Elastic Stack을 활용한 Data Dashboard 만들기

Week 2 - Kibana를 더 잘 사용해보자

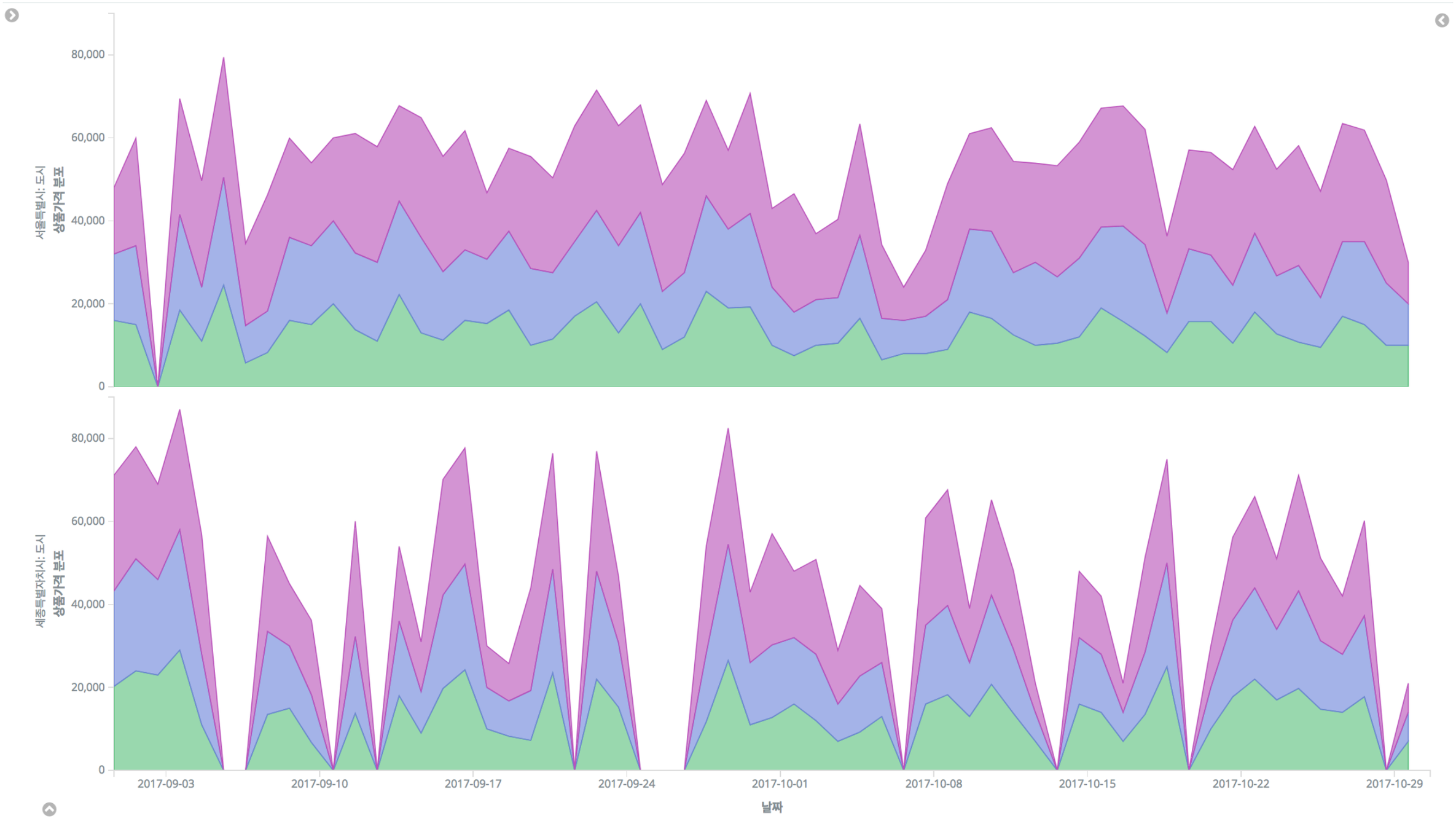
Fast Campus

목차

- 실전 Visualize
  - Area Map 4
  - Gauge 8
  - Heat Map 12
  - Data Table 16
- 실전 Dashboard 20
- 몇 가지 고민들
  - data format 변경 31
  - csv 출력 56
  - JSON Input 61
  - Scripted Field 79
- 데이터 조회
  - Filtering by Field 101
  - Lucene Query 114

문제 1 - Area Map ✍ 👑

Visualize - Area Map



Visualize - Area Map

데이터

- Index : shopping
- Time Range : 2017년 9월 1일 ~ 2017년

문제

- “상품가격”의 합이 가장 큰
- “고객주소_시도” 2개의
- “상품가격”의 25분위, 50분위, 95분위를
- “주문시간”을 기준으로 daily로 표시하고
- week2_{id}_area로 저장하세요

사용한 Aggregation

- Metrics (Y-Axis) : **Percentiles** Aggregation
- Buckets
  - Split Chart (Rows) : **Terms** Aggregation
  - X-Axis : **Date Histogram** Aggregation

순서에 주의!!

순서가 바뀌면 결과가 왜 바뀌는지 생각해보자

사용 필드

- 상품가격
- 고객주소_시도
- 주문시간

Visualize - Area Map

metrics

Y-Axis

Aggregation

Percentiles

Field

상품가격

Percents

25

+ Add

Custom Label

상품가격 분포

buckets

Split Chart

Rows | Columns

Aggregation

Field

고객주소_시도

Order By

Custom Metric

Aggregation

Field

상품가격

Advanced

Order **Size**

Descending

Custom Label

X-Axis

Sub Aggregation

Date Histogram

Field

Interval

Custom Label

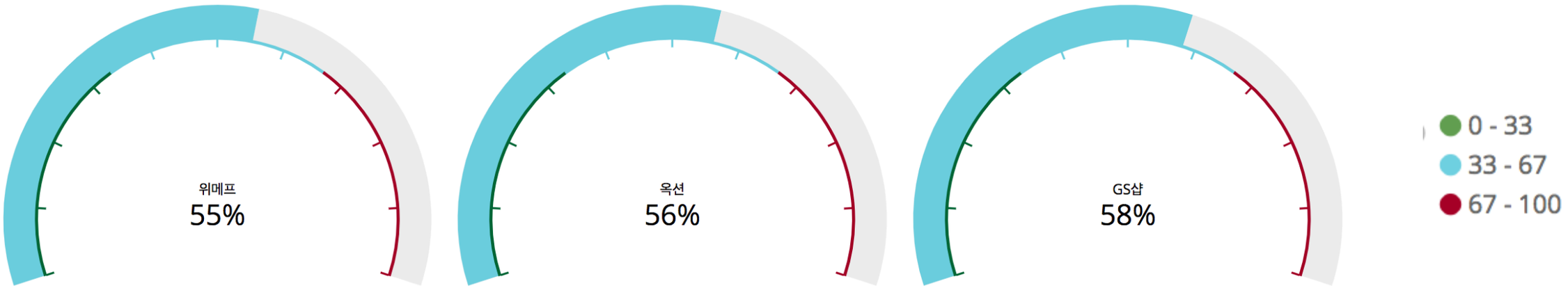
날짜

Advanced

Add sub-buckets

문제 2 - Gauge ✍ 👑

Visualize - Gauge



일별 상품가격 평균의 전체 평균

Visualize - Gauge

데이터

- Index : shopping
- Time Range : 2017년 9월 1일 ~ 2017년

문제

- “판매자평점”의 평균이 큰
- “구매사이트” 3개를
- “주문시간” 을 기준으로 daily로 나누고
- daily “상품가격”의 평균을 낸 후
- daily “상품가격”의 평균의 전체 평균을 표시하고
- week2_{id}_gauge로 저장하세요

사용한 Aggregation

- Metrics
  - Aggregation : Average Bucket Aggregation
  - Bucket : Date Histogram Aggregation
  - Metric : Average Aggregation
- Buckets
  - Aggregation : Terms Aggregation

사용 필드

- 주문시간
- 상품가격
- 구매사이트
- 판매자평점

단계별 목표 구간

0 ~ 9999

10000 ~ 19999

20000 ~ 29999

단순히 Average Aggregation 사용하는 것과 뭐가 다를까?

또는 어떤 상황에서 두 값의 차이가 크게 날까?

Visualize - Gauge

metrics

Metric

Aggregation

Average Bucket

Bucket

Aggregation

Field

주문시간

Interval

Daily

Custom Label

Advanced

Metric

Aggregation

Field

상품가격

Custom Label

Advanced

Custom Label

일별 상품가격 평균의 전체 평균

buckets

Split Group

Aggregation

Terms

Field

Order By

Custom Metric

Aggregation

Average

Field

Advanced

Order **Size**

3

Custom Label

Data Options

Gauge Type Arc

Percentage Mode

Vertical Split

Show Legend

Show Labels

Sub Text

Auto Extend Range

Ranges

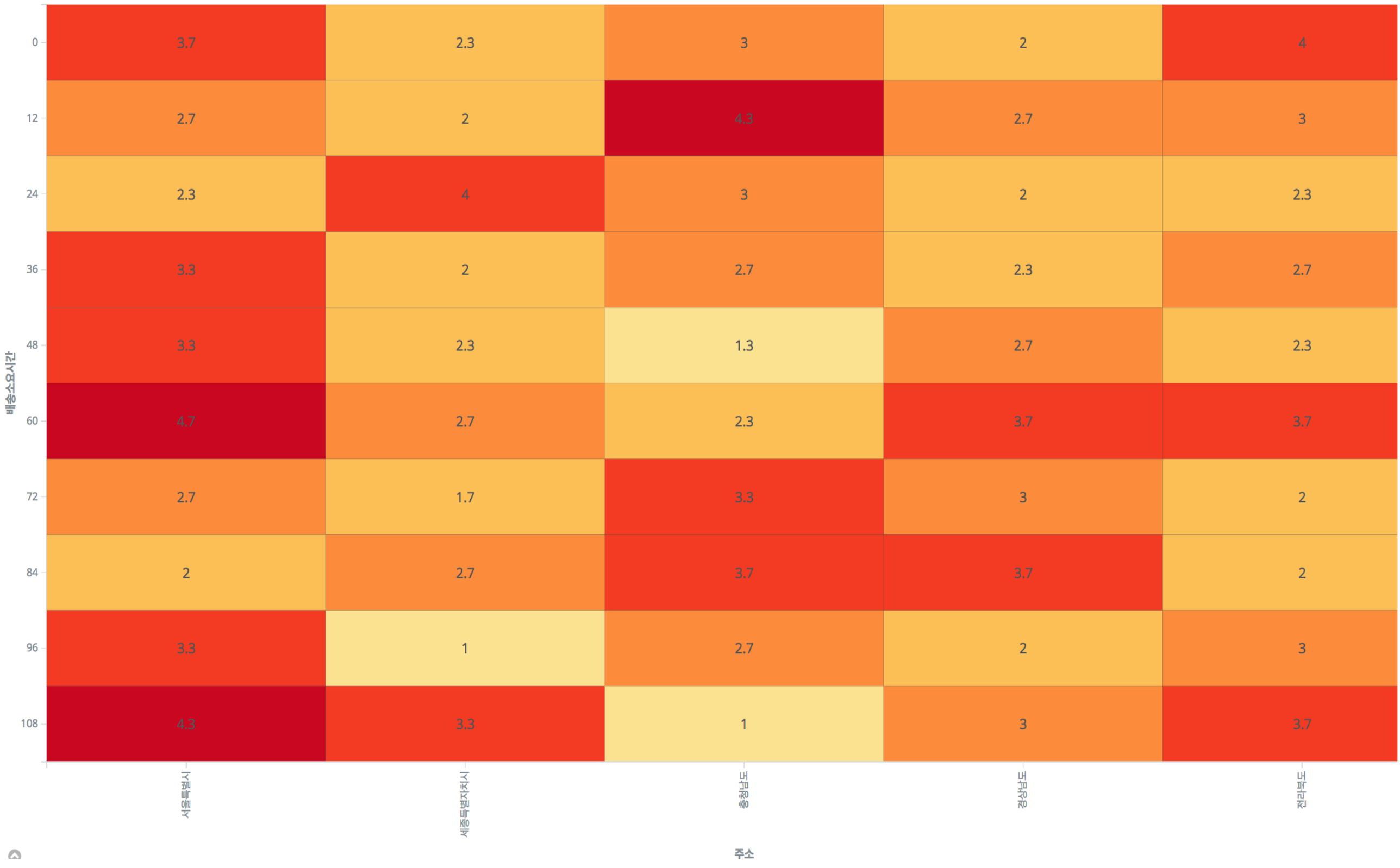
From	To
0	
10000	19999
	29999

Add Range

Note: colors can be changed in the legend

문제 3 - Heat Map ✍ 👑

Visualize - Heat Map



Visualize - Heat Map

데이터

- Index : shopping
- Time Range : 2017년 9월 1일 ~ 2017년

문제

- “상품가격”의 합이 큰 “고객주소_시도” 5개를 X축으로 하고
- “배송소요시간”을 12시간 간격으로 나누어 Y축으로 한 후
- “고객나이”가 가장 작은 3명의
- “판매자평점”의 평균을 표시하고,
- week2_{id}_heatmap으로 저장하세요

사용한 Aggregation

- Metrics
  - Aggregation : Top Hit Aggregation
- Buckets
  - X-Axis : Terms Aggregation, Sum Aggregation
  - Y-Axis : Histogram Aggregation

사용 필드

- 판매자평점
- 고객주소_시도
- 배송소요시간
- 상품가격
- 고객나이

Visualize - Heat Map

shopping

Data Options

metrics

Value

Aggregation

Top Hit

Field

Aggregate With **Size**

Average

Sort On

고객나이

Order

Custom Label

Advanced

shopping

Data Options

Advanced

buckets

X-Axis

Aggregation

Field

고객주소_시도

Order By

Custom Metric

Aggregation

Field

상품가격

Advanced

Order **Size**

5

Custom Label

주소

Advanced

Y-Axis

Sub Aggregation

Histogram

Field

Interval

12

shopping

Data Options

Basic Settings

Show Tooltips

Highlight

Legend Position right

Heatmap Settings

Color Schema Yellow to Red

Reverse Color Schema

Color Scale linear

Scale to Data Bounds

Percentage Mode

Number of colors 6

Custom Ranges

Show Labels

문제 4 - Data Table ✍ 👑

Visualize - Data Table

날짜	인기 Top 3	매출	매출 증감	매출 누적
09월01일	가디건, 스웨터, 원피스	331,000	-	331,000
09월02일	티셔츠, 가디건, 가디건	419,000	88,000	750,000
09월03일	원피스, 청바지, 블라우스	407,000	-12,000	1,157,000
09월04일	팬츠, 청바지, 니트	443,000	36,000	1,600,000
09월05일	셔츠, 가디건, 남방	433,000	-10,000	2,033,000
09월06일	점퍼, 남방, 니트	341,000	-92,000	2,374,000
09월07일	코트, 셔츠, 가디건	360,000	19,000	2,734,000
09월08일	셔츠, 니트, 남방	428,000	68,000	3,162,000
09월09일	가디건, 남방, 가디건	358,000	-70,000	3,520,000
09월10일	가디건, 코트, 셔츠	364,000	6,000	3,884,000

Visualize - Data Table

데이터

- Index : shopping
- Time Range : 2017년 9월 1일 ~ 2017년

문제

“상품가격”이 가장 비쌌던 “상품분류” 3개를 “주문시간” 기준으로 일별로 선별하고 다음을 각각 표시하고 week2_{id}_datatable로 저장하자.

- “상품가격”의 합
- “상품가격”의 합의 일별 증감
- “상품가격"의 일별 누적합

사용한 Aggregation

- Metrics
  - Aggregation
    - Top Hit Aggregation
    - Sum Aggregation
    - Derivative Aggregation
    - Cumulative Sum Aggregation
- Buckets
  - (Split Rows) Aggregation : Date Histogram Aggregation

사용 필드

- 상품가격
- 상품분류
- 고객나이

Visualize - Data Table

shopping

Data | Options

Metric

Aggregation: Top Hit

Field:

Aggregate With: Concatenate

Size:

Sort On:

Order:

Custom Label: 인기 Top 3

Metric

Aggregation: Sum

Field:

Custom Label: 매출

Metric

Aggregation: Derivative

Metric:

Custom Label: 매출 증감

Advanced

Metric

Aggregation: Cumulative Sum

Metric:

Custom Label: 매출 누적

buckets

Split Rows

Aggregation: Date Histogram

Field:

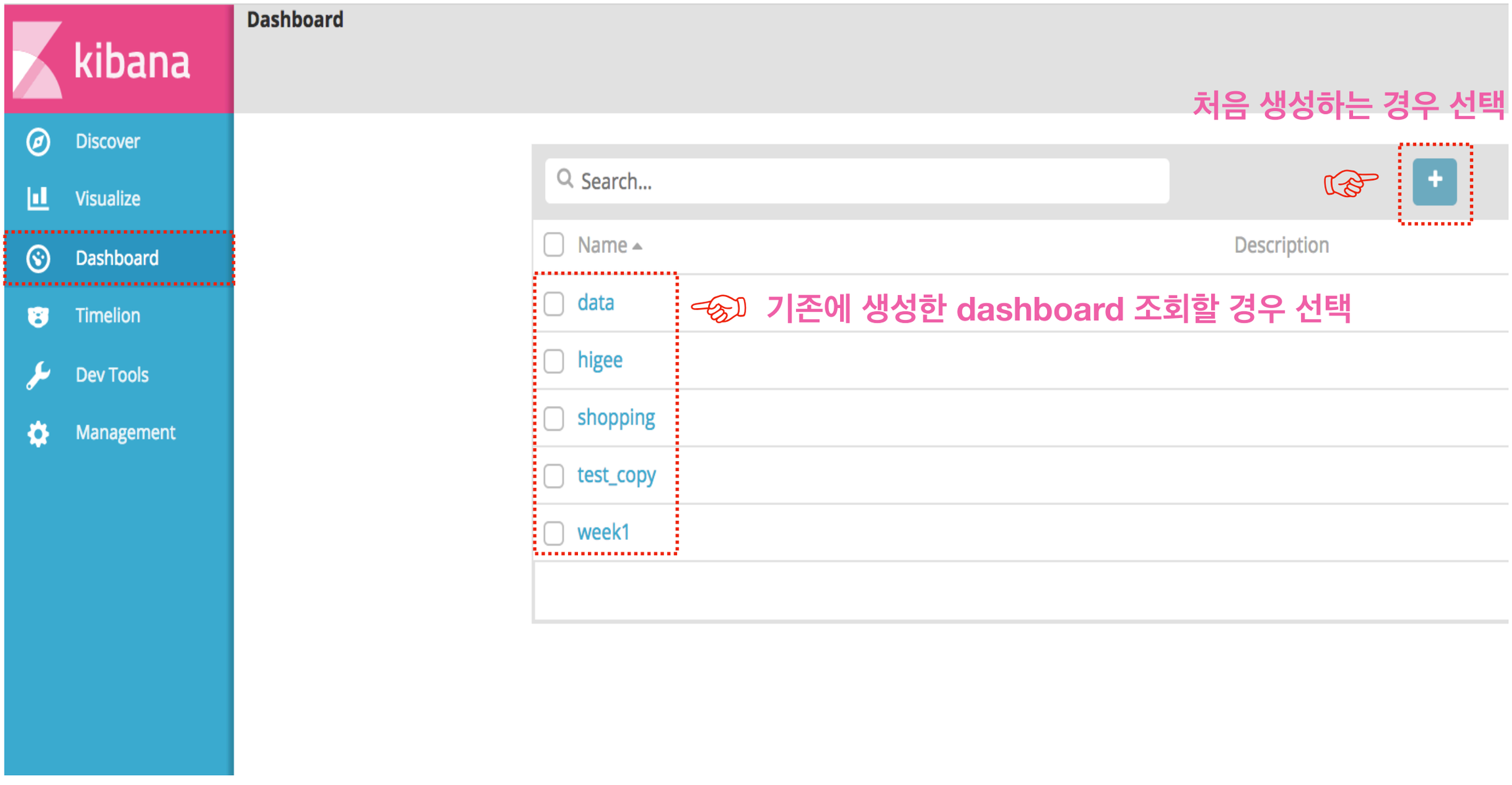
Interval:

Custom Label: 날짜


Dashboard

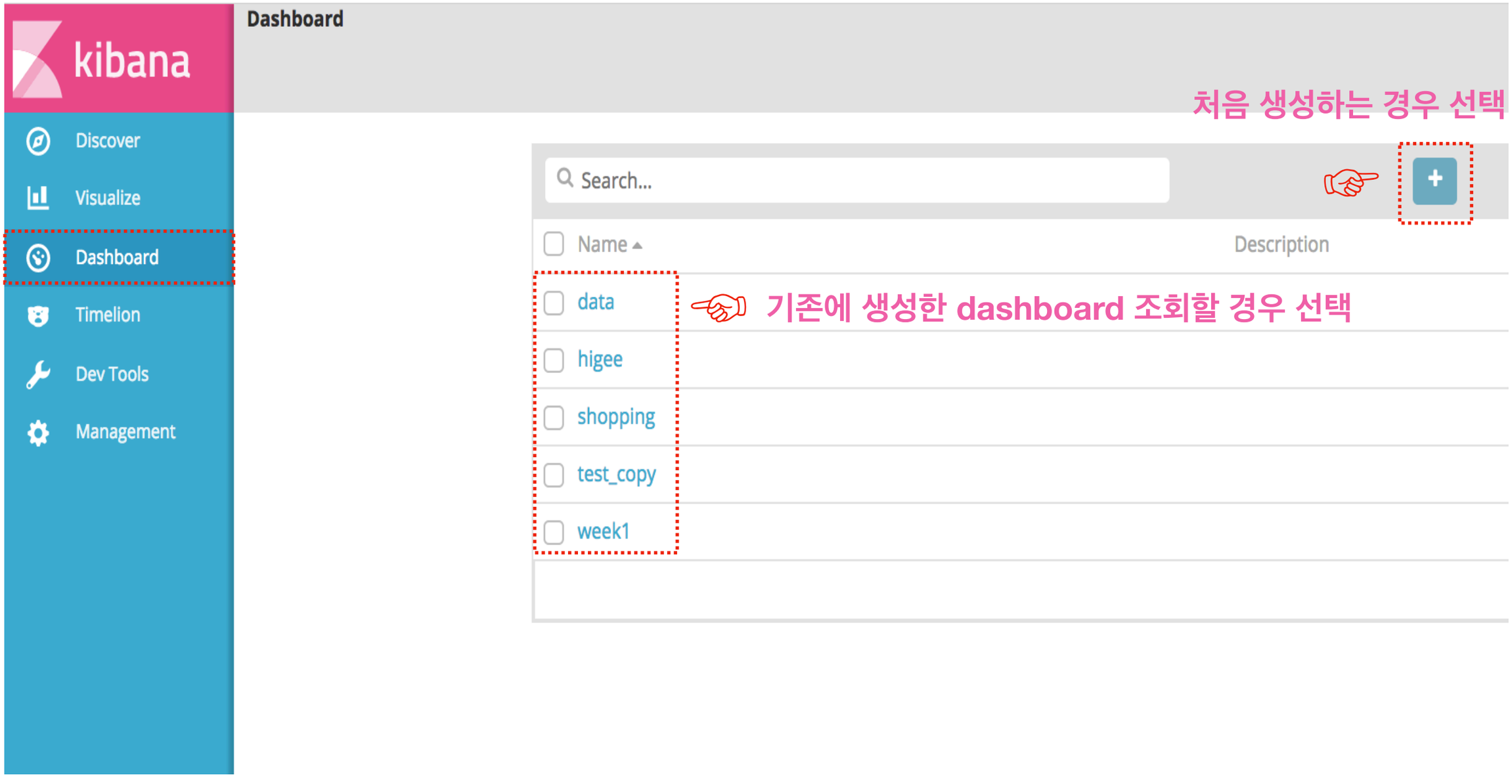
Kibana - dashboard

Kibana에 접속해서 Dashboard 화면으로 가보자 !



Kibana - dashboard

Kibana에 접속해서 Dashboard 화면으로 가보자 !

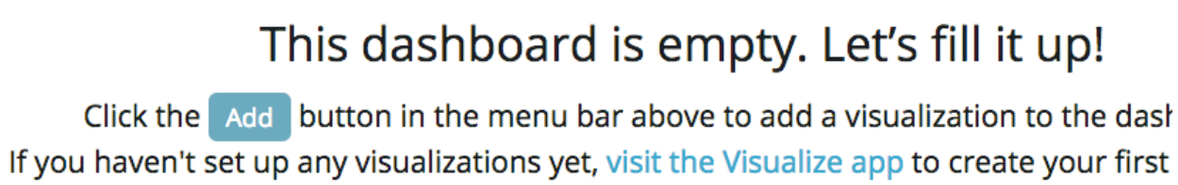


Kibana - dashboard

기존에 만든 visualization을 추가하자

2. 모두 추가한 후 week2_{id}로 저장하자

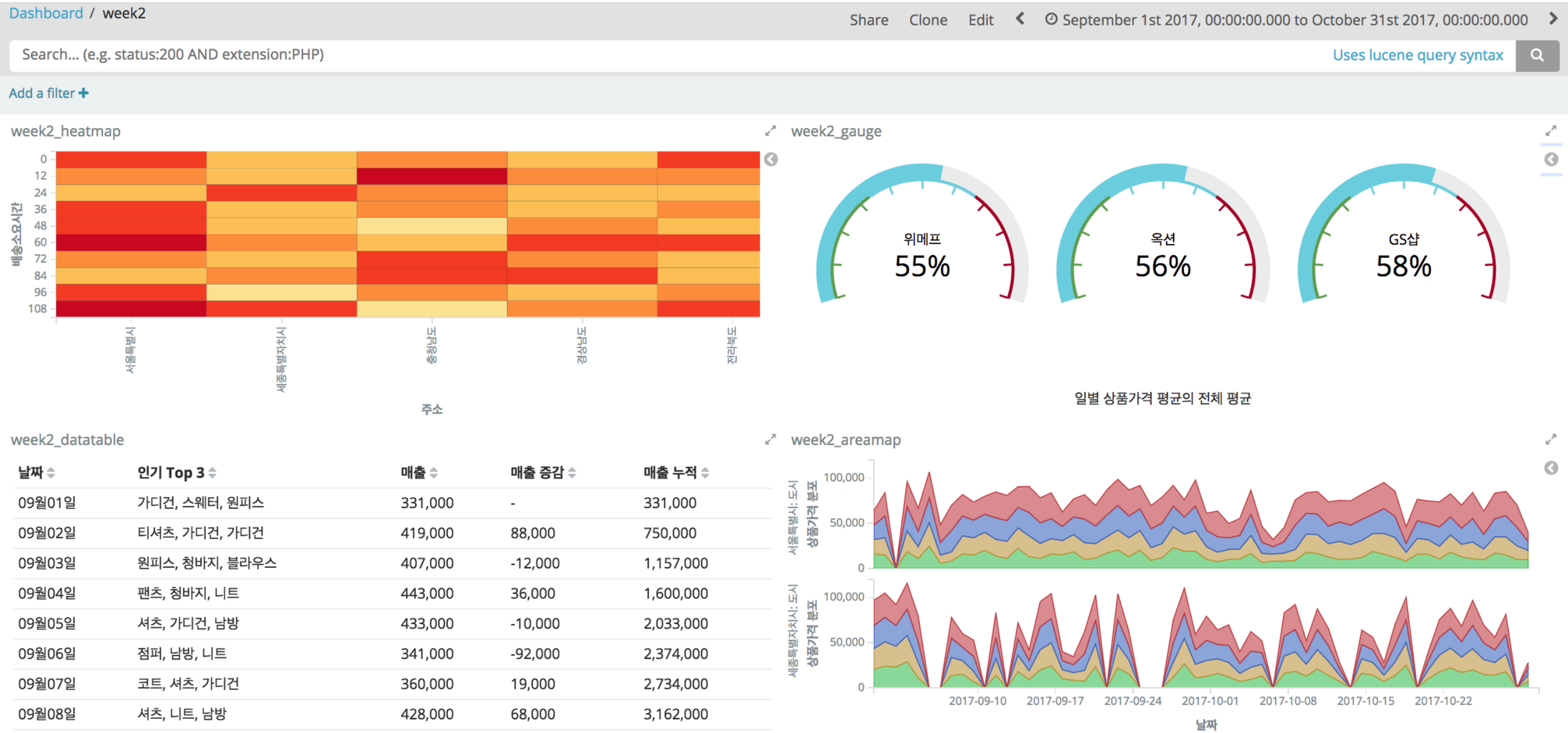
1. Add를 선택하고 visualization을 하나씩 추가하자 ✍



- week2_{id}_area
- week2_{id}_gauge
- week2_{id}_heatmap
- week2_{id}_datatable

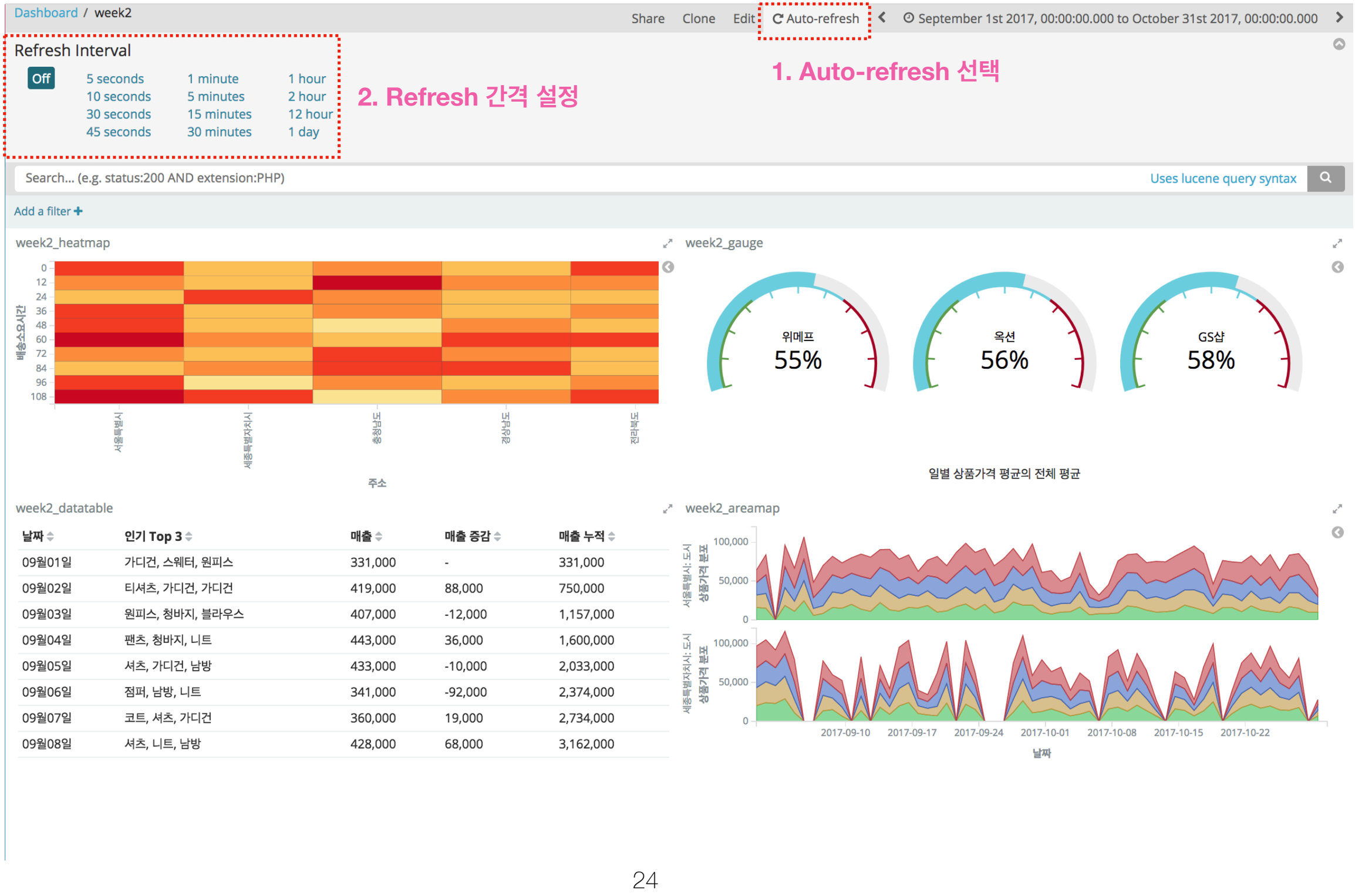
Kibana - dashboard

앞에서 Add를 했다면 아래와 같은 형태로 생성될 것이다



Kibana - dashboard

Auto Refresh 기능을 이용하자



날짜	인기 Top 3	매출	매출 증감	매출 누적
09월01일	가디건, 스웨터, 원피스	331,000	-	331,000
09월02일	티셔츠, 가디건, 가디건	419,000	88,000	750,000
09월03일	원피스, 청바지, 블라우스	407,000	-12,000	1,157,000
09월04일	팬츠, 청바지, 니트	443,000	36,000	1,600,000
09월05일	셔츠, 가디건, 남방	433,000	-10,000	2,033,000
09월06일	점퍼, 남방, 니트	341,000	-92,000	2,374,000
09월07일	코트, 셔츠, 가디건	360,000	19,000	2,734,000
09월08일	셔츠, 니트, 남방	428,000	68,000	3,162,000

Kibana - dashboard

Interactive Dashboard 기능을 이용하자

Dashboard / week2 Share Clone Edit Auto-refresh September 1st 2017, 00:00:00.000 to October 31st 2017, 00:00:00.000

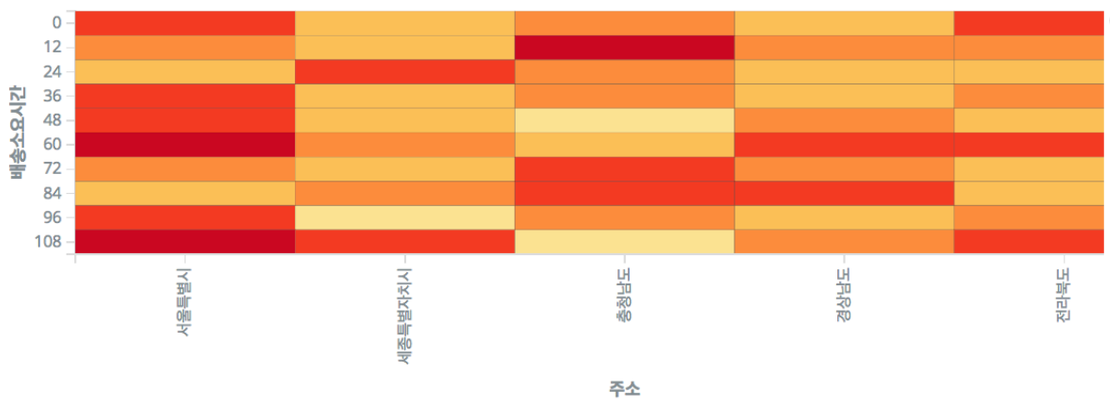
Refresh Interval

Off | 5 seconds | 10 seconds | 30 seconds | 45 seconds | 1 minute | 5 minutes | 15 minutes | 30 minutes | 1 hour | 2 hour | 12 hour | 1 day

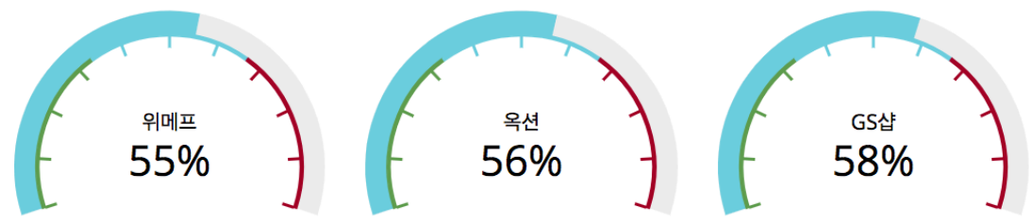
Search... (e.g. status:200 AND extension:PHP) Uses lucene query syntax

Add a filter +

week2_heatmap



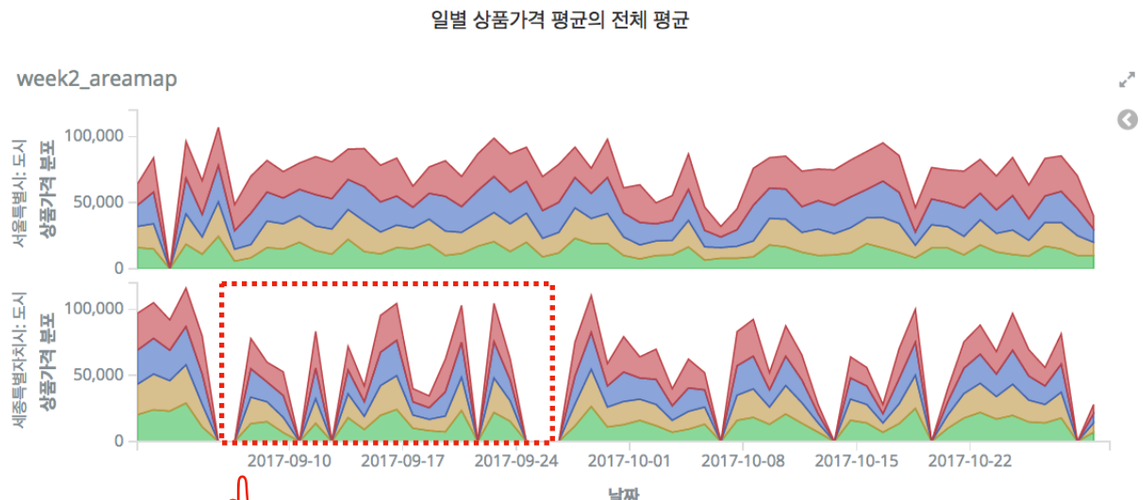
week2_gauge



week2_datatable

날짜	인기 Top 3	매출	매출 증감	매출 누적
09월01일	가디건, 스웨터, 원피스	331,000	-	331,000
09월02일	티셔츠, 가디건, 가디건	419,000	88,000	750,000
09월03일	원피스, 청바지, 블라우스	407,000	-12,000	1,157,000
09월04일	팬츠, 청바지, 니트	443,000	36,000	1,600,000
09월05일	셔츠, 가디건, 남방	433,000	-10,000	2,033,000
09월06일	점퍼, 남방, 니트	341,000	-92,000	2,374,000
09월07일	코트, 셔츠, 가디건	360,000	19,000	2,734,000
09월08일	셔츠, 니트, 남방	428,000	68,000	3,162,000

week2_areamap



1. 적당한 범위를 드래그 하자

Kibana - dashboard

선택한 조건에 맞게 dashboard에 filter가 적용되었다

2. 날짜 filter를 적용했기에 날짜가 변경되었다

Dashboard / week2

Share Clone Edit

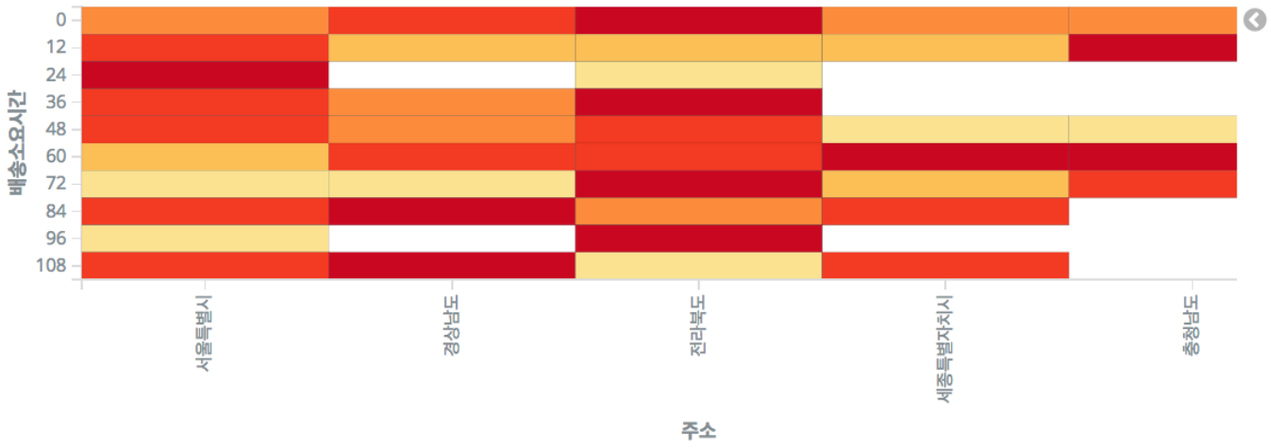
September 10th 2017, 14:21:19.752 to September 19th 2017, 20:55:16.228

Search... (e.g. status:200 AND extension:PHP)

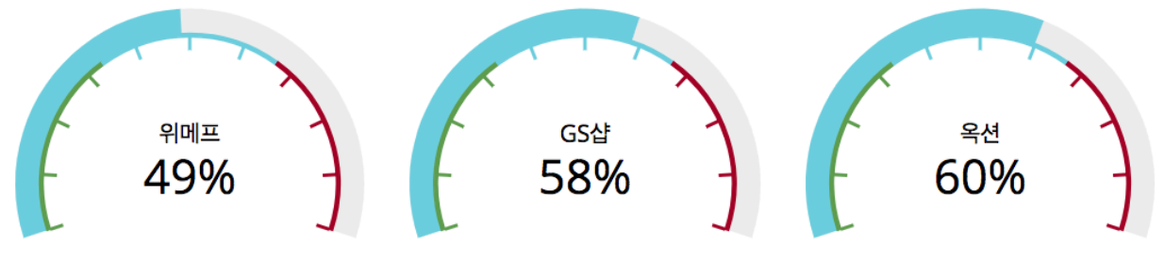
Uses lucene query syntax

Add a filter +

week2_heatmap



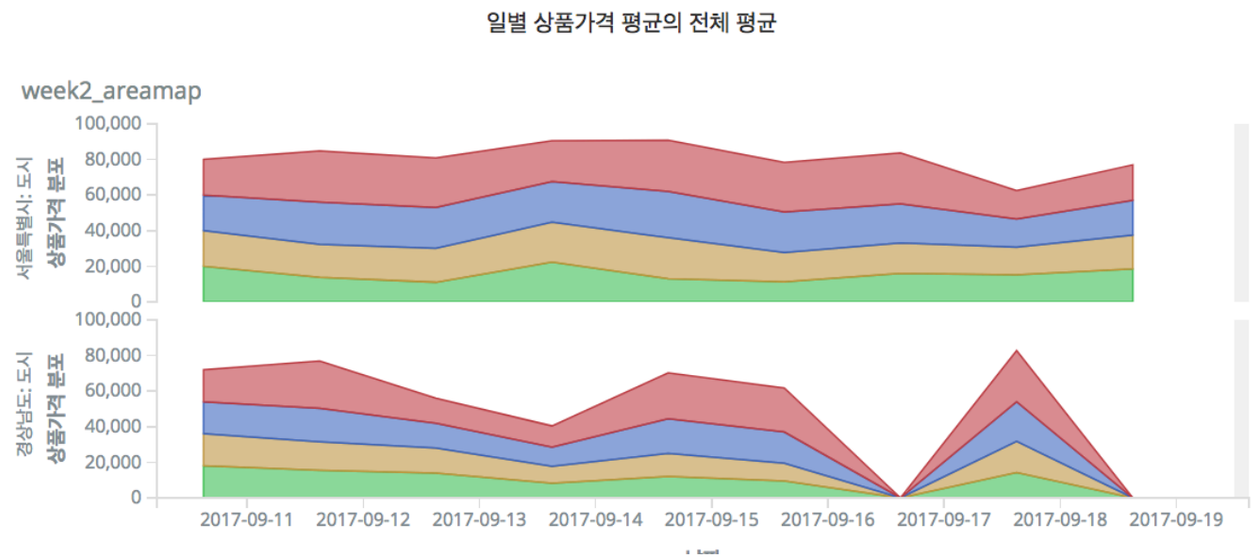
week2_gauge



week2_datatable

날짜	인기 Top 3	매출	매출 증감	매출 누적
09월10일	코트, 스웨터, 스커트	141,000	-	141,000
09월11일	티셔츠, 청바지, 팬츠	222,000	81,000	363,000
09월12일	스웨터, 자켓, 팬츠	438,000	216,000	801,000
09월13일	가디건, 셔츠, 베스트	268,000	-170,000	1,069,000
09월14일	팬츠, 자켓, 남방	288,000	20,000	1,357,000
09월15일	티셔츠, 코트, 티셔츠	299,000	11,000	1,656,000
09월16일	니트, 가디건, 수트	321,000	22,000	1,977,000
09월17일	청바지, 자켓, 가디건	279,000	-42,000	2,256,000

week2_areamap

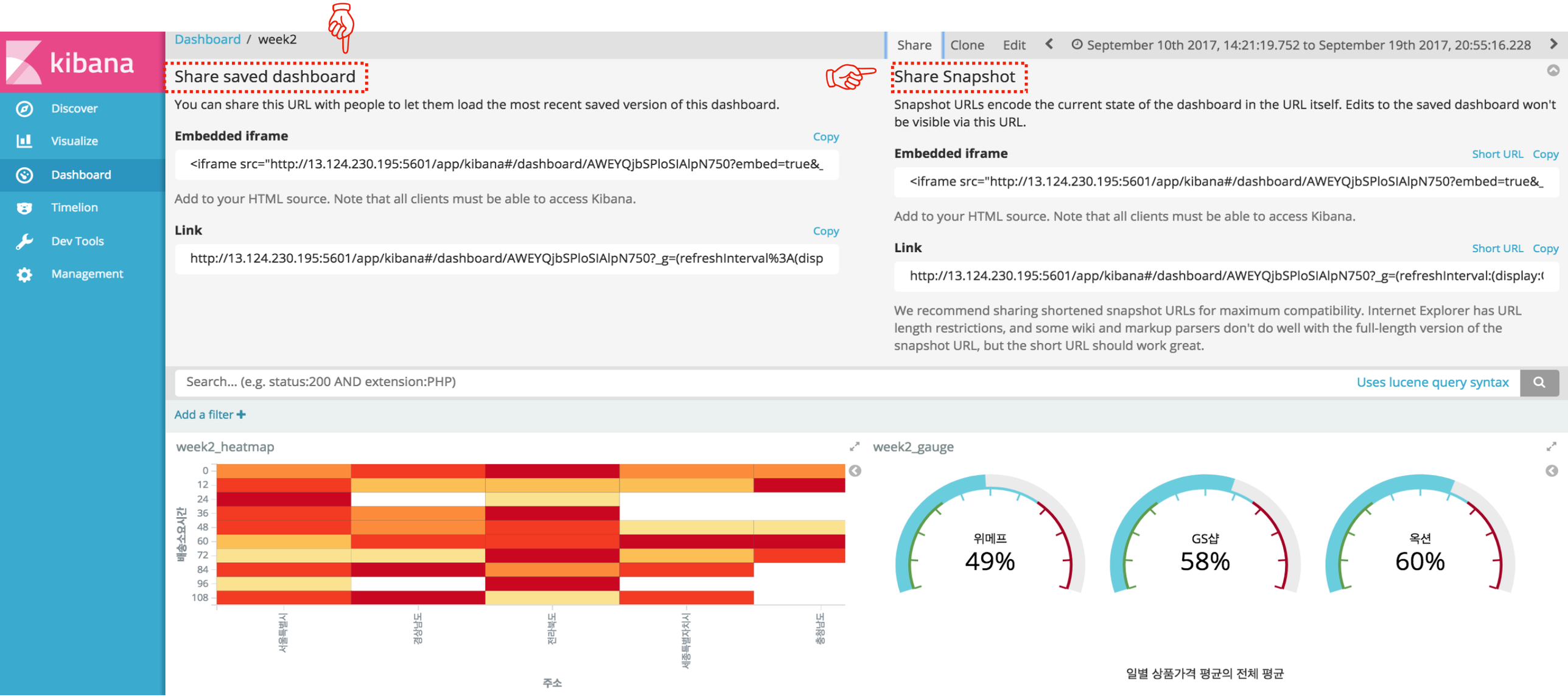


Kibana - dashboard

dashboard를 공유하자

	공유 전 변경	공유 후 변경
saved dashboard	반영 o	반영 o
Snapshot	반영 o	반영 x

Share 선택

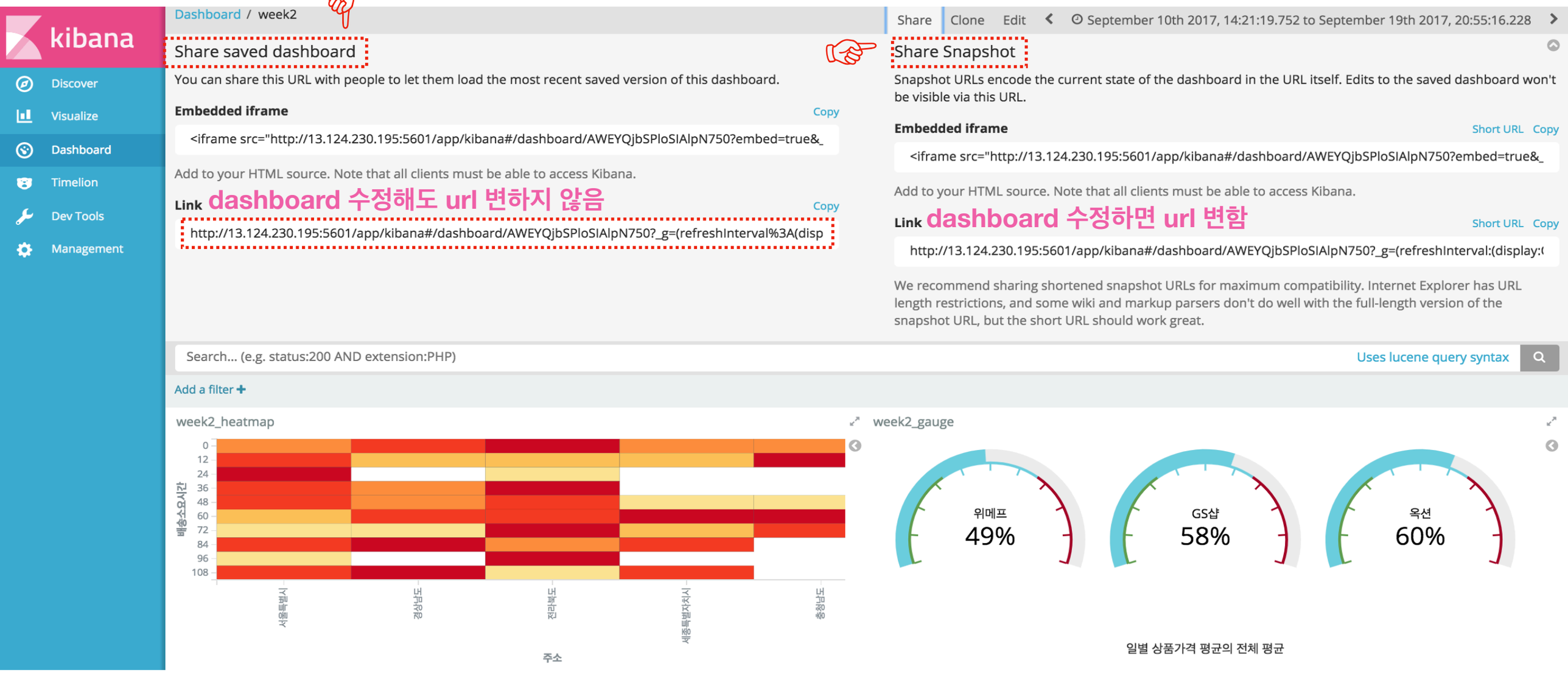


Kibana - dashboard

dashboard를 공유하자

	공유 전 변경	공유 후 변경
saved dashboard	반영 o	반영 o
snapshot	반영 o	반영 x

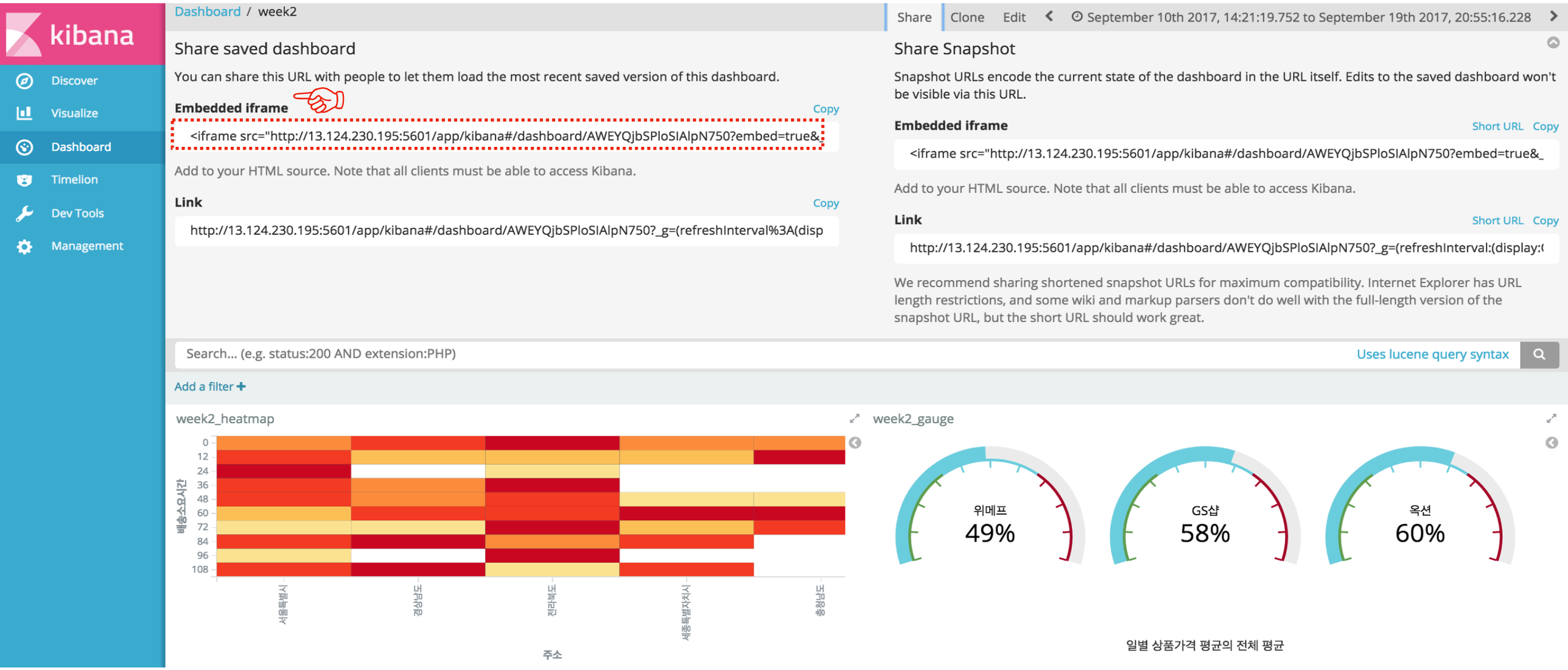
1. Share 선택



Kibana - dashboard

Embedded iFrame을 공유하자

1. 아래 url를 활용한 간단한 html 파일을 만들자 👑

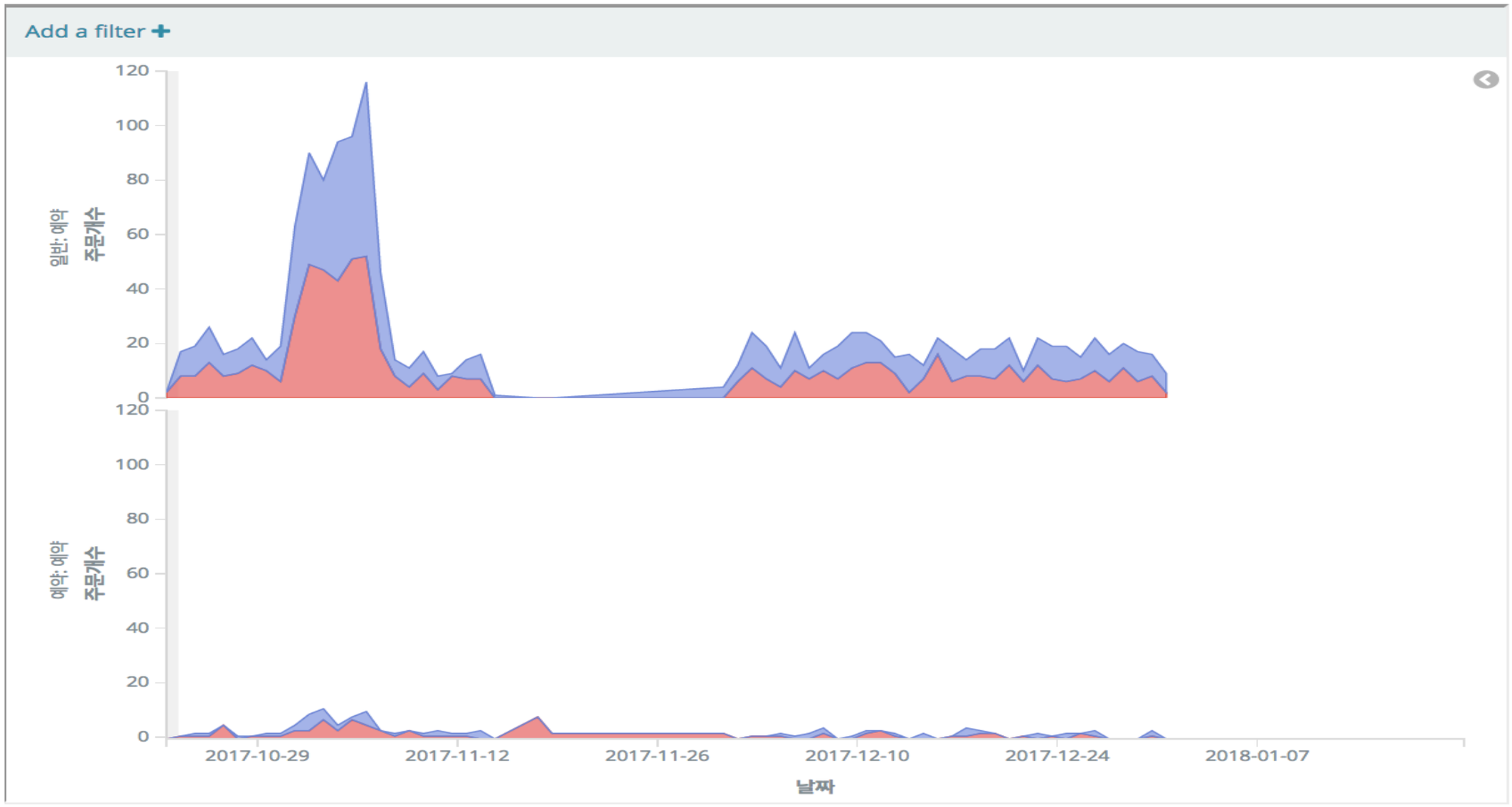


Kibana - dashboard

Embedded iFrame을 공유하자

2. html 파일을 실행해보자 (local 또는 web-server 선택)

Test Visualization



Managing Field ♕

Kibana - management

Date를 다르게 표현할 수 없나?

- 2017년 11월 20일 17시 44분
- 1511199899
- 11/20/2017 5:44pm

String를 다르게 보여줄 수 없나?

- www.fastcampus.co.kr/data_camp_dsbd
- www.fastcampus.co.kr/data_camp_dsbd 링크 형식
- WWW.FASTCAMPUS.CO.KR/DATA_CAMP_DSDB

Number를 다르게 보여줄 수 없나?

- 1,000
- 100,000 %
- 1KB
- 17 minutes

주의!!

Data Format이 변하는 것이지 **Data Type**이 변하는 것이 아니다.

그러므로 Elasticsearch에 저장된 데이터 자체는 변하지 않는다!

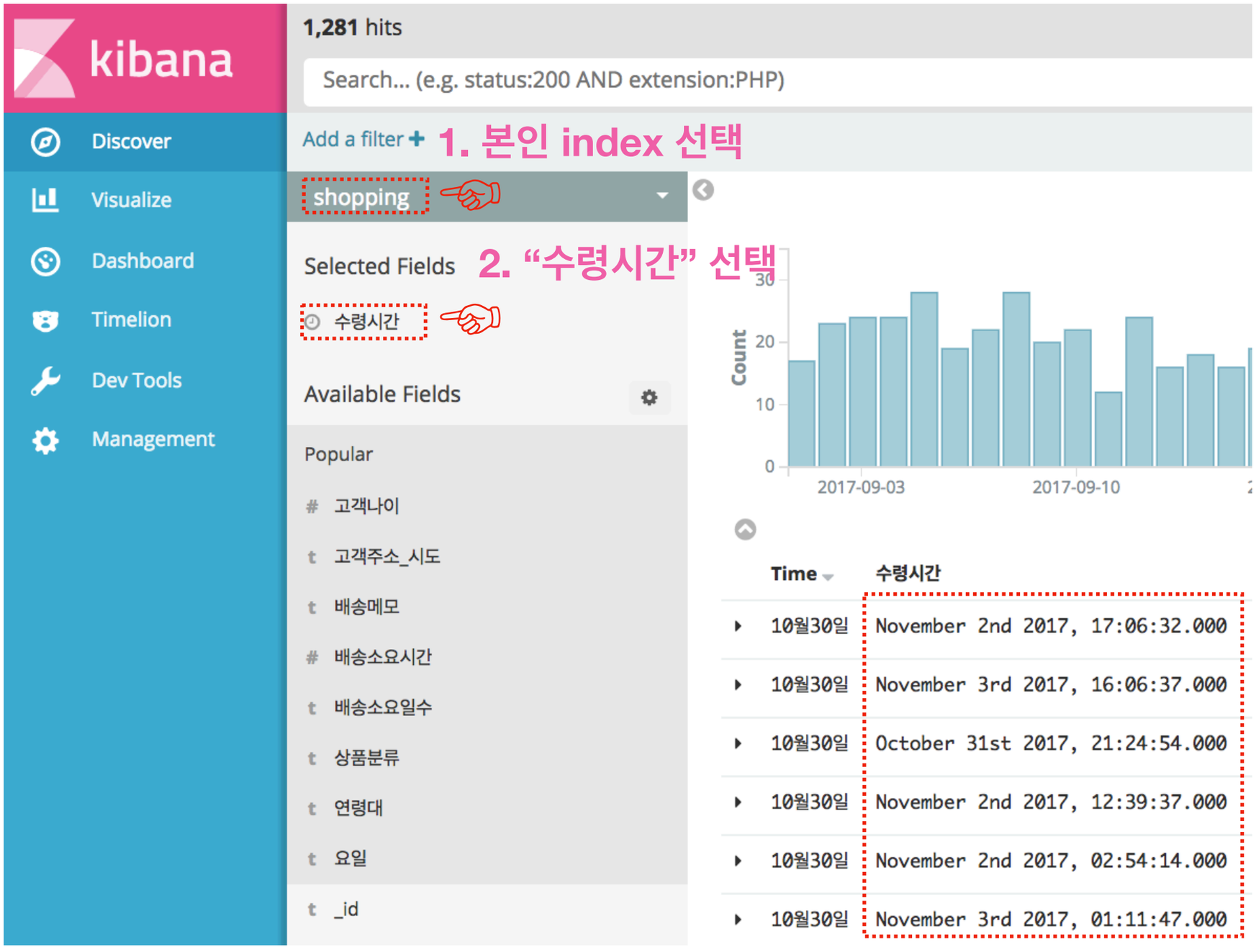
Date Type의 Format 전환 실습을 위해 Index를 등록하자 👑

Index : week2_{id}

Time Filter Field : 주문시간

Kibana - management

그리고 Discover에 가서 Index 선택 후 “수령시간”을 표시해보자

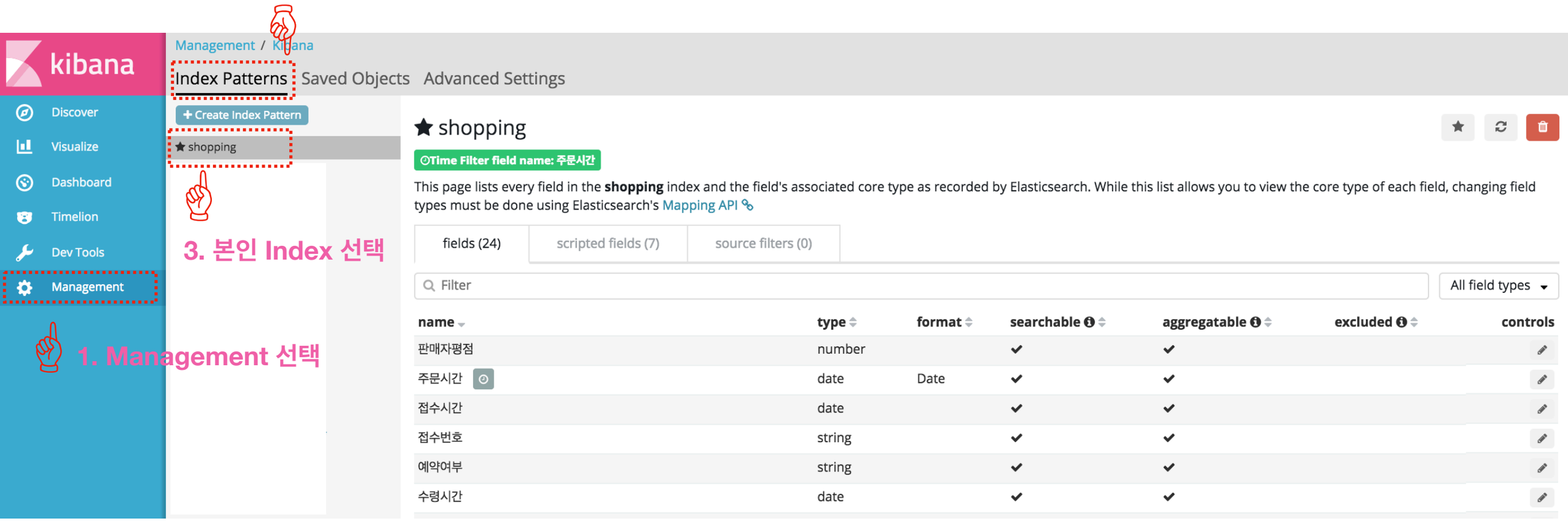


mission : Date Format 변경

Kibana - management

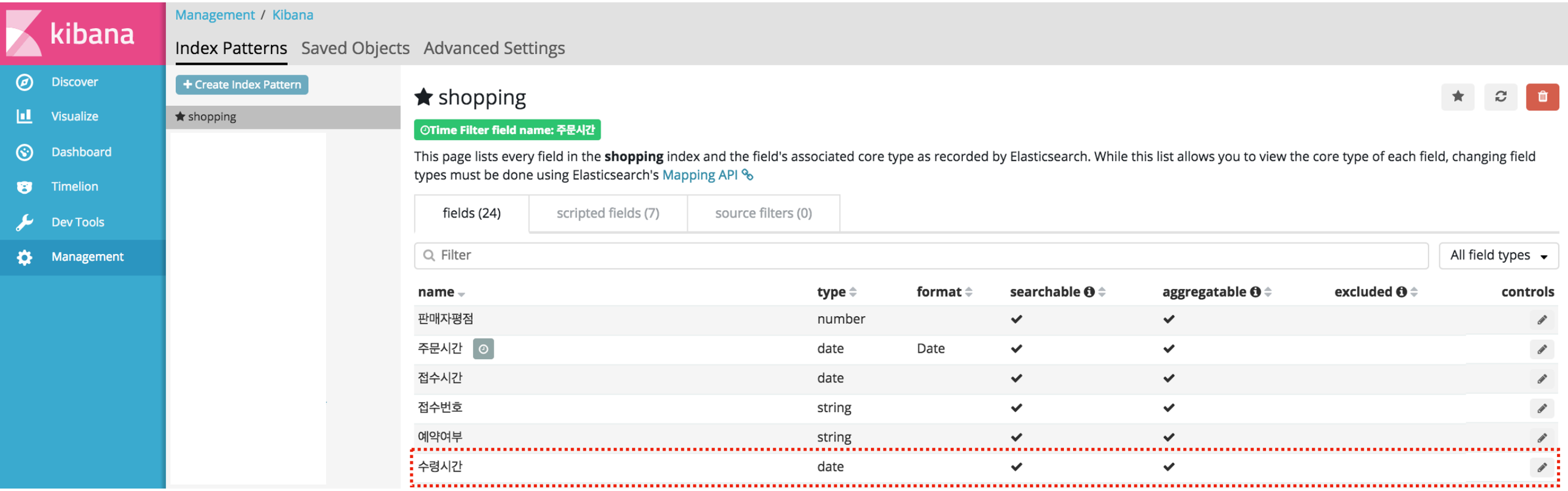
Management - Index Patterns - Index 선택하자

2. Index Patterns 선택

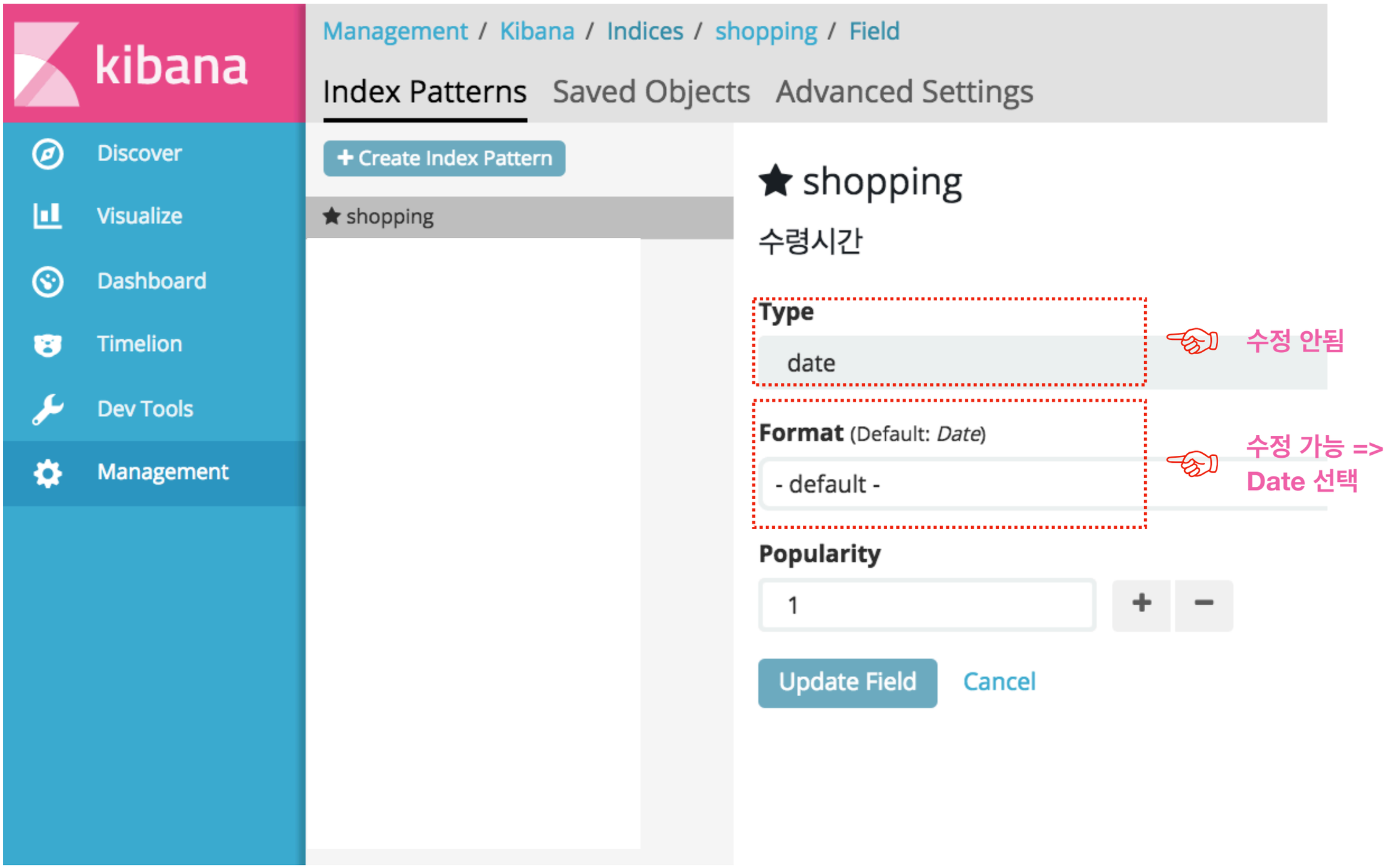


Kibana - management

Date Format 변경하려는 Field의 Control 선택

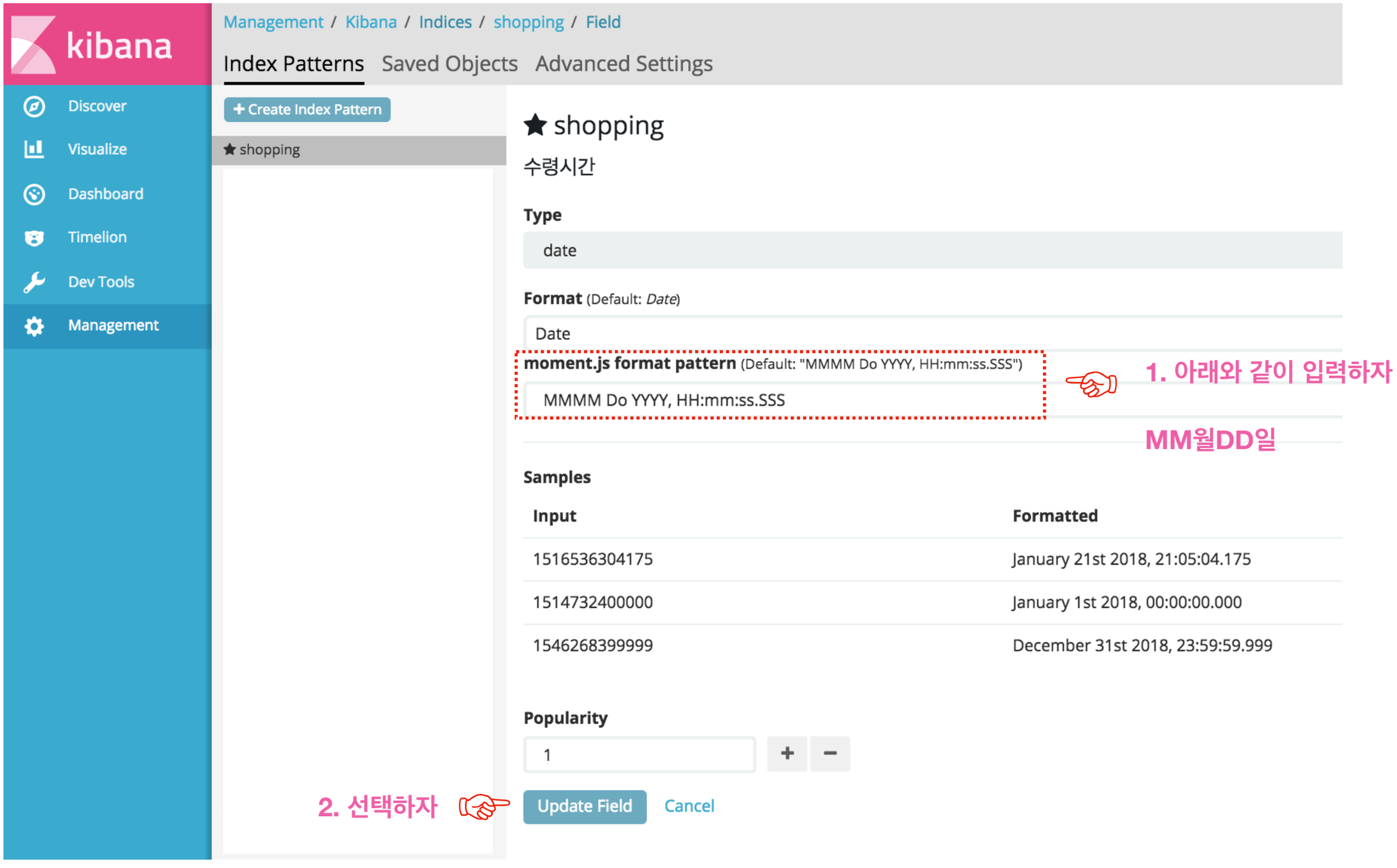


Format을 변경하자



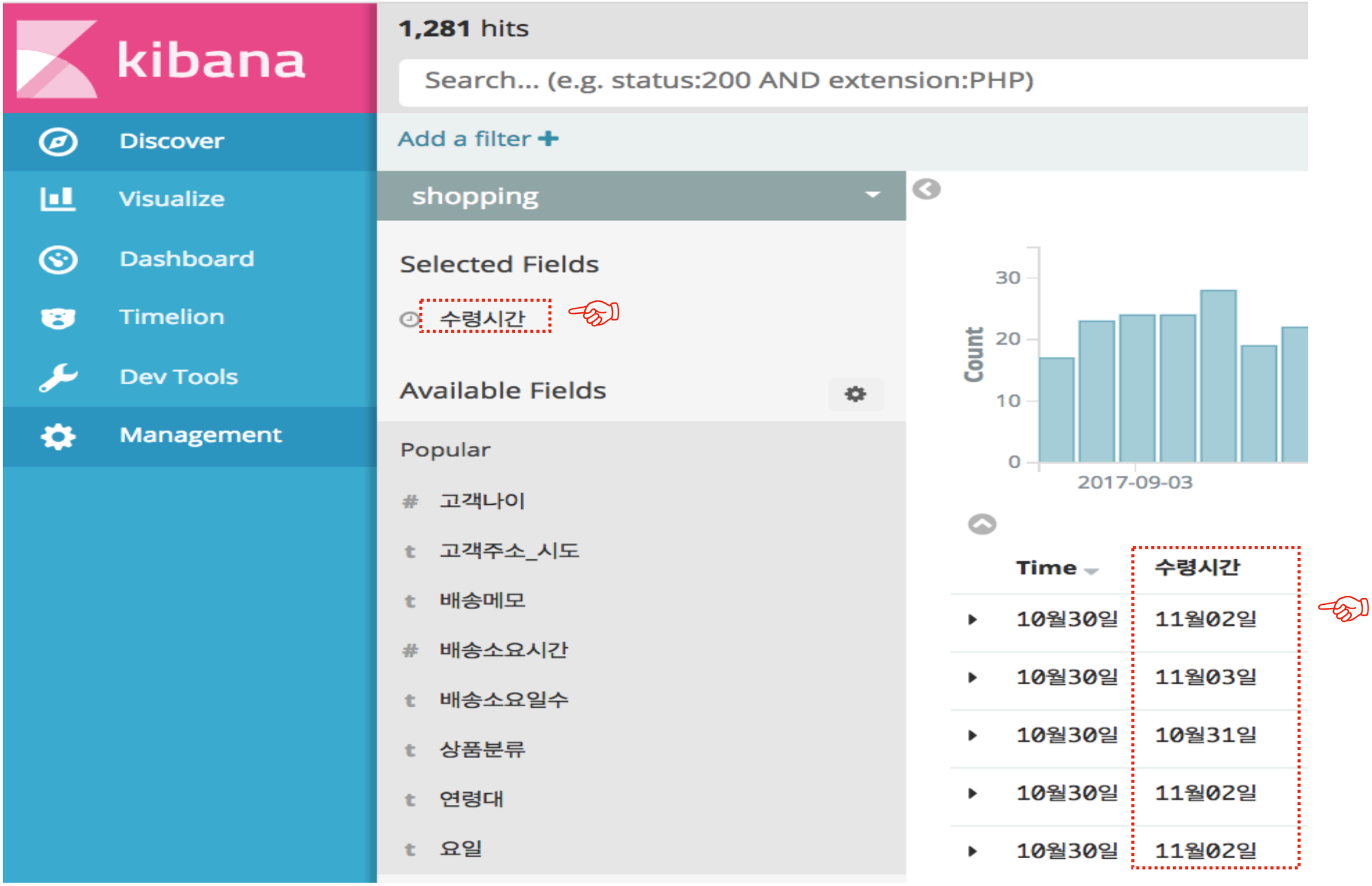
Kibana - management

적절한 Format Pattern으로 수정하자



Kibana - management

Discover에 돌아가서 확인하자



아래를 참고해서 “주문시간”을 Unix Timestamp (Millisecond)로 표시해보자 ✍ 👑

날짜 단위	문법	예시	설명
Year	YYYY	2014	4자리 표시
Year	YY	14	2자리 표시
Month	M	1	1~2자리 표시
Month	MM	01	2자리 표시
Day	D	1	1~2자리 표시
Day	DD	01	2자리 표시
Day	Do	1st	며칠째인지 표시
Hour	H	1	1자리 표시 (24시)
Hour	HH	01	1~2자리 표시 (24시)
Hour	h	1	1자리 표시 (12시)
Hour	hh	01	1~2자리 표시 (12시)
a	h	am/pm	소문자 표시
A	hh	AM/PM	대문자 표시
Minute	m	1	1자리 표시
Minute	mm	01	1~2자리 표시
Second	s	1	1자리 표시
Second	ss	01	1~2자리 표시
Second	X	1410715640.579	Unix Timestamp 초
Millisecond	x	1410715640579	Unix Timestam 밀리초

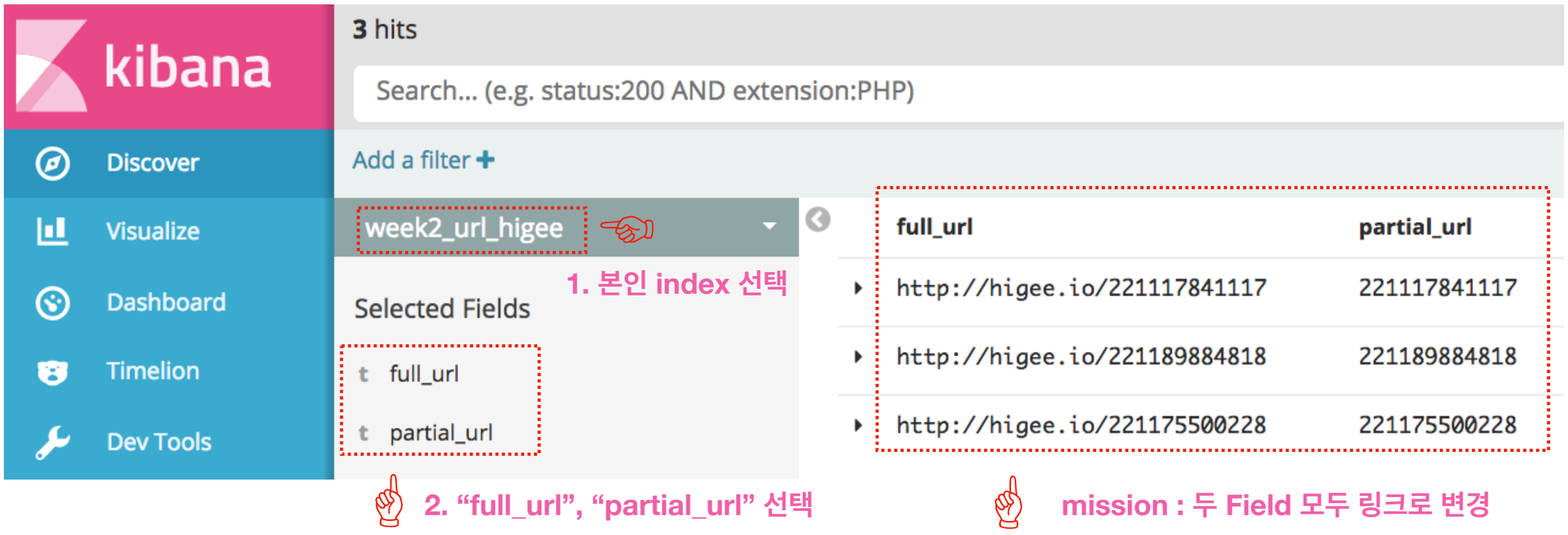
String Type의 Format 전환 실습을 위해 Index를 등록하자 👑

Index : week2__url_{id}

Time Filter Field : 없음

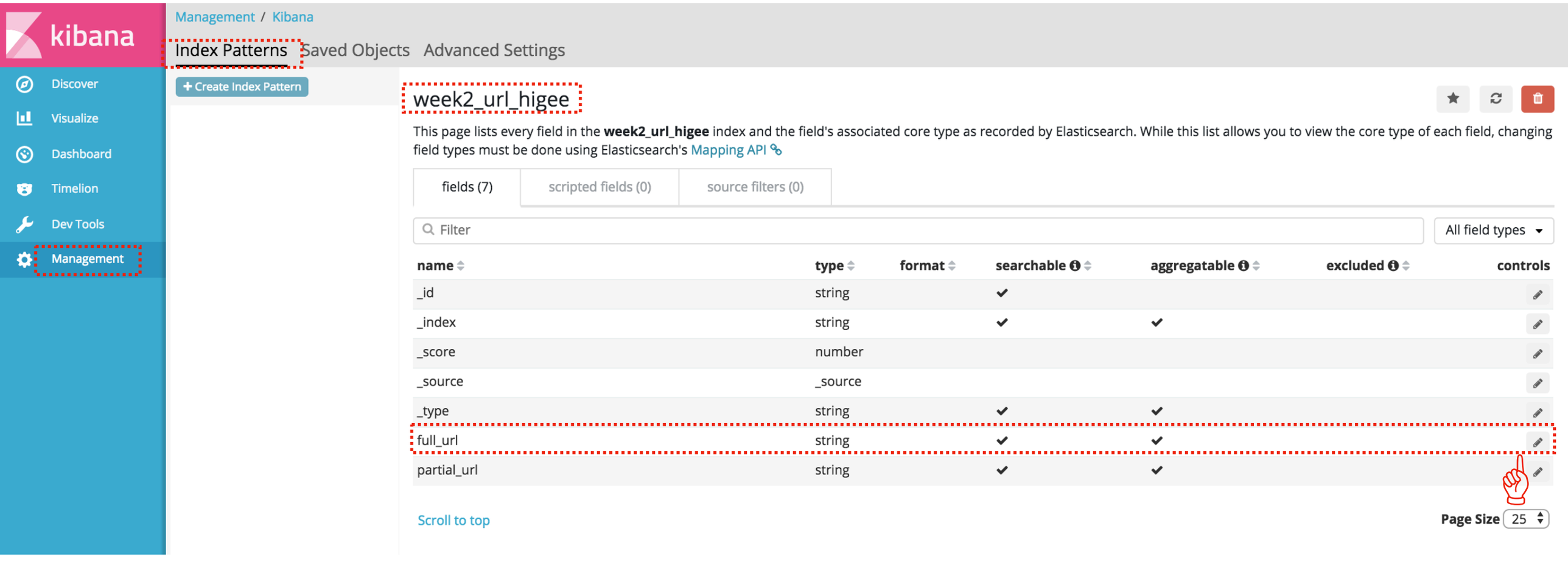
Kibana - management

그리고 Discover에 가서 “full_url”과 “partial_url”을 표시해보자



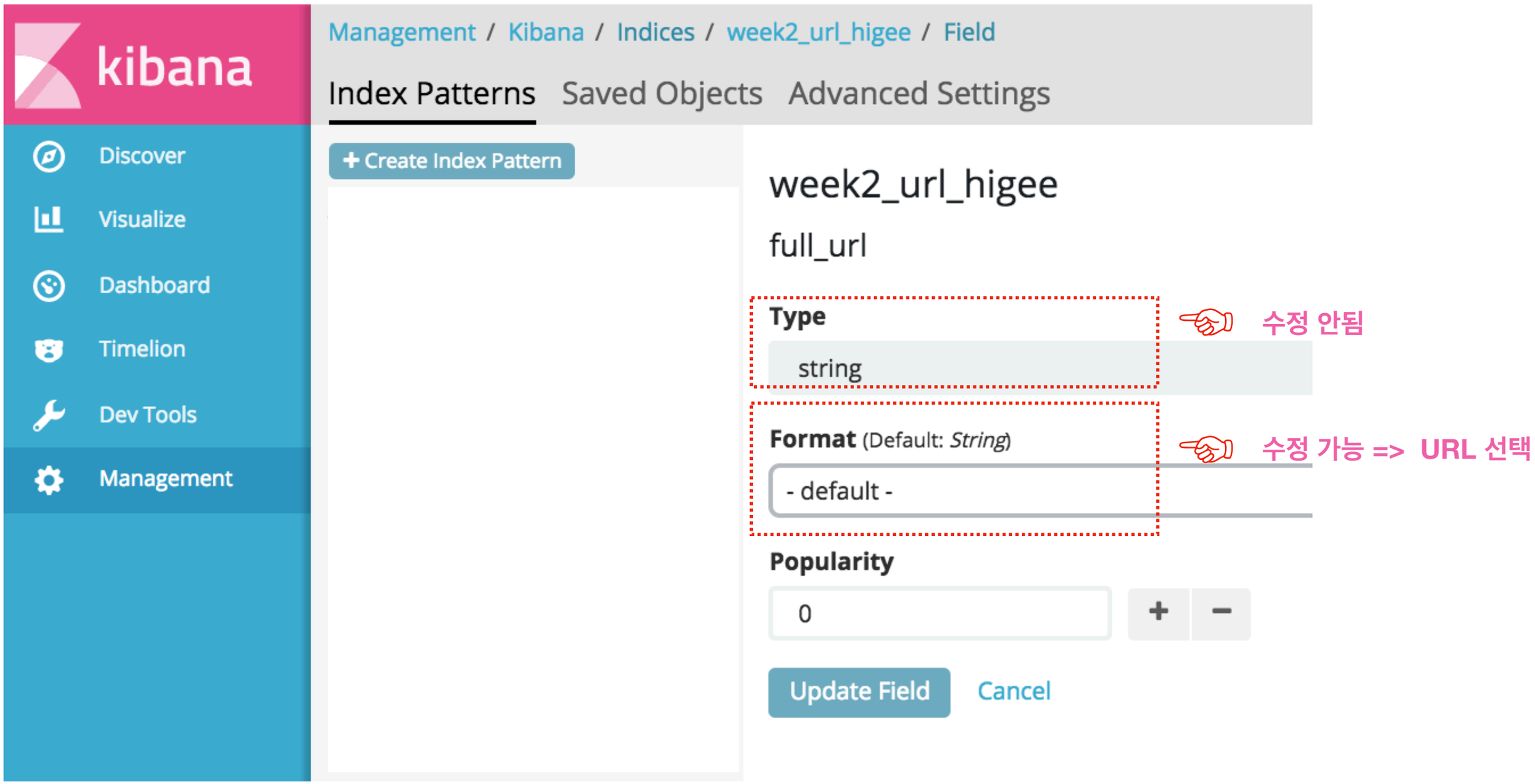
Kibana - management

(앞에서 했던 것처럼) Date Format 변경하려는 Field의 Control 선택



"http://higee.io/221117841117" 와 같이 full_url 값을 가진 field를 url화 하는 작업

Format을 변경하자



Url Templates에 적절한 값을 입력하자

week2_url_higee

full_url

Type

string

Format (Default: *String*) ⚠ Warning

Url

Type

Link

Url Template ℹ Url Template Help

{{rawValue}}

1. 다음과 같이 입력하자

Label Template ℹ Label Template Help

{{rawValue}}

Samples

Input	Formatted
john	john
/some/pathname/asset.png	/some/pathname/asset.png
1234	1234

Popularity

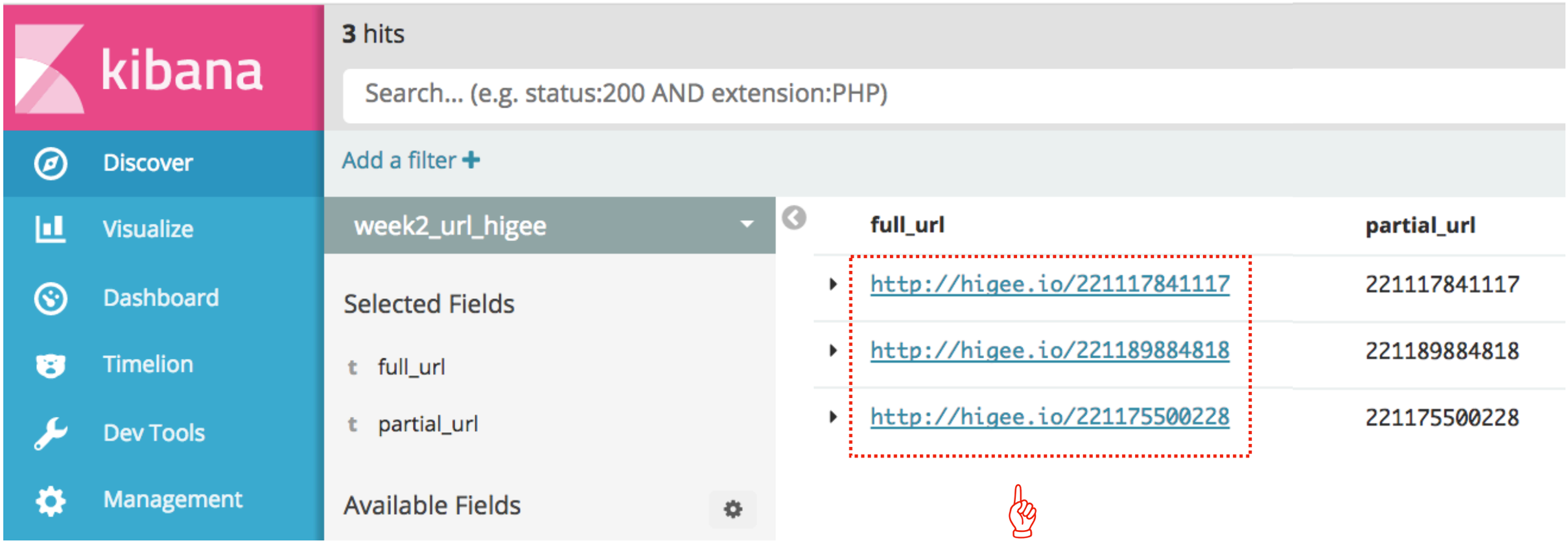
0 + −

Update Field Cancel

2. Update Field 선택

Kibana - management

Discover에 돌아가서 확인하자

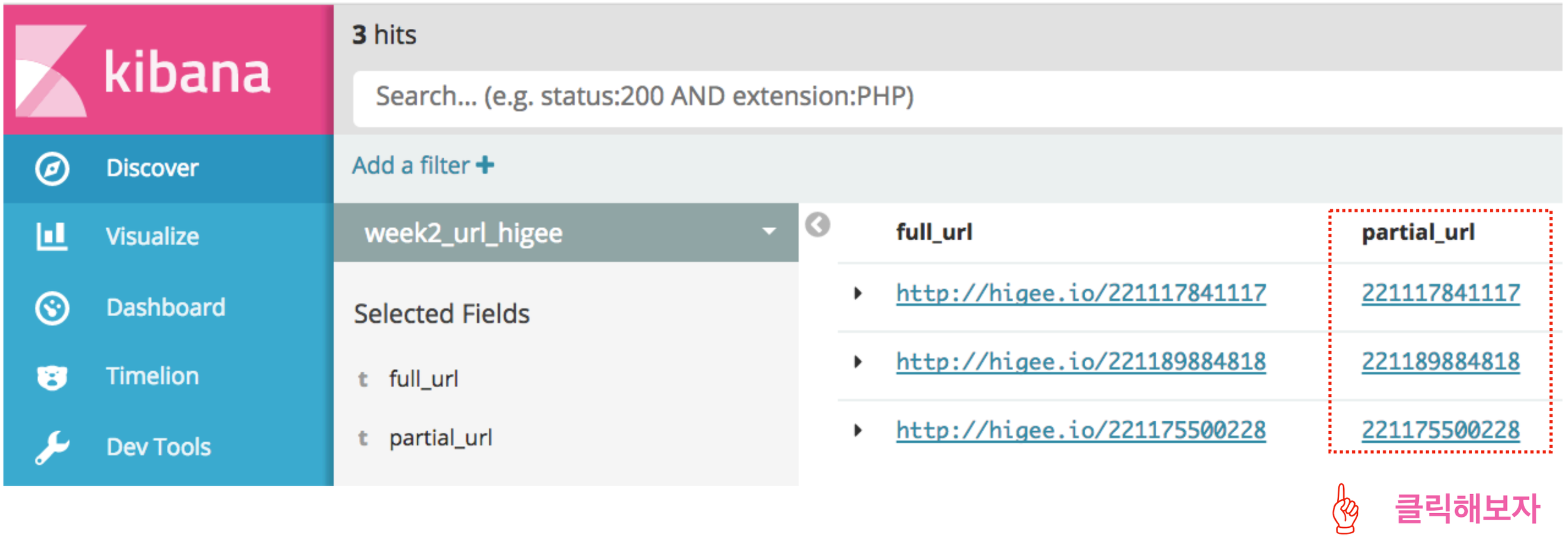


그렇다면 이번에는 “partial_url” field를 url data format으로 변형해보자.
단, url_templaet 부분에는 다음과 같이 입력해야 한다.

http://higee.io/{{value}}

Kibana - management ✍

Discover에 돌아가서 확인하자



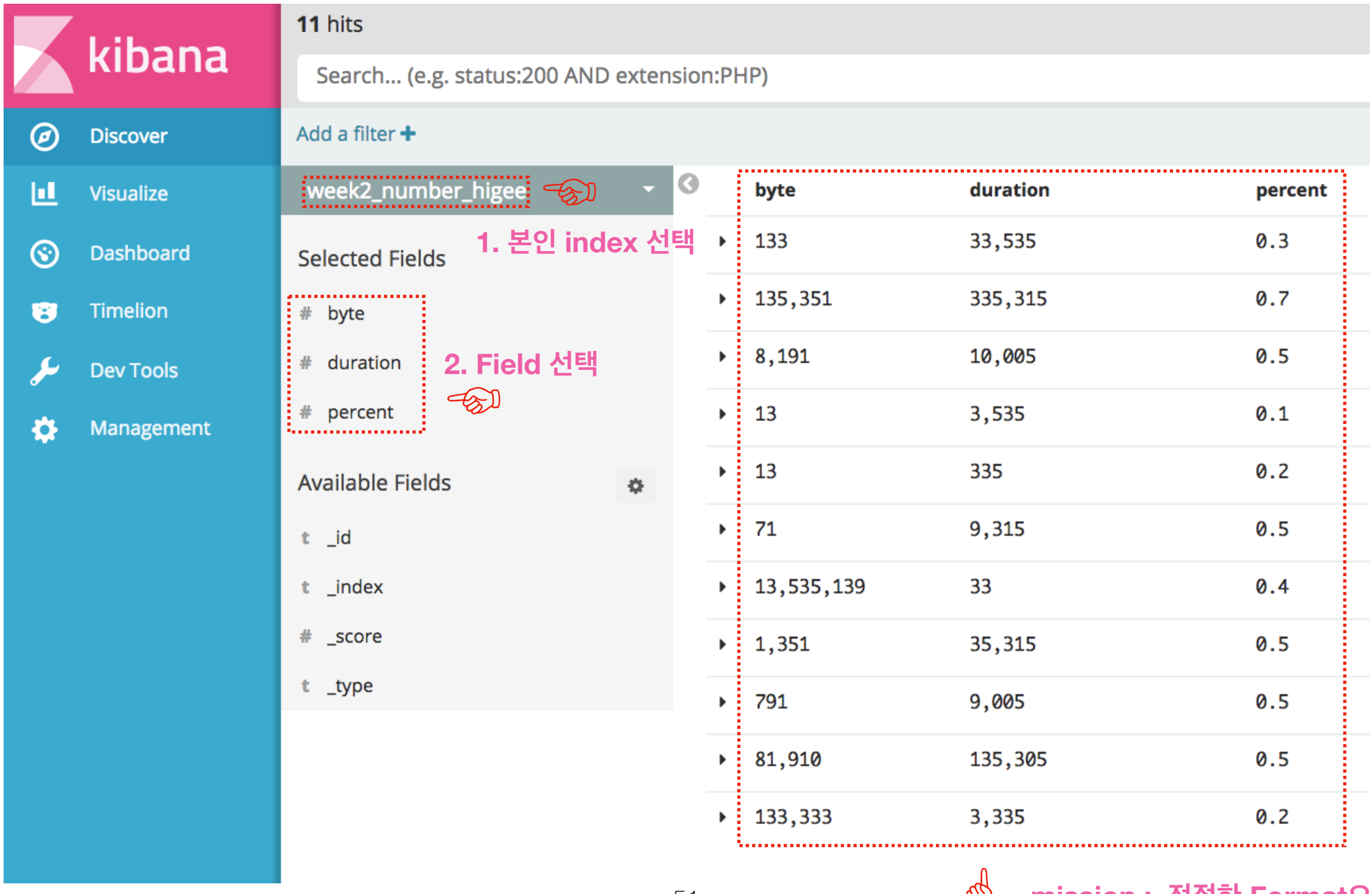
Number Type의 Format 전환 실습을 위해 Index를 등록하자 👑

Index : week2__number_{id}

Time Filter Field : 없음

Kibana - management

그리고 Discover에 가서 “percent”, “bytes”, “duration”을 표시해보자



mission : 적절한 Format으로 변경!

duration Field Format을 변경하자

week2_number_higee

duration

Type

number

Format (Default: *Number*)

Duration 선택

Input Format

Milliseconds 원본 시간 단위

Output Format

Seconds 변환하려는 시간 단위

Decimal Places

2 소수점 자리수

Samples

Input	Formatted
-123	-0.12
1	0.00
12	0.01
123	0.12
658	0.66
1988	1.99
3857	3.86
123292	123.29
923528271	923528.27

Popularity

1 + −

Update Field Cancel

선택

Kibana - management

Discover에 돌아가서 확인하자

kibana

Discover
Visualize
Dashboard
Timelion
Dev Tools
Management

11 hits

Search... (e.g. status:200 AND extension:PHP)

Add a filter +

week2_number_higee

Selected Fields

byte
duration
percent

Available Fields

t _id
t _index
_score
t _type

byte	duration	percent
133B	33.53	0.3
132.18KB	335.31	0.7
8KB	10.01	0.5
13B	3.54	0.1
13B	0.34	0.2
71B	9.31	0.5
12.91MB	0.03	0.4
1.32KB	35.31	0.5
791B	9.01	0.5
79.99KB	135.31	0.5
130.21KB	3.33	0.2

seconds 단위로 변환됐다!

access log 👑 수집할 때 유용하다

percent Field Format을 변경하자

week2_number_higee

percent

Type

number

Format (Default: *Number*)

Percentage 선택

Numeral.js format pattern (Default: "0,0.[000]%")

0,0.[000]% 입력

Samples

Input	Formatted
0.1	10%
0.99999	99.999%
1	100%
100	10,000%
1000	100,000%

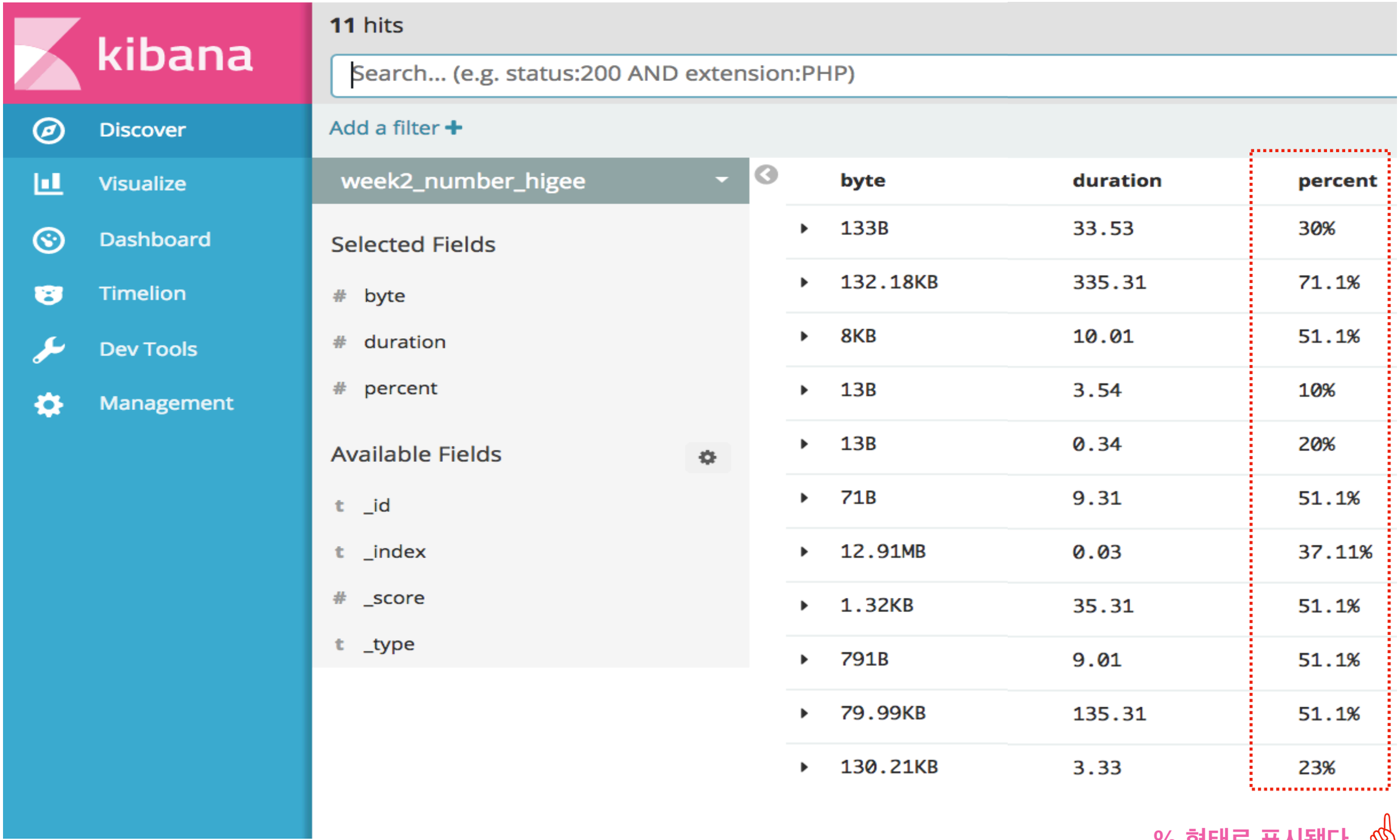
Popularity

1 + −

Update Field Cancel

선택

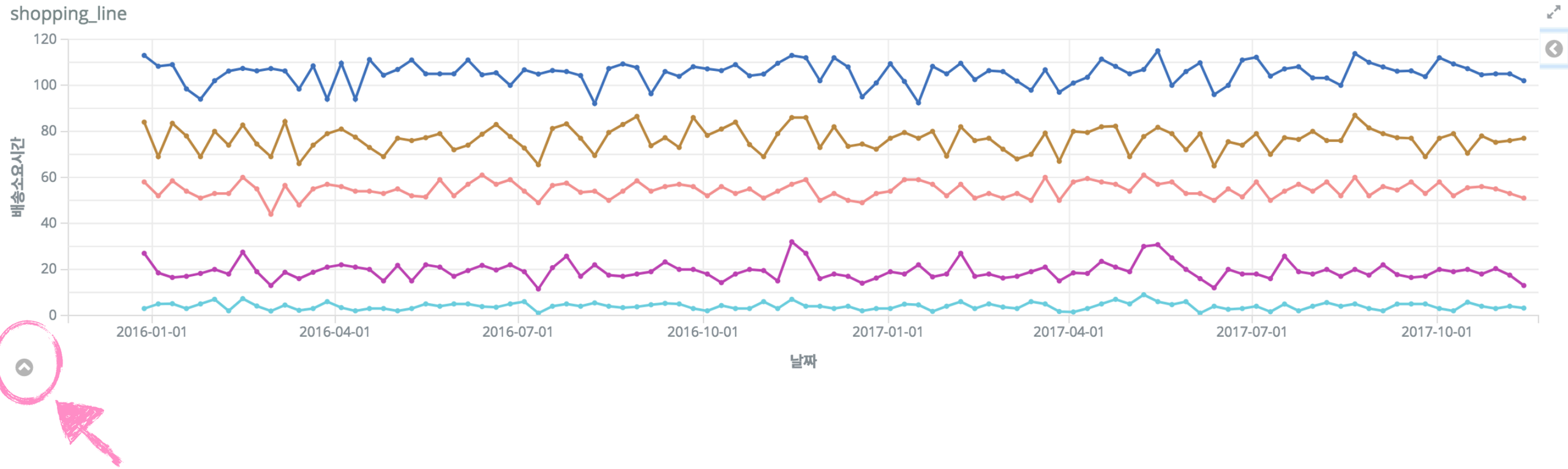
Discover에 돌아가서 확인하자



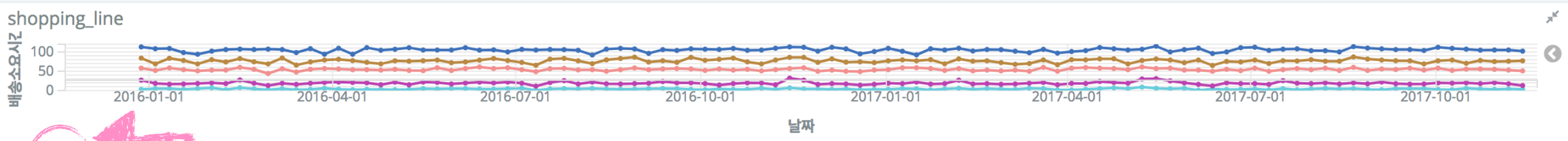
Visualization Spy

Visualization 결과를 csv로 다운 받을 수 없나?

Visualization Spy



Visualization Spy



Table

날짜	5th percentile of 배송소요시간	25th percentile of 배송소요시간	50th percentile of 배송소요시간	75th percentile of 배송소요시간	95th percentile of 배송소요시간
12월28일	3	27	58	84	113
01월04일	5	18.5	52	69	108.3
01월11일	5.1	16.5	58.5	83.5	109
01월18일	3	17	54	78	98.45
01월25일	5	18.25	51	69	94
02월01일	7	20	53	80	102
02월08일	2	18	53	74	106.2
02월15일	7.3	27.5	60	82.75	107.35
02월22일	4.1	19	55	74.5	106.3
02월29일	1.9	13	44	69	107.3

Export: Raw Formatted

1 2 3 4 5 ...10 »

Page Size 10

Visualization Spy

Raw

	A	B	C
1	날짜	th percentile of 배송소요시간	25th percentile of 배송소요시간
2	1.45126E+12	3	27
3	1.45187E+12	5	18.5
4	1.45247E+12	5.1	16.5
5	1.45308E+12	3	17
6	1.45368E+12	5	18.25
7	1.45428E+12	7	20
8	1.45489E+12	2	18
9	1.45549E+12	7.3	27.5
10	1.4561E+12	4.1	19
11	1.4567E+12	1.9	13
12	1.45731E+12	4.5	18.75
13	1.45791E+12	2.2	16
14	1.45852E+12	3	18.75
15	1.45912E+12	6	21
16	1.45973E+12	3.4	22

Elasticsearch에 저장된 Data Format

Format

	A	B	C
1	날짜	5th percentile of 배송소요시간	25th percentile of 배송소요시간
2	12월28일	3	27
3	01월04일	5	18.5
4	01월11일	5.1	16.5
5	01월18일	3	17
6	01월25일	5	18.25
7	02월01일	7	20
8	02월08일	2	18
9	02월15일	7.3	27.5
10	02월22일	4.1	19
11	02월29일	1.9	13
12	03월07일	4.5	18.75
13	03월14일	2.2	16
14	03월21일	3	18.75
15	03월28일	6	21
16	04월04일	3.4	22

Kibana에서 변경된 Data Format

JSON Input 👑

JSON Input

missing data가 있을 때 특정한 값으로 대체할 수 없나?

이동평균을 구할 때 window size를 변경할 수 없나?

Term Aggregation 시 5개 이하 Bucket은 제외할 수 없나?

:

JSON Input

missing data가 있을 때 특정한 값으로 대체할 수 없나?

이동평균을 구할 때 window size를 변경할 수 없나?

Term Aggregation 시 5개 이하 Bucket은 제외할 수 없나?

JSON Input ✍

다음과 같은 Visualization을 만들자

데이터

- Index : shopping
- Time Range : 2017년 11월 18일 03시~ 2017년 11월 18일 23시

문제

- “주문시간”을 기준으로 시간당
- “상품가격”의 합을 표시하자

사용한 Visualization

- Data Table

사용한 Aggregation

- Metrics : Sum Aggregation
- Buckets : Date Histogram

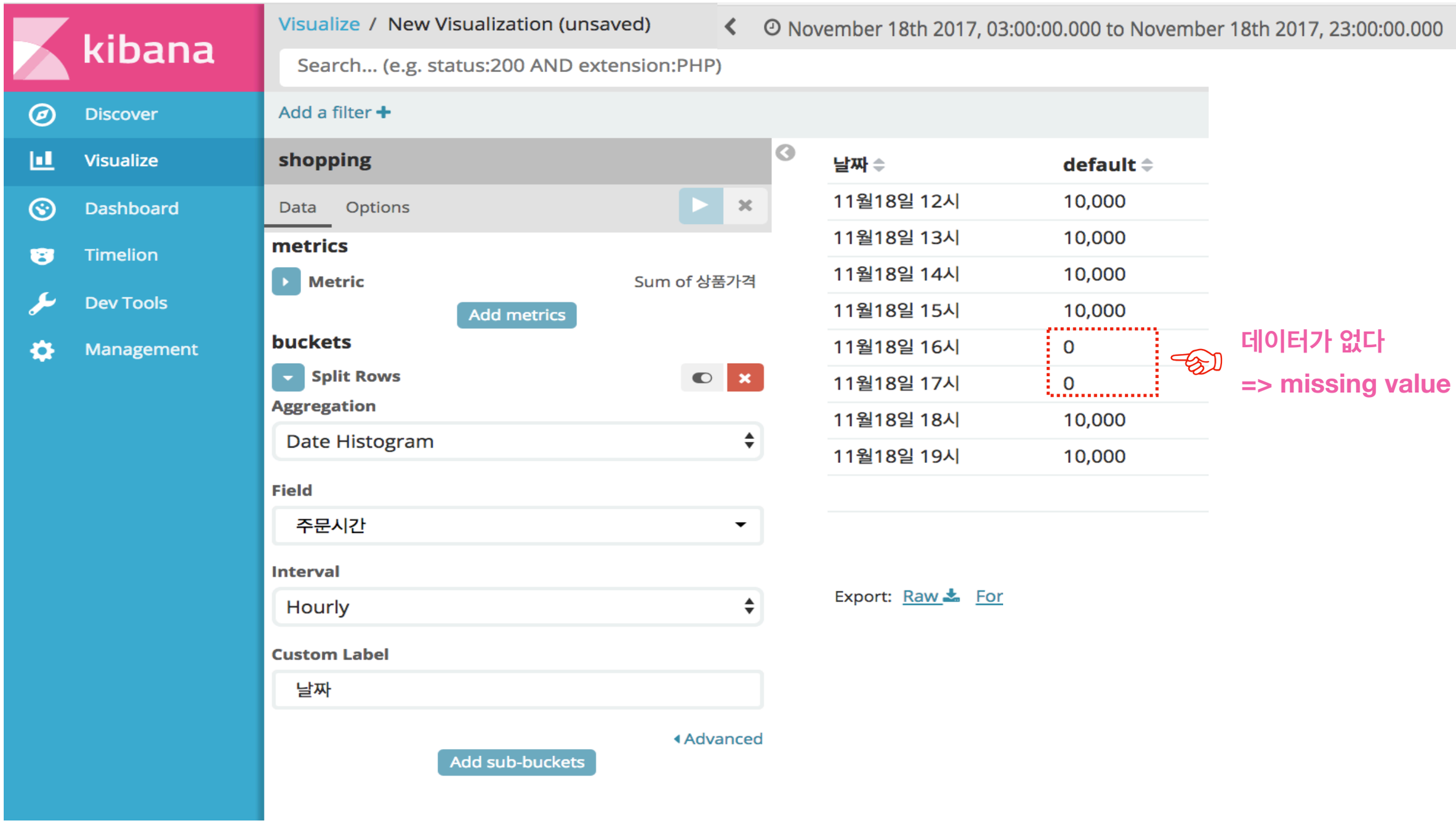
사용 필드

- 상품가격
- 주문시간

~~힌트는 다음 페이지~~

JSON Input

Missing Value 확인



JSON Input

Missing Value 처리

kibana

Visualize / New Visualization (unsaved)

November 18th 2017, 03:00:00.000 to November 18th 2017, 23:00:00.000

Search... (e.g. status:200 AND extension:PHP)

Add a filter +

- Discover
- Visualize
- Dashboard
- Timelion
- Dev Tools
- Management

shopping

3. 선택

Data | Options

metrics

Metric Sum of 상품가격

Metric

Aggregation

Sum

Field

상품가격

Custom Label

missing value 처리

1. 선택

Advanced

JSON Input

{"missing":100}

Add metrics

2. 다음과 같이 입력

{“missing” : 100}

날짜	default	missing value 처리
11월18일 12시	10,000	10,000
11월18일 13시	10,000	10,000
11월18일 14시	10,000	10,000
11월18일 15시	10,000	10,000
11월18일 16시	0	100
11월18일 17시	0	100
11월18일 18시	10,000	10,000
11월18일 19시	10,000	10,000

Export: Raw

4. 확인

같은 bucket인데 이번에는 100이라는 값을 갖는다

JSON Input

무슨 일이 벌어진걸까? Visualization Spy의 Request를 보자

Visualize / New Visualization (unsaved) Save Share Refresh November 18th 2017, 03:00:00.000 to November 18th 2017, 23:00:00.000

Search... (e.g. status:200 AND extension:PHP) Uses lucene query syntax

Add a filter +

shopping

Data Options

metrics

Metric Sum of 상품가격

Metric

Aggregation

Sum

Field

상품가격

Custom Label

missing value 처리

Advanced

JSON Input

{"missing":100}

Add metrics

buckets

Split Rows 주문시간 per hour

Add sub-buckets

날짜	default	missing value 처리
11월18일 12시	10,000	10,000
11월18일 13시	10,000	10,000
11월18일 14시	10,000	10,000
11월18일 15시	10,000	10,000
11월18일 16시	0	100
11월18일 17시	0	100
11월18일 18시	10,000	10,000
11월18일 19시	10,000	10,000

Export: Raw Formatted

Request

2. Request 선택

Elasticsearch request body

```
  "excludes": []
},
"aggs": {
  "2": {
    "date_histogram": {
      "field": "주문시간",
      "interval": "1h",
      "time_zone": "Asia/Tokyo",
      "min_doc_count": 1
    },
    "aggs": {
      "1": {
        "sum": {
          "field": "상품가격"
        }
      },
      "3": {
        "sum": {
          "field": "상품가격",
          "missing": 100
        }
      }
    }
  }
}
}
```

3. Sum Aggregation Parameter에 반영된 거 확인

정확히 무슨 역할을 하는지는 공식 문서 확인 👑

1. 선택

JSON Input

Q. 어떤 Aggregation에 어떤 parameter가 쓰이는지 외워야 되나?

A.

1) 사용하려는 Aggregation 검색 👑 (ex: Moving Average)
2) Parameter List 확인
3) Kibana JSON Input에 비슷하게 넣어보기
4) Visualization Spy - Request에서 확인하고 디버깅하기

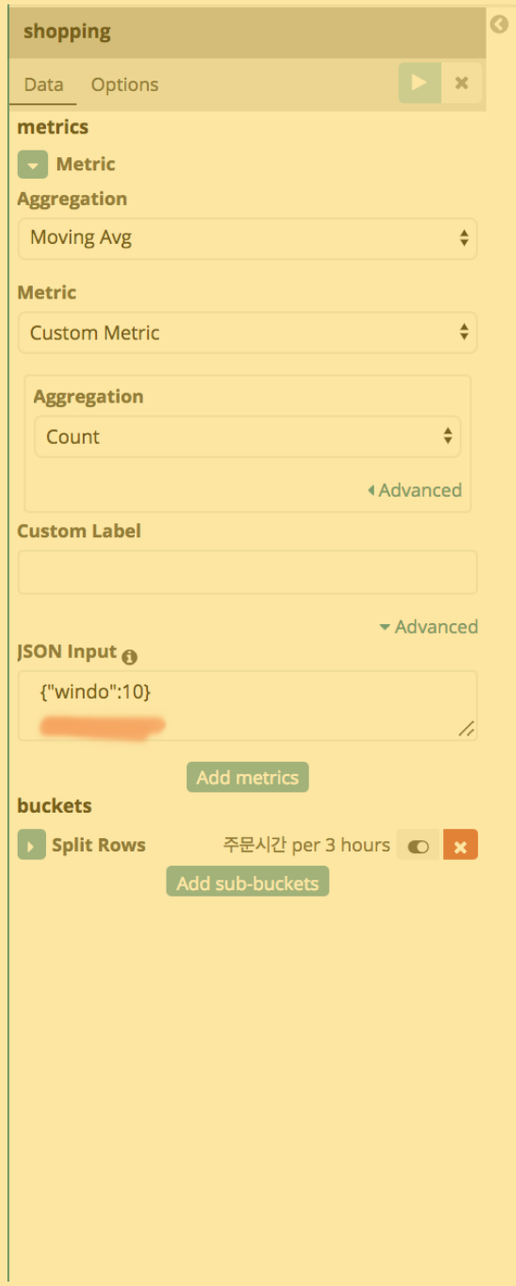
JSON Input

JSON Input 넣고 확인하기

Elasticsearch

```
POST /_search
{
    "size": 0,
    "aggs": {
        "my_date_histo":{
            "date_histogram":{
                "field":"date",
                "interval":"1M"
            },
            "aggs":{
                "the_sum":{
                    "sum":{ "field": "price" }
                },
                "the_movavg":{
                    "moving_avg":{
                        "buckets_path": "the_sum",
                        "window" : 30,
                        "model" : "simple"
                    }
                }
            }
        }
    }
}
```

Kibana JSON



JSON Input

missing data가 있을 때 특정한 값으로 대체할 수 없나?

이동평균을 구할 때 window size를 변경할 수 없나?

Term Aggregation 시 5개 이하 Bucket은 제외할 수 없나?

JSON Input ✍

다음과 같은 Visualization을 만들자

데이터

- Index : shopping
- Time Range : 2017년 10월 1일~ 2017년 10월 10일

문제

- “주문시간”을 기준으로 일별로(daily)
- “상품가격”의 합과,
- “상품가격”의 합의 이동평균을
- 각각 표시하자

사용한 Visualization

- Data Table

사용한 Aggregation

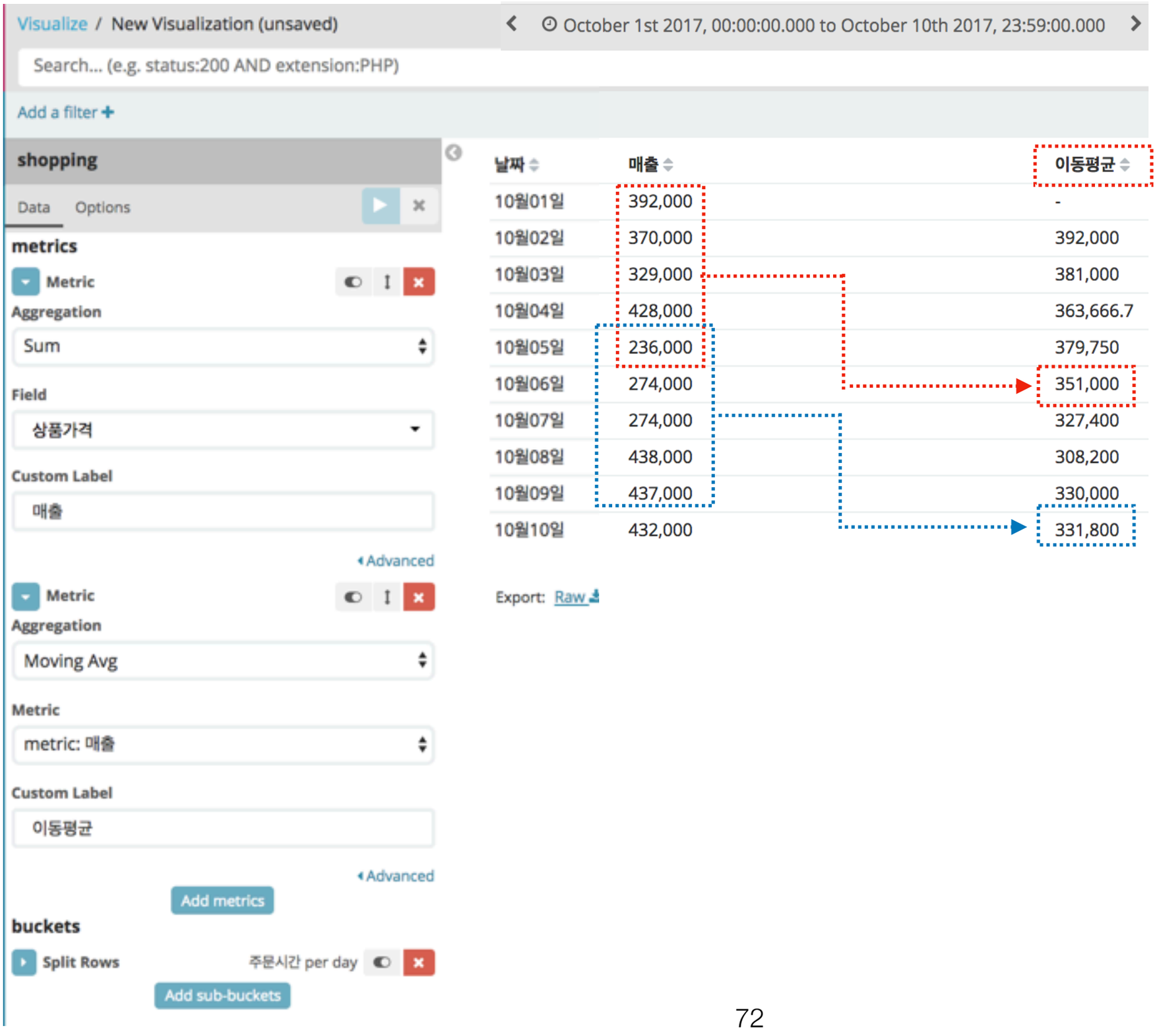
- Metrics : Sum Aggregation, Moving Average Aggregation
- Buckets : Date Histogram

사용 필드

- 상품가격
- 주문시간

JSON Input

Moving Average 확인

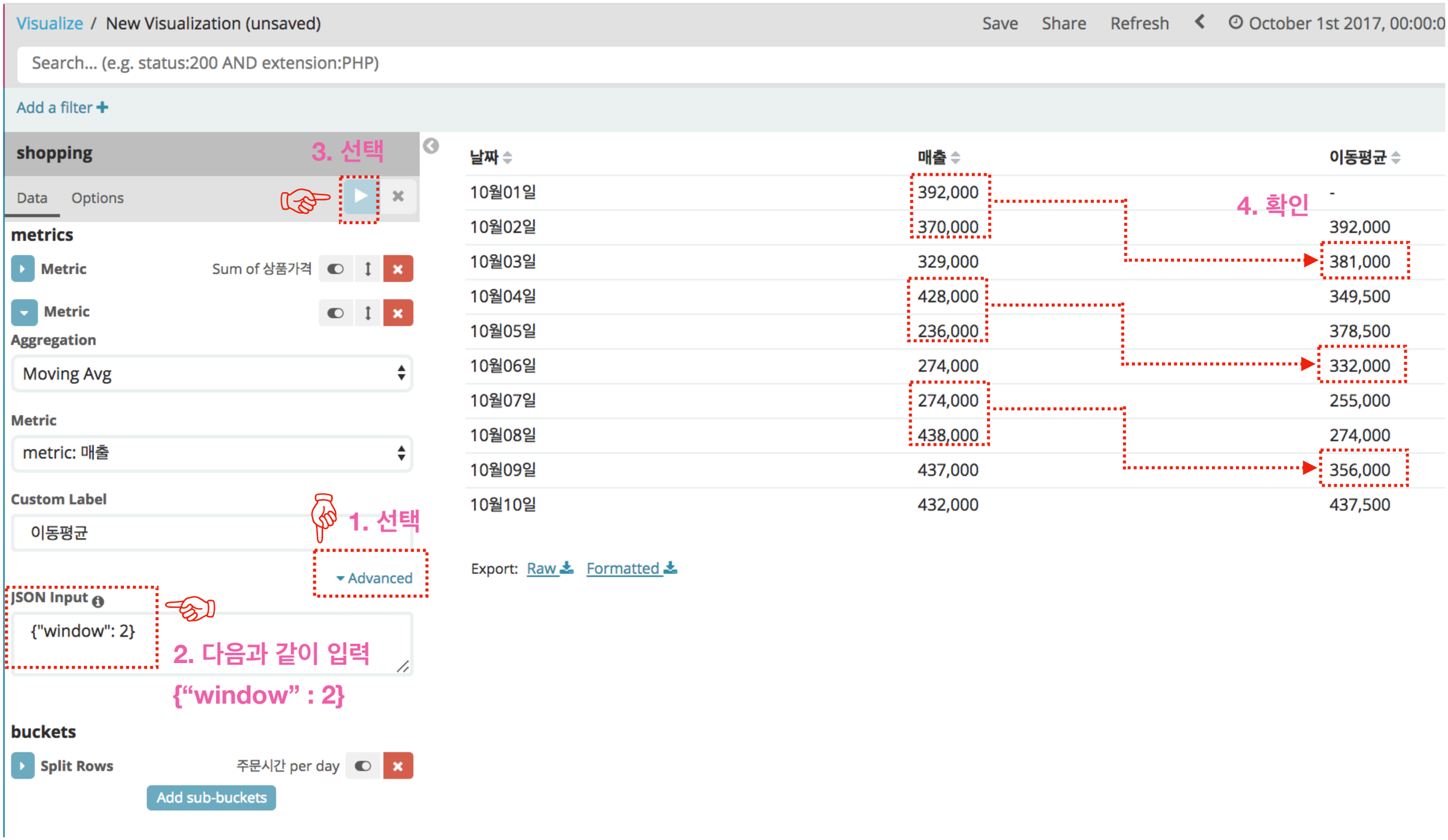


이동평균은 어떻게 구할까? 👑

=> default window size : 5

JSON Input

Moving Average window size 변경



JSON Input

missing data가 있을 때 특정한 값으로 대체할 수 없나?

이동평균을 구할 때 window size를 변경할 수 없나?

Term Aggregation 시 5개 이하 Bucket은 제외할 수 없나?

JSON Input ✍

다음과 같은 Visualization을 만들자

데이터

- Index : shopping
- Time Range : 2017년 10월 1일~ 2017년 10월 10일

문제

- 개수가 가장 많은 “상품분류”를 10개 선별하고
- 그 개수를 바탕으로 Tag Size를 표시하자

사용한 Visualization

- Tag Cloud

사용한 Aggregation

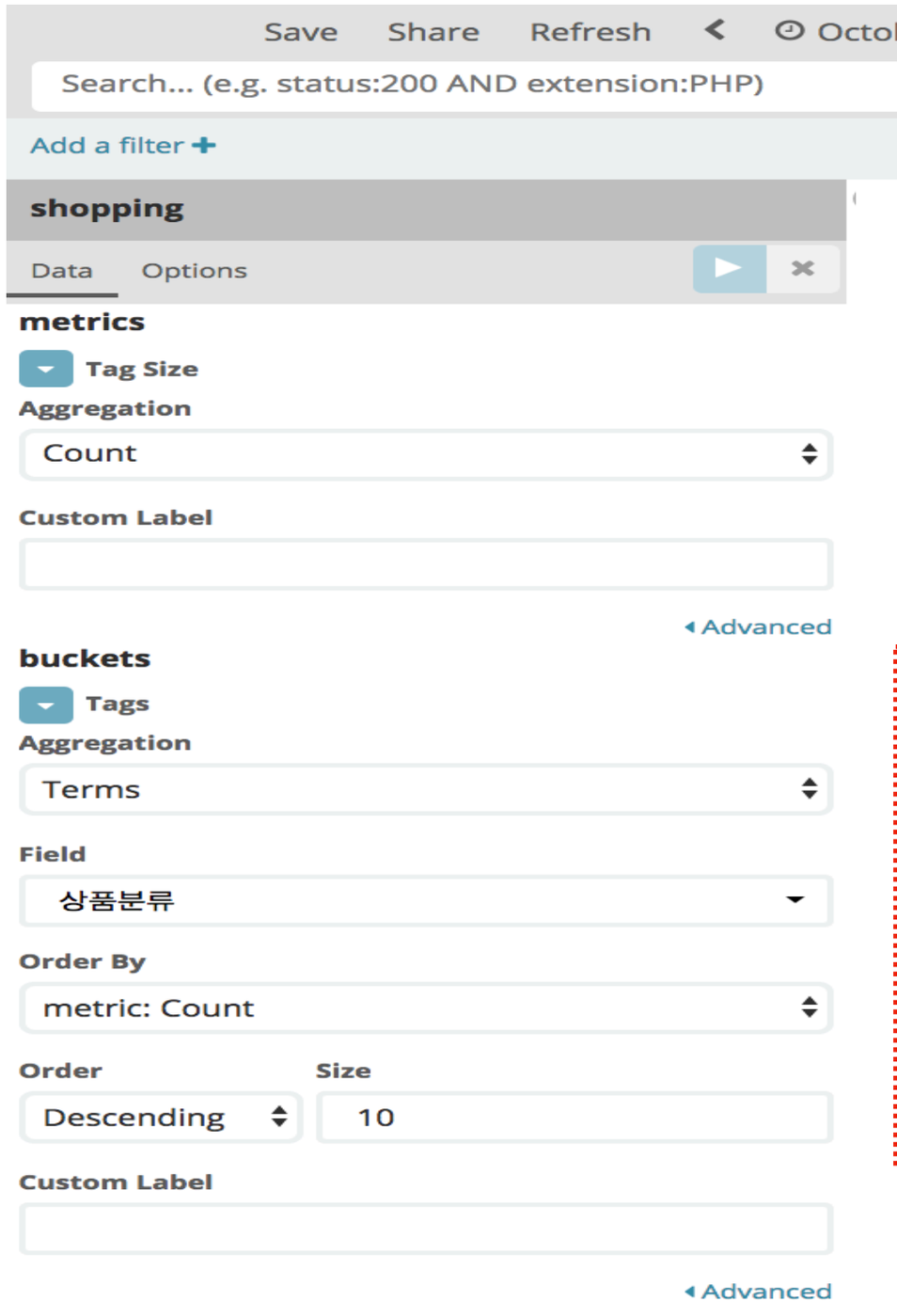
- Metrics : Count Aggregation
- Buckets : Terms Aggregation

사용 필드

- 상품분류

JSON Input

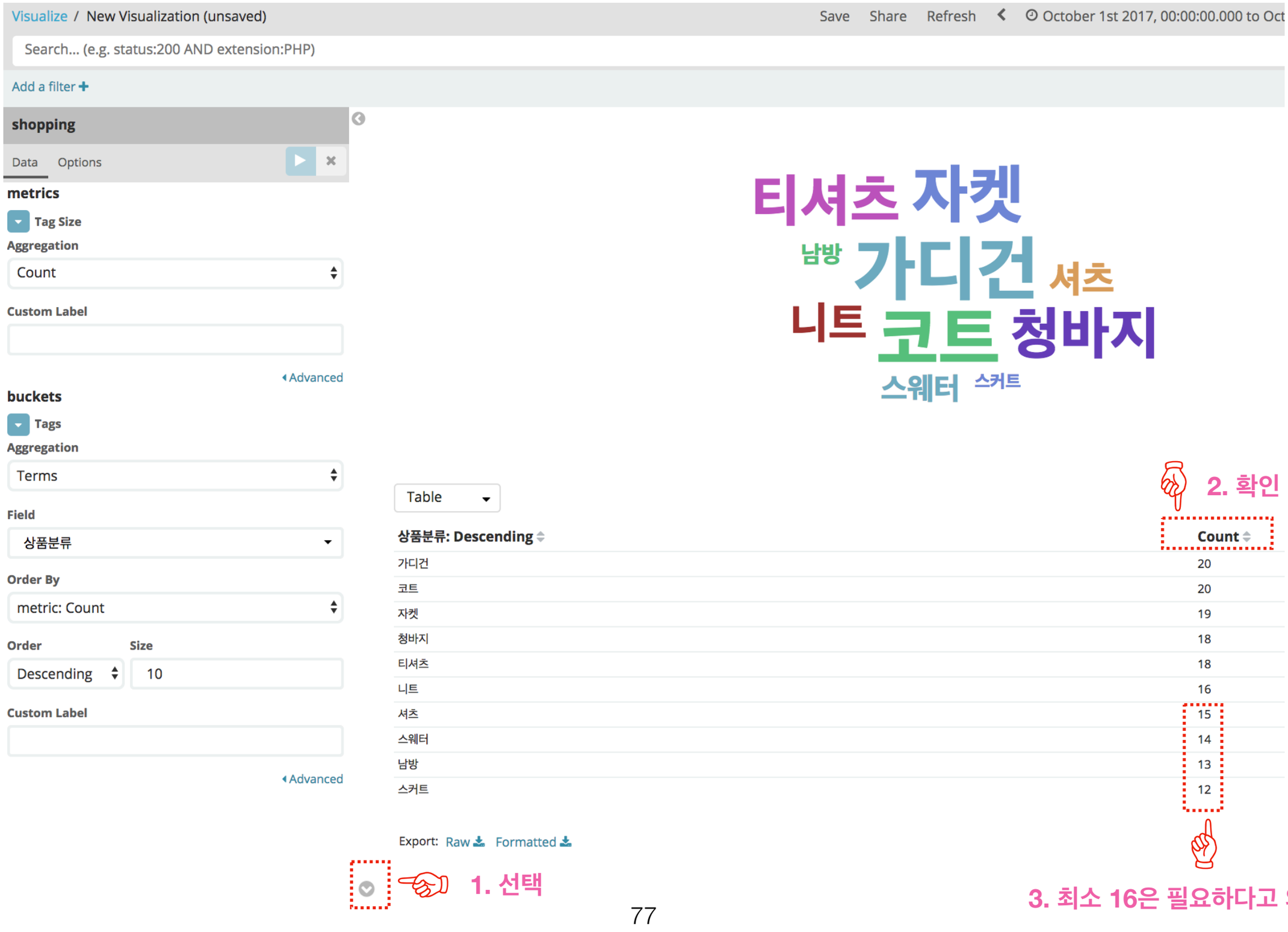
Visualization 확인



만약 이 중에서 Count가 너무 작은 상품이 있다면?
즉, Count의 최소값을 설정하고 싶다면?

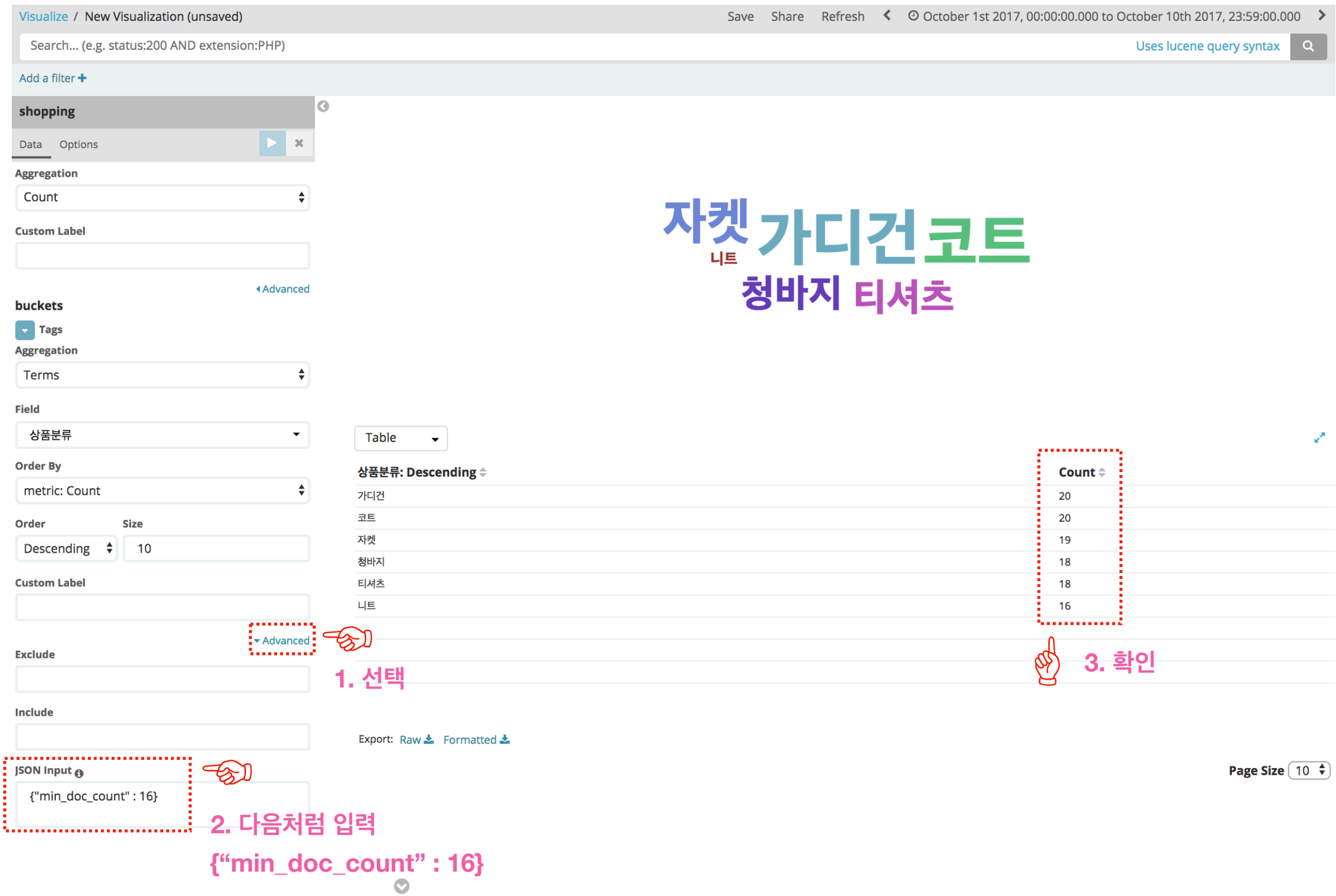
JSON Input

Data 확인



JSON Input

Bucket 별 최소값 설정




Scripted Field

Scripted Field

String Concatenation 연산은 안되나?

Field 간 연산은 안되나?

Date Field에서 연/월/일/시/분/초 등 접근 안되나?

기존 Field에 조건을 적용해서 새로운 Field를 만들 수 없나?

Scripted Field - String Concat

String Concatenation 연산은 안되나? 

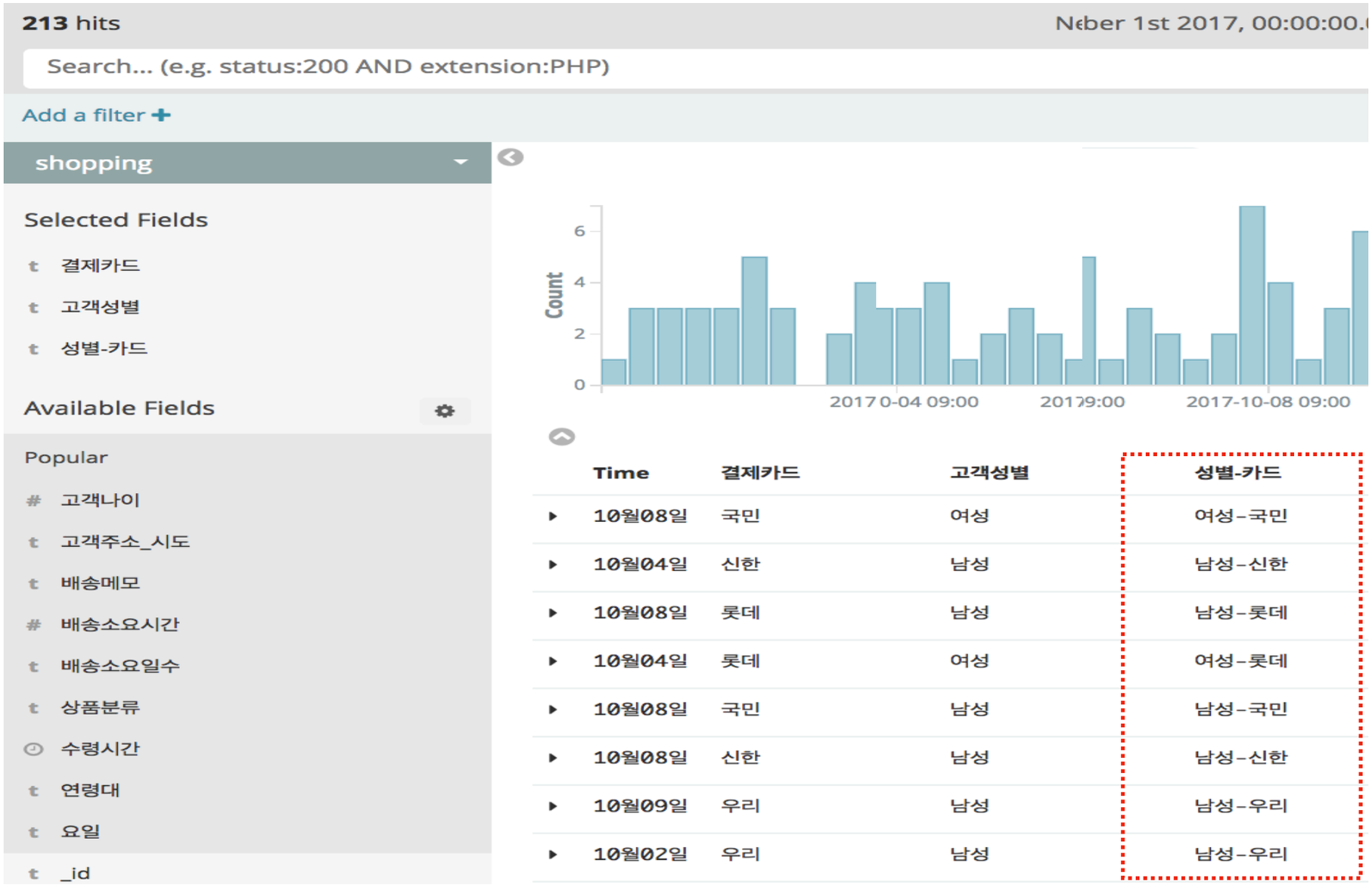
Field 간 연산은 안되나?

Date Field에서 연/월/일/시/분/초 등 접근 안되나?

기존 Field에 조건을 적용해서 새로운 Field를 만들 수 없나?

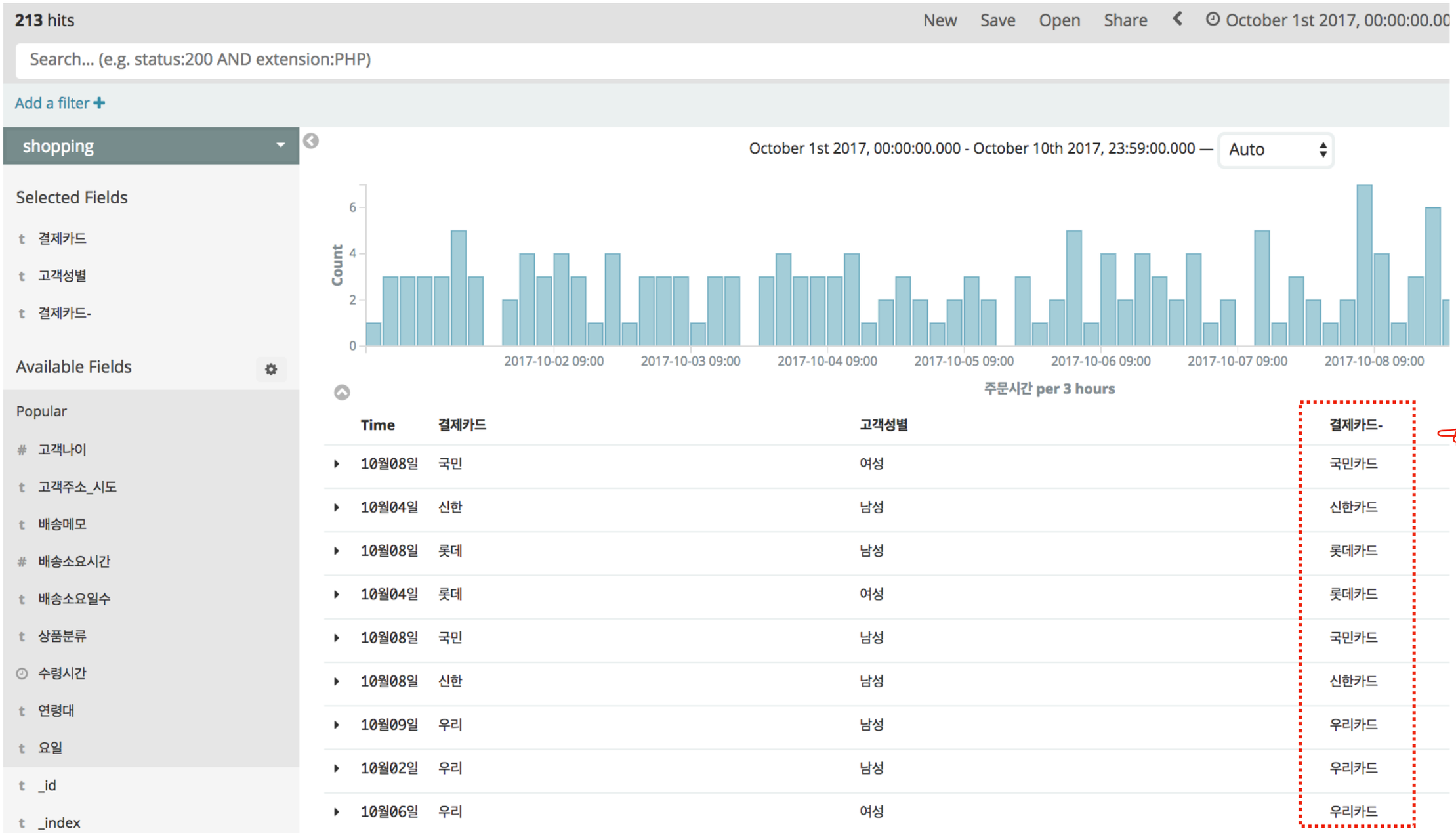
Scripted Field - String Concat

특정한 두 개 혹은 그 이상의 Field를 합쳐서 하나의 Field를 만들고 싶으면 어떻게 해야할까?



Scripted Field - String Concat

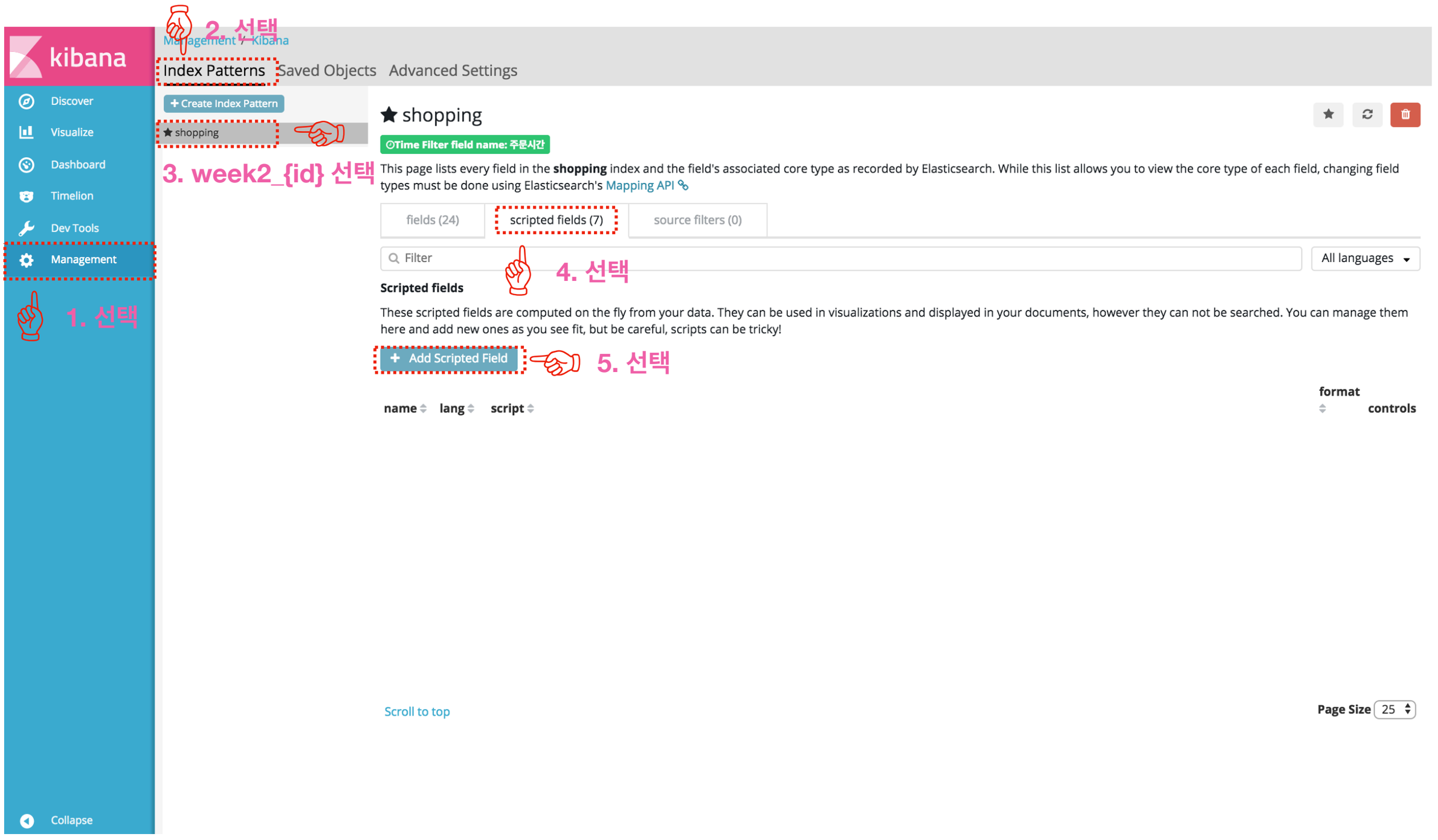
혹은 특정 Field에 글자를 덧붙이고 싶다면?



Time	결제카드	고객성별	결제카드-
10월08일	국민	여성	국민카드
10월04일	신한	남성	신한카드
10월08일	롯데	남성	롯데카드
10월04일	롯데	여성	롯데카드
10월08일	국민	남성	국민카드
10월08일	신한	남성	신한카드
10월09일	우리	남성	우리카드
10월02일	우리	남성	우리카드
10월06일	우리	여성	우리카드

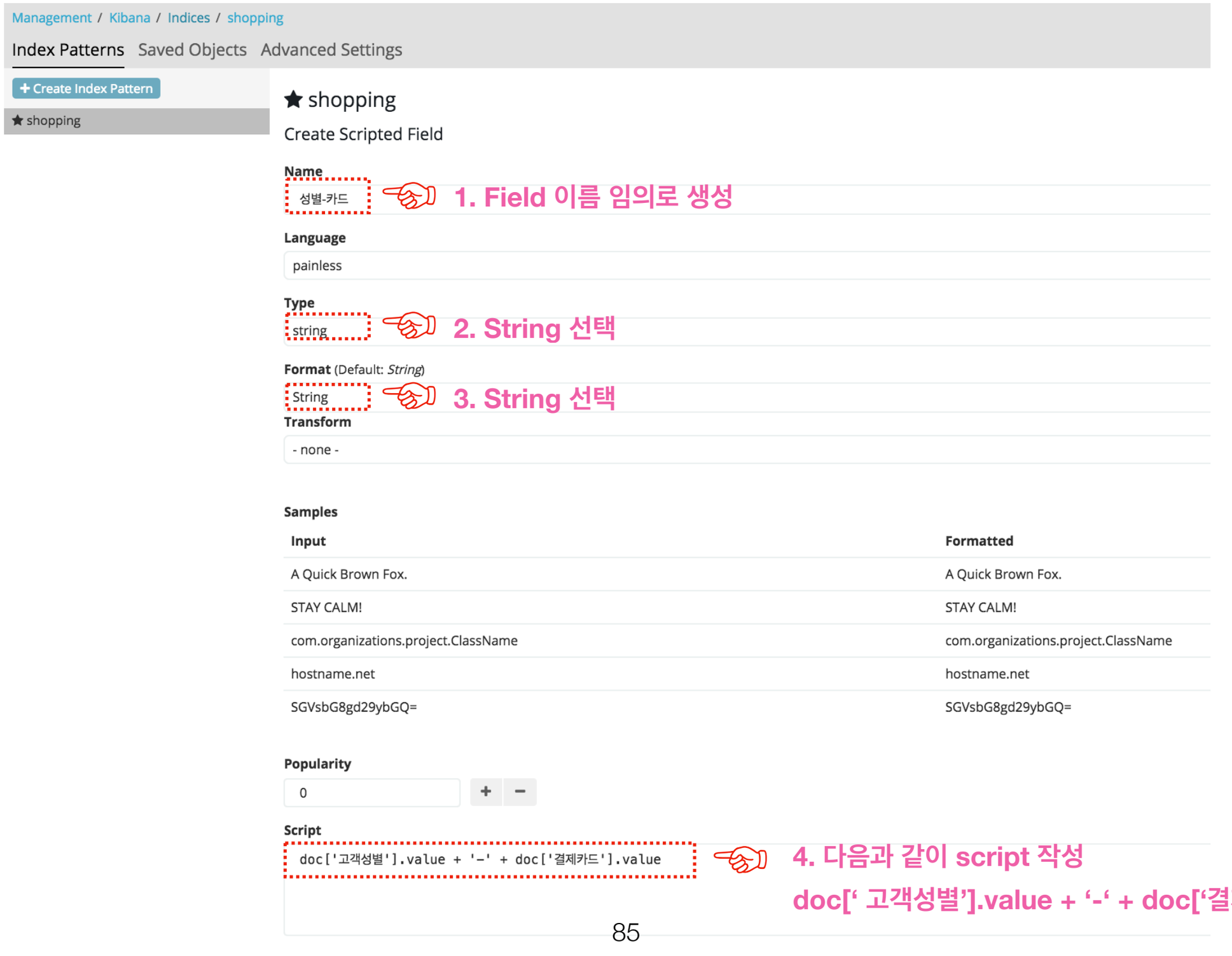
Scripted Field - String Concat

scripted field를 추가하자



Scripted Field - String Concat

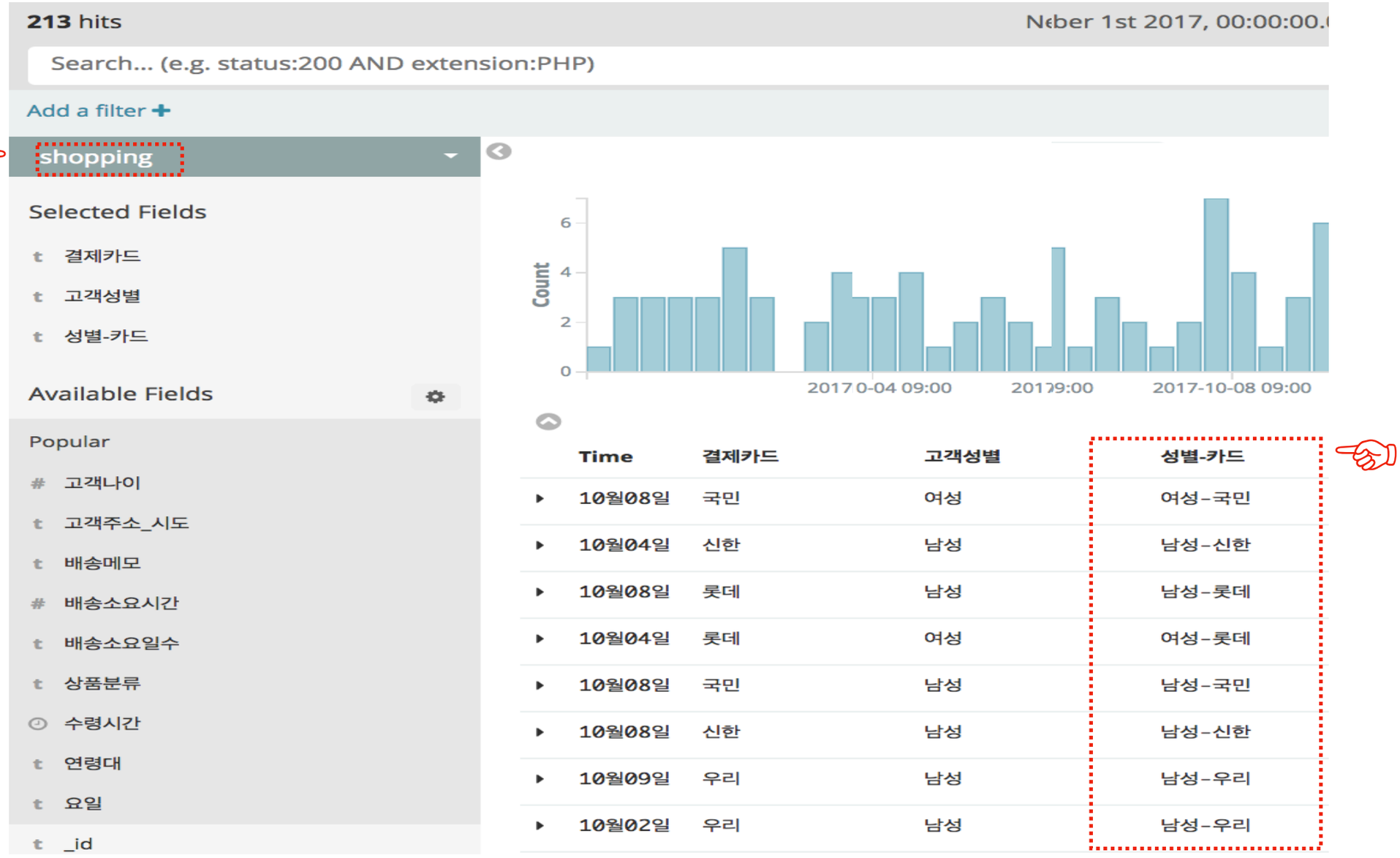
script를 작성하자



Scripted Field - String Concat

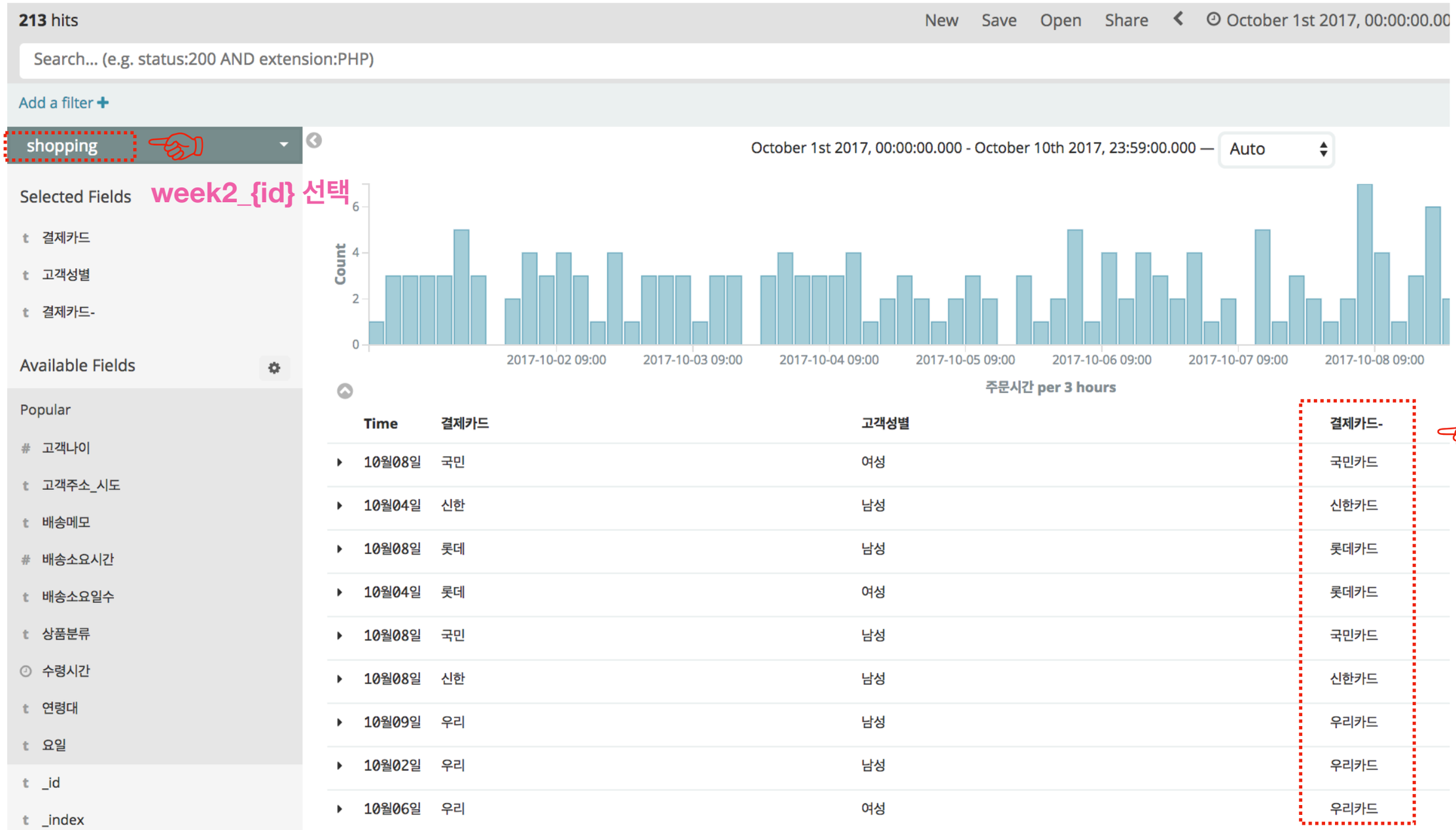
Discover에 가서 확인하자

week2_{id} 선택



Scripted Field - String Concat

이번에는 아래와 같은 Field를 생성해보자 ✍



Scripted Field - arithmetic operation

String Concatenation 연산은 안되나?

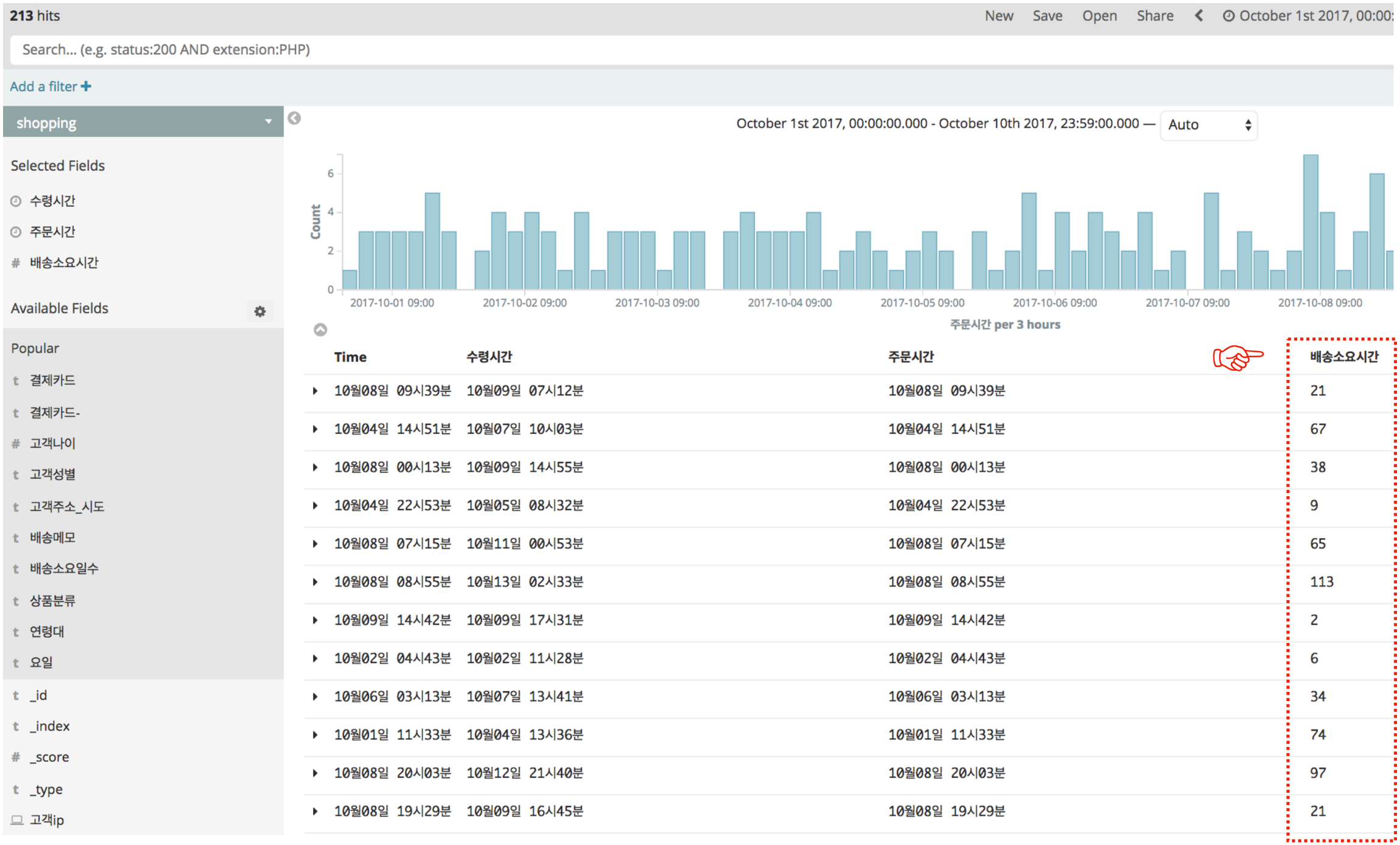
Field 간 연산은 안되나?

Date Field에서 연/월/일/시/분/초 등 접근 안되나?

기존 Field에 조건을 적용해서 새로운 Field를 만들 수 없나?

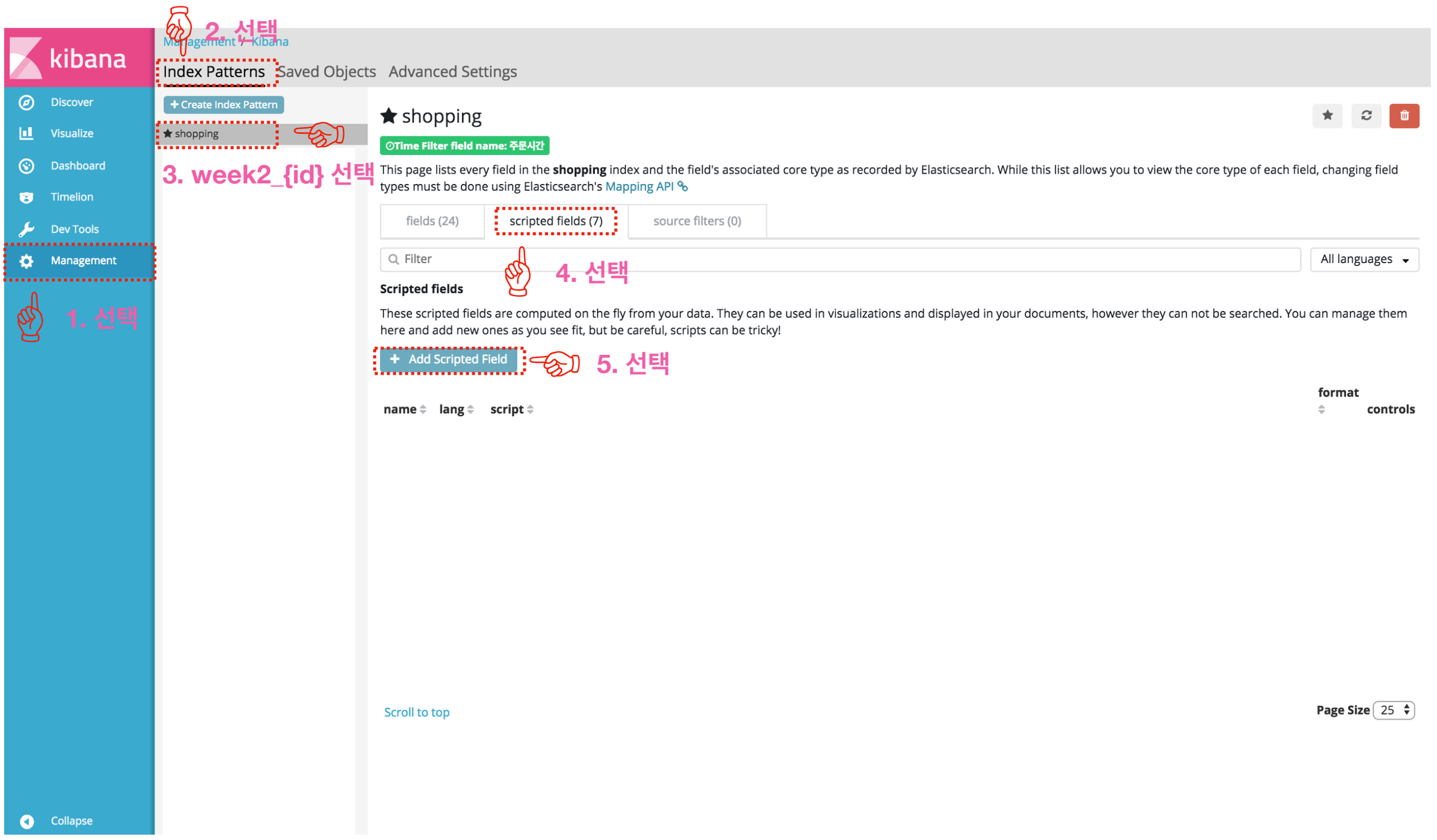
Scripted Field - arithmetic operation

2개의 Date Field의 차이를 통해 특정한 event의 소요시간을 구하고 싶다면?



Scripted Field - arithmetic operation

scripted field를 추가하자



Scripted Field - arithmetic operation

script를 작성하자

Management / Kibana / Indices / shopping

Index Patterns | Saved Objects | Advanced Settings

+ Create Index Pattern

★ shopping

★ shopping

Create Scripted Field

Name

배송소요시간-

1. Field 이름 임의로 생성

Language

painless

Type

number

2. Number 선택

Format (Default: *Number*)

Number

3. Number 선택

Numeral.js format pattern (Default: "0,0.[0]")

0,0.[0]

4. 적당한 Format 입력

Samples

Input	Formatted
10000	10,000
12.345678	12.3
-1	-1
-999	-999
0.52	0.5

Popularity

0 + −

Script

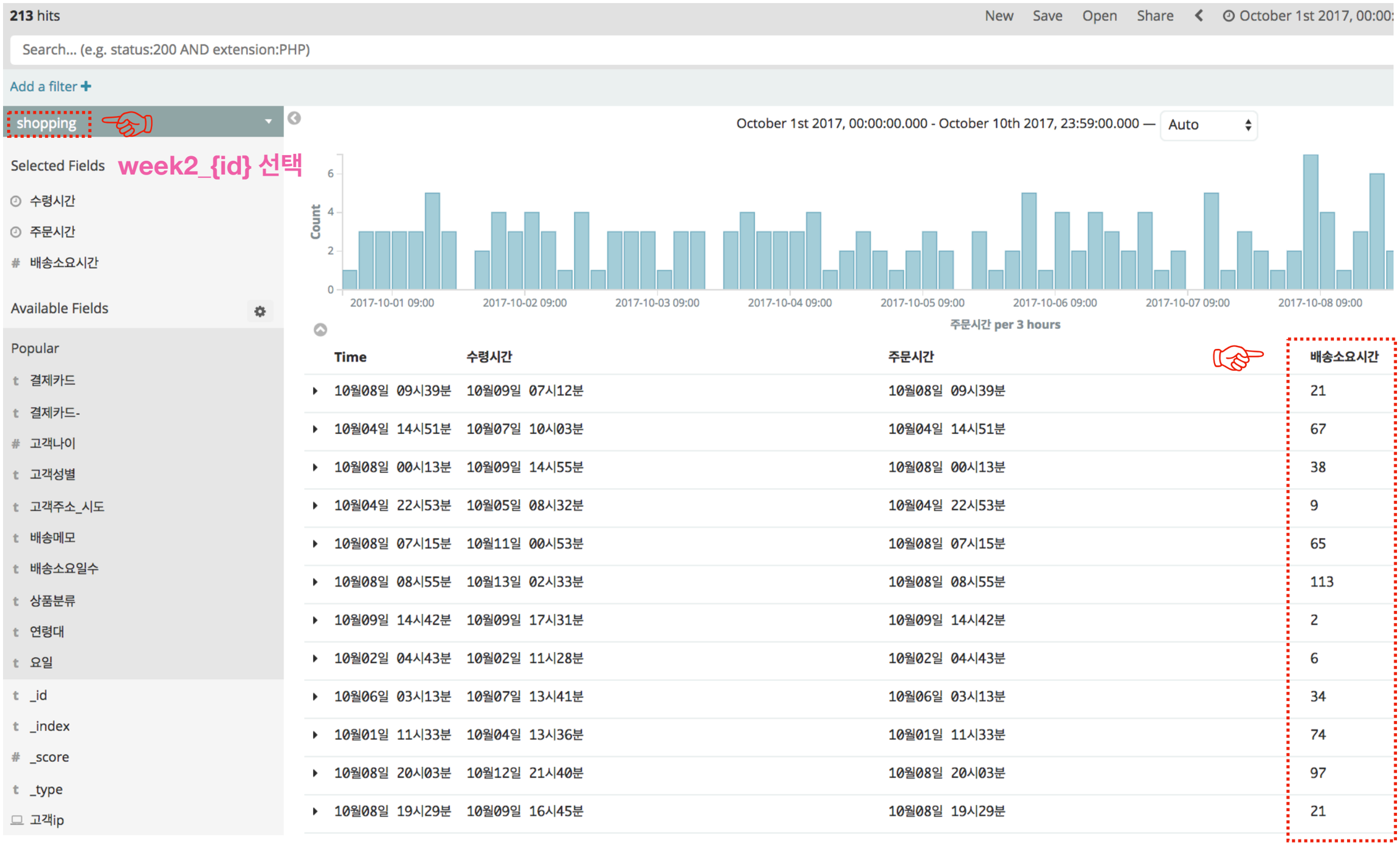
```
(doc['수령시간'].value-doc['주문시간'].value)/1000/60/60
```

5. 다음과 같이 script 작성

(doc[‘수령시간’].value-doc[‘주문시간’].value)/1000/60/60

Scripted Field - arithmetic operation

Discover에 가서 확인하자



Scripted Field - Date

String Concatenation 연산은 안되나?

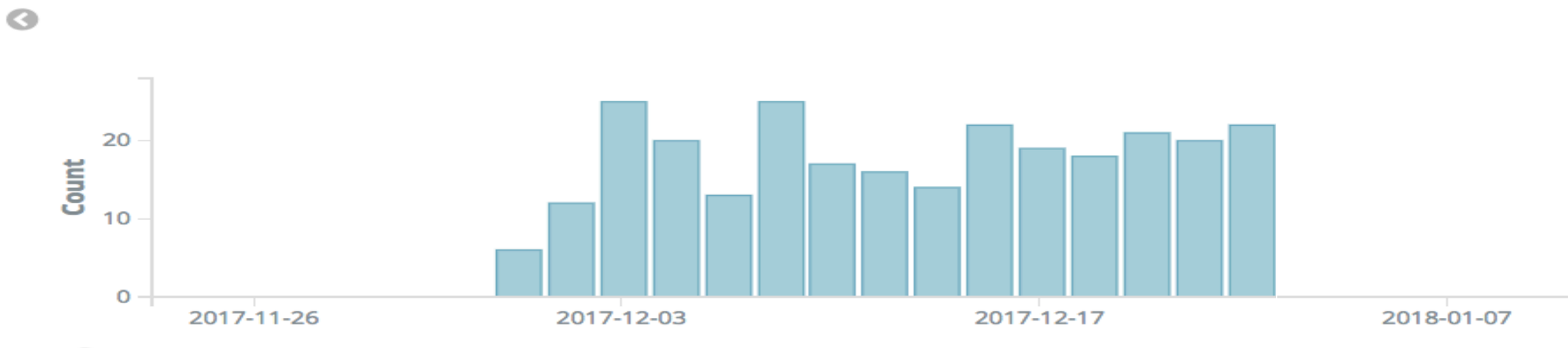
Field 간 연산은 안되나?

Date Field에서 연/월/일/시/분/초 등 접근 안되나?

기존 Field에 조건을 적용해서 새로운 Field를 만들 수 없나?

Scripted Field - Date

Date Field에서 요일, 시간대 등을 뽑고 싶으면?



If 사용해야 한다

Time	요일	요일-	시간대
01월01일 08시58분	일	7	23
01월01일 08시51분	일	7	23
01월01일 06시47분	일	7	21
01월01일 05시43분	일	7	20
01월01일 05시30분	일	7	20
01월01일 05시18분	일	7	20
01월01일 04시11분	일	7	19
01월01일 01시48분	일	7	16
01월01일 01시47분	일	7	16

Scripted Field - Date

아래를 참고해서 “주문시간”의 요일과 시간대를 추출하자 ✍

Expression	Description
`doc['field_name'].date.centuryOfEra`	Century (1-2920000)
`doc['field_name'].date.dayOfMonth`	Day (1-31), e.g. `1` for the first of the month.
`doc['field_name'].date.dayOfWeek`	Day of the week (1-7), e.g. `1` for Monday.
`doc['field_name'].date.dayOfYear`	Day of the year, e.g. `1` for January 1.
`doc['field_name'].date.era`	Era: `0` for BC, `1` for AD.
`doc['field_name'].date.hourOfDay`	Hour (0-23).
`doc['field_name'].date.millisOfDay`	Milliseconds within the day (0-86399999).
`doc['field_name'].date.millisOfSecond`	Milliseconds within the second (0-999).
`doc['field_name'].date.minuteOfDay`	Minute within the day (0-1439).
`doc['field_name'].date.minuteOfHour`	Minute within the hour (0-59).
`doc['field_name'].date.monthOfYear`	Month within the year (1-12), e.g. `1` for January.
`doc['field_name'].date.secondOfDay`	Second within the day (0-86399).
`doc['field_name'].date.secondOfMinute`	Second within the minute (0-59).
`doc['field_name'].date.year`	Year (-292000000 - 292000000).
`doc['field_name'].date.yearOfCentury`	Year within the century (1-100).
`doc['field_name'].date.yearOfEra`	Year within the era (1-292000000).

Scripted Field - If Condition

String Concatenation 연산은 안되나?

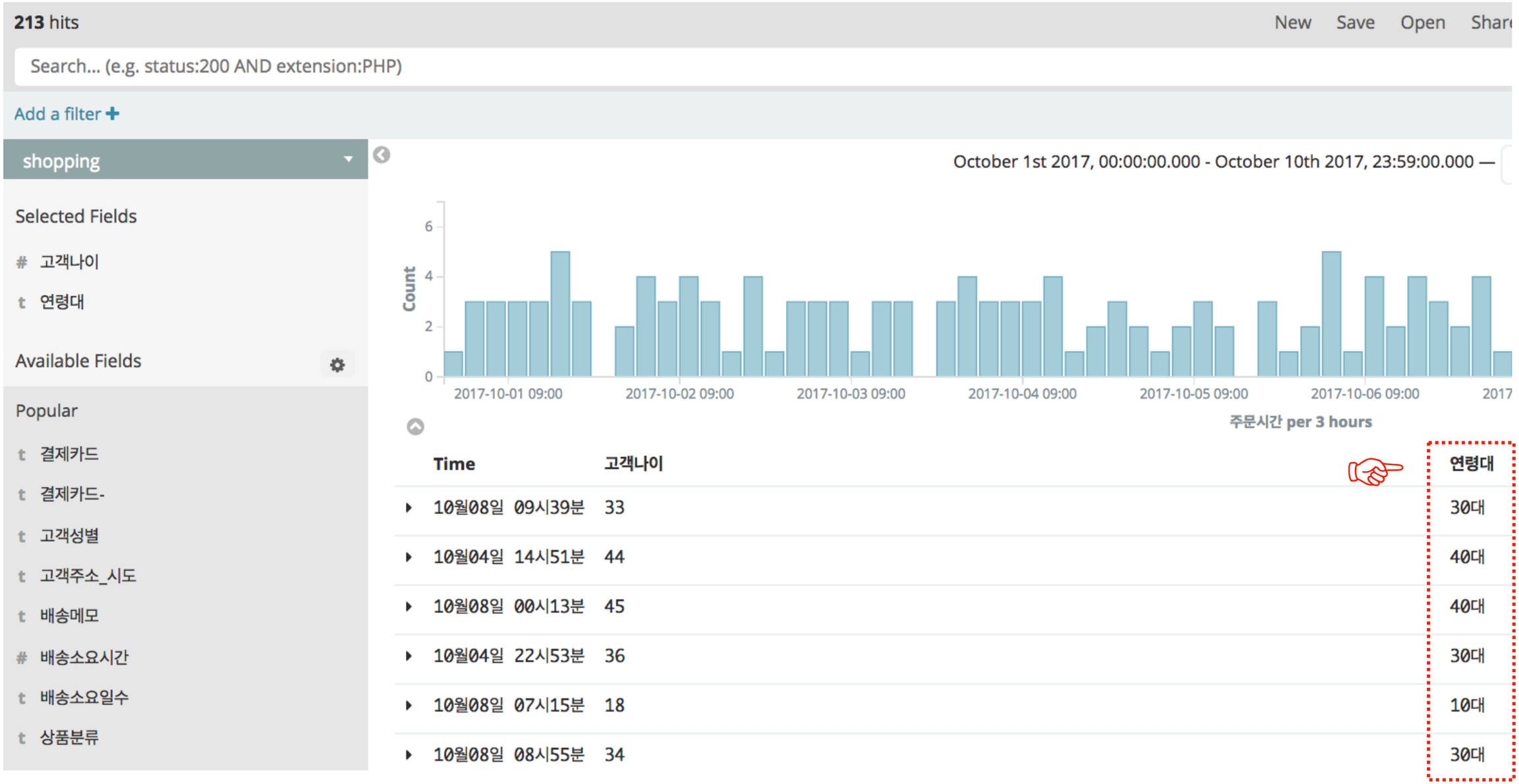
Field 간 연산은 안되나?

Date Field에서 연/월/일/시/분/초 등 접근 안되나?

기존 Field에 조건을 적용해서 새로운 Field를 만들 수 없나? 👉

Scripted Field - If Condition

특정 조건에 따라 Value를 설정하고 싶다면?



Scripted Field - If Condition

script를 작성하자 👑

★ shopping

Create Scripted Field

Name

연령대-

1. Field 이름 임의로 생성

Language

painless

Type

string

2. String 선택

Format (Default: *String*)

String

3. String 선택

Transform

Popularity

0 + −

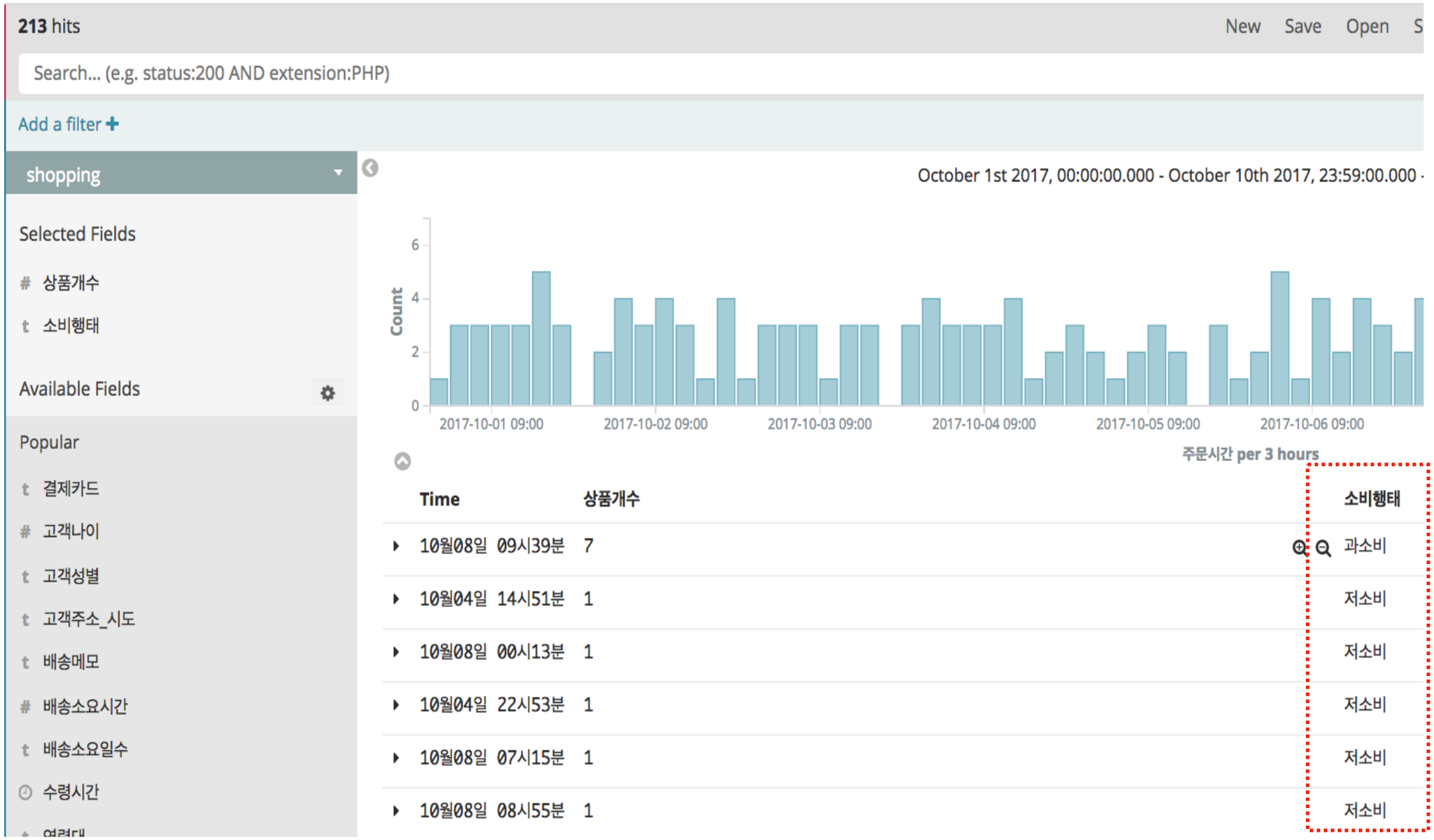
Script

```
if (doc['고객나이'].value < 20) {
  return "10대"
}
else if (doc['고객나이'].value < 30) {
  return "20대"
}
else if (doc['고객나이'].value < 40) {
  return "30대"
}
else if (doc['고객나이'].value < 50) {
  return "40대" }
else {
  return "50대 이상"
}
```

4. Script 작성

Scripted Field - If Condition

아래와 같은 Field를 생성해보자 ✍



"상품개수"에 따라 카테고리화 하기

- 1~2 : 저소비
- 3~5 : 평균
- 6~7 : 과소비

Scripted Field

단, Scripted Field는 **Search**는 안된다!!

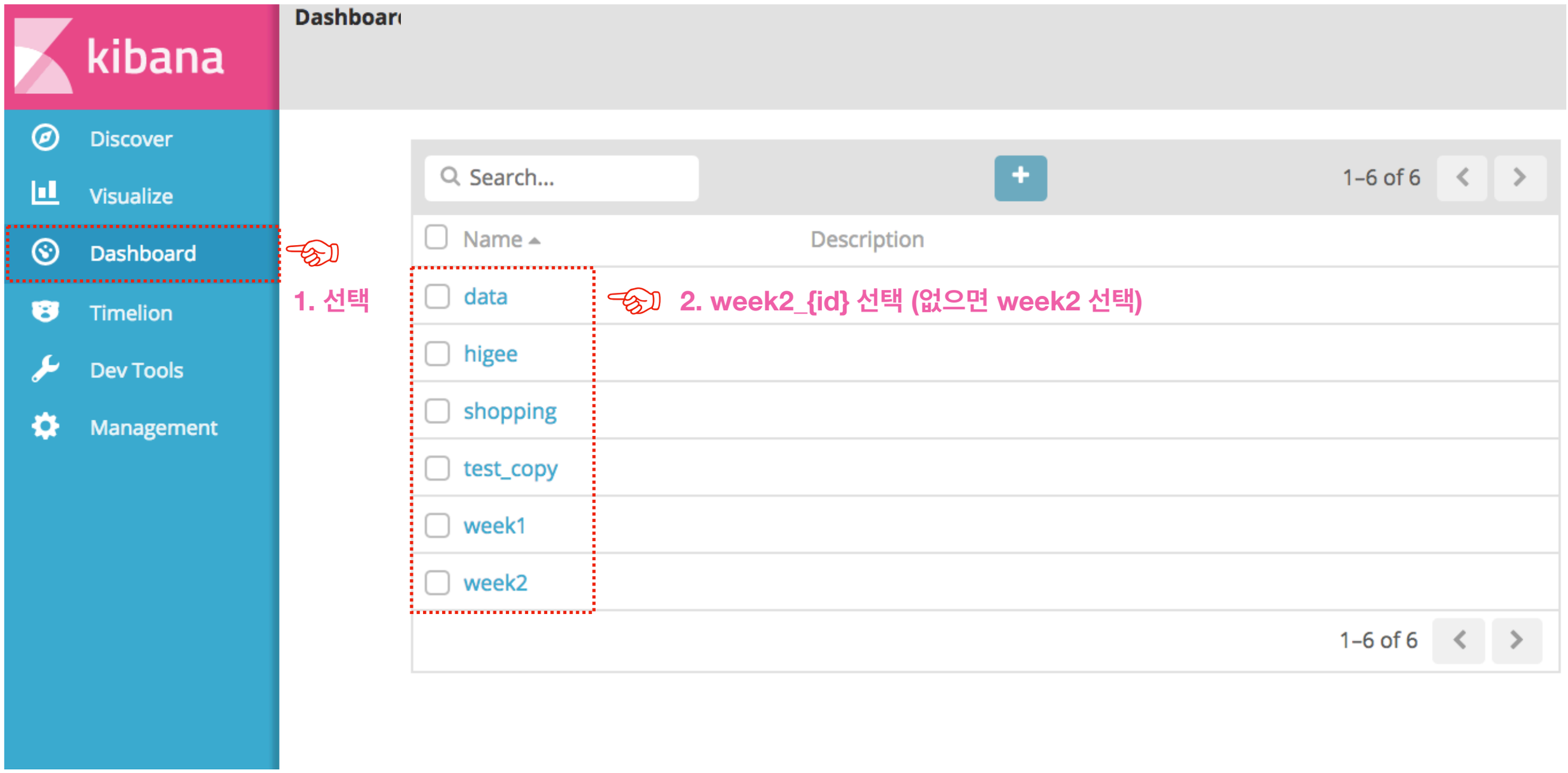
그러므로 Scripted Field를 검색/조회할 때는 **Filter**를 이용해야 한다

Filtering by Field ♕

Dashboard는 만들었는데
원하는 조건의 데이터만 보고 싶으면?

Filter by Field

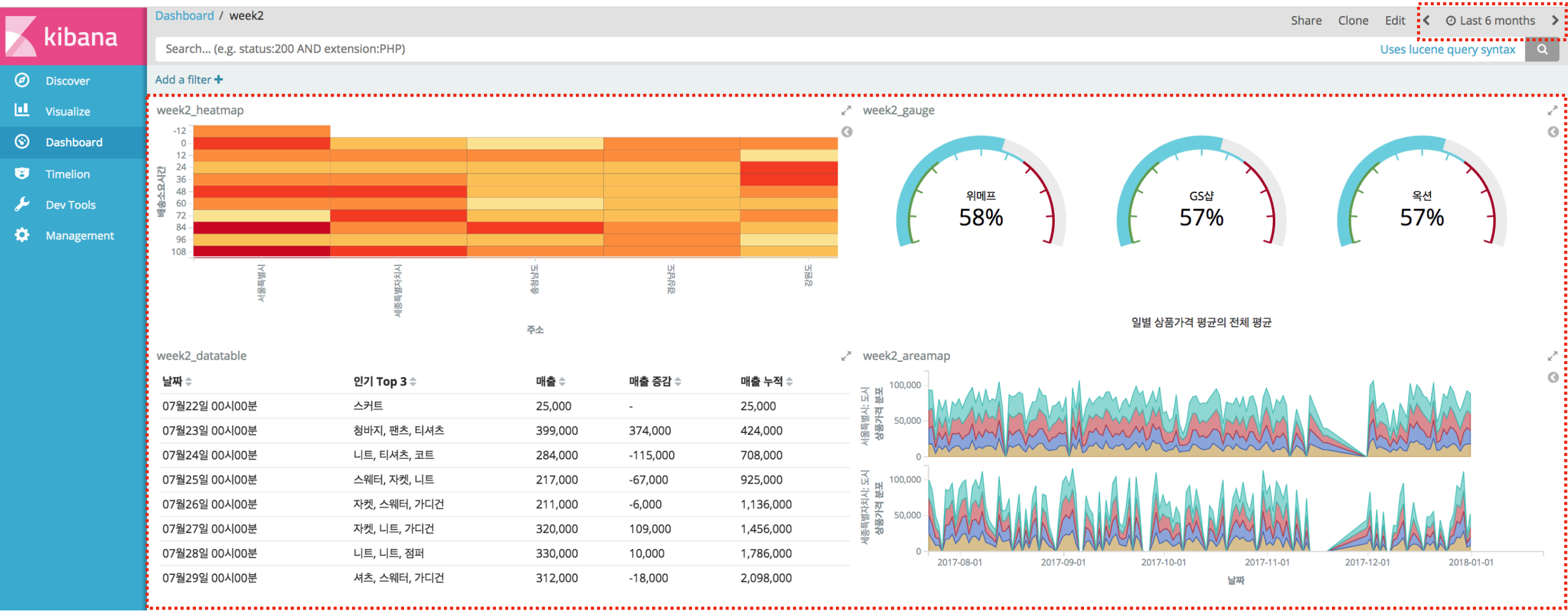
우선 Dashboard를 열자



Filter by Field

Dashboard를 확인하자

Time Picker



날짜	인기 Top 3	매출	매출 증감	매출 누적
07월22일 00시00분	스커트	25,000	-	25,000
07월23일 00시00분	청바지, 팬츠, 티셔츠	399,000	374,000	424,000
07월24일 00시00분	니트, 티셔츠, 코트	284,000	-115,000	708,000
07월25일 00시00분	스웨터, 자켓, 니트	217,000	-67,000	925,000
07월26일 00시00분	자켓, 스웨터, 가디건	211,000	-6,000	1,136,000
07월27일 00시00분	자켓, 니트, 가디건	320,000	109,000	1,456,000
07월28일 00시00분	니트, 니트, 점퍼	330,000	10,000	1,786,000
07월29일 00시00분	셔츠, 스웨터, 가디건	312,000	-18,000	2,098,000

전체 Documents 중에서 **Time Picker 구간에 속한** Documents만 보여준다.

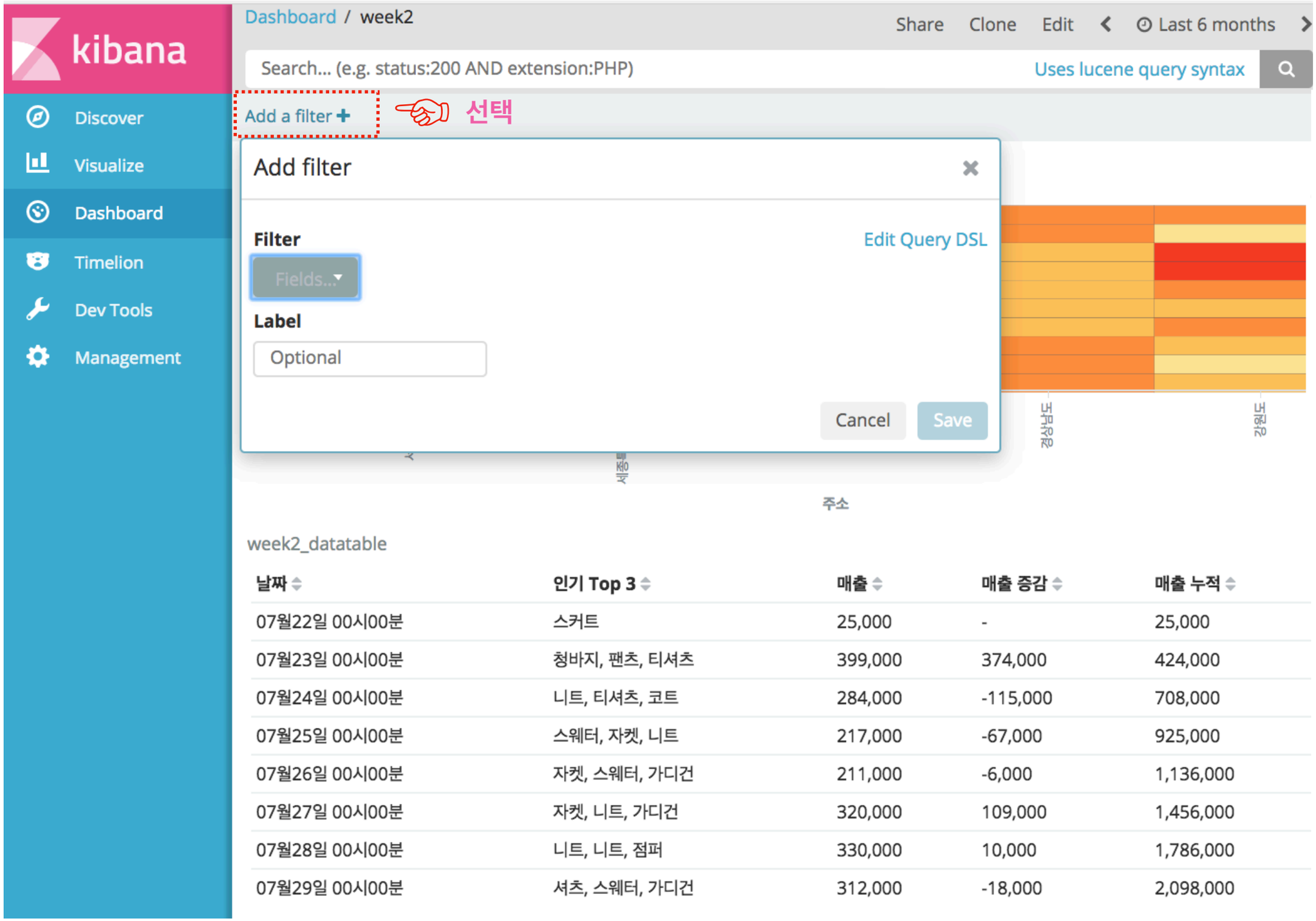
만약에 다른 조건을 추가하고 싶다면? 예를 들어, 20대의 데이터만 보고 싶으면 어떻게 할까?

Filter by Field

Filter를 이용하면 **특정 조건을 만족**하는
데이터만 선별하여 Dashboard에 시각화할 수 있다.

Filter by Field

Filter를 실행하자

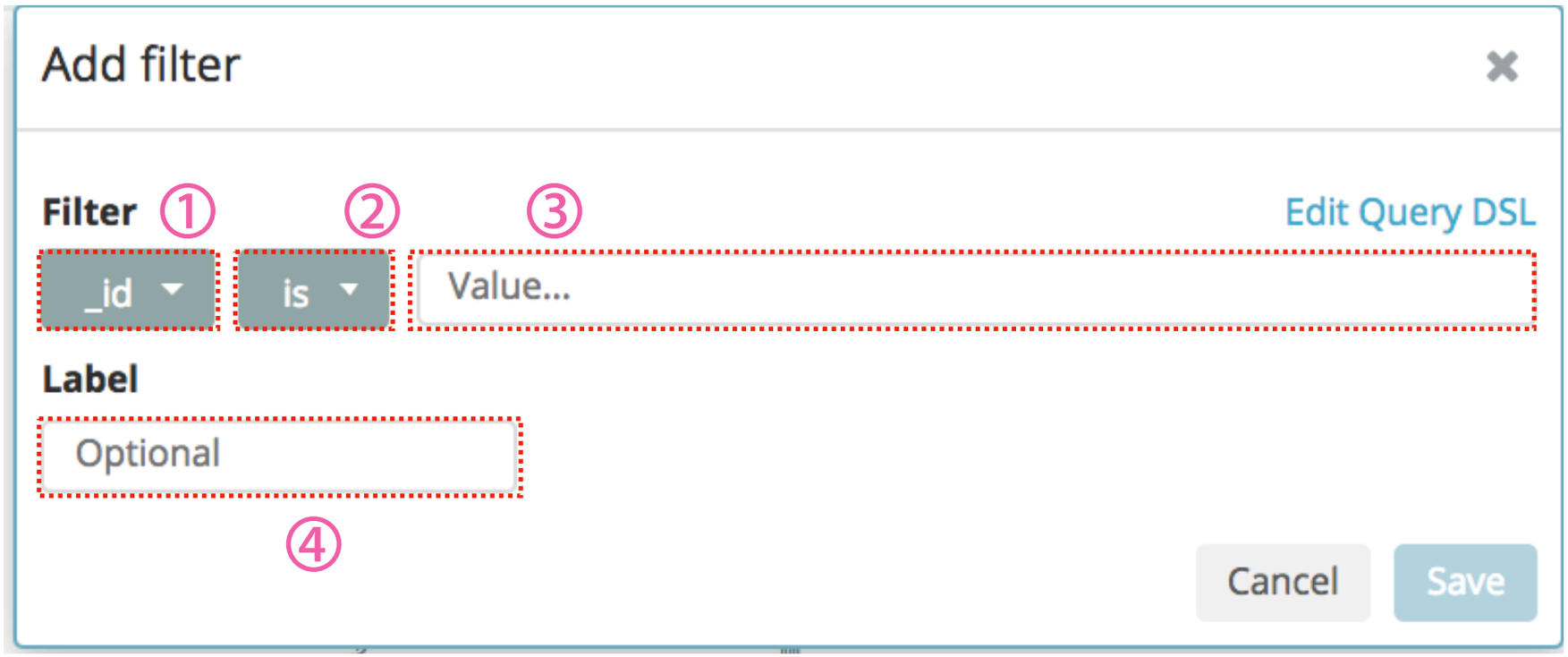


week2_datatable

날짜	인기 Top 3	매출	매출 증감	매출 누적
07월22일 00시00분	스커트	25,000	-	25,000
07월23일 00시00분	청바지, 팬츠, 티셔츠	399,000	374,000	424,000
07월24일 00시00분	니트, 티셔츠, 코트	284,000	-115,000	708,000
07월25일 00시00분	스웨터, 자켓, 니트	217,000	-67,000	925,000
07월26일 00시00분	자켓, 스웨터, 가디건	211,000	-6,000	1,136,000
07월27일 00시00분	자켓, 니트, 가디건	320,000	109,000	1,456,000
07월28일 00시00분	니트, 니트, 점퍼	330,000	10,000	1,786,000
07월29일 00시00분	셔츠, 스웨터, 가디건	312,000	-18,000	2,098,000

Filter by Field

Filter의 사용법을 익히자



① Filter 적용할 Field 선택

② 적용할 Operator 선택 (다음 페이지 참조)

③ Filter에 적용하려는 Value 입력

④ (여러 Filter 구분하기 위한) 이름 입력

Filter by Field

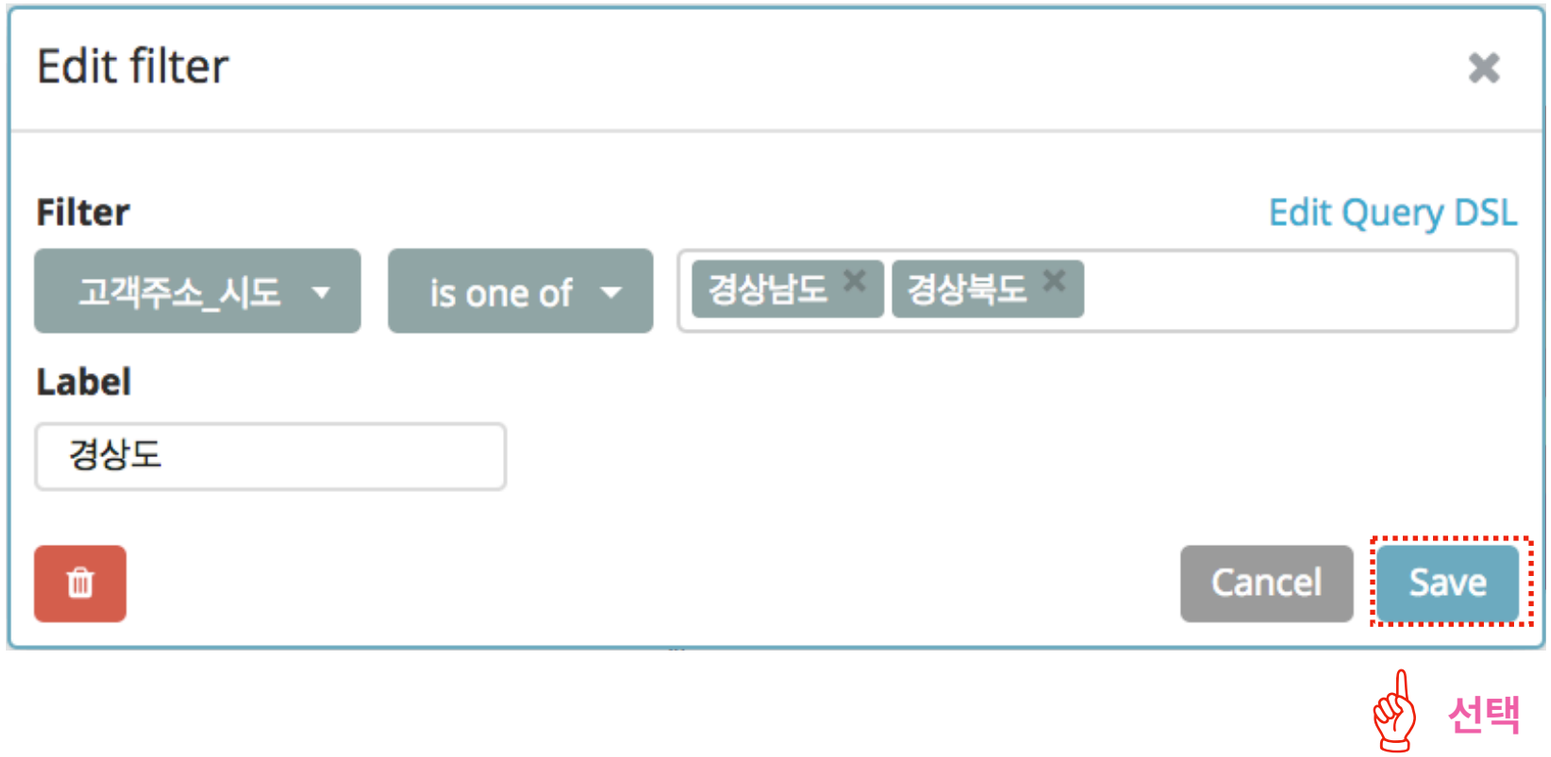
Operator 설명

Operator	역할
is	Field의 Value가 입력한 값과 일치하는 Documents 선택
is not	Field의 Value가 입력한 값과 일치하지 않는 Documents 선택
is one of	Field의 Value가 입력한 값 중에 존재하는 Documents 선택
is not one of	Field의 Value가 입력한 값 중에 존재하지 않는 Documents 선택
exists	Field가 적어도 한 개의 non-null 값을 가지는 Documents 선택
does not exist	Field가 존재하지 않거나 null 값만 가지는 Documents 선택
is between	Field의 Value가 입력한 값 사이에 존재하는 Documents 검색
is not between	Field의 Value가 입력한 값 사이에 존재하지 않는 Documents 검색

Filter by Field

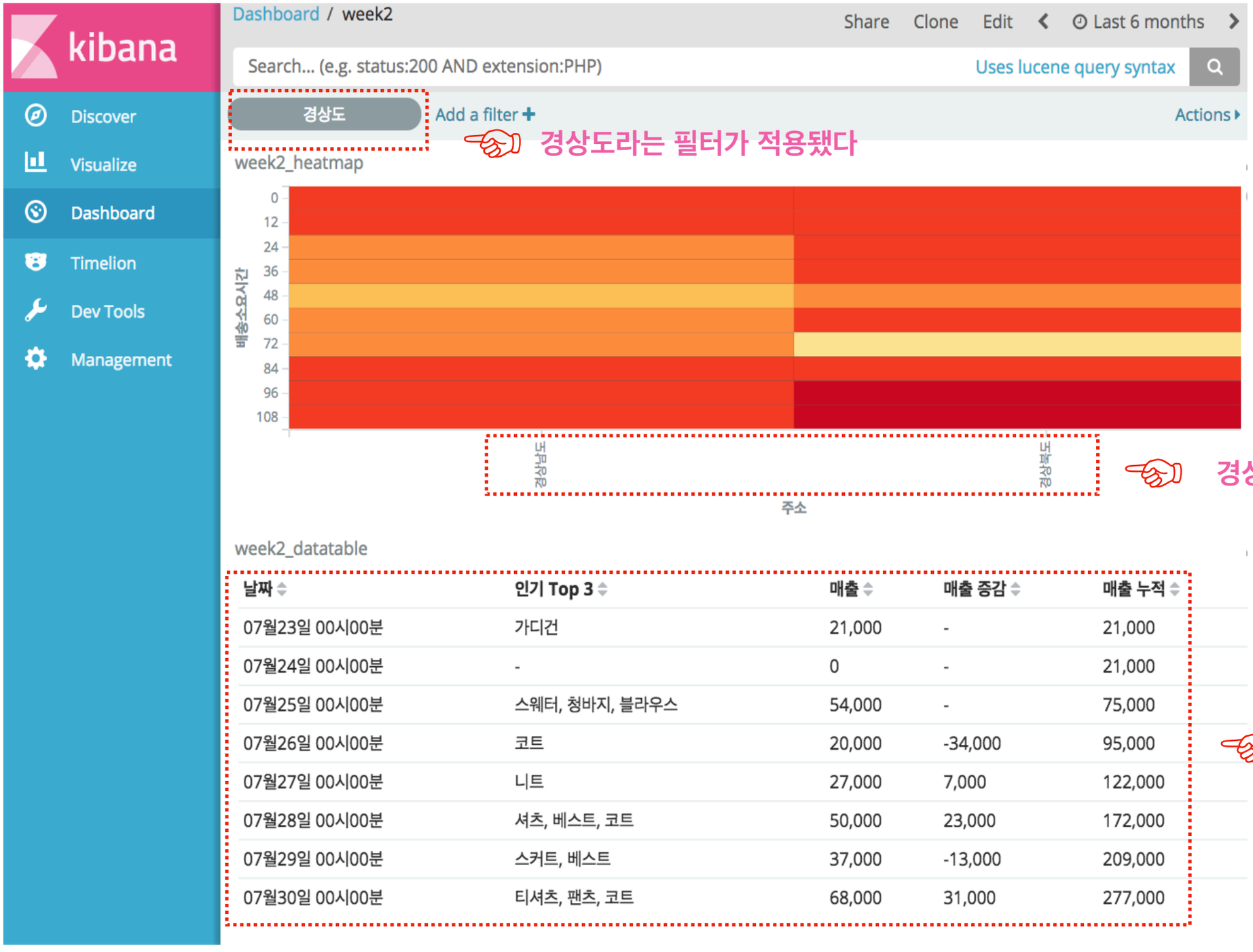
실 사례를 보자 ✍

“고객주소_시도” 가 “경상남도” 또는 “경상북도” 인 Documents를 선택하는 Filter를 만들고 “경상도”라고 하자



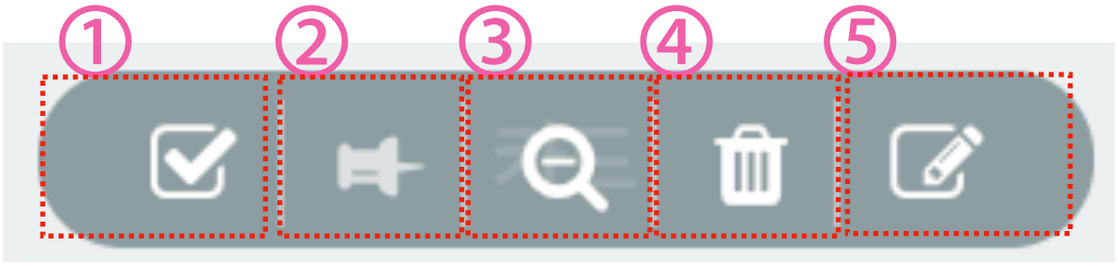
Filter by Field

다시 Dashboard를 보자



Filter by Field

필터에 마우스오버하면...



① 필터 적용 <=> 필터 적용 해제

② 필터 고정 (Discover, Visualize, Dashboard)

③ 필터 효과 적용 <=> 필터 효과 반대 적용

④ 필터 삭제

⑤ 필터 수정

Filter by Field

Discover에서 Filter를 적용해보자

Filter 적용

Filter된 Documents 하나하나를 볼 수 있다

Filter by Field

아래와 같은 Fiter를 Dashboard에 적용해보자 ✍

Dashboard : week2_{id}

Time Range : Last 6 months

1. “고객성별”이 여성인 Data
2. “결제카드”가 우리 또는 국민인 Data
3. “고객성별”이 남성이면서 “연령대”가 20대
4. “구매사이트”가 쿠팡이거나 “상품개수”가 1~3인 Data
5. “결제카드”가 “우”로 시작하는 모든 Data
6. “구매사이트”가 22번가(오타 아니에요)와 유사한 Data

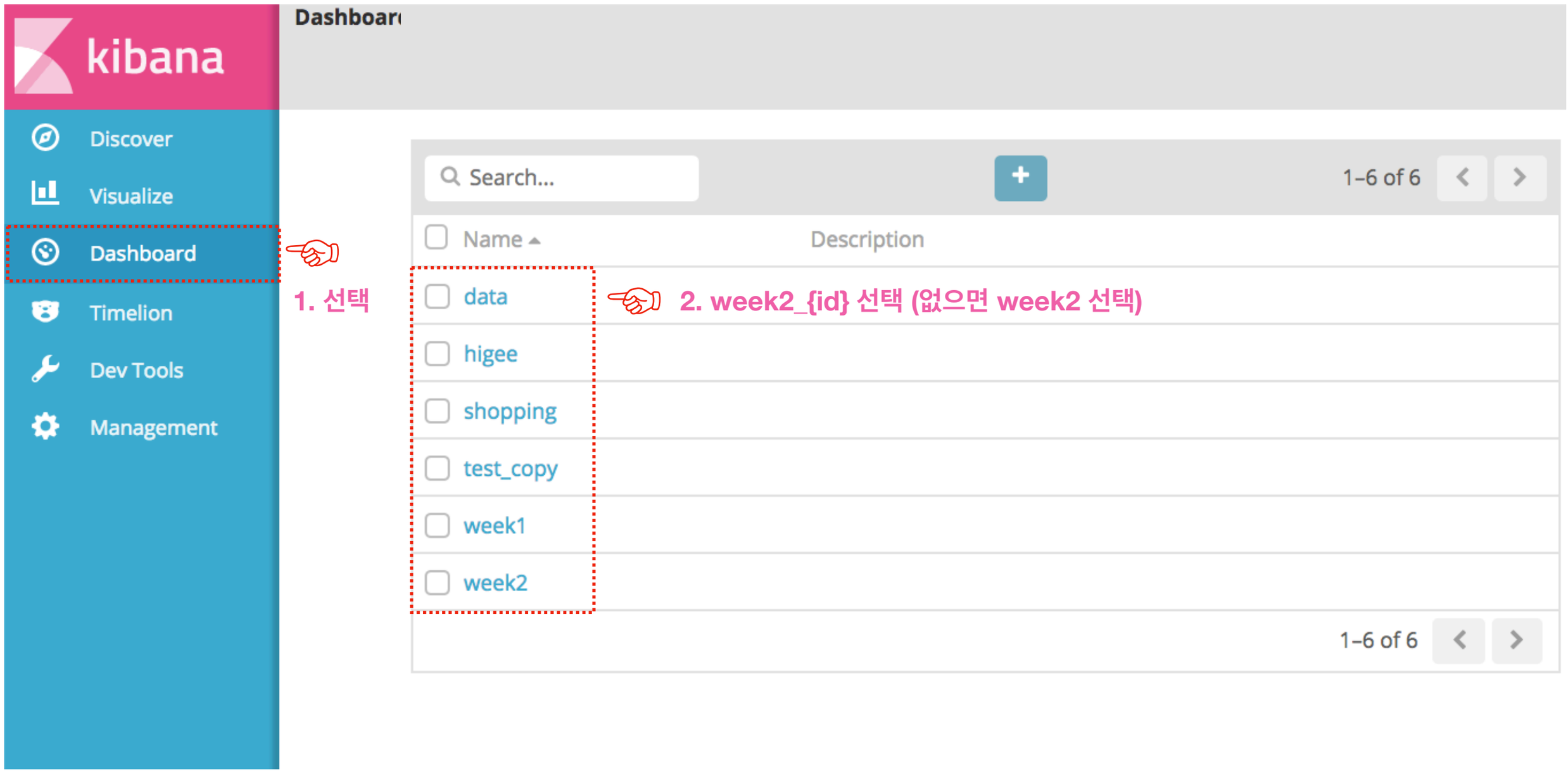
Lucene Query ♕

Filter는 사용하기 간단하나 기능이 제한적이다

+

더 스마트한 검색!!

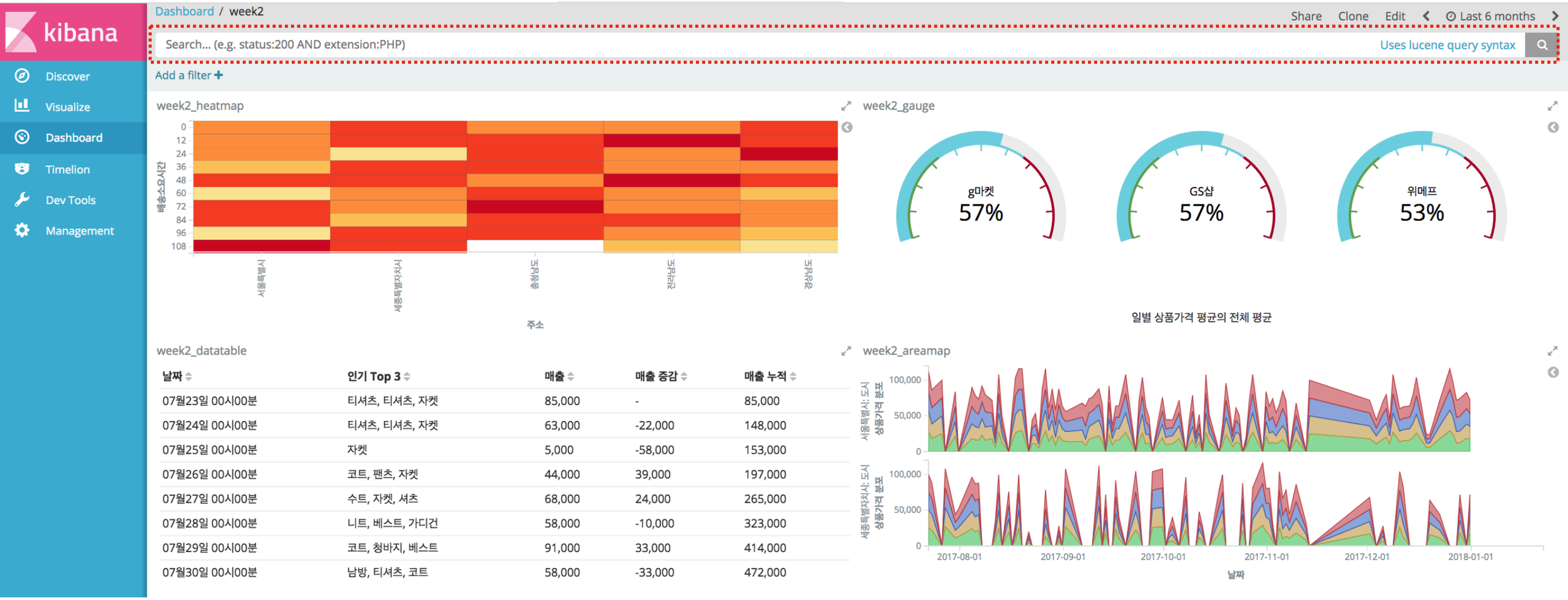
우선 Dashboard를 열자



Lucene Query

Query Bar를 확인하자

Query Bar



날짜	인기 Top 3	매출	매출 증감	매출 누적
07월23일 00시00분	티셔츠, 티셔츠, 자켓	85,000	-	85,000
07월24일 00시00분	티셔츠, 티셔츠, 자켓	63,000	-22,000	148,000
07월25일 00시00분	자켓	5,000	-58,000	153,000
07월26일 00시00분	코트, 팬츠, 자켓	44,000	39,000	197,000
07월27일 00시00분	수트, 자켓, 셔츠	68,000	24,000	265,000
07월28일 00시00분	니트, 베스트, 가디건	58,000	-10,000	323,000
07월29일 00시00분	코트, 청바지, 베스트	91,000	33,000	414,000
07월30일 00시00분	남방, 티셔츠, 코트	58,000	-33,000	472,000

Query Bar에 뭐라고 검색을 해야될까?

Lucene Query

Lucene Query의 사용법을 익히자 ✍

종류	기능	Query 예시
Keyword 검색	Field에 상관없이 Value 일치하는 Documents 검색	여성
Field Match 검색	특정 Field의 Value가 일치하는 Documents 검색	고객성별:여성
Exact Field Match 검색	특정 Value가 정확히 모두 일치하는 Documents 검색	배송메모:"상품 이상"
Must be 검색	특정 Field가 존재하는 Documents 검색	_exists_:구매사이트
Must not be present 검색	특정 Field가 존재하지 않는 Documents 검색	_missing_:구매사이트
AND 검색	특정 조건들을 모두 만족하는 Documents 검색	고객성별:여성 AND 상품분류:셔츠
OR 검색	특정 조건들 중 적어도 1개를 만족하는 Documents 검색	고객성별:남성 OR 상품분류:셔츠
NOT 검색	특정 조건을 만족하지 않는 Documents 검색	NOT 구매사이트:옥션
Term 검색	조건 중 적어도 하나라도 만족하는 Documents 검색	상품분류: (니트 코트)
Fuzzy 검색	검색어와 유사한 Documents 검색	경상복도~
Proximity 검색	검색어의 순서를 변경해서 찾을 수 있는 Documents 검색	배송메모:"내에 시간 배송 못함"~2
Numeric Value 검색	Numeric Field Value로 Documents 검색	상품가격:>5000
Range 검색	Field의 Value가 입력한 값 사이에 존재하는 Documents 검색	고객나이 : [10 TO 30]
Wildcard ? 검색	Wildcard ?(한글자)를 활용해서 Documents 검색	서?특별시
Wildcard * 검색	Wildcard *(모든글자)를 활용해서 Documents 검색	쿠*

아래와 같은 Query를 Dashboard에서 검색해보자 ✍

Dashboard : week2_{id}

Time Range : Last 6 months

1. “고객성별”이 여성인 Data
2. “결제카드”가 우리 또는 국민인 Data
3. “고객성별”이 남성이면서 “연령대”가 20대
4. “구매사이트”가 쿠팡이거나 “상품개수”가 1~3인 Data
5. “결제카드”가 “우”로 시작하는 모든 Data
6. “구매사이트”가 22번가(오타 아니에요)와 유사한 Data

Filter와 Search를 비교해보자

		Filter	Search
“고객성별”이 여성인 Data		✓	✓
“결제카드”가 우리 또는 국민인 Data		✓	✓
“고객성별”이 남성이면서 “연령대”가 20대	**SCRIPTED FIELD**	✓	
“구매사이트”가 쿠팡이거나 “상품개수”가 1~3인 Data	**OR 연산**		✓
“결제카드”가 “우”로 시작하는 모든 Data	**WILDCARD 검색**		✓
“구매사이트”가 22번가(오타 아니에요)와 유사한 Data	**FUZZY / PROXIMITY 검색**		✓

질문 및 Feedback은

gshock94@gmail.com로 주세요