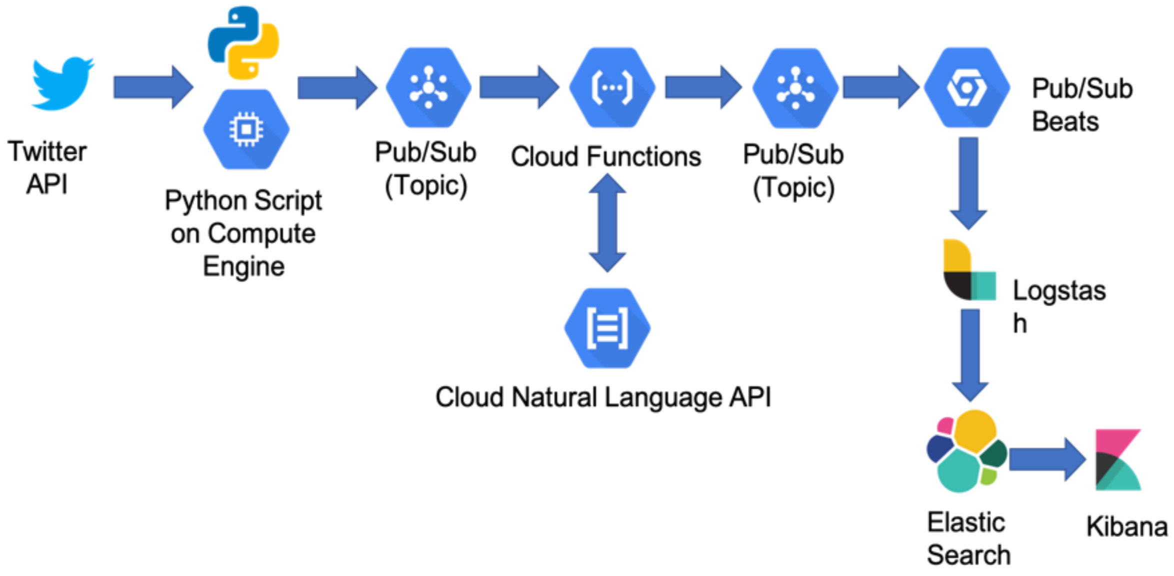
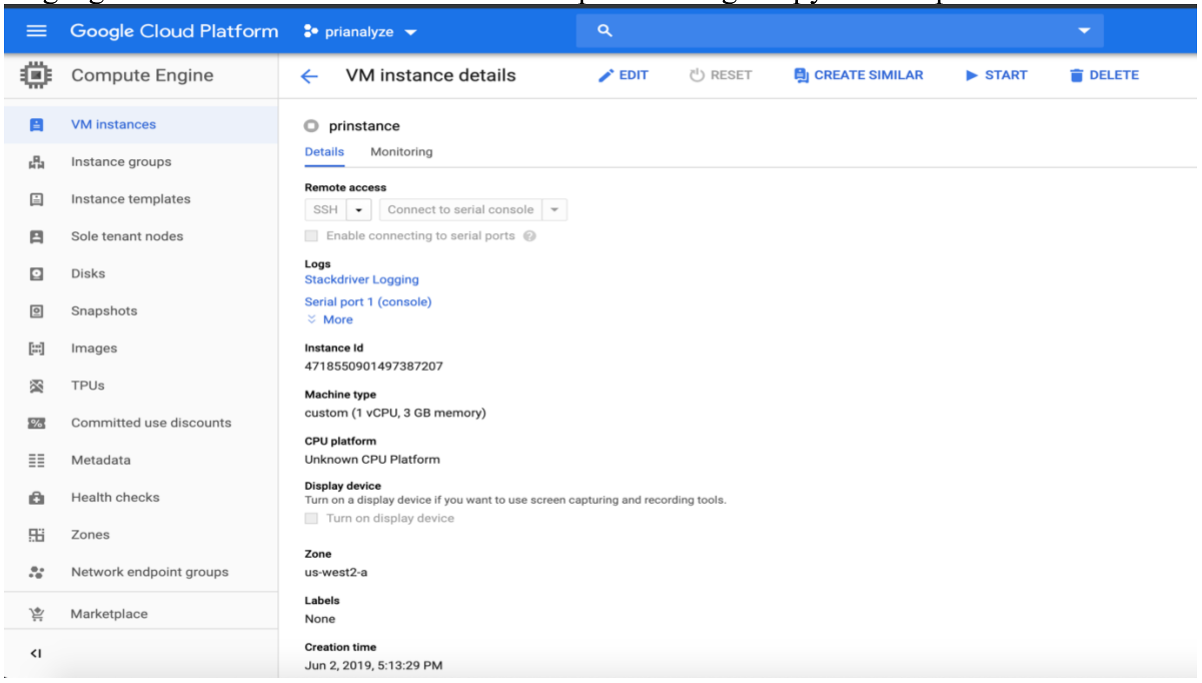
Google Cloud Setup for Sentiment Analysis



Implementation and Steps

Data Collection and the Instance: Using Twitter Developer Portal, create Access token & access token secret. The Instance is setup with custom configuration to python script. Python3 is installed on the instance using pip command. Python Script containing the Twitter API Tokens runs on the Google Cloud Instance (Compute Engine). The real-time tweets are streamed and English language tweets with a filter condition are captured using the python script.



prinstance

Remote access

Logs

Instance Id
4718550901497387207

Machine type
custom (1 vCPU, 3 GB memory)

CPU platform
Unknown CPU Platform

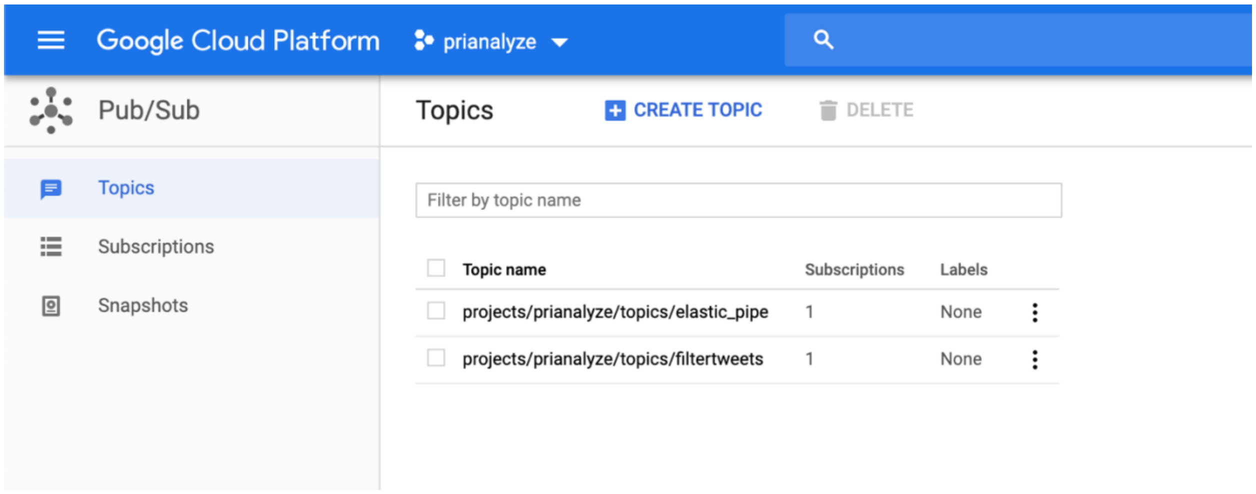
Display device
Turn on a display device if you want to use screen capturing and recording tools.

Zone
us-west2-a

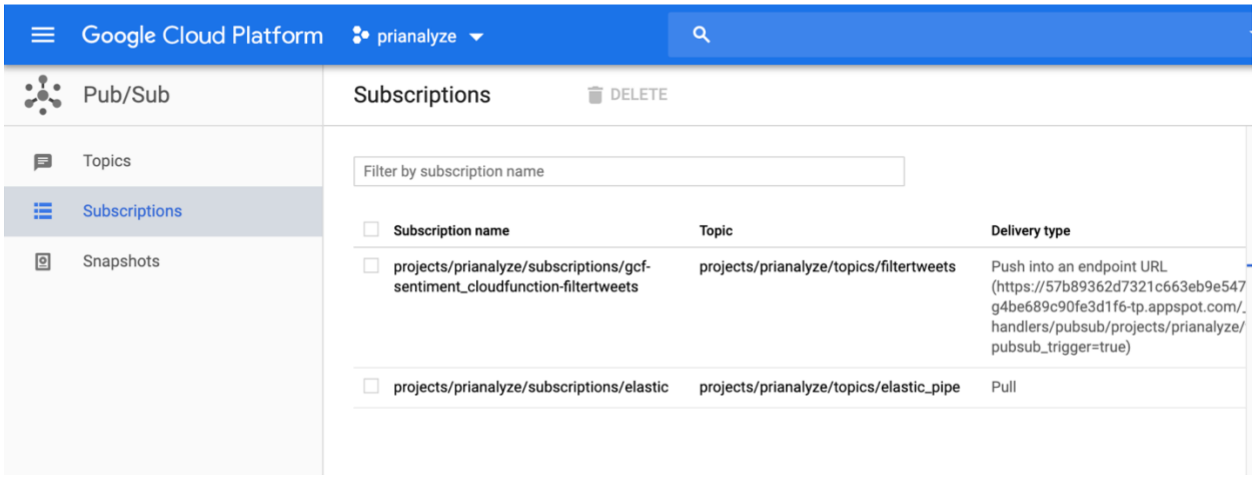
Labels
None

Creation time
Jun 2, 2019, 5:13:29 PM

Data Ingestion: Cloud Pub/Sub provides staging location for data based on event (tweets streaming) on its journey towards storage as logs. A publisher application creates and publishes messages to a topic. Subscriber applications create a subscription to a topic to receive messages from it. Communication can be one-to-many (fan-out), many-to-one (fan-in), and many-to-many. A topic "filtertweets" is created in Pub/Sub. The tweets are published to this topic based on the event of their arrival. A subscription "sentiment_cloudfunction-filtertweets" is created where the cloud function receives this data.

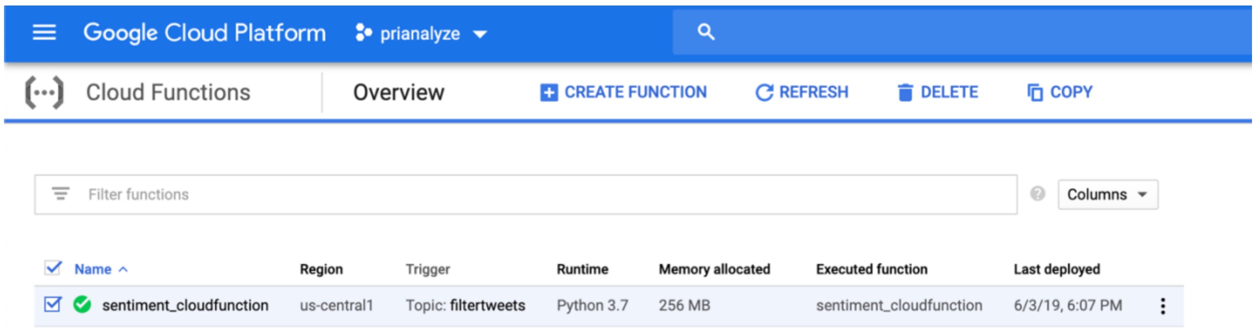


Topic name	Subscriptions	Labels
projects/prianalyze/topics/elastic_pipe	1	None
projects/prianalyze/topics/filtertweets	1	None

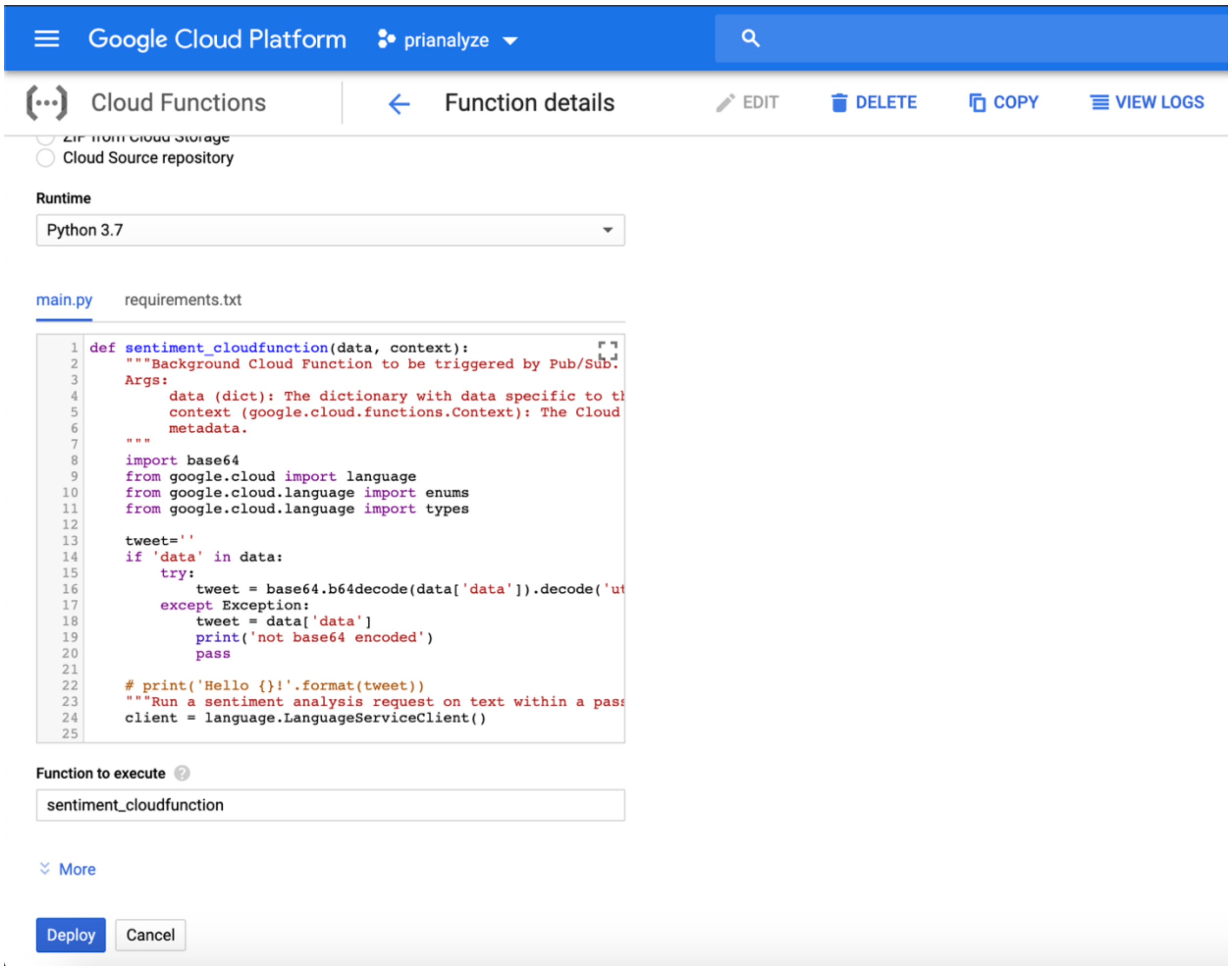


Subscription name	Topic	Delivery type
projects/prianalyze/subscriptions/gcf-sentiment_cloudfunction-filtertweets	projects/prianalyze/topics/filtertweets	Push into an endpoint URL (https://57b89362d7321c663eb9e547 g4be689c90fe3d1f6-tp.appspot.com/_ handlers/pubsub/projects/prianalyze/ pubsub_trigger=true)
projects/prianalyze/subscriptions/elastic	projects/prianalyze/topics/elastic_pipe	Pull

Sentiment Analysis: Google Cloud Function is an event- driven serverless execution environment where the user's function is attached to the event of tweets streamed to Pub/Sub. Cloud function invokes Cloud Natural Language API where staged data is converted to a structured json format with fields "tweet", "score", "magnitude". The score can range from -1 to +1, but the user has the flexibility to define a custom range. Google defines 0.25 to 1.0 as Positive, -0.25 - +0.25 as neutral, and -1 - -0.25 as negative, however the user can achieve more granularity by classifying the tweets as Highly Positive, Positive, Neutral, Negative and Highly Negative.



Name	Region	Trigger	Runtime	Memory allocated	Executed function	Last deployed
sentiment_cloudfunction	us-central1	Topic: filtertweets	Python 3.7	256 MB	sentiment_cloudfunction	6/3/19, 6:07 PM



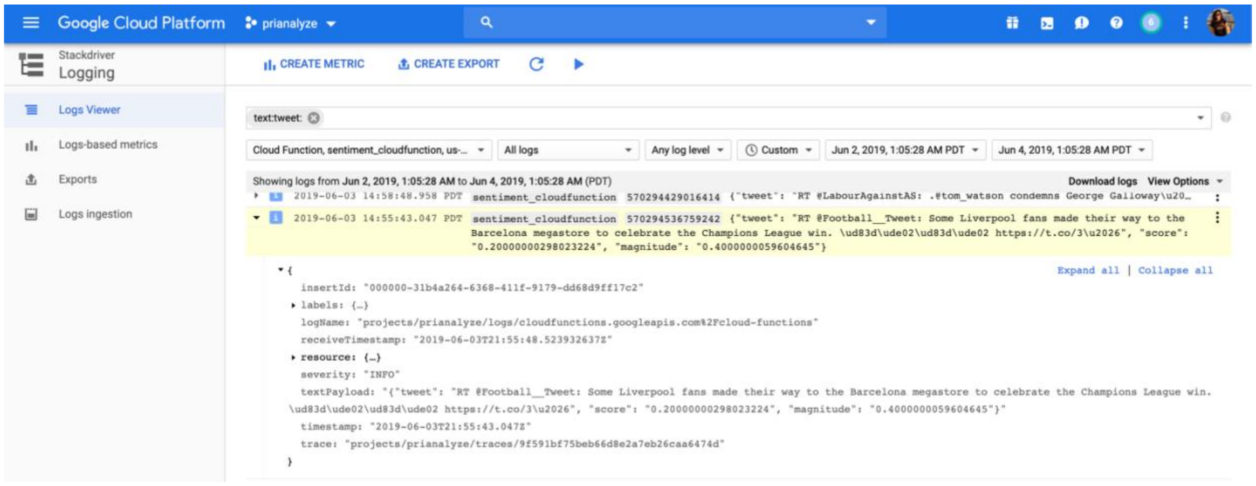
Runtime: Python 3.7

```
def sentiment_cloudfunction(data, context):
    """Background Cloud Function to be triggered by Pub/Sub.
    Args:
         data (dict): The dictionary with data specific to tl
         context (google.cloud.functions.Context): The Cloud
         metadata.
    """
    import base64
    from google.cloud import language
    from google.cloud.language import enums
    from google.cloud.language import types

    tweet=''
    if 'data' in data:
        try:
            tweet = base64.b64decode(data['data']).decode('ut
        except Exception:
            tweet = data['data']
            print('not base64 encoded')
            pass

    # print('Hello {}!'.format(tweet))
    """Run a sentiment analysis request on text within a pass
    client = language.LanguageServiceClient()
```

Function to execute: sentiment_cloudfunction

Storage and Visualization: Logstash is server- side data processing pipeline subscribed to Pub/Sub. Logstash Ingests logging data from Pub/Sub and transforms it and sends the data to Elasticsearch. Kibana could be used to visualize the data present in Elasticsearch. However, the nested json response from Google cloud platform's logging service was incompatible with

Kibana's indexing mechanism. Google Cloud Platform's Stackdriver Monitoring provides visibility into the performance, uptime, and overall health of cloud-powered applications and could not be used for visualization.

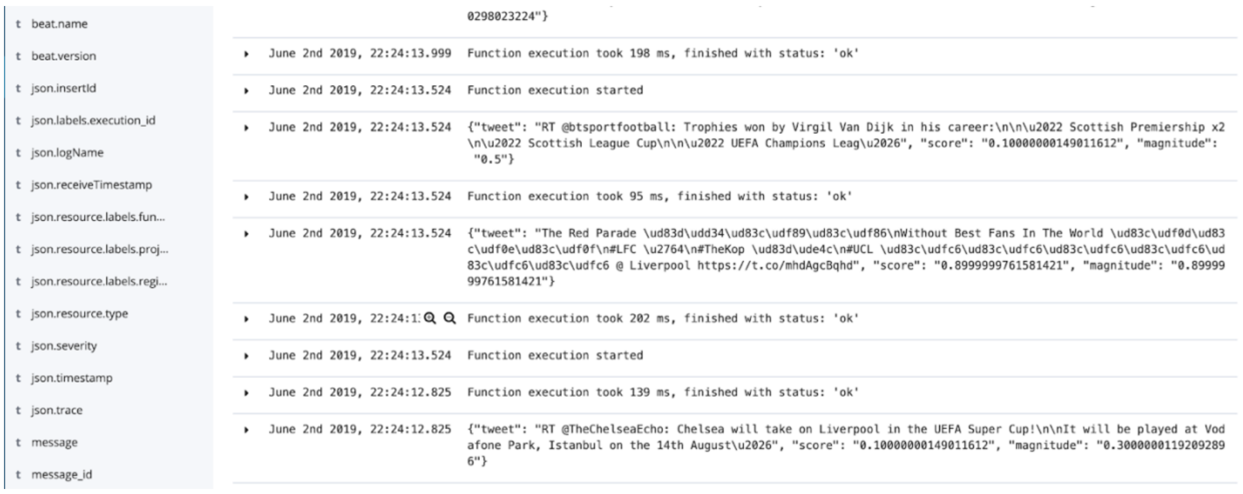


text:tweet:

Cloud Function, sentiment_cloudfunction, us-... All logs Any log level Custom Jun 2, 2019, 1:05:28 AM PDT Jun 4, 2019, 1:05:28 AM PDT

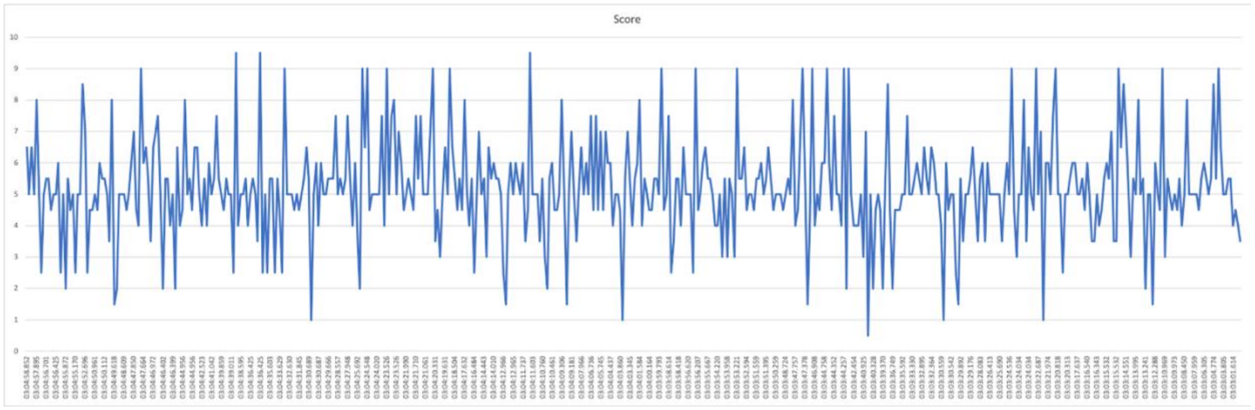
Showing logs from Jun 2, 2019, 1:05:28 AM to Jun 4, 2019, 1:05:28 AM (PDT)

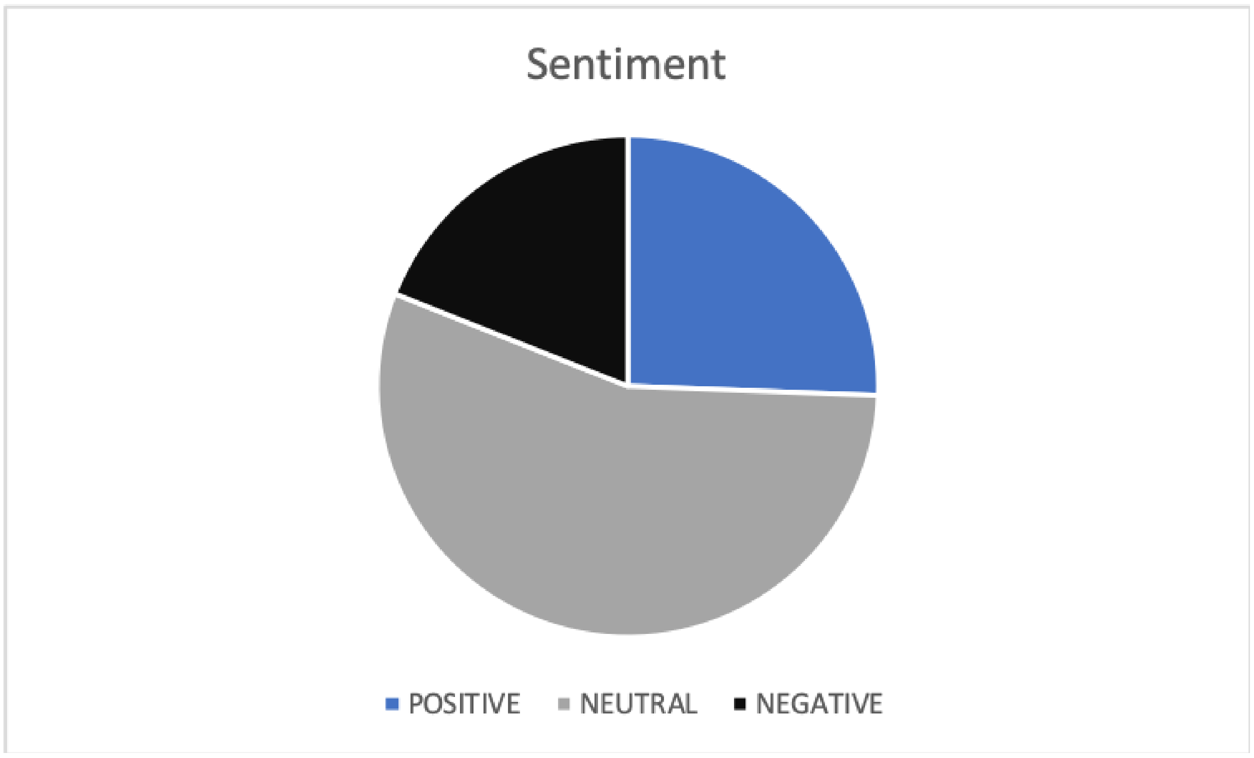
```
2019-06-03 14:58:48.958 PDT sentiment_cloudfunction 570294429016414 {"tweet": "RT @LabourAgainstAS: .@tom_watson condemns George Galloway\u20…
2019-06-03 14:55:43.047 PDT sentiment_cloudfunction 570294536759242 {"tweet": "RT @Football__Tweet: Some Liverpool fans made their way to the
                                                    Barcelona megastore to celebrate the Champions League win. \ud83d\ude02\ud83d\ude02 https://t.co/3\u2026", "score":
                                                    "0.20000000298023224", "magnitude": "0.4000000059604645"}
  {
    insertId: "000000-31b4a264-6368-411f-9179-dd68d9ff17c2"
   labels: {…}
    logName: "projects/prianalyze/logs/cloudfunctions.googleapis.com%2Fcloud-functions"
    receiveTimestamp: "2019-06-03T21:55:48.523932637Z"
   resource: {…}
    severity: "INFO"
    textPayload: "{"tweet": "RT @Football__Tweet: Some Liverpool fans made their way to the Barcelona megastore to celebrate the Champions League win.
  \ud83d\ude02\ud83d\ude02 https://t.co/3\u2026", "score": "0.20000000298023224", "magnitude": "0.4000000059604645"}"
    timestamp: "2019-06-03T21:55:43.047Z"
    trace: "projects/prianalyze/traces/9f591bf75beb66d8e2a7eb26caa6474d"
  }
```



```
t beat.name
t beat.version
t json.insertId
t json.labels.execution_id
t json.logName
t json.receiveTimestamp
t json.resource.labels.fun...
t json.resource.labels.proj...
t json.resource.labels.regi...
t json.resource.type
t json.severity
t json.timestamp
t json.trace
t message
t message_id
```

```
                            0298023224"}
June 2nd 2019, 22:24:13.999  Function execution took 198 ms, finished with status: 'ok'
June 2nd 2019, 22:24:13.524  Function execution started
June 2nd 2019, 22:24:13.524  {"tweet": "RT @btsportfootball: Trophies won by Virgil Van Dijk in his career:\n\n\u2022 Scottish Premiership x2
                             \n\u2022 Scottish League Cup\n\n\u2022 UEFA Champions Leag\u2026", "score": "0.10000000149011612", "magnitude":
                              "0.5"}
June 2nd 2019, 22:24:13.524  Function execution took 95 ms, finished with status: 'ok'
June 2nd 2019, 22:24:13.524  {"tweet": "The Red Parade \ud83d\udd34\ud83c\udf89\ud83c\udf86\nWithout Best Fans In The World \ud83c\udf0d\ud83
                             c\udf0e\ud83c\udf0f\n#LFC \u2764\n#TheKop \ud83d\ude4c\n#UCL \ud83c\udfc6\ud83c\udfc6\ud83c\udfc6\ud83c\udfc6\ud
                             83c\udfc6\ud83c\udfc6 @ Liverpool https://t.co/mhdAgcBqhd", "score": "0.8999999761581421", "magnitude": "0.89999
                             99761581421"}
June 2nd 2019, 22:24:1       Function execution took 202 ms, finished with status: 'ok'
June 2nd 2019, 22:24:13.524  Function execution started
June 2nd 2019, 22:24:12.825  Function execution took 139 ms, finished with status: 'ok'
June 2nd 2019, 22:24:12.825  {"tweet": "RT @TheChelseaEcho: Chelsea will take on Liverpool in the UEFA Super Cup!\n\nIt will be played at Vod
                             afone Park, Istanbul on the 14th August\u2026", "score": "0.10000000149011612", "magnitude": "0.3000000119209289
                             6"}
```




Sentiment
POSITIVE
NEUTRAL
NEGATIVE