

07 - Vyhledávání v Kibaně

V Kibaně je možné v sekci **Discover** psát jednoduché full textové dotazy. Všechny nalezené dokumenty pak budou zobrazeny v tabulce. Díky našeptávání je pak psaní dotazů opravdu snadné.

V UI Kibany je na výběr z dvou různých syntax - Kibana Query Language (KQL) Query String (respektive Lucene query). Ta první má jednodušší syntax a je lépe integrována do Kibany, druhá zase nabí o něco více funkcionalit.

Kibana Query Language

KQL dokumentace: <https://www.elastic.co/guide/en/kibana/current/kuery-query.html>

Nejsnazším způsobem vyhledávání je napsat dotaz do vyhledávacího pole:

The screenshot shows the Kibana Discover interface. The search bar at the top contains the query 'betty@cortez-family.zzz'. The left sidebar shows the 'kibana_sample_data_ecommerce' index pattern. The main view displays 111 hits in a table. The table has columns: Time, order_id, email, geoip.city_name, and taxful_total_price. The first few rows show results for 'diane@cortez-family.zzz', 'betty@cortez-family.zzz', 'betty@moran-family.zzz', 'betty@rowe-family.zzz', 'betty@hansen-family.zzz', 'betty@cross-family.zzz', 'betty@tran-family.zzz', and 'samir@cortez-family.zzz'. A bar chart is visible above the table, showing the distribution of hits over time.

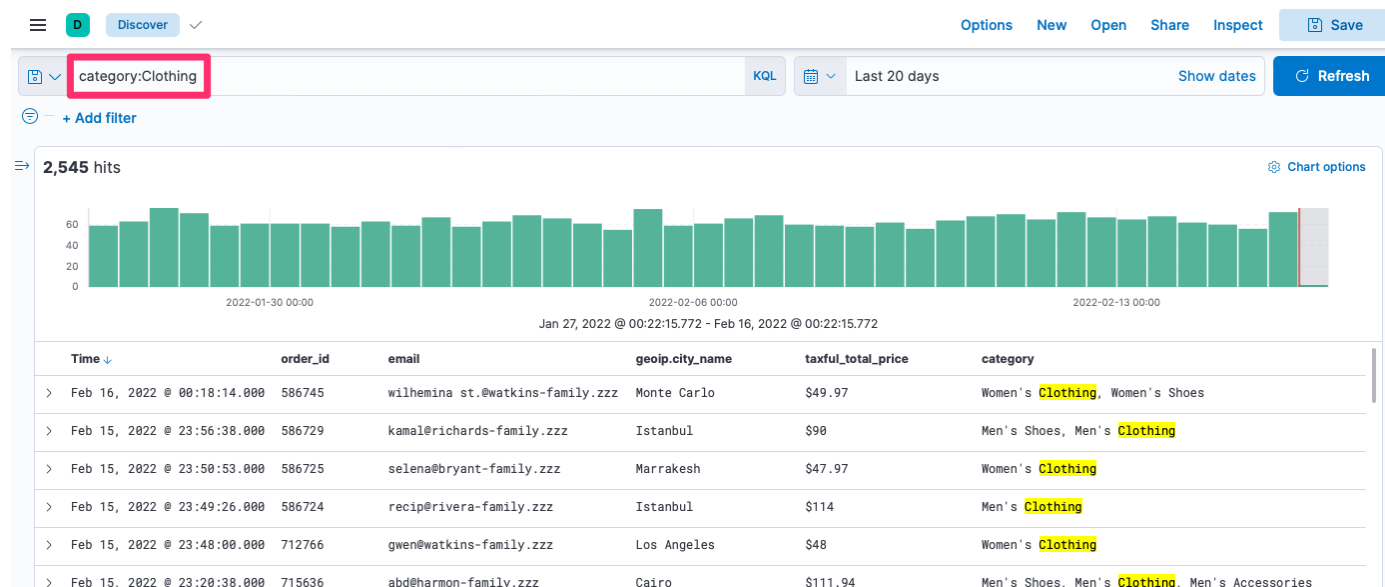
Time	order_id	email	geoip.city_name	taxful_total_price
Feb 15, 2022 @ 22:48:58.000	586676	diane@cortez-family.zzz	-	\$45.97
Feb 15, 2022 @ 21:57:07.000	712647	betty@cortez-family.zzz	New York	\$48
Feb 15, 2022 @ 21:03:50.000	586564	betty@moran-family.zzz	New York	\$53.97
Feb 15, 2022 @ 18:55:41.000	586452	betty@rowe-family.zzz	New York	\$60.97
Feb 15, 2022 @ 17:20:38.000	586366	betty@hansen-family.zzz	New York	\$58.97
Feb 15, 2022 @ 10:51:50.000	585990	betty@cross-family.zzz	New York	\$46.97
Feb 15, 2022 @ 03:21:07.000	712795	betty@tran-family.zzz	New York	\$48
Feb 15, 2022 @ 02:07:41.000	585543	samir@cortez-family.zzz	Dubai	\$90

Pokud není uvedeno, vyhledává se ve všech polích. Pro vyhledávání pouze v určitých polích je nutné je uvést:

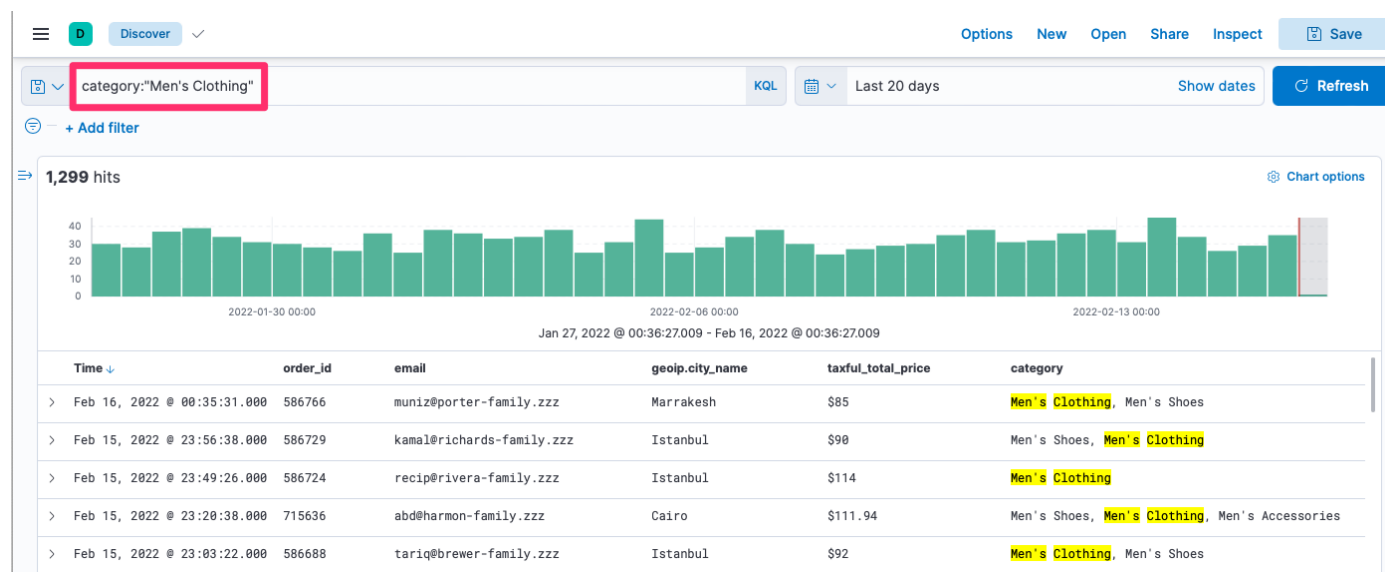
The screenshot shows the Kibana Discover interface with the search query 'email:betty@cortez-family.zzz'. The left sidebar is the same. The main view displays 1 hit in a table. The table has columns: Time, order_id, email, geoip.city_name, and taxful_total_price. The single row shows results for 'betty@cortez-family.zzz' with order_id 712647 and taxful_total_price \$48. A bar chart is visible above the table, showing a single bar for the hit.

Time	order_id	email	geoip.city_name	taxful_total_price
Feb 15, 2022 @ 21:57:07.000	712647	betty@cortez-family.zzz	New York	\$48

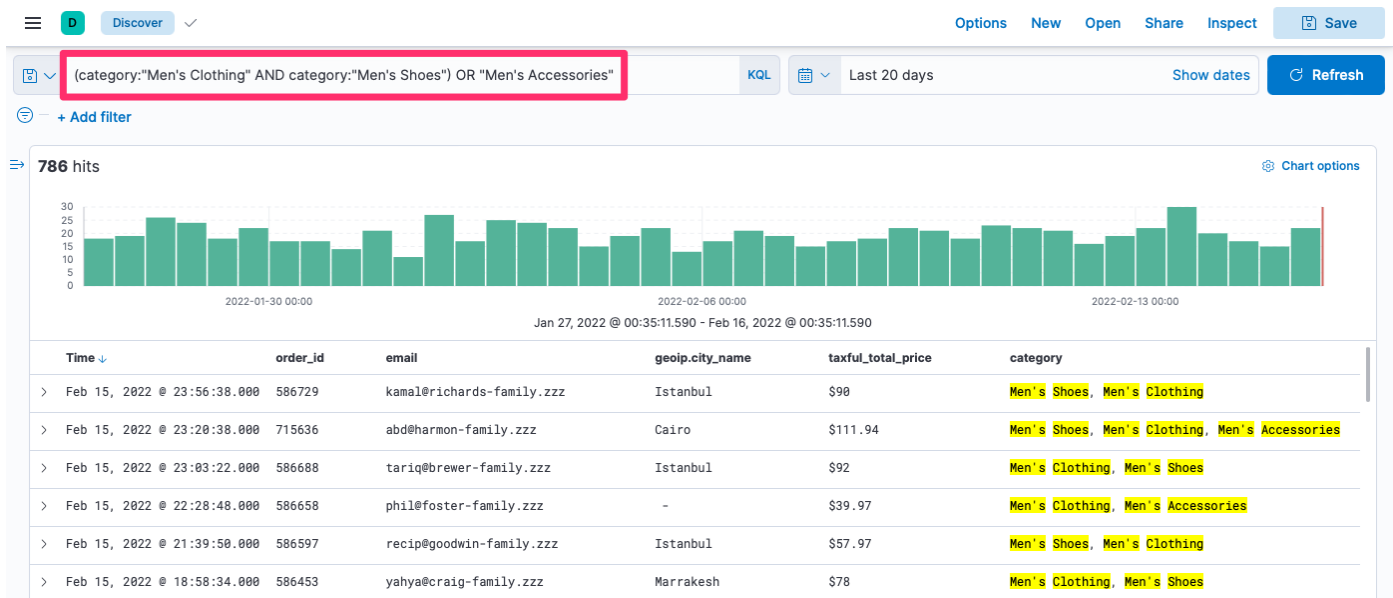
Pokud je pole uloženo jako typ `keyword`, je nutné při vyhledávání uvést jeho přesnou podobu. Pokud je však uvedeno jako `keyword`, stačí shoda v libovolném slově. Například na poli `category`:



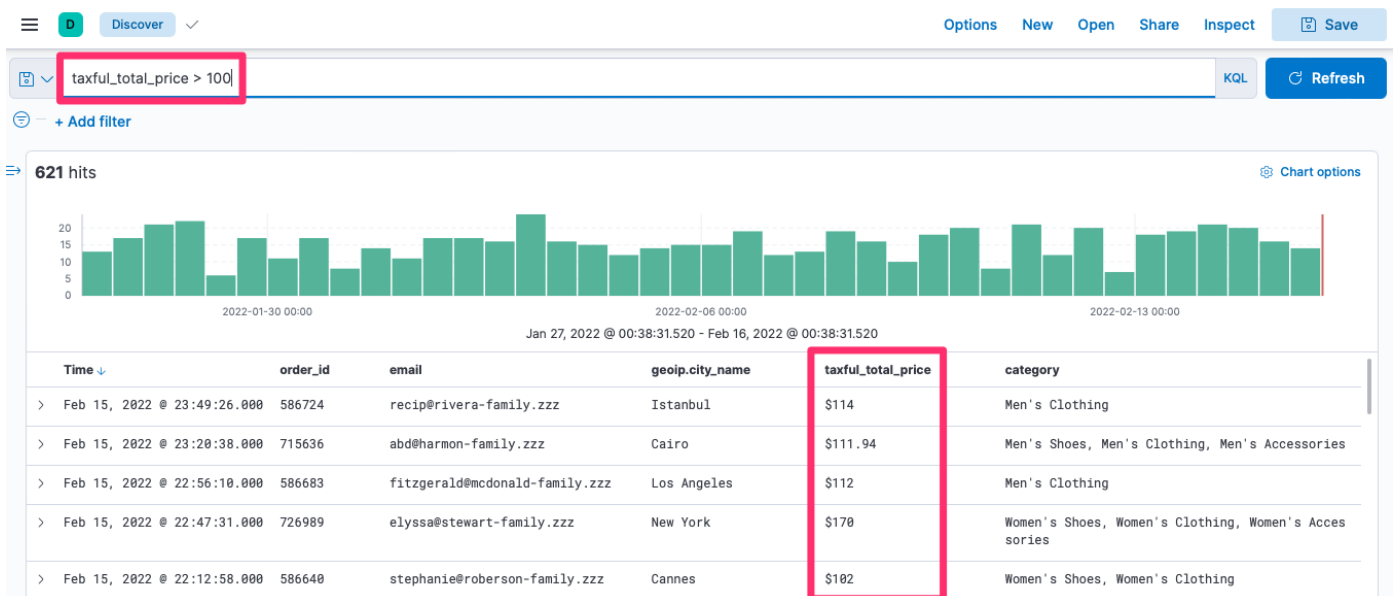
U polí typu `text` není nutné psát všechna slova v shodném pořadí. Pokud ale na dodržení pořadí trváme, lze použít frázové vyhledávání pomocí uvozovek:



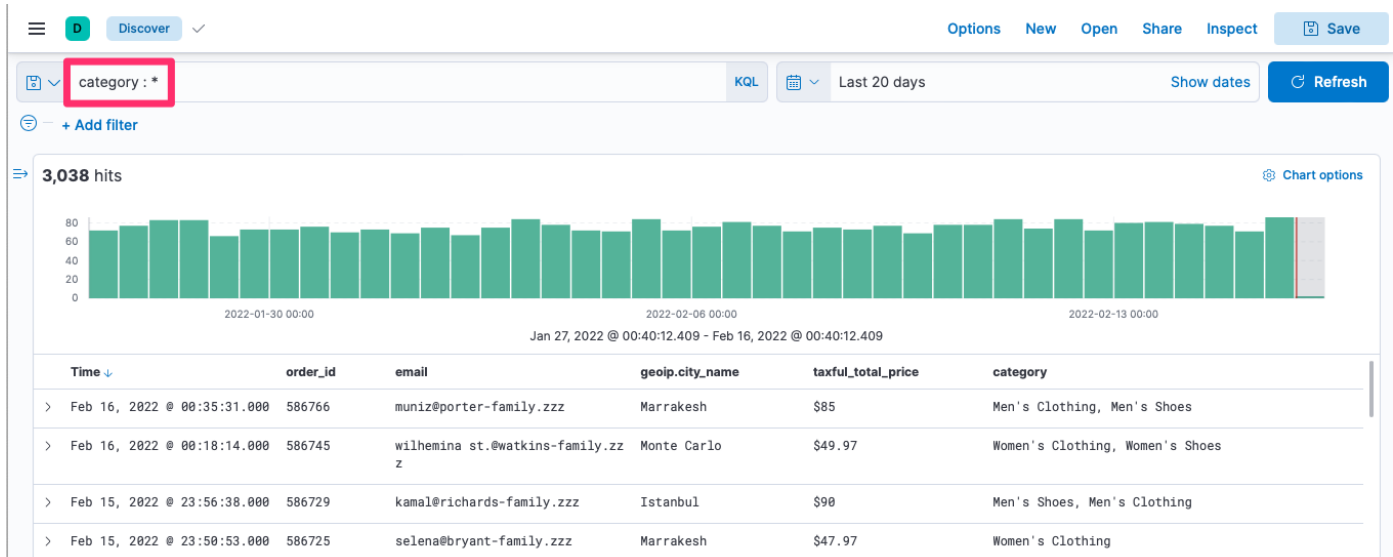
Dotazy lze spojovat pomocí spojek `AND`, `OR` a `NOT`. Priority lze určit kulatými závorkami:



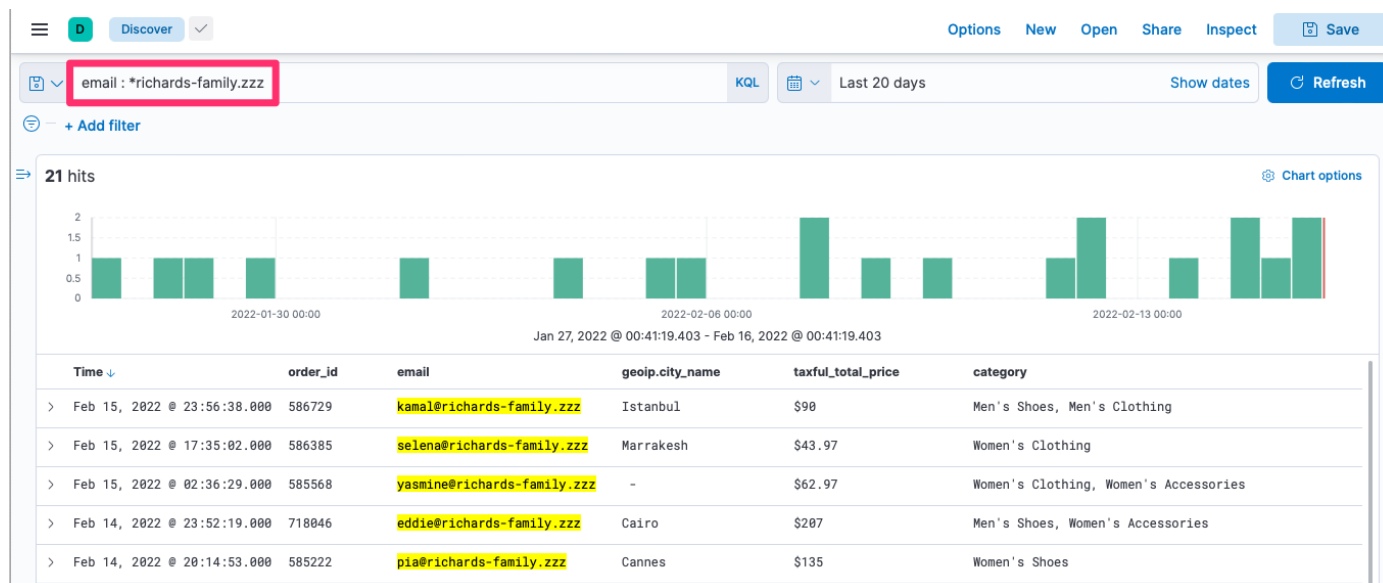
Filtrovat číselné hodnoty (nebo datum) lze pomocí $>$, $<$, $>=$, $<=$:



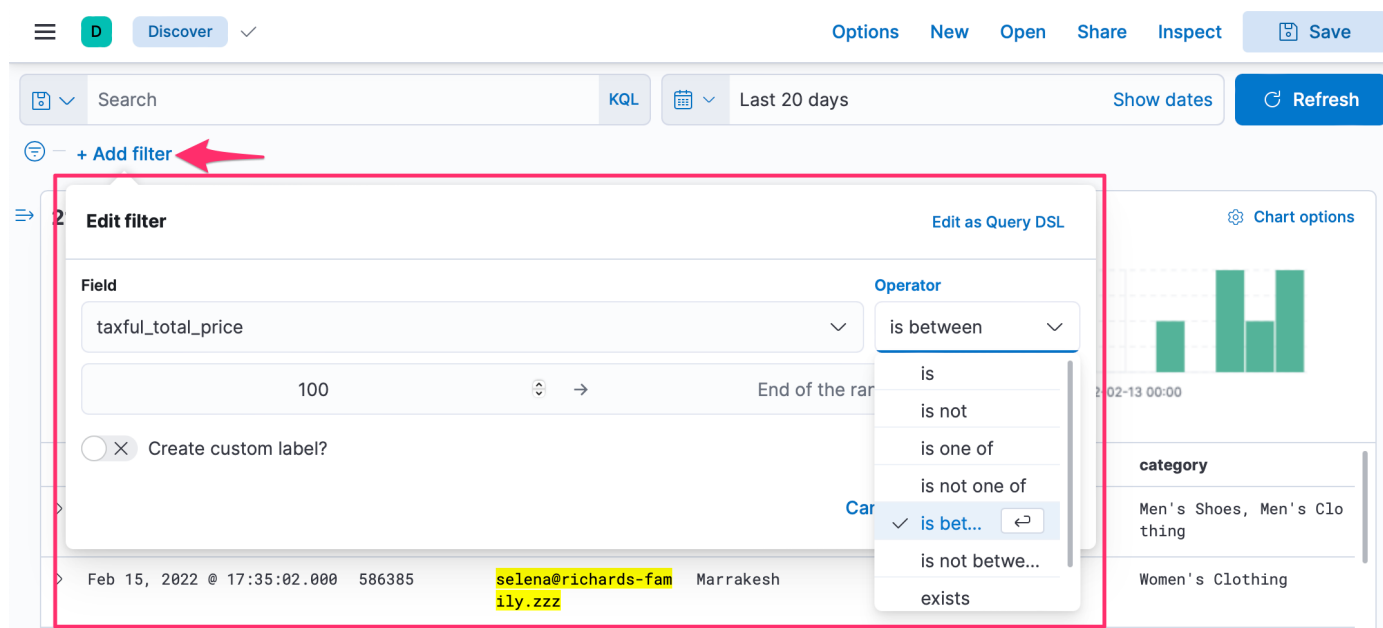
Pro vyhledání všech dokumentů, které obsahují dané pole s libovolnou (neprázdnou) hodnotou lze využít znak $*$:



Hvězdičku lze také využít pro částečnou shodu, což funguje perfektně s typem `keyword`. Například pro zobrazení všech e-mailů z domény `richards-family.zzz`:



Další možností filtrace dat v tabulce je využití filtrů - **filters**. Filtr lze vytvořit pomocí tlačítka plus (respektive minus) vedle každé hodnoty v tabulce. Nebo jej lze ručně definovat kliknutím na `+ Add filter`:



Filtry a full textové dotazy lze kombinovat. Každý filtr může být upraven, dočasně zakázán, nebo může být invertována jeho logika:

Search

taxful_total_price: \$100 to +∞ × + Add filter

518 hits

order_date ⌚ ↓ Document

Feb 19, 2022 @ 17:36:29.000 category: Men's Clo

Jakmile jsou query a filtry hotové, lze si je uložit pro pozdější použití:

Discover ✓ Options New Open Share Inspect Save

email: *richards-family.zzz KQL Last 20 days Show dates Refresh

Saved Queries

There are no saved queries. Save query text and filters that you want to use again.

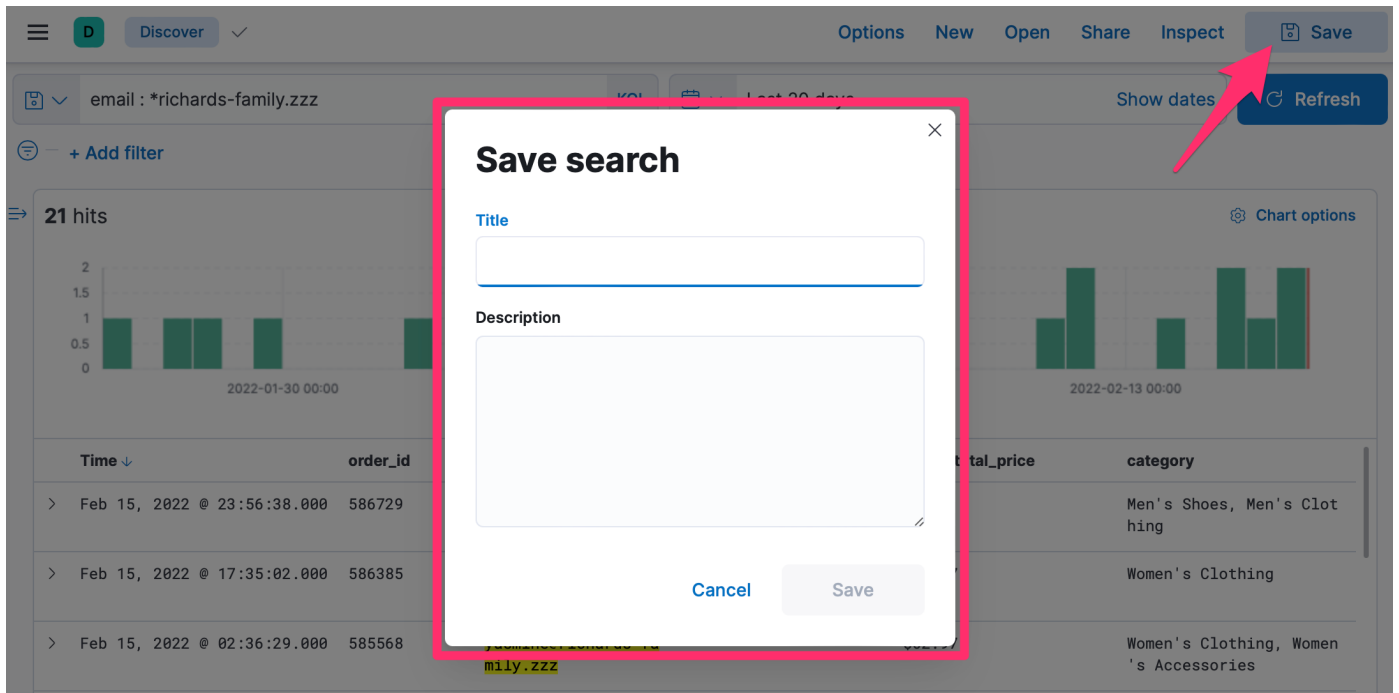
Save current query

2022-01-30 00:00 2022-02-06 00:00 2022-02-13 00:00

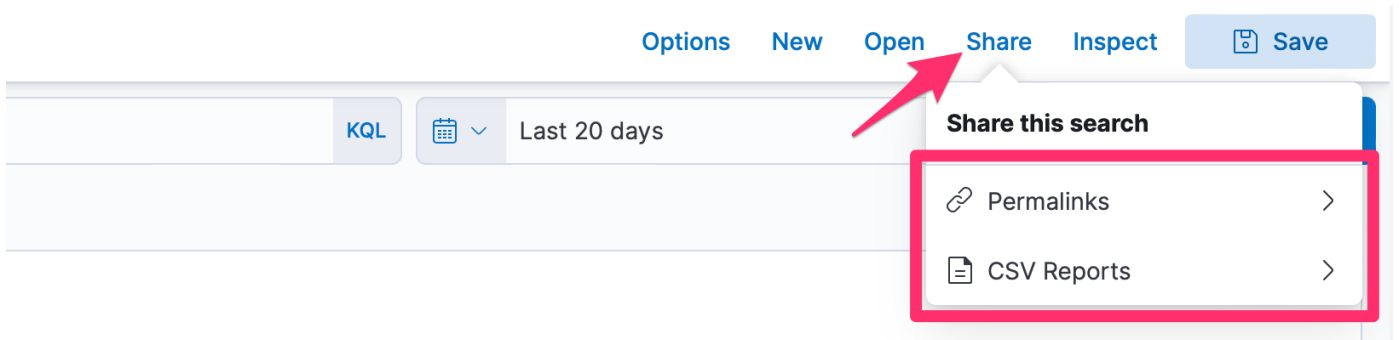
Jan 27, 2022 @ 00:41:19.403 - Feb 16, 2022 @ 00:41:19.403

Time ↓	order_id	email	geoip.city_name	taxful_total_price	category
> Feb 15, 2022 @ 23:56:38.000	586729	kamal@richards-family.zzz	Istanbul	\$90	Men's Shoes, Men's Clothing

Lze také uložit kompletně celou tabulku (včetně vyhledávání):

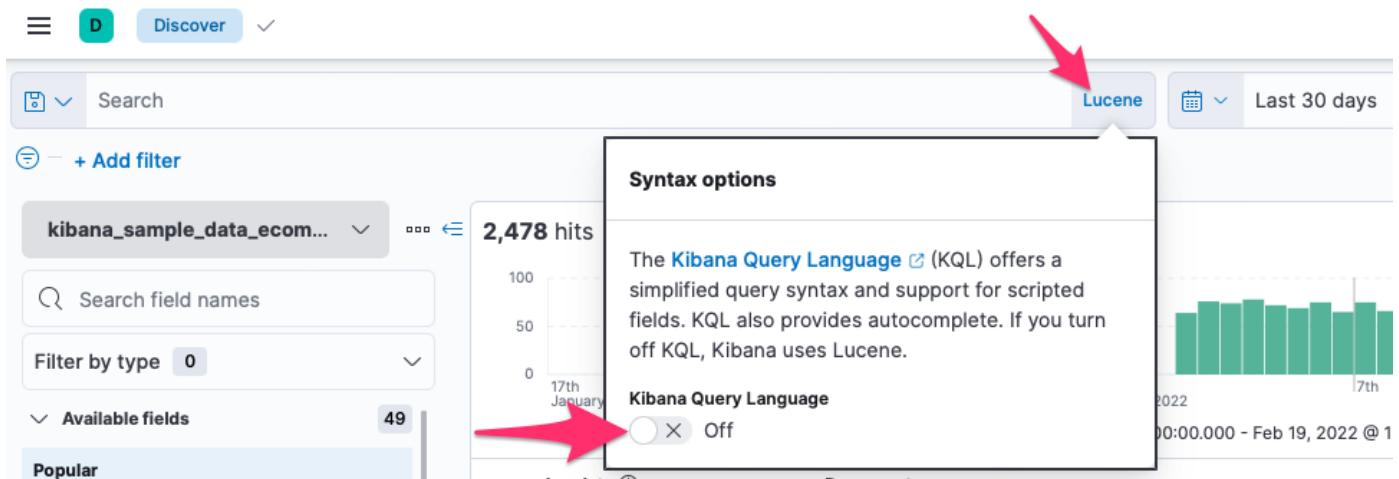


V neposlední řadě je také možné tabulku sdílet pomocí odkazu, nebo si data stáhnout jako CSV soubor:



Query String

Query String (nebo také Lucene Query) je alternativa k KQL. Lucene nám umožňuje psát o něco pokročilejší dotazy - například povolit vyhledávání s překlepy. Na druhou stranu však nefunguje našeptávání. Do módu Lucene se přepnete vypnutím KQL:



[Syntax](#) je až na pár odlišností podobná KQL

- Vyhledávání ve všech polích: `Shoes`
- Frázové hledání: `"Men's Shoes"`
- Logické spojky (AND, OR): `Man OR Women`
- Negace: `NOT Shoes`
- Udávání priorit závorkami: `(Man OR Women) AND Shoes`
- Kontrola existence pole: `_exists_:category`
- Wildcard pro jeden znak (`?`) a více znaků (`*`): `M?n Sh*`
- [Regulární výrazy](#) se píšou mezi lomítky: `/Shoes?/`
- Fuzzy vyhledávání (překlepy): `Shoes~`
- Rozsahy:
 - včetně hraničních hodnot: `price:[0 TO 100]`
 - vyjma hraničních hodnot: `price:{0 TO 100}`
 - datum - vše před rokem 2022: `date:{* TO 2022-01-01}`

Query string může být také součástí URL requestu, takže jej lze použít v dev tools, nebo také externí aplikaci dotazující se Elasticsearch. V tomto případě bude query uvedena jako parametr `q` endpointu `_search`:

```
GET kibana_sample_data_ecommerce/_search?q=Shoes
```

Úkol: Vyhledávání v Kibaně

1. Uložte následující dokumenty do Elasticsearch:

```
POST reviews/_doc
{
  "customer_id": 12932123,
  "email": "tom.k@gmail.com",
  "title": "8 Ball Pool",
  "comment": "It is a fun game but sometimes hard to control the cue.",
  "rating": 4,
  "date": "2022-02-21T15:29:50+02:00"
}

POST reviews/_doc
{
  "customer_id": 13937931,
  "email": "jane.f@gmail.com",
  "title": "8 Ball Pool",
  "comment": "Cannot download even though it says I purchased.",
}
```

```
"rating": 1,  
"date": "2022-02-18T12:22:40+02:00"  
}
```

POST reviews/_doc

```
{  
  "customer_id": 13937931,  
  "email": "jane.f@gmail.com",  
  "title": "Yes Chef!",  
  "comment": "Love it!",  
  "rating": 5,  
  "date": "2022-02-18T11:13:34+02:00"  
}
```

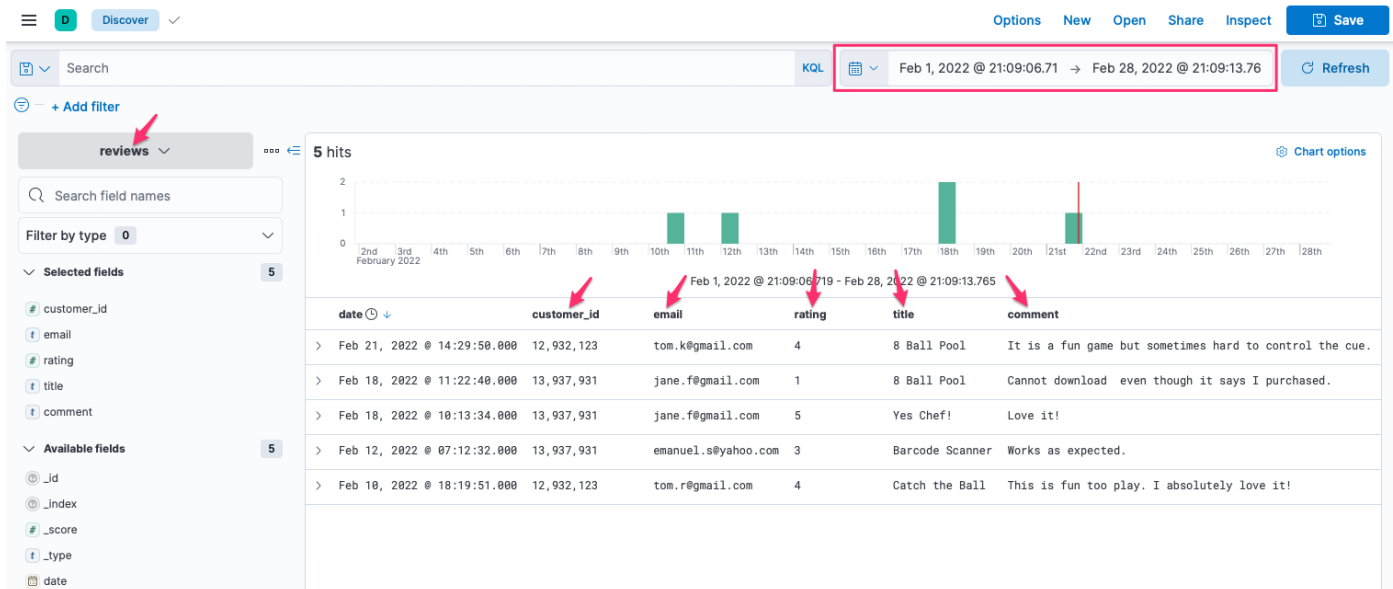
POST reviews/_doc

```
{  
  "customer_id": 13937931,  
  "email": "emanuel.s@yahoo.com",  
  "title": "Barcode Scanner",  
  "comment": "Works as expected.",  
  "rating": 3,  
  "date": "2022-02-12T08:12:32+02:00"  
}
```

POST reviews/_doc

```
{  
  "customer_id": 12932123,  
  "email": "tom.r@gmail.com",  
  "title": "Catch the Ball",  
  "comment": "This is fun too play. I absolutely love it!",  
  "date": "2022-02-10T19:19:51+02:00"  
}
```

2. Vytvořte **Data view** pro index `reviews`. použijte `date` jako timestamp field.
3. Vytvořte tabulku v **Discover** s použitím data view `reviews` s použitím následujících sloupců (nezapomeňte změnit vhodně time range):



4. Napište vyhledávací dotazy, které vyfiltrují data:

1. Vyhledejte dokumenty obsahující slovo `Ball` v titulku (pole `title`)
2. vyhledejte dokumenty s `customer_id` větším než `12932200` a jejichž e-mailová adresa (pole `email`) končí `gmail.com`
3. Vyhledejte dokumenty vytvořené před `Feb 15 2022` a zároveň jejichž `title` není `Barcode Scanner`
4. Vyhledejte dokumenty, jejichž `title` je přesně `8 Ball Pool`
5. Vyhledejte dokumenty, které neobsahují pole `rating`
6. Vyhledejte dokumenty obsahující `sometimes` v poli `comment` — s povolenými překlady