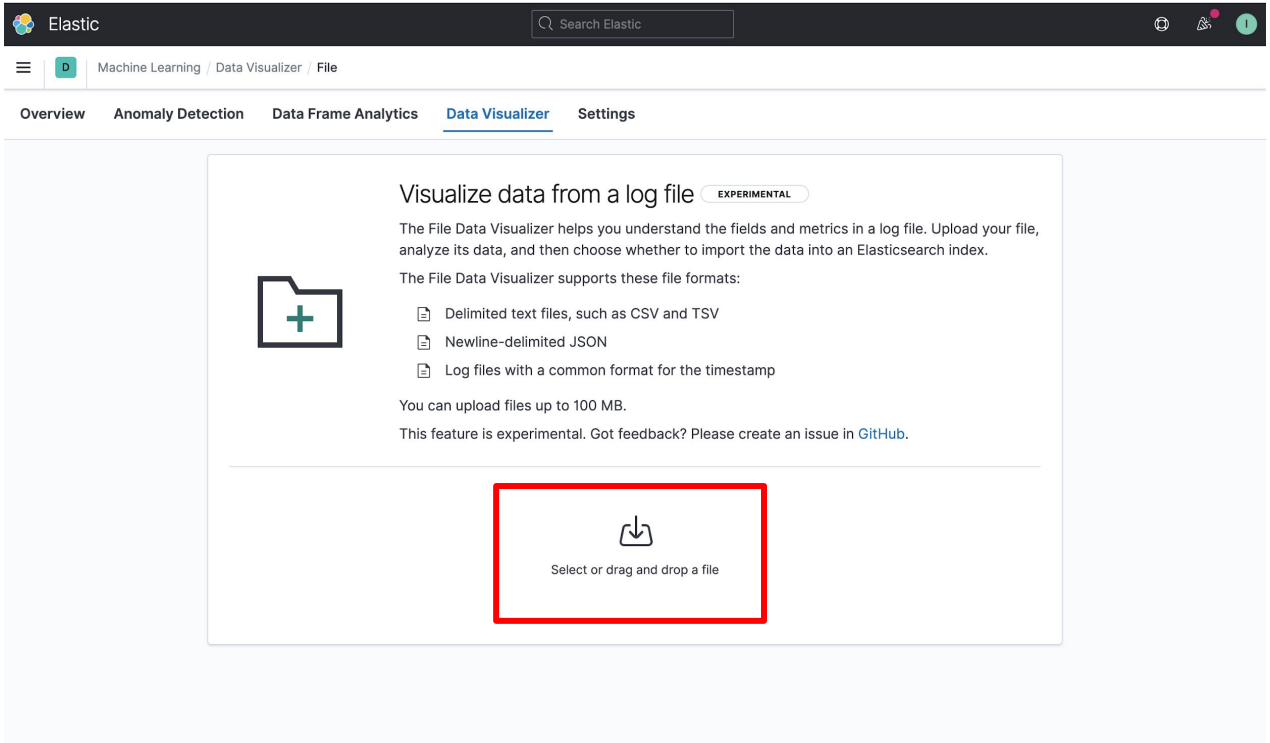
elastic

Download and unzip News Category Dataset from Kaggle



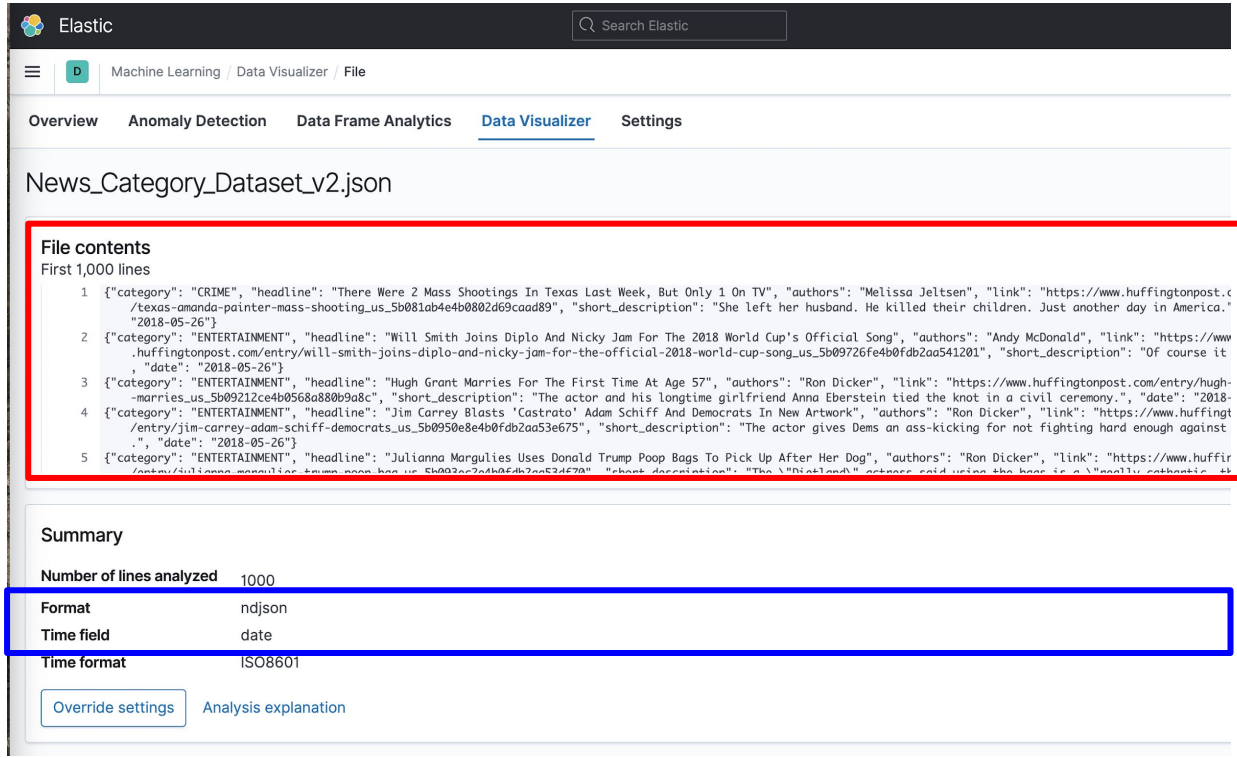
elastic

Drag and drop a file you want to upload.



elastic

Kibana will give you an analysis of the first 1000 lines of your data and give you a summary of your dataset.

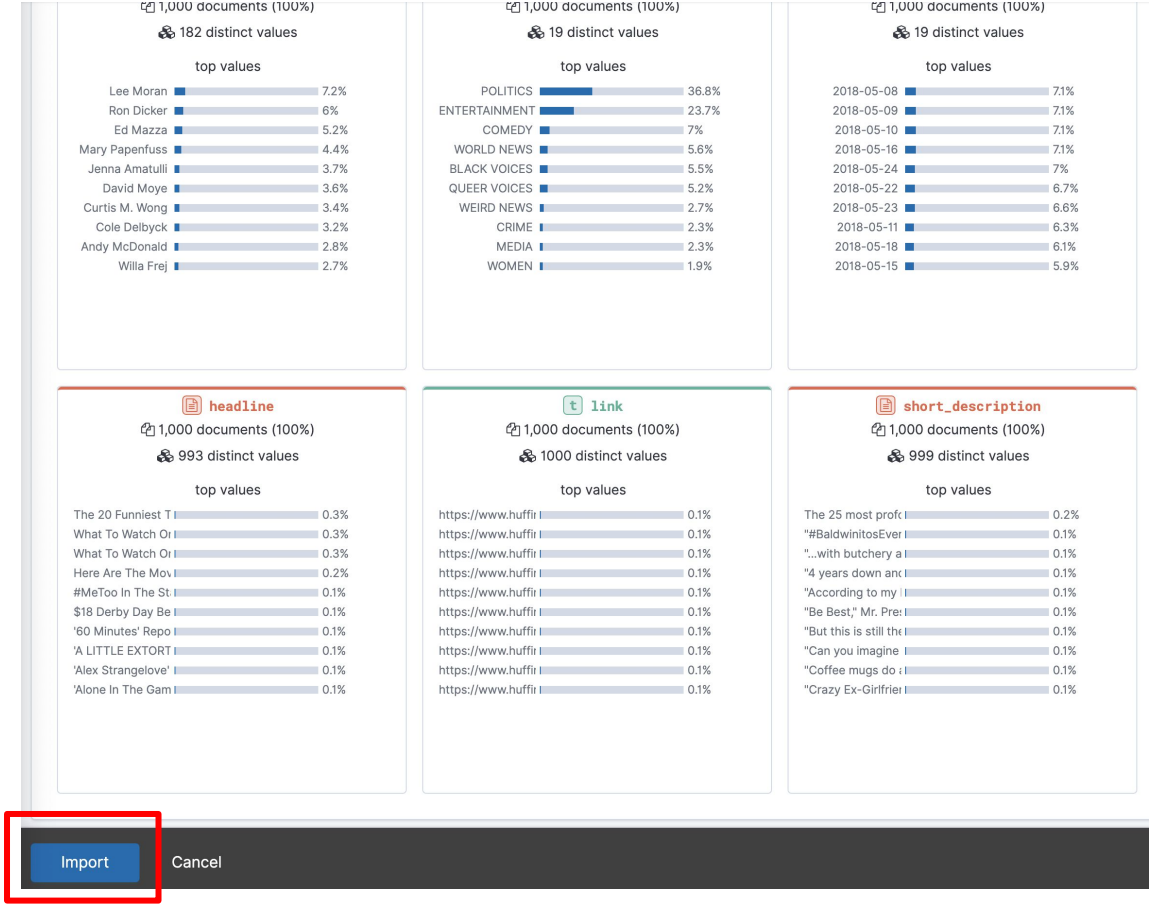


elastic

Field section displays fields identified, high level statistics, and top occuring values

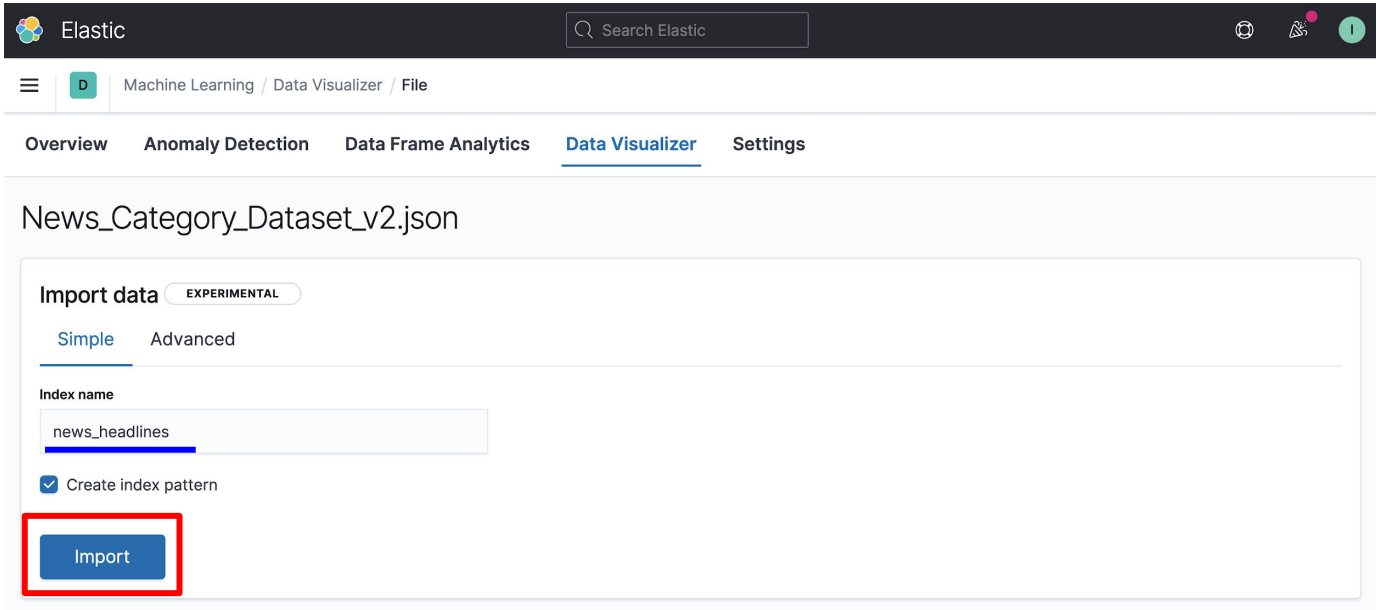
elastic

Click on import button

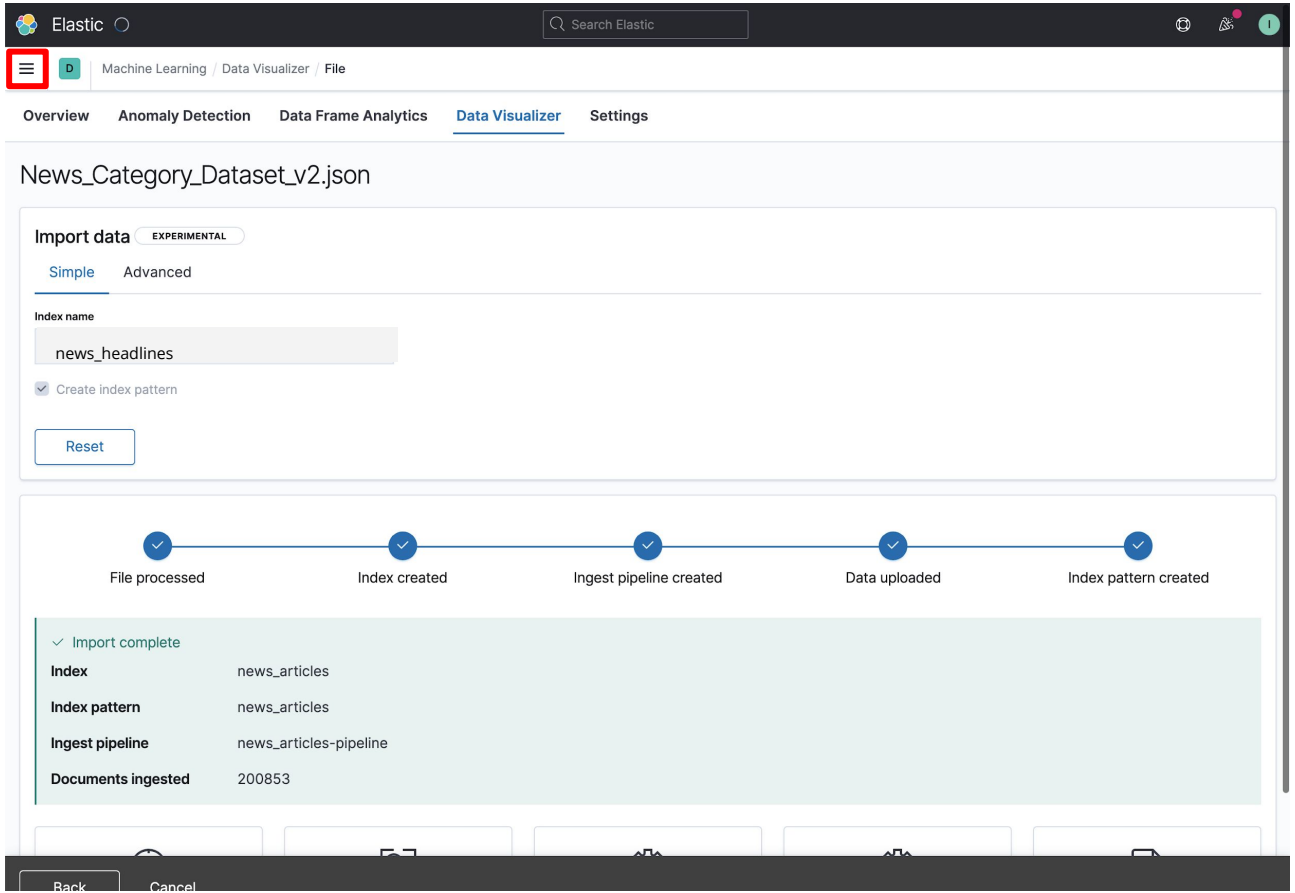


elastic

Name your index and click on import.

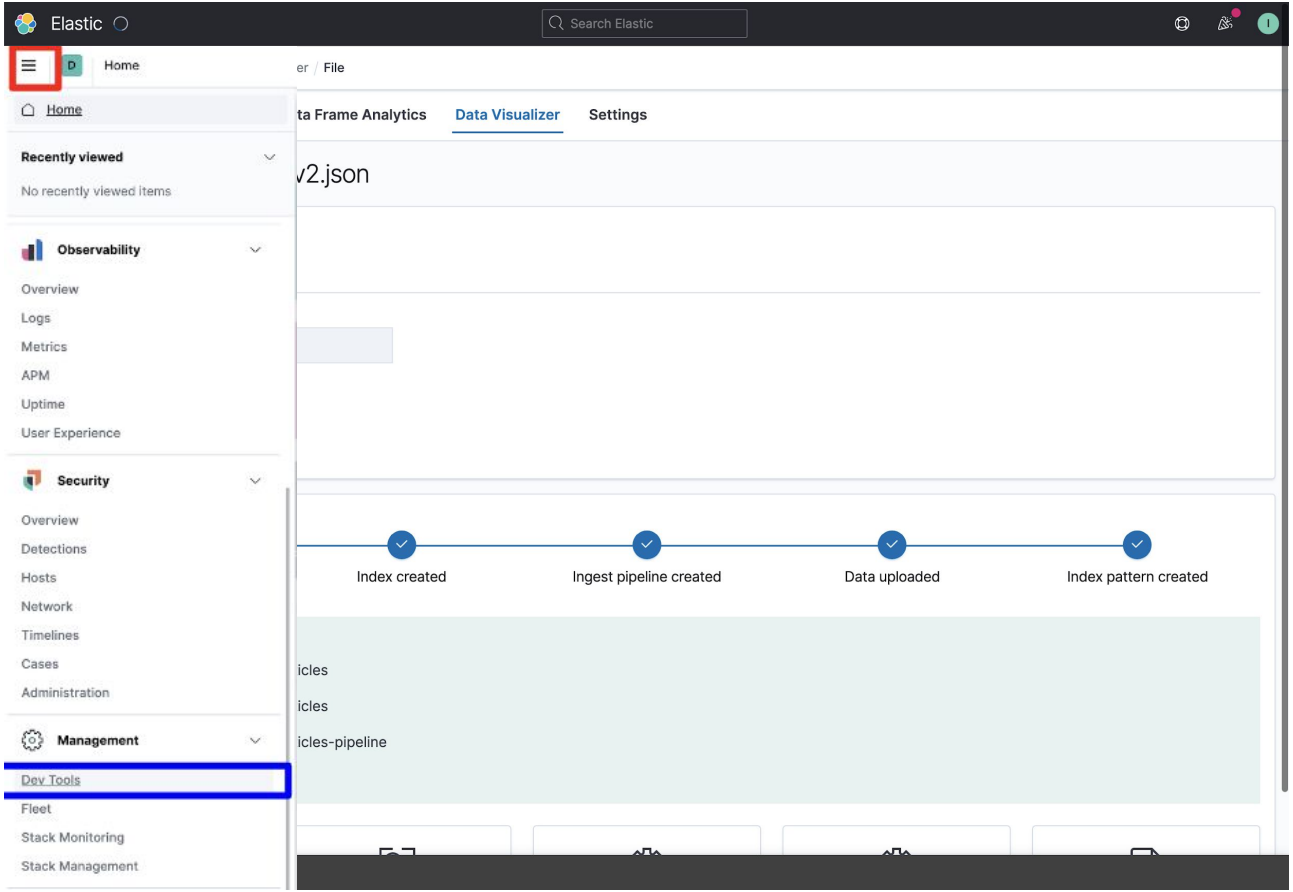


Then Elasticsearch will take of the rest!



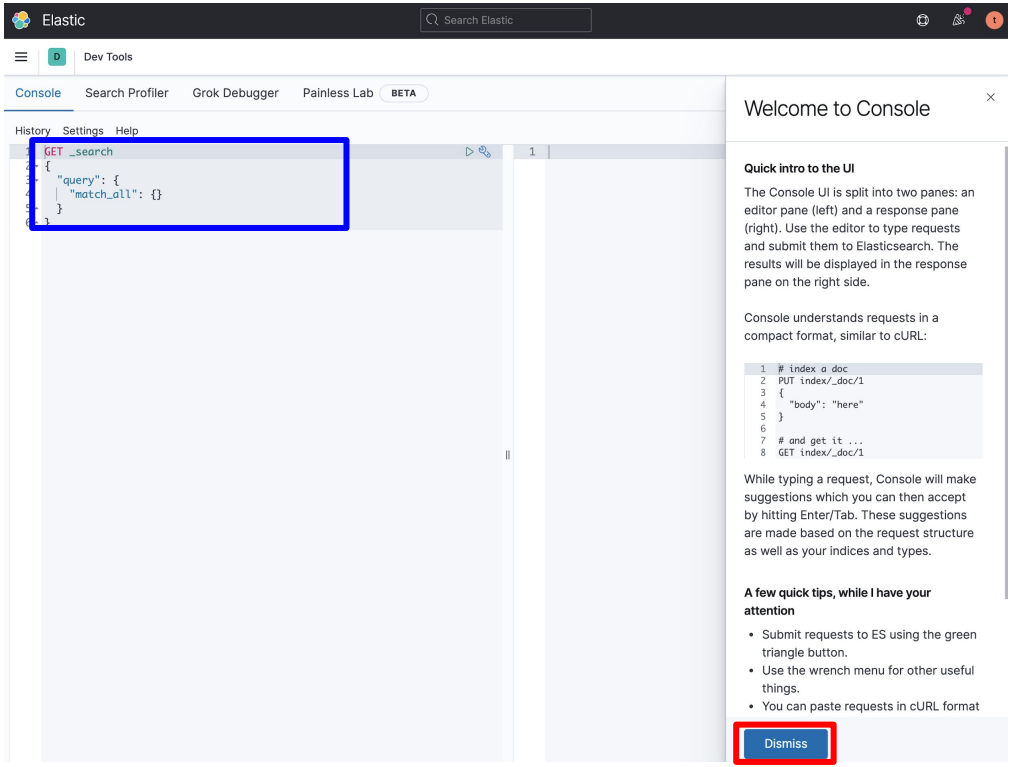
elastic

Click on menu icon, and open Dev Tools.



elastic

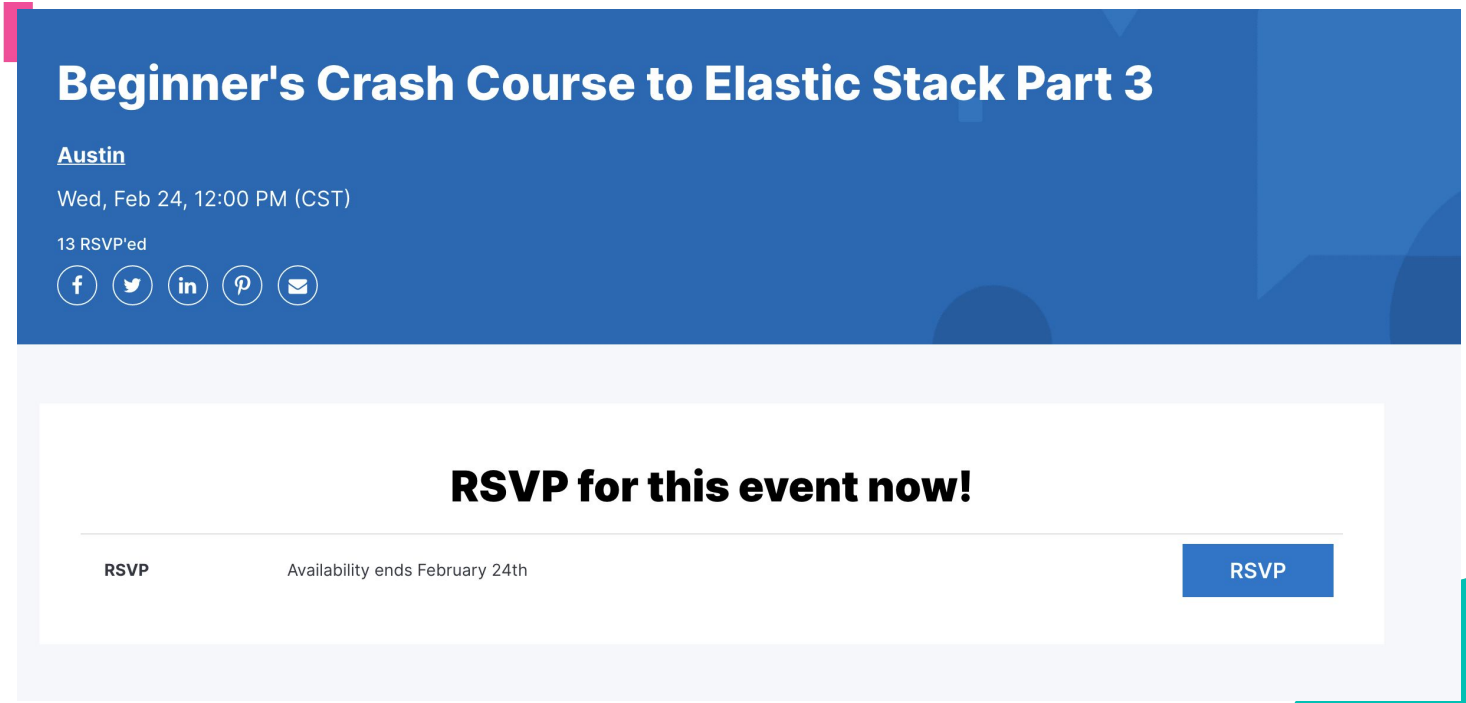
Click on dismiss and delete the default query.



Fine tuning precision or recall using Elasticsearch and Kibana

Questions?

Join us for the part 3 of the workshop series!




Lisa Jung

Developer Advocate @Elastic

Discussion forum: https://discuss.elastic.co/

Blog: https://dev.to/lisahjung

Twitter: @LisaHJung