

UAS TAKEHOME

Kibana dan ElasticSearch



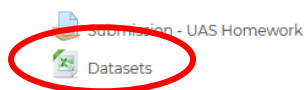
KOMPUTER DAN SISTEM INFORMASI

FEBRIYAN YOGA PRATAMA

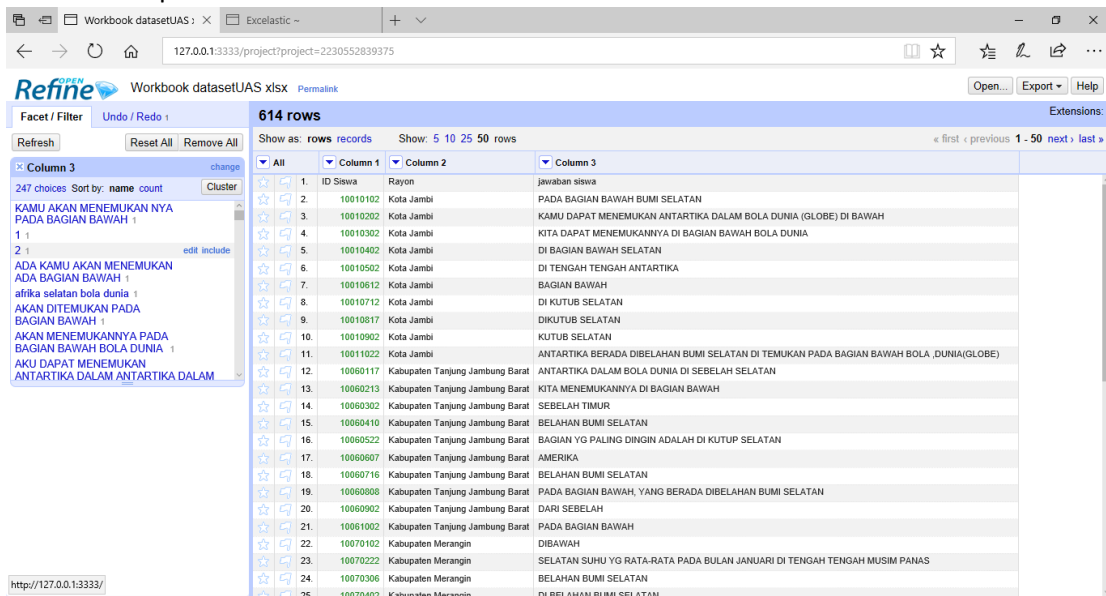
15/384468/SV/08825

1. Menyiapkan data yang akan diolah dengan cara download Dataset di <http://elok.ugm.ac.id>.

- Buat Dataset tersebut menjadi sandbox (kumpulan data yang akan di olah)
- Index konten dataset tersebut dan masukan dalam Elastic Search
- Visualisasikan data yang telah di- index dengan mengeluarkan beberapa parameter sebagai berikut:
 - Kelompok kalimat yang ada pada dataset tersebut dan frekuensi masing-masing kalimat
 - Urutkan kelompok pada poin i dari yang terbanyak ke yang terkecil
 - Identifikasi outlier yang dapat anda identifikasi
 - Buat analisis dan pendapat anda
- Soal kode B ini dapat dilaporkan dalam bentuk artikel di word, powerpoint dan bentuk lainnya.
- Submit hasil pengerjaan anda melalui menu assignment pada bagian Dataset ini sesuai batas waktu yang ditentukan.
- Batas waktu submission maksimal 4 hari dari tanggal ujian akhir dijalankan.

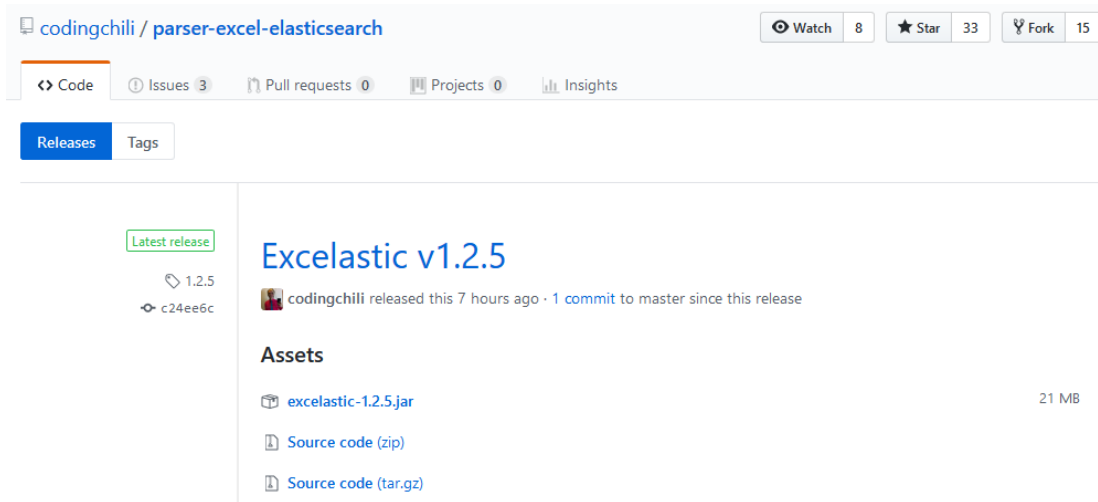


2. Setelah melihat data yang akan diolah ternyata data tersebut membutuhkan cleaning sebelum di import ke ElasticSearch agar tidak terjadi redudansi data dan memiliki format yang sama, oleh karena itu kita melakukan cleaning terlebih dahulu menggunakan tools open refine. Dengan menggunakan berbagai metode pada open refine maka dihasilkan data yang sudah clean dan siap diolah.



ID Siswa	Rayon	
10010102	Kota Jambi	PADA BAGIAN BAWAH BUMI SELATAN
10010202	Kota Jambi	KAMU DAPAT MENEMUKAN ANTARTIKA DALAM BOLA DUNIA (GLOBE) DI BAWAH
10010302	Kota Jambi	KITA DAPAT MENEMUKANNYA DI BAGIAN BAWAH BOLA DUNIA
10010402	Kota Jambi	DI BAGIAN BAWAH SELATAN
10010502	Kota Jambi	DI TENGAH TENGAH ANTARTIKA
10010612	Kota Jambi	BAGIAN BAWAH
10010712	Kota Jambi	DI KUTUB SELATAN
10010817	Kota Jambi	DI KUTUB SELATAN
10010902	Kota Jambi	KUTUB SELATAN
10011022	Kota Jambi	ANTARTIKA BERADA DIBELAHAN BUMI SELATAN DI TEMUKAN PADA BAGIAN BAWAH BOLA, DUNIA(GLOBE)
10060117	Kabupaten Tanjung Jambung Barat	ANTARTIKA DALAM BOLA DUNIA DI SEBELAH SELATAN
10060213	Kabupaten Tanjung Jambung Barat	KITA MENEMUKANNYA DI BAGIAN BAWAH
10060302	Kabupaten Tanjung Jambung Barat	SEBELAH TIMUR
10060410	Kabupaten Tanjung Jambung Barat	BELAHAN BUMI SELATAN
10060522	Kabupaten Tanjung Jambung Barat	BAGIAN YG PALING DINGIN ADALAH DI KUTUB SELATAN
10060607	Kabupaten Tanjung Jambung Barat	AMERIKA
10060716	Kabupaten Tanjung Jambung Barat	BELAHAN BUMI SELATAN
10060808	Kabupaten Tanjung Jambung Barat	PADA BAGIAN BAWAH, YANG BERADA DIBELAHAN BUMI SELATAN
10060902	Kabupaten Tanjung Jambung Barat	DARI SEBELAH
10061002	Kabupaten Tanjung Jambung Barat	PADA BAGIAN BAWAH
10070102	Kabupaten Merangin	DIBAWAH
10070222	Kabupaten Merangin	SELATAN SUHU YG RATA-RATA PADA BULAN JANUARI DI TENGAH TENGAH MUSIM PANAS
10070306	Kabupaten Merangin	BELAHAN BUMI SELATAN
10070407	Kabupaten Merangin	DI RFI ANJANI RUMI RFI STAN

3. Siapkan tools bernama parser-excel-elasticsearch yang dapat didownload di <https://github.com/codingchili/parser-excel-elasticsearch/releases> untuk import dataset berbentuk excel ke ElasticSearch.



The screenshot shows the GitHub interface for the repository `codingchili / parser-excel-elasticsearch`. The top navigation bar includes links for `Code`, `Issues` (3), `Pull requests` (0), `Projects` (0), and `Insights`. On the right, there are buttons for `Watch` (8), `Star` (33), and `Fork` (15). Below the navigation bar, the `Releases` tab is selected, showing the latest release, `Excelastic v1.2.5`, which was released 7 hours ago. The release details include a commit hash `c24ee6c` and a note that `codingchili` released this version. Under the `Assets` section, three files are listed: `excelastic-1.2.5.jar` (21 MB), `Source code (zip)`, and `Source code (tar.gz)`.

codingchili / parser-excel-elasticsearch

Watch 8 Star 33 Fork 15

Code Issues 3 Pull requests 0 Projects 0 Insights

Releases Tags

Latest release

1.2.5
c24ee6c

Excelastic v1.2.5

codingchili released this 7 hours ago · 1 commit to master since this release

Assets

excelastic-1.2.5.jar 21 MB

Source code (zip)

Source code (tar.gz)

4. Jalankan kibana dan Elasticsearch dengan mengklik batch file masing masing tools di folder bin masing masing tools.

Elasticsearch sudah berjalan dan siap digunakan.

```
Administrator: Elasticsearch 5.6.3
[2017-12-15T08:27:45,023][INFO ][o.e.n.Node               ] [] initializing ...
[2017-12-15T08:27:46,014][INFO ][o.e.e.NodeEnvironment ] [fLXMJiW] using [1] data paths, mounts [(D:)], net usable_
space [116.5gb], net total_space [182.1gb], spins? [unknown], types [NTFS]
[2017-12-15T08:27:46,015][INFO ][o.e.e.NodeEnvironment ] [fLXMJiW] heap size [1.9gb], compressed ordinary object poin
ters [true]
[2017-12-15T08:27:46,640][INFO ][o.e.n.Node               ] node name [fLXMJiW] derived from node ID [fLXMJiW4Q0GUh5QRVB
C6Vg]; set [node.name] to override
[2017-12-15T08:27:46,642][INFO ][o.e.n.Node               ] version[5.6.3], pid[13980], build[1a2f265/2017-10-06T20:33:3
9.012Z], OS[Windows 10/10.0/amd64], JVM[Oracle Corporation/Java HotSpot(TM) 64-Bit Server VM/1.8.0_131/25.131-b11]
[2017-12-15T08:27:46,655][INFO ][o.e.n.Node               ] JVM arguments [-Xms2g, -Xmx2g, -XX:+UseConcMarkSweepGC, -XX:
CMSInitiatingOccupancyFraction=75, -XX:+UseCMSInitiatingOccupancyOnly, -XX:+AlwaysPreTouch, -Xss1m, -Djava.awt.headless=
true, -Dfile.encoding=UTF-8, -Djna.nosys=true, -Djdk.io.permissionsUseCanonicalPath=true, -Dio.netty.noUnsafe=true, -Dio
.netty.noKeySetOptimization=true, -Dio.netty.recycler.maxCapacityPerThread=0, -Dlog4j.shutdownHookEnabled=false, -Dlog4j
2.disable.jmx=true, -Dlog4j.skipJansi=true, -XX:+HeapDumpOnOutOfMemoryError, -Delasticsearch, -Des.path.home=D:\KOMSI\Se
mester 5\teori STKI\elasticsearch-5.6.3]
[2017-12-15T08:27:55,466][INFO ][o.e.p.PluginsService     ] [fLXMJiW] loaded module [aggs-matrix-stats]
[2017-12-15T08:27:55,467][INFO ][o.e.p.PluginsService     ] [fLXMJiW] loaded module [ingest-common]
[2017-12-15T08:27:55,467][INFO ][o.e.p.PluginsService     ] [fLXMJiW] loaded module [lang-expression]
[2017-12-15T08:27:55,468][INFO ][o.e.p.PluginsService     ] [fLXMJiW] loaded module [lang-groovy]
[2017-12-15T08:27:55,468][INFO ][o.e.p.PluginsService     ] [fLXMJiW] loaded module [lang-mustache]
[2017-12-15T08:27:55,468][INFO ][o.e.p.PluginsService     ] [fLXMJiW] loaded module [lang-painless]
[2017-12-15T08:27:55,468][INFO ][o.e.p.PluginsService     ] [fLXMJiW] loaded module [parent-join]
[2017-12-15T08:27:55,469][INFO ][o.e.p.PluginsService     ] [fLXMJiW] loaded module [percolator]
[2017-12-15T08:27:55,469][INFO ][o.e.p.PluginsService     ] [fLXMJiW] loaded module [reindex]
[2017-12-15T08:27:55,469][INFO ][o.e.p.PluginsService     ] [fLXMJiW] loaded module [transport-netty3]
[2017-12-15T08:27:55,470][INFO ][o.e.p.PluginsService     ] [fLXMJiW] loaded module [transport-netty4]
[2017-12-15T08:27:55,471][INFO ][o.e.p.PluginsService     ] [fLXMJiW] no plugins loaded
[2017-12-15T08:28:08,846][INFO ][o.e.d.DiscoveryModule ] [fLXMJiW] using discovery type [zen]
[2017-12-15T08:28:10,731][INFO ][o.e.n.Node               ] initialized
[2017-12-15T08:28:10,732][INFO ][o.e.n.Node               ] [fLXMJiW] starting ...
```

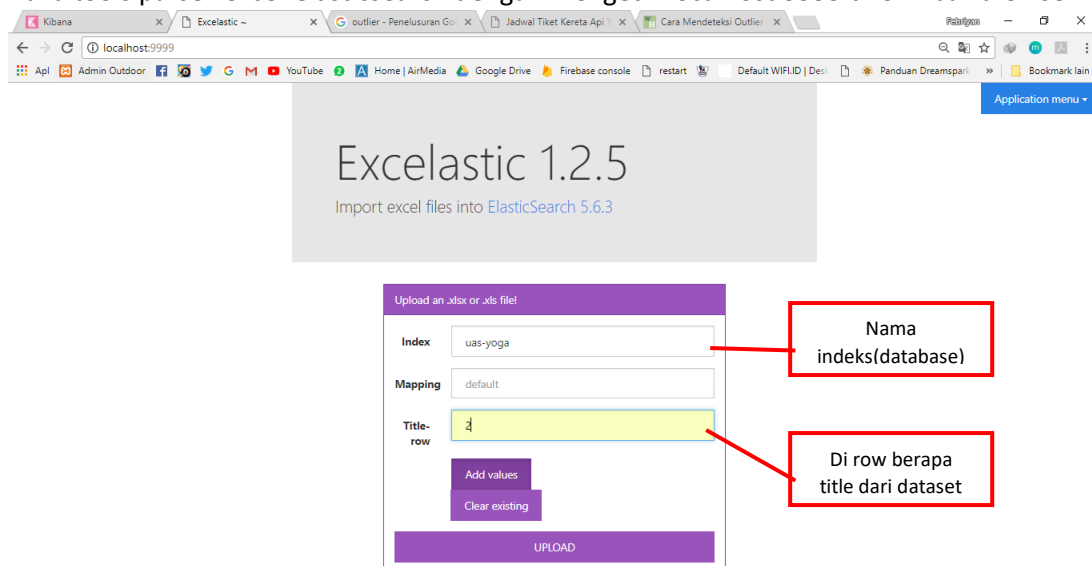
Kibana sudah berjalan dan siap digunakan.

```
Administrator: Kibana Server
log [01:29:06.407] [info][status][plugin:kibana@5.6.3] Status changed from uninitialized to green - Ready
log [01:29:06.809] [info][status][plugin:elasticsearch@5.6.3] Status changed from uninitialized to yellow - Waiting
for Elasticsearch
log [01:29:06.878] [info][status][plugin:console@5.6.3] Status changed from uninitialized to green - Ready
log [01:29:06.909] [info][status][plugin:metrics@5.6.3] Status changed from uninitialized to green - Ready
log [01:29:10.856] [error][status][plugin:elasticsearch@5.6.3] Status changed from yellow to red - Request Timeout a
fter 3000ms
log [01:29:10.860] [info][status][plugin:timelion@5.6.3] Status changed from uninitialized to green - Ready
log [01:29:10.875] [info][listening] Server running at http://localhost:5601
log [01:29:10.878] [error][status][ui settings] Status changed from uninitialized to red - Elasticsearch plugin is r
ed
log [01:29:13.865] [info][status][plugin:elasticsearch@5.6.3] Status changed from red to green - Kibana index ready
log [01:29:13.871] [info][status][ui settings] Status changed from red to green - Ready
```

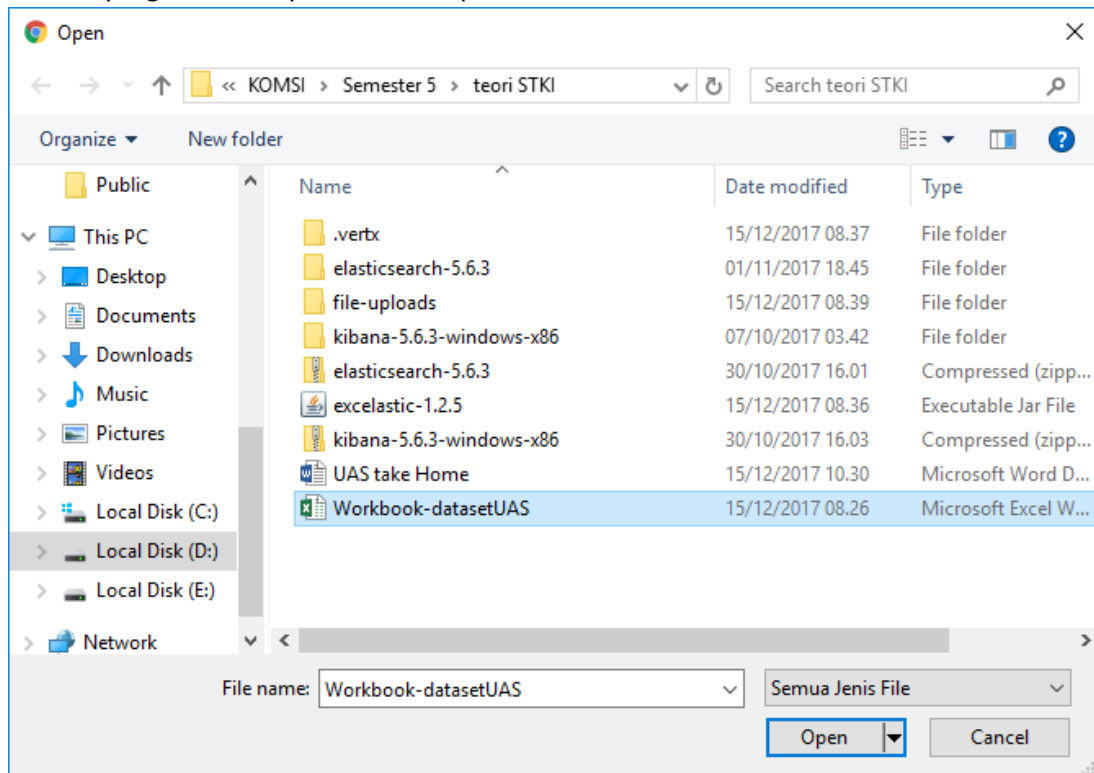
5. Jalankan tools parser-excel-elasticsearch dengan cara eksekusi file .jar tersebut dengan windows power shell.

```
Windows PowerShell
PS D:\KOMSI\Semester 5\teori STKI> java -jar excelastic-1.2.5.jar
Des 15, 2017 8:37:51 AM com.codingchilli.ApplicationLauncher <init>
INFO: Starting excelastic 1.2.5..
Des 15, 2017 8:37:51 AM com.codingchilli.ApplicationLauncher <init>
INFO: to import files without the web interface use: <filename> <index> <mapping>
Des 15, 2017 8:37:52 AM com.codingchilli.Model.ElasticWriter start
INFO: Started elastic writer
Des 15, 2017 8:37:54 AM com.codingchilli.Model.Configuration getConfiguration
INFO: configuration file configuration.json is not present, using defaults.
Des 15, 2017 8:37:54 AM io.vertx.core.impl.BlockingThreadChecker
WARNING: Thread Thread[vert.x-eventloop-thread-0,5,main] has been blocked for 2662 ms, time limit is 2000
Des 15, 2017 8:37:54 AM io.vertx.core.impl.BlockingThreadChecker
WARNING: Thread Thread[vert.x-eventloop-thread-1,5,main] has been blocked for 2645 ms, time limit is 2000
Des 15, 2017 8:37:54 AM com.codingchilli.Controller.Website lambda$start$7
INFO: Started website on port 9999
Des 15, 2017 8:37:54 AM com.codingchilli.ApplicationLauncher lambda$new$1
INFO: Successfully started application
Des 15, 2017 8:37:55 AM com.codingchilli.Model.ElasticWriter lambda$null$9
INFO: Connected to elasticsearch server 5.6.3 at localhost:9200
Des 15, 2017 8:37:55 AM com.codingchilli.ApplicationLauncher lambda$null$10
INFO: Attempting to open browser.. [ES connected=true]
Des 15, 2017 8:38:33 AM io.vertx.core.impl.ContextImpl
SEVERE: Unhandled exception
io.vertx.core.json.DecodeException: Failed to decode:null
    at io.vertx.core.json.Json.decodeValue(Json.java:172)
    at io.vertx.core.json.JsonObject.fromBuffer(JsonObject.java:958)
    at io.vertx.core.json.JsonObject.<init>(JsonObject.java:78)
    at io.vertx.core.buffer.impl.BufferImpl.toJsonObject(BufferImpl.java:82)
    at com.codingchilli.Controller.website.lambda$null$3(Website.java:93)
    at io.vertx.core.http.impl.WebSocketImplBase.handleFrame(WebSocketImplBase.java:262)
    at io.vertx.core.http.impl.ServerConnection.handleWsFrame(ServerConnection.java:336)
    at io.vertx.core.http.impl.ServerConnection.processMessage(ServerConnection.java:452)
    at io.vertx.core.http.impl.ServerConnection.handleMessage(ServerConnection.java:156)
    at io.vertx.core.http.impl.HttpServerImpl$ServerHandlerWithWebSockets.handleMessage(HttpServerImpl.java:671)
    at io.vertx.core.http.impl.HttpServerImpl$ServerHandlerWithWebSockets.handleMessage(HttpServerImpl.java:614)
    at io.vertx.core.net.impl.VertxHandler.lambda$channelRead$1(VertxHandler.java:150)
    at io.vertx.core.impl.ContextImpl.lambda$wrapTask$2(ContextImpl.java:342)
    at io.vertx.core.impl.ContextImpl.executeFromIO(ContextImpl.java:200)
    at io.vertx.core.net.impl.VertxHandler.channelRead(VertxHandler.java:148)
    at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:362)
    at io.netty.channel.AbstractChannelHandlerContext.invokeChannelRead(AbstractChannelHandlerContext.java:348)
    at io.netty.channel.AbstractChannelHandlerContext.fireChannelRead(AbstractChannelHandlerContext.java:340)
    at io.netty.handler.codec.ByteToMessageDecoder.fireChannelRead(ByteToMessageDecoder.java:310)
```

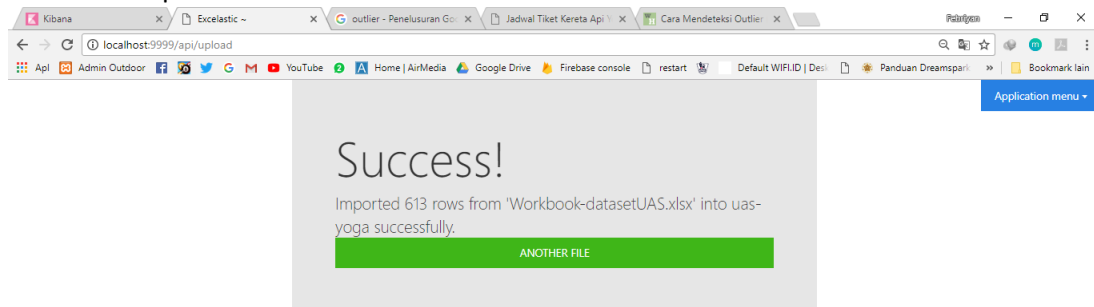
6. Buka tools parser-excel-elasticsearch dengan mengetik localhost:9999 di URL bar browser



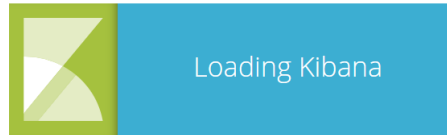
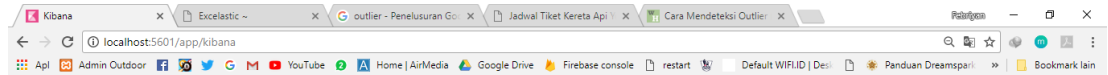
Kemudian isi index dan Title-row nya sesuai yang dibutuhkan. Kemudian klik upload, pilih dataset yang akan di import dan klik open.



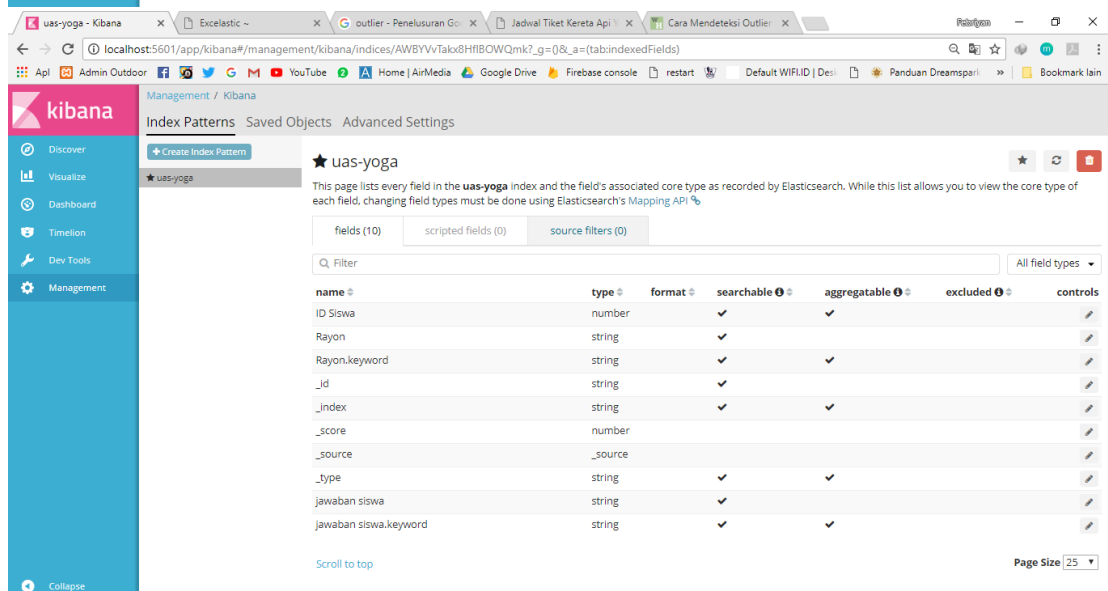
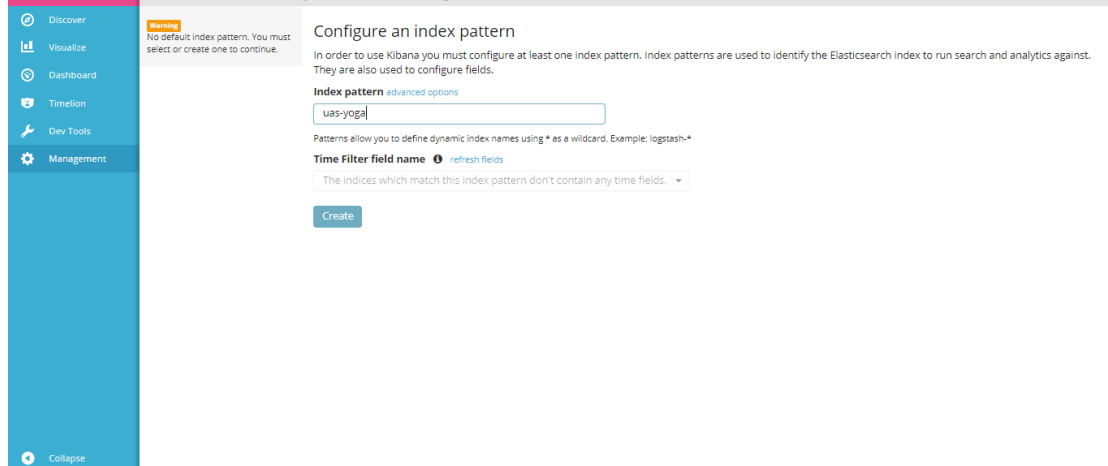
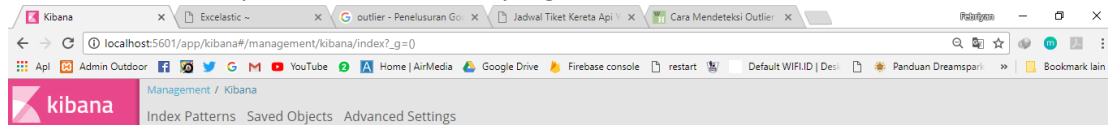
Setelah proses upload terdapat tampilan seperti berikut yang menandakan bahwa dataset berhasil diimport ke Elasticsearch.



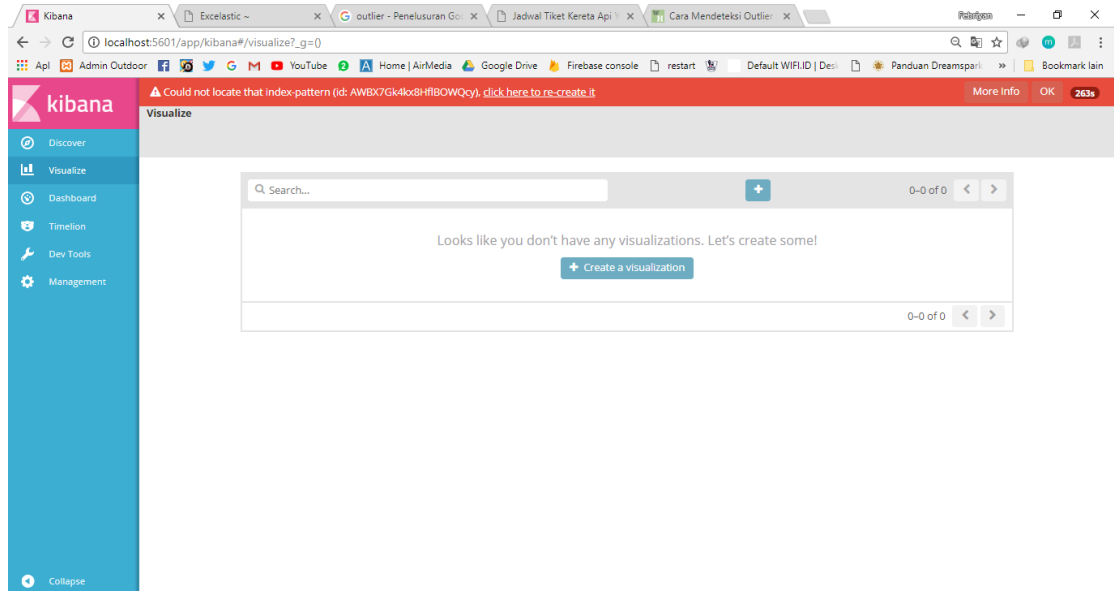
7. Buka kibana dengan mengetik localhost:5601 di URL bar browser.



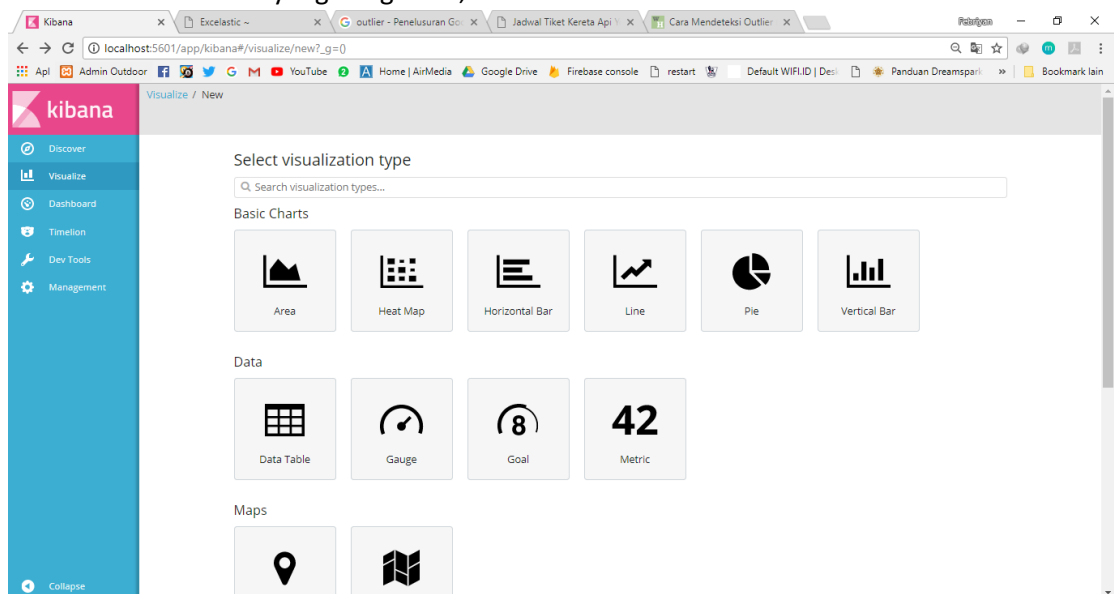
Kemudian isi index-pattern sesuai indeks yang dibuat tadi dan klik create.



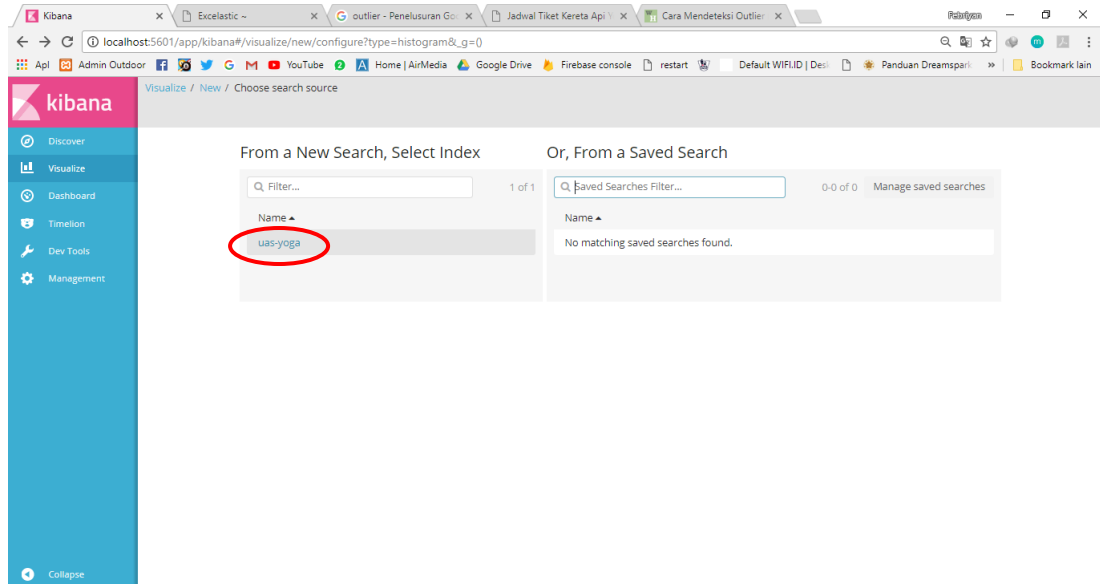
8. Membuat visualisasi dengan klik visualization kemudian klik create a visualization



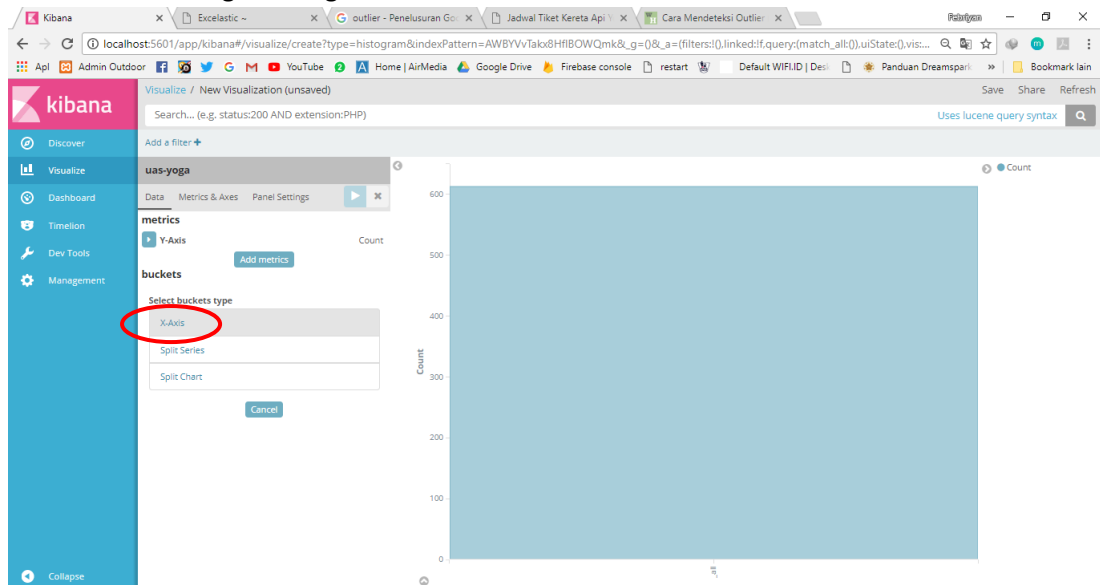
Pilih model visualisasi yang diinginkan, misal vertical Bar.



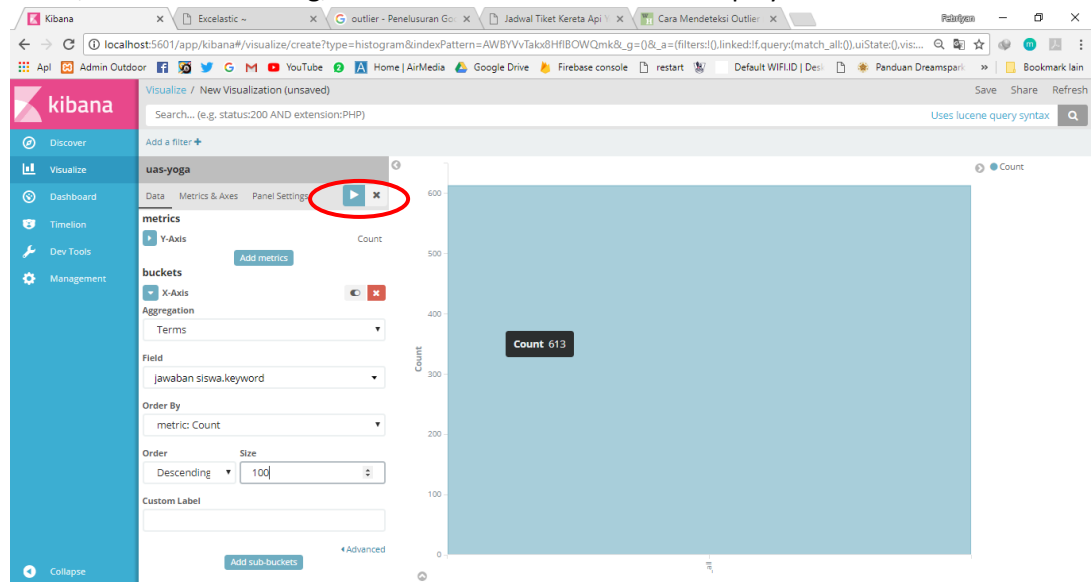
Kemudian pilih indeks-patern yang telah dibuat.



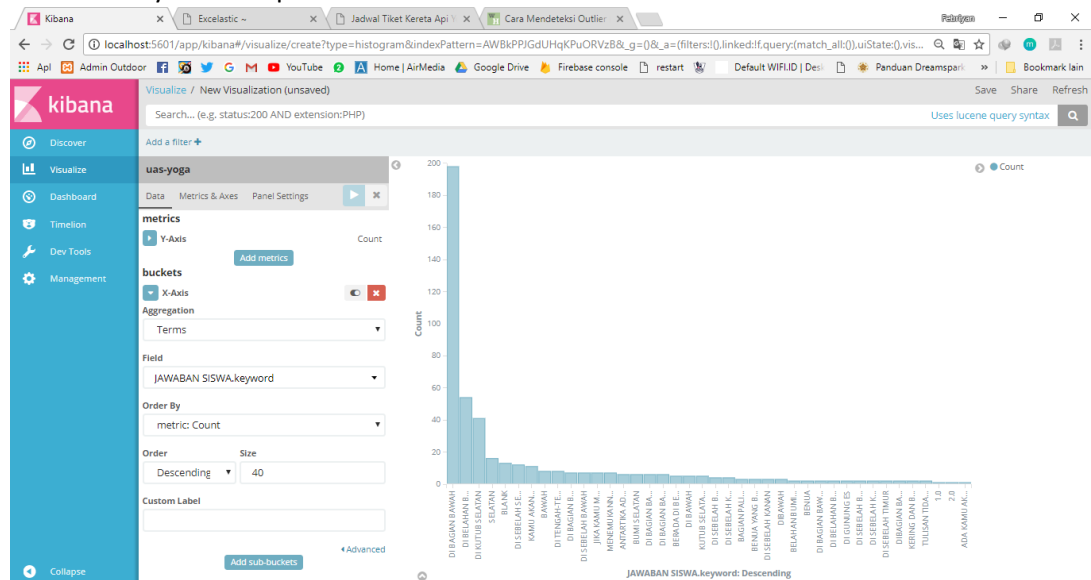
Atur sumbu X dengan mengklik X-Axis .



Kemudian atur misal agregations : terms, fields : jawaban siswa.keyword, order by : metric: count, Order : Descending, Size misal 40. Kemudian klik icon play disebelah kanan atas.



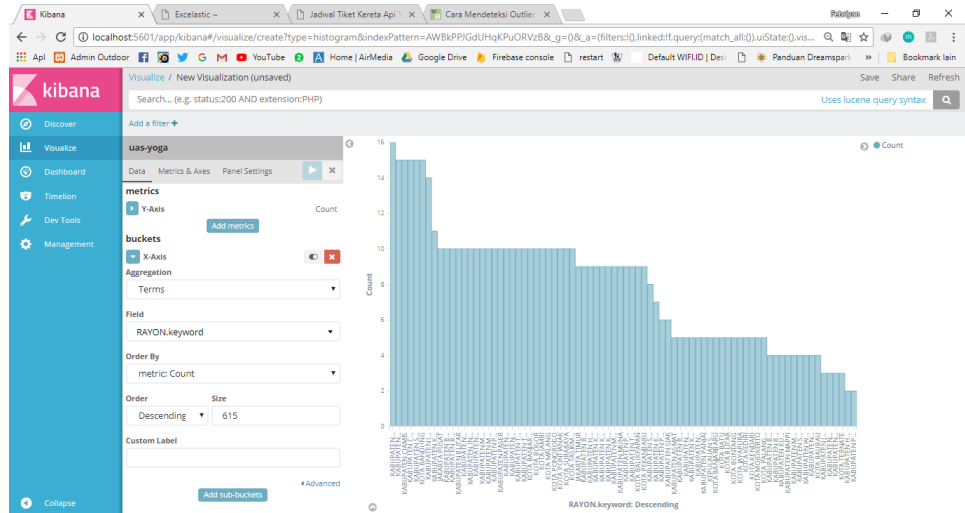
Maka hasilnya akan seperti berikut.



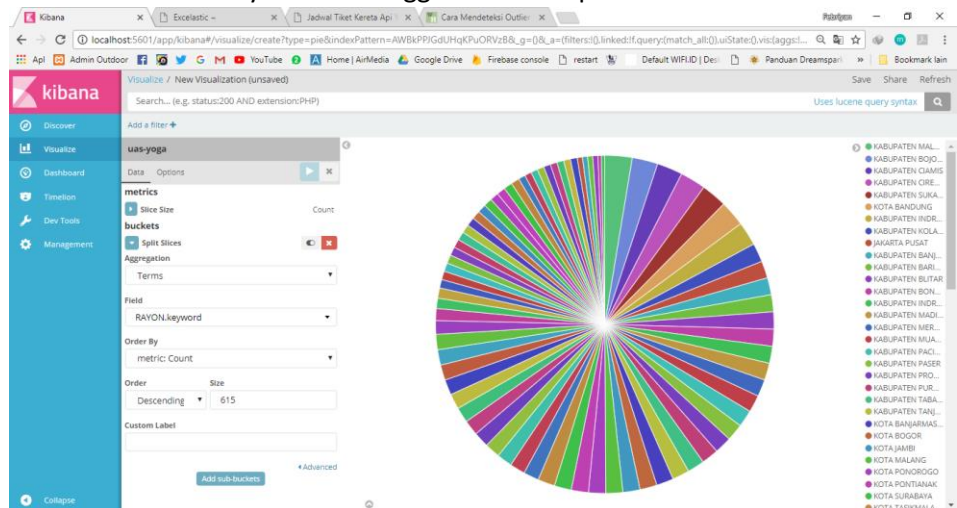
Jawaban Soal UAS Homework

- i. Visualisasi kelompok kalimat yang ada pada dataset dan frekuensi masing-masing serta diurutkan dari terbanyak ke terkecil

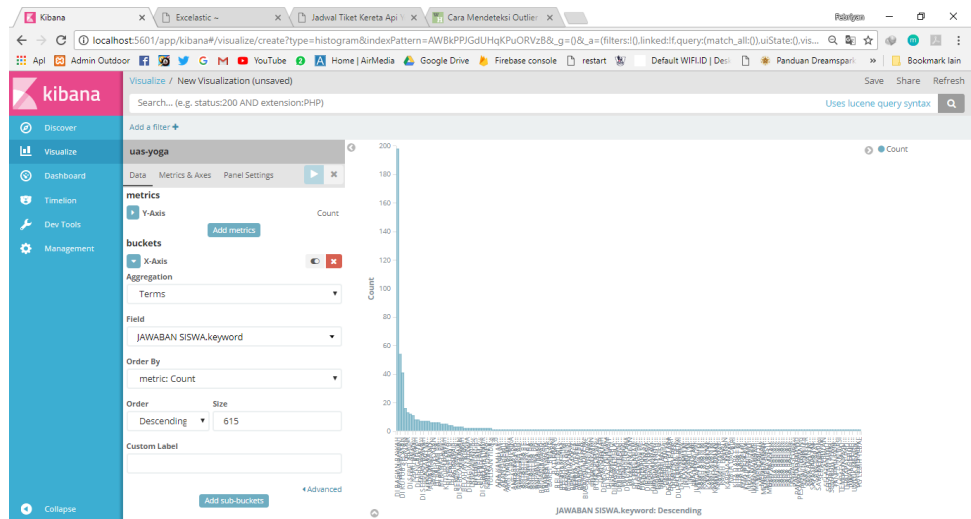
- Berdasarkan field rayon dan menggunakan cart bar



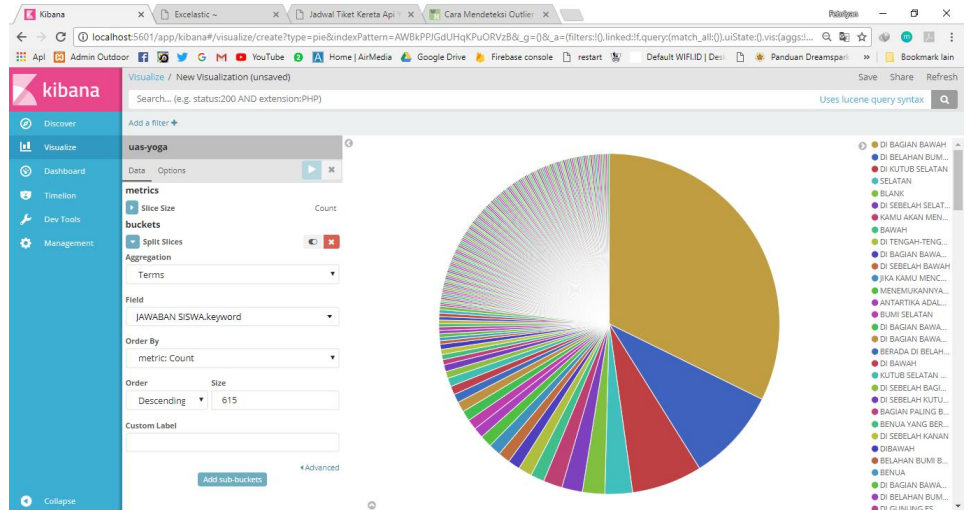
- Berdasarkan field rayon dan menggunakan cart pie



- Berdasarkan field rayon dan menggunakan chart bar



- Berdasarkan field rayon dan menggunakan chart pie

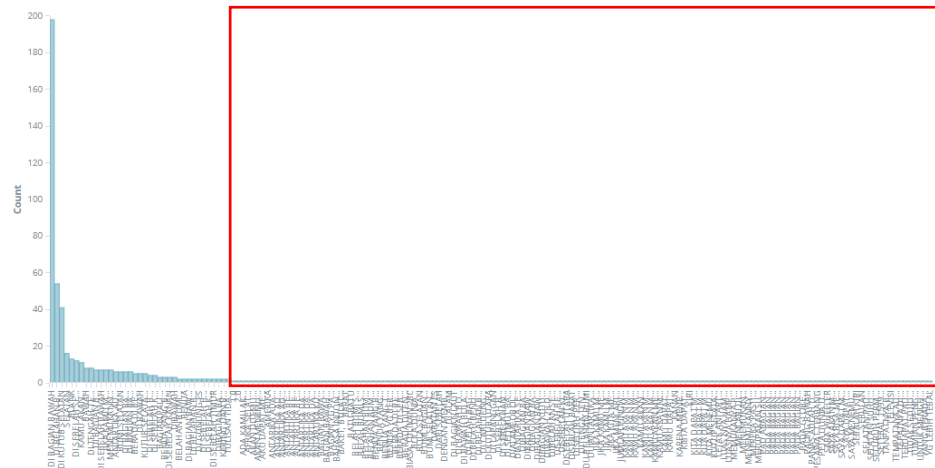


ii. Menjadi satu dengan visualisasi nomor i

iii. Outlier

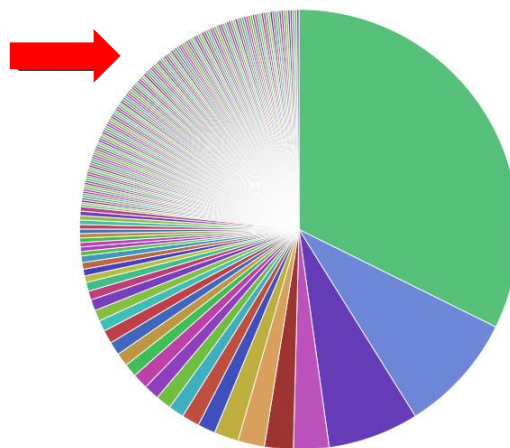
Outliers merupakan data yang memiliki pola yang unik. Dalam kasus ini ditemukan pola unik di bagian Jawaban Siswa. Seperti yang terlihat dalam chart bar dan chart pie yang telah dibuat.

1. Pada chart bar



Menurut analisa saya data pada kotak merah tersebut merupakan outlier karena data tersebut berbeda dari mayoritas data yang ada.

2. Pada chart pie



Pada chart pie juga data outlier ditunjukkan oleh warna putih tersebut (perhatikan panah merah).

iv. Analisis dan pendapat

Dari data yang telah dioalah tersebut dapat diambil sebuah informasi bahwa sangat beragam sekali jenis jawaban yang siswa berikan, baik itu dari konteks kalimatnya maupun diksi yang digunakan para siswa. Dari data tersebut juga dapat di informasikan jenis keberagaman jawaban siswa berdasarkan rayon.

Outlier pada data tersebut berisi jawaban siswa yang jumlahnya hanya 1 setiap jawaban karena tidak adanya kesamaan jawaban dengan yang lain, jawaban terebut menurut saya dapat disebut sebagai outlier data tersebut, mengingat lebih banyak siswa yang memiliki jawaban lain yang banyak memiliki kemiripan.