



Exploring Data Insights with Elasticsearch, Kibana, and Bandcamp

July 2024

Justin Castilla - Senior Developer Advocate



Scan the code to access this **deck and repository**



Written Documentation

- This Slide Deck
- Walkthrough of each aggregation
- Links to further reading and suggested exploration



Code and Script Samples - Python

- Connecting to Elastic
- Consuming Bandcamp API data
- Creating Documents in Elasticsearch



What audience this talk is **Intended** for:

- Developers new to Elastic, yet comfortable with document-based data stores, APIs, and data
- Developers who throw everything into ES, yet haven't done anything other than search
- People who **love** music and code



What is Bandcamp?

The screenshot shows the Bandcamp homepage with a dark blue header featuring the Bandcamp logo and a search bar. Below the header are navigation links for Genres, Vinyl, CDs, Cassettes, T-shirts, Radio, Bandcamp Daily, and Best of's. A message at the top states: "We've updated our [Terms of Use](#) to reflect our new entity name and address. You can review the changes [here.](#)"

A large image of a band is displayed, followed by a "Bandcamp Weekly" section for July 9, 2024, featuring Brijean guests hosts sharing tracks from their new album 'Macro'. Below this is a "SELLING RIGHT NOW" section showing various album covers and a pink button labeled "WANNA BE MY LOVER".

Fans have paid artists **\$1.33 billion** using Bandcamp, and **\$193 million** in the last year.

<https://bandcamp.com/>

Bandcamp is a web platform for musicians and bands to share, sell, and promote their music directly to fans.

It allows artists to retain a higher percentage of revenue from their sales while providing fans with a diverse range of independent music.

Bandcamp also has a great API endpoint of sales data

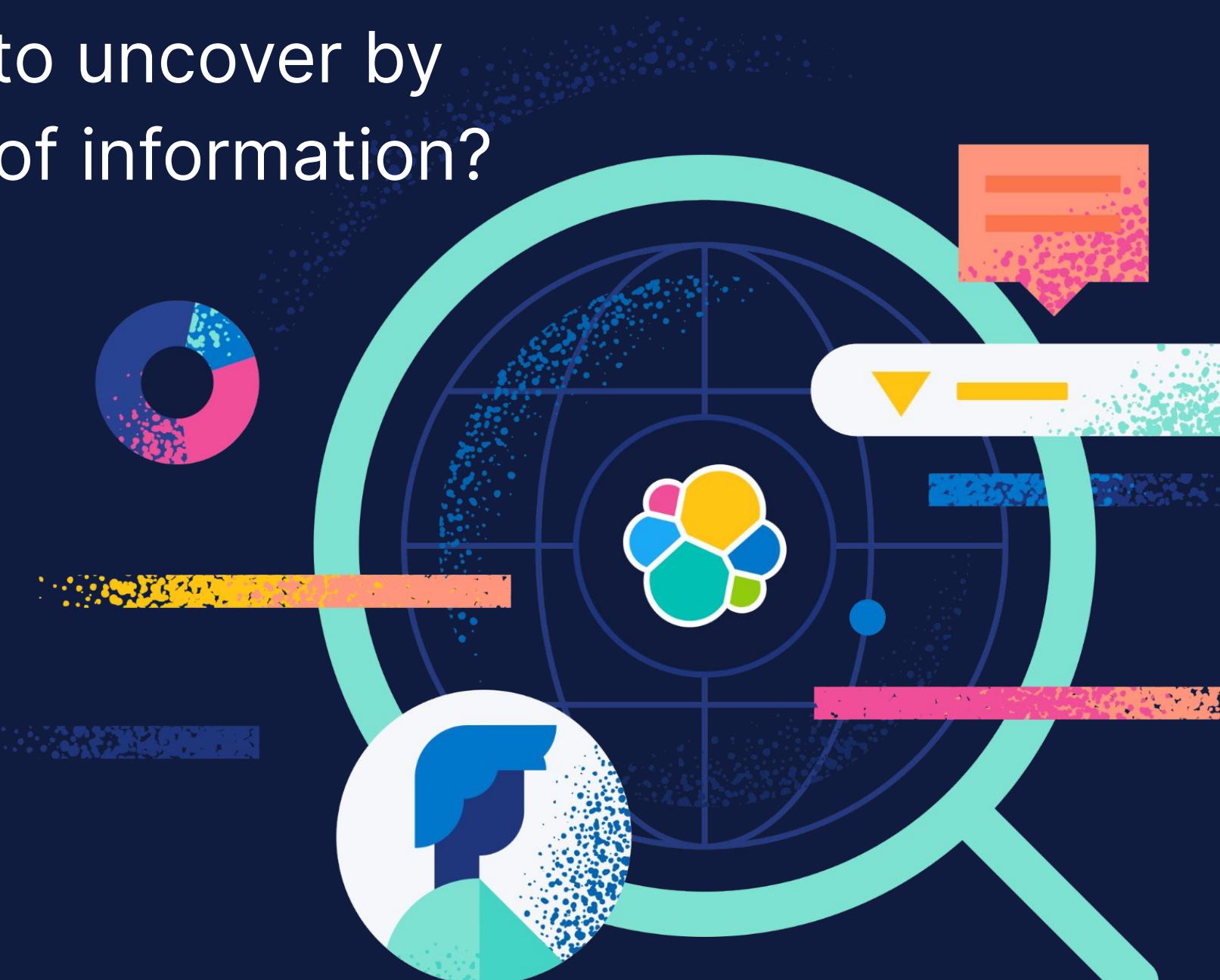
Let's take a peek with Postman!

<https://bandcamp.com/api/salesfeed/1/get>

Challenge

Given a live stream of publicly-visible purchases on bandcamp with interesting sales data what are we able to observe and glean over time?

What new insights and trends are we able to uncover by grouping data and creating new “buckets” of information?



Solution

1. Export the **Bandcamp** data into an **Elasticsearch** index
2. Create **Aggregations** over the pertinent fields
3. **Discover** information not readily available
4. Visualize new results in **Kibana**



1. Export the **Bandcamp** data into an **Elasticsearch** index

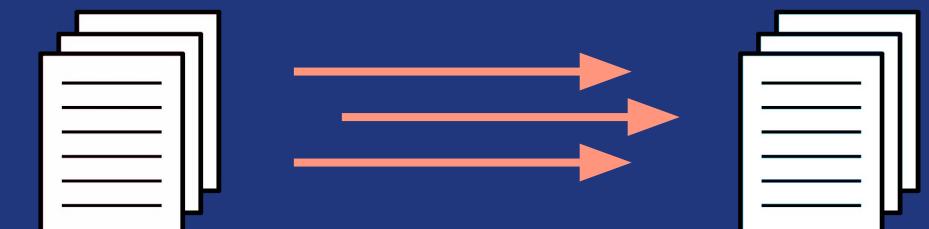


Bandcamp API

Automated Python Script

Fetch sales data from Bandcamp
every two minutes

Light formatting & filtering



Insert documents into Elastic via
Elasticsearch client in a purchases index



Elastic Search



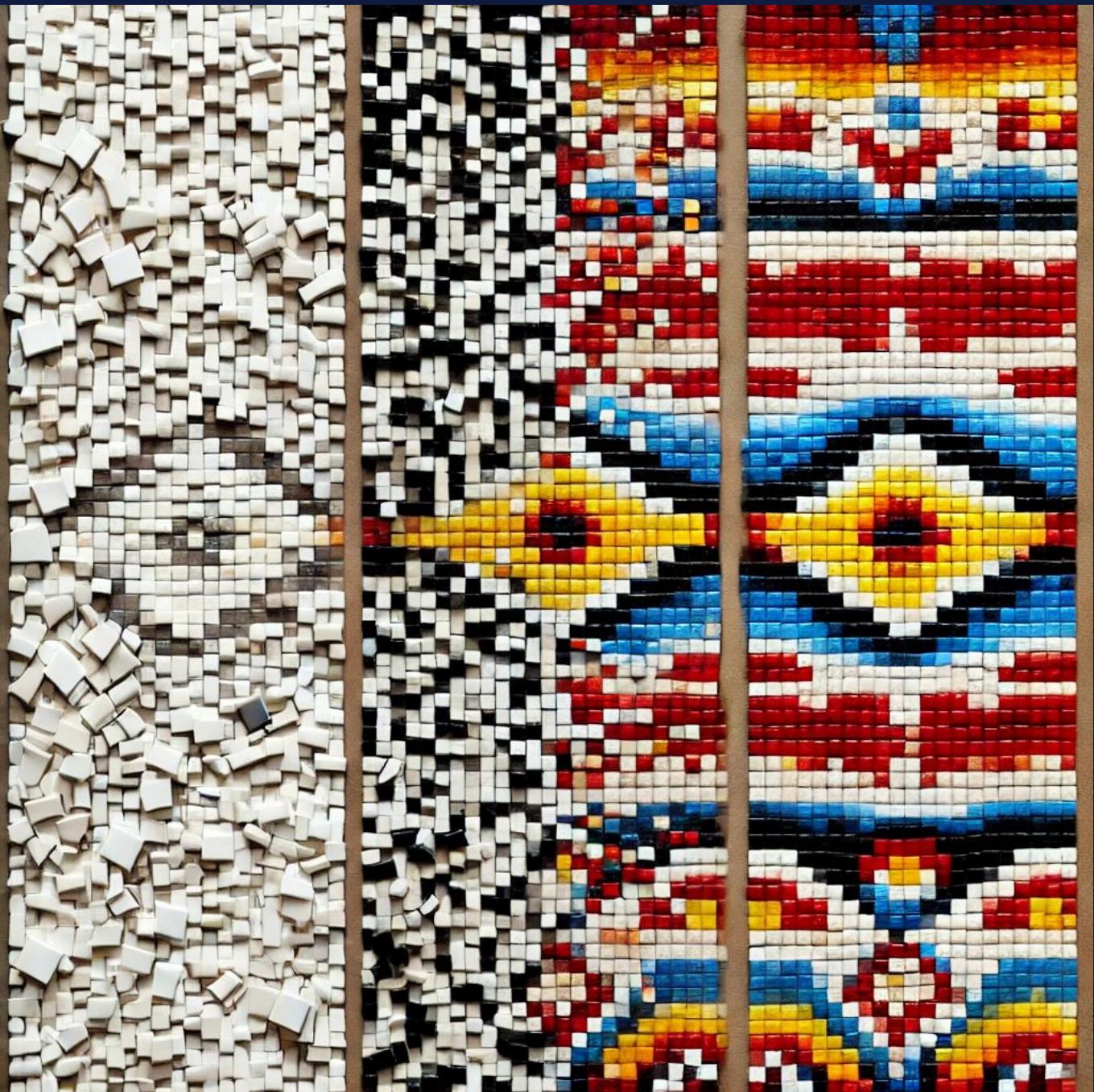
1. Export the **Bandcamp** data into an **Elasticsearch** index

Let's look at the code!

2. What are Aggregations over the pertinent fields

Aggregations summarize, analyze, and extract new data from your existing data by grouping and performing calculations on the indexed documents.

It works similarly to SQL's GROUP BY clauses but offers greater flexibility and nesting capabilities.



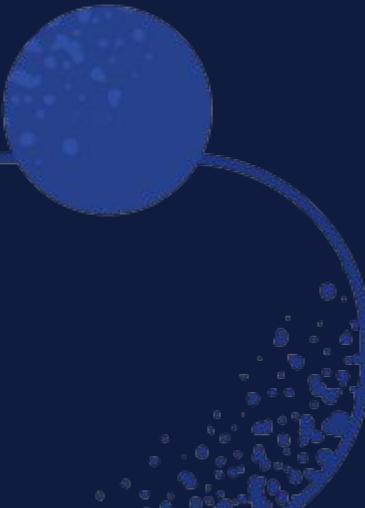
2. Create **Aggregations** over the pertinent fields

Real world **Example**

"I am a musician with a touring band ready to go out on the road. We keep a database of our merchandise and music being sold. We focus on the sales amount and location of every transaction ever recorded. By finding the top ten cities that were the location of the majority of our sales, we have a good idea of where we may sell more tickets than anywhere else."

2. Create **Aggregations** over the pertinent fields

To create Aggregations, we first need an index.
Why?

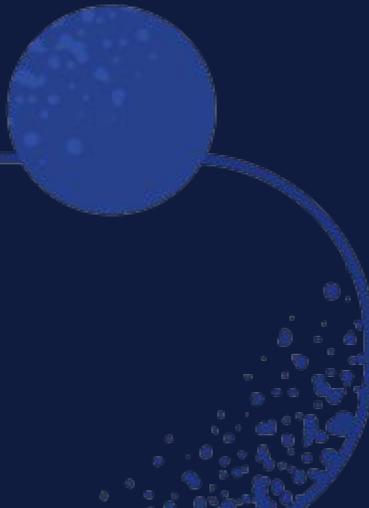


2. Create **Aggregations** over the pertinent fields

Structured Data Storage



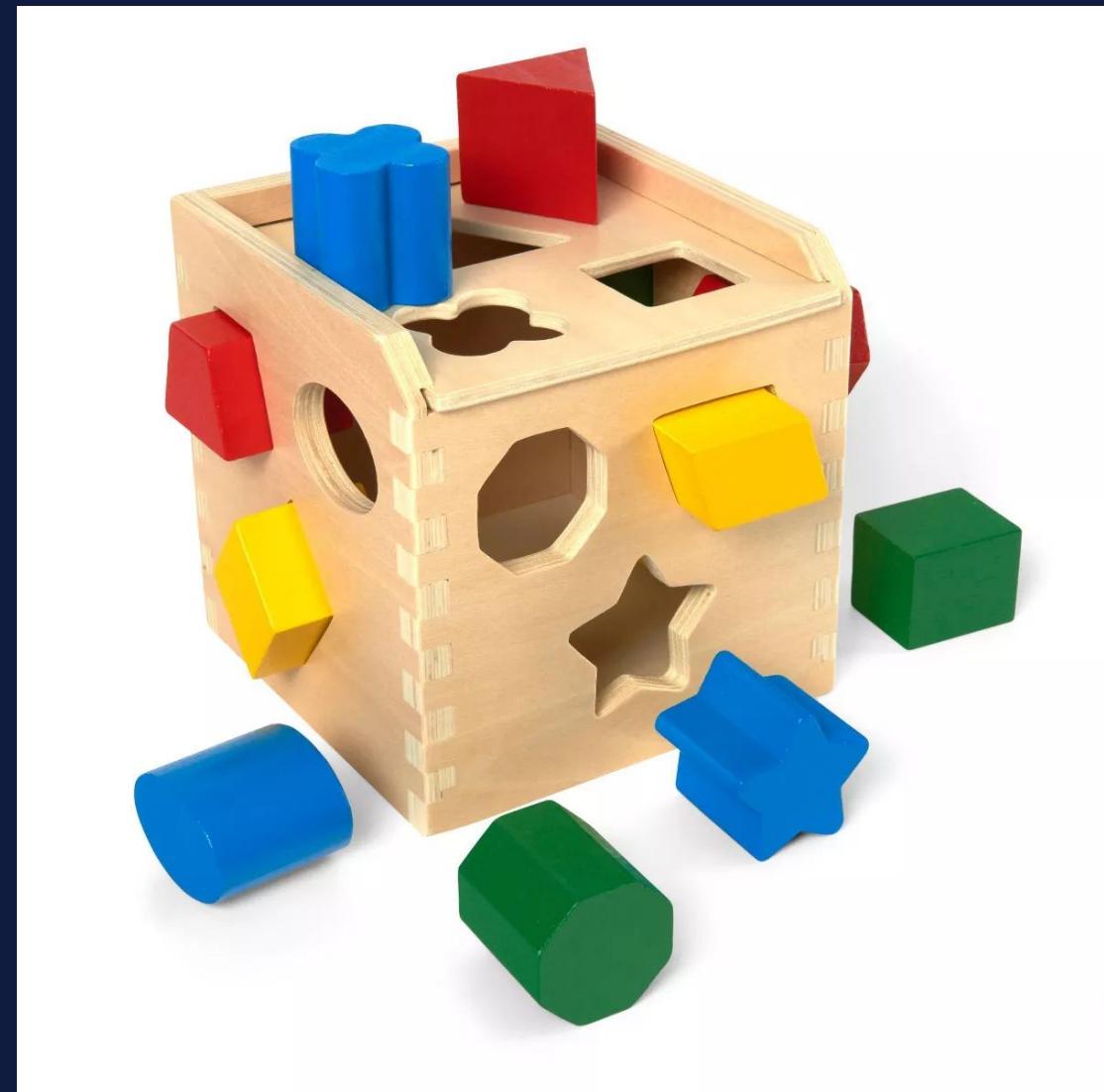
An Elasticsearch index is similar to a database in traditional SQL systems, where you store and organize your documents for efficient searching and analysis.



2. Create **Aggregations** over the pertinent fields

Each index has a mapping that defines the structure of the documents, specifying field types and how they should be indexed and stored.

Schema Definition



2. Create **Aggregations** over the pertinent fields

Scalability & Performance



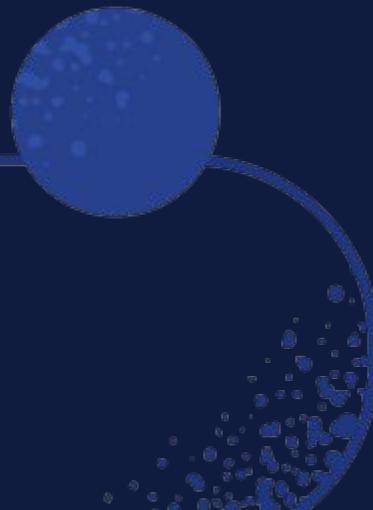
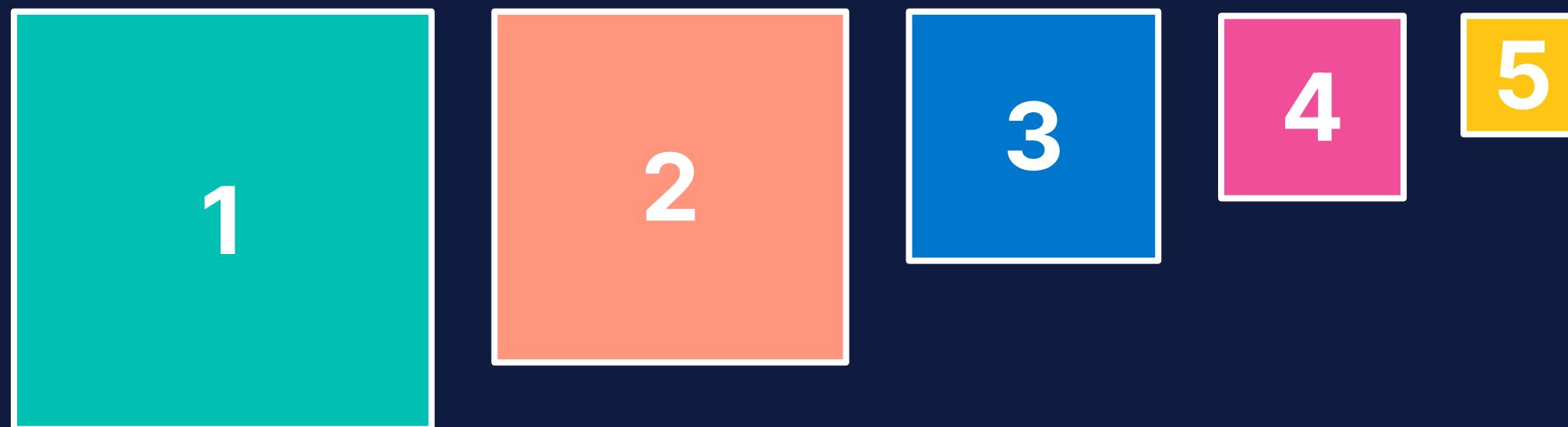
Elasticsearch indexes are designed to handle large volumes of data, offering fast search and retrieval capabilities through distributed architecture and advanced indexing techniques

2. Create **Aggregations** over the pertinent fields

What are the top 5 selling accounts within a given time span?

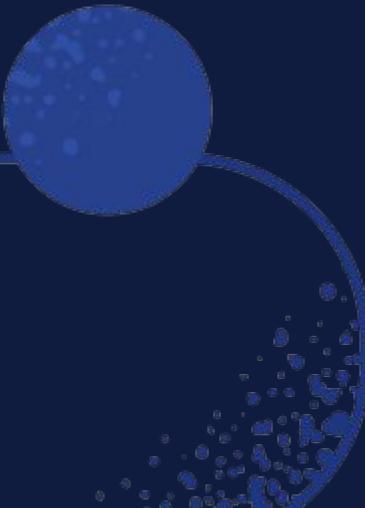
Clues:

- A time span means between two timestamps
- Top selling means a sum of all sales for each account
- Five results returned



2. Create **Aggregations** over the pertinent fields

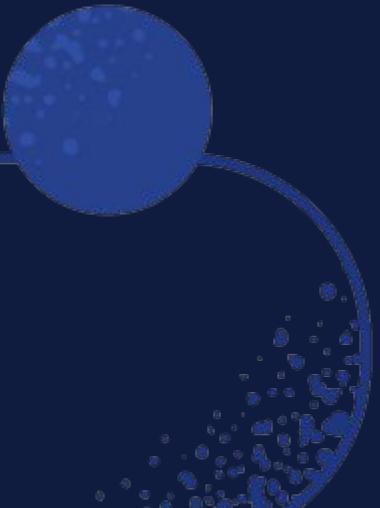
**Use the Elasticsearch Dev Tools Console to
execute API calls!**



3. Discover information not readily available

account_name	sales_quantity	total_amount_sold
aphextwin	956	\$28031.28
xtheband	613	\$16772.00
xclub	497	\$12109.96
mjlenderman	571	\$10783.27
iliantape	1052	\$ 8991.99

4. Visualize new results in **Kibana**



Ok what now?

- Interested in presenting at an in-person or virtual meetup? Email us at meetups@elastic.co
- Find the nearest Elastic User Group chapter and subscribe to stay up to date on upcoming events:
<https://www.meetup.com/pro/elastic/>



Elastic Contributors

This community program is designed to recognize and reward the hard work of our awesome contributors. Join this friendly competition and earn points for:

Code contributions | Presentations | Video tutorials | Event organization
Translations | Technical Q&A | Written content | Content validation

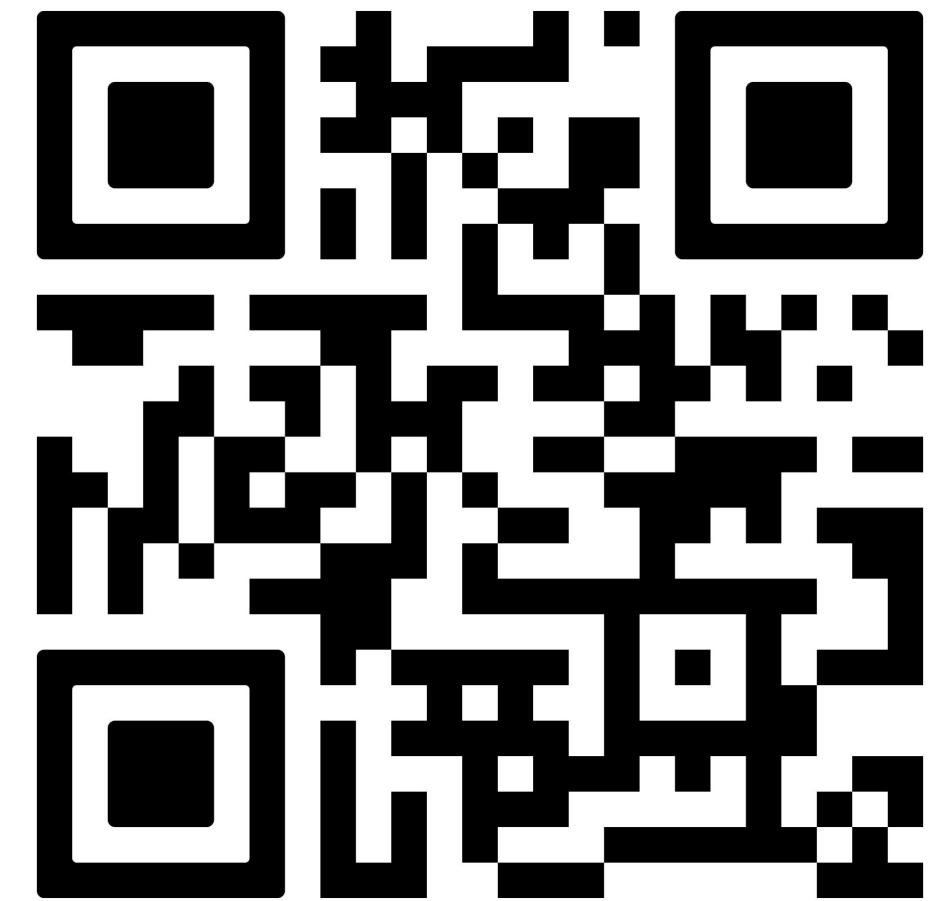
The top contributors will win cool prizes such as **Macbooks**, **Elastic trainings**, **certificate exams** and more. Check out the rules on our [website](#) and start submitting today!



elastic.co/community/contributor

The current cycle ends on January 31st 2023, but contributions made after February 1st 2022 can still be submitted.

Thank you!



Do you have Elastic-related questions? We have answers.

The screenshot shows the official Elastic Community YouTube channel. It features a blue header with the 'elastic' logo and the word 'Community'. Below the header, there's a section for the 'Elastic Community Conference 2022 - English' with a 'Play all' button. Underneath, there are sections for 'Beginner's Crash Course to Elastic Stack' and 'Official Elastic Community' with various video thumbnails and descriptions.

The screenshot shows the Elastic Community forum interface. On the left, there's a sidebar with 'Topics', 'Drafts', 'More', 'Categories' (which is expanded to show 'Announcements', 'Elastic Stack', 'Elastic Search', 'Elastic Observability', 'Elastic Security'), 'Tags' (with items like 'filebeat', 'docker', 'elastic-stack-security', 'metricbeat', 'elastic-stack-alerting'), and 'All tags'. The main area has tabs for 'categories', 'tags', 'Categories', 'Latest', and 'Top'. It displays several categories: 'Announcements' (1/week), 'Elastic Stack' (173/week), 'Elastic Search' (6/week), and 'Elastic Observability' (9/week). Each category lists recent posts with details like title, author, and timestamp.

The screenshot shows the '#announcements' channel in the Elastic Community Slack workspace. It has a pinned message about Elasticsearch 8.13. The channel lists several messages from users like 'Notes on Using These Forums', 'Forcemerger of index with vectors takes long, many hours or forever', 'Not getting any results for a field that requires exact match', 'Kibana behind firewall', 'Unable to install sample Web Logs data set in strigo lab for Practice Exam', 'Logstash with Docker (Unable to connect to database)', and 'Text_embedding configured for model but rejected for query'. At the bottom, there are reaction counts for each message.

Community Youtube Channel

Discuss

Elastic Community Slack Workspace