MinhVo-FinalProject-Filtering

March 11, 2023

0.1 Big Data Platforms - Winter 2023

0.2 Final Project Education

0.2.1 Twitter Analysis (Data Filtering)

Minh Vo

[1]:
```
import os
import time
import re
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from itertools import islice
# import sh
from pyspark.sql.functions import *
from pyspark.sql.types import *
from itertools import compress

pd.set_option('display.max_colwidth', None)
pd.reset_option('display.max_rows')
warnings.filterwarnings(action='ignore')
```

[2]:
```
from google.cloud import storage
```

[3]:
```
path = "gs://msca-bdp-tweets/final_project"
```

[4]:
```
!hadoop fs -ls "gs://msca-bdp-tweets/final_project" | tail -10
```

```
-rwx------   3 root root    9410904 2023-02-08 13:57 gs://msca-bdp-
tweets/final_project/part-50685-aa6d3cb4-7022-4df2-9921-218307589ce2-c000.json
-rwx------   3 root root   13046317 2023-02-08 13:57 gs://msca-bdp-
tweets/final_project/part-50686-aa6d3cb4-7022-4df2-9921-218307589ce2-c000.json
-rwx------   3 root root   10826130 2023-02-08 13:57 gs://msca-bdp-
tweets/final_project/part-50687-aa6d3cb4-7022-4df2-9921-218307589ce2-c000.json
-rwx------   3 root root    9099590 2023-02-08 13:57 gs://msca-bdp-
tweets/final_project/part-50688-aa6d3cb4-7022-4df2-9921-218307589ce2-c000.json
-rwx------   3 root root    9860829 2023-02-08 13:57 gs://msca-bdp-
```

```
tweets/final_project/part-50689-aa6d3cb4-7022-4df2-9921-218307589ce2-c000.json
-rwx------   3 root root   11562361 2023-02-08 13:57 gs://msca-bdp-
tweets/final_project/part-50690-aa6d3cb4-7022-4df2-9921-218307589ce2-c000.json
-rwx------   3 root root    9132693 2023-02-08 13:57 gs://msca-bdp-
tweets/final_project/part-50691-aa6d3cb4-7022-4df2-9921-218307589ce2-c000.json
-rwx------   3 root root   15376390 2023-02-08 13:57 gs://msca-bdp-
tweets/final_project/part-50692-aa6d3cb4-7022-4df2-9921-218307589ce2-c000.json
-rwx------   3 root root    8586044 2023-02-08 13:57 gs://msca-bdp-
tweets/final_project/part-50693-aa6d3cb4-7022-4df2-9921-218307589ce2-c000.json
-rwx------   3 root root   14317778 2023-02-08 13:57 gs://msca-bdp-
tweets/final_project/part-50694-aa6d3cb4-7022-4df2-9921-218307589ce2-c000.json
```

0.2.2 Data Loading

[5]:
```
spark.conf.set("spark.sql.repl.eagerEval.enabled",True)
```

Load the collection of Twitter data, which consists of around 100 million Tweets (~500GB)

[ ]:
```
%%time
twitter_df = spark.read.json(path)
twitter_df.count()
```

```
23/03/09 06:43:10 WARN org.apache.spark.scheduler.cluster.YarnScheduler: Initial
job has not accepted any resources; check your cluster UI to ensure that workers
are registered and have sufficient resources
23/03/09 06:43:25 WARN org.apache.spark.scheduler.cluster.YarnScheduler: Initial
job has not accepted any resources; check your cluster UI to ensure that workers
are registered and have sufficient resources
23/03/09 06:44:35 WARN
org.apache.spark.sql.execution.datasources.SharedInMemoryCache: Evicting cached
table partition metadata from memory due to size constraints
(spark.sql.hive.filesourcePartitionFileCacheSize = 262144000 bytes). This may
impact query planning performance.
23/03/09 06:50:56 WARN org.apache.spark.sql.catalyst.util.package: Truncated the
string representation of a plan since it was too large. This behavior can be
adjusted by setting 'spark.sql.debug.maxToStringFields'.
[Stage 4:>                                                          (0 + 1) / 1]

CPU times: user 2.84 s, sys: 525 ms, total: 3.36 s
Wall time: 12min 6s
```

[ ]:
```
99994342
```

[ ]:
```
#twitter_df.describe()
```

[ ]:
```
twitter_df.printSchema()
```

```
root
 |-- coordinates: struct (nullable = true)
 |    |-- coordinates: array (nullable = true)
 |    |    |-- element: double (containsNull = true)
 |    |-- type: string (nullable = true)
 |-- created_at: string (nullable = true)
 |-- display_text_range: array (nullable = true)
 |    |-- element: long (containsNull = true)
 |-- entities: struct (nullable = true)
 |    |-- hashtags: array (nullable = true)
 |    |    |-- element: struct (containsNull = true)
 |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |-- text: string (nullable = true)
 |    |-- media: array (nullable = true)
 |    |    |-- element: struct (containsNull = true)
 |    |    |    |-- additional_media_info: struct (nullable = true)
 |    |    |    |    |-- description: string (nullable = true)
 |    |    |    |    |-- embeddable: boolean (nullable = true)
 |    |    |    |    |-- monetizable: boolean (nullable = true)
 |    |    |    |    |-- title: string (nullable = true)
 |    |    |    |-- description: string (nullable = true)
 |    |    |    |-- display_url: string (nullable = true)
 |    |    |    |-- expanded_url: string (nullable = true)
 |    |    |    |-- id: long (nullable = true)
 |    |    |    |-- id_str: string (nullable = true)
 |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |-- media_url: string (nullable = true)
 |    |    |    |-- media_url_https: string (nullable = true)
 |    |    |    |-- sizes: struct (nullable = true)
 |    |    |    |    |-- large: struct (nullable = true)
 |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |-- medium: struct (nullable = true)
 |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |-- small: struct (nullable = true)
 |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |-- thumb: struct (nullable = true)
 |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |-- source_status_id: long (nullable = true)
```

```
|    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |-- type: string (nullable = true)
|    |    |    |-- url: string (nullable = true)
|    |-- symbols: array (nullable = true)
|    |    |-- element: struct (containsNull = true)
|    |    |    |-- indices: array (nullable = true)
|    |    |    |    |-- element: long (containsNull = true)
|    |    |    |-- text: string (nullable = true)
|    |-- urls: array (nullable = true)
|    |    |-- element: struct (containsNull = true)
|    |    |    |-- display_url: string (nullable = true)
|    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |-- indices: array (nullable = true)
|    |    |    |    |-- element: long (containsNull = true)
|    |    |    |-- url: string (nullable = true)
|    |-- user_mentions: array (nullable = true)
|    |    |-- element: struct (containsNull = true)
|    |    |    |-- id: long (nullable = true)
|    |    |    |-- id_str: string (nullable = true)
|    |    |    |-- indices: array (nullable = true)
|    |    |    |    |-- element: long (containsNull = true)
|    |    |    |-- name: string (nullable = true)
|    |    |    |-- screen_name: string (nullable = true)
|-- extended_entities: struct (nullable = true)
|    |-- media: array (nullable = true)
|    |    |-- element: struct (containsNull = true)
|    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |-- title: string (nullable = true)
|    |    |    |-- description: string (nullable = true)
|    |    |    |-- display_url: string (nullable = true)
|    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |-- id: long (nullable = true)
|    |    |    |-- id_str: string (nullable = true)
|    |    |    |-- indices: array (nullable = true)
|    |    |    |    |-- element: long (containsNull = true)
|    |    |    |-- media_url: string (nullable = true)
|    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- medium: struct (nullable = true)
```

```
|    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |-- type: string (nullable = true)
|    |    |    |-- url: string (nullable = true)
|    |    |    |-- video_info: struct (nullable = true)
|    |    |    |    |-- aspect_ratio: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- duration_millis: long (nullable = true)
|    |    |    |    |-- variants: array (nullable = true)
|    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |-- bitrate: long (nullable = true)
|    |    |    |    |    |    |-- content_type: string (nullable = true)
|    |    |    |    |    |    |-- url: string (nullable = true)
|-- extended_tweet: struct (nullable = true)
|    |-- display_text_range: array (nullable = true)
|    |    |-- element: long (containsNull = true)
|    |-- entities: struct (nullable = true)
|    |    |-- hashtags: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- text: string (nullable = true)
|    |    |-- media: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
```

```
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |-- video_info: struct (nullable = true)
|    |    |    |    |    |-- aspect_ratio: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- duration_millis: long (nullable = true)
|    |    |    |    |    |-- variants: array (nullable = true)
|    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
|    |    |    |    |    |    |    |-- content_type: string (nullable = true)
|    |    |    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |-- symbols: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- text: string (nullable = true)
|    |    |    |-- urls: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |-- user_mentions: array (nullable = true)
```

```
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- name: string (nullable = true)
|    |    |    |    |-- screen_name: string (nullable = true)
|    |-- extended_entities: struct (nullable = true)
|    |    |-- media: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |-- video_info: struct (nullable = true)
```

```
|    |    |    |    |    |-- aspect_ratio: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- duration_millis: long (nullable = true)
|    |    |    |    |    |-- variants: array (nullable = true)
|    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
|    |    |    |    |    |    |    |-- content_type: string (nullable = true)
|    |    |    |    |    |    |    |-- url: string (nullable = true)
|    |-- full_text: string (nullable = true)
|-- favorite_count: long (nullable = true)
|-- favorited: boolean (nullable = true)
|-- filter_level: string (nullable = true)
|-- geo: struct (nullable = true)
|    |-- coordinates: array (nullable = true)
|    |    |-- element: double (containsNull = true)
|    |-- type: string (nullable = true)
|-- id: long (nullable = true)
|-- id_str: string (nullable = true)
|-- in_reply_to_screen_name: string (nullable = true)
|-- in_reply_to_status_id: long (nullable = true)
|-- in_reply_to_status_id_str: string (nullable = true)
|-- in_reply_to_user_id: long (nullable = true)
|-- in_reply_to_user_id_str: string (nullable = true)
|-- is_quote_status: boolean (nullable = true)
|-- lang: string (nullable = true)
|-- place: struct (nullable = true)
|    |-- bounding_box: struct (nullable = true)
|    |    |-- coordinates: array (nullable = true)
|    |    |    |-- element: array (containsNull = true)
|    |    |    |    |-- element: array (containsNull = true)
|    |    |    |    |    |-- element: double (containsNull = true)
|    |    |-- type: string (nullable = true)
|    |-- country: string (nullable = true)
|    |-- country_code: string (nullable = true)
|    |-- full_name: string (nullable = true)
|    |-- id: string (nullable = true)
|    |-- name: string (nullable = true)
|    |-- place_type: string (nullable = true)
|    |-- url: string (nullable = true)
|-- possibly_sensitive: boolean (nullable = true)
|-- quote_count: long (nullable = true)
|-- quoted_status: struct (nullable = true)
|    |-- coordinates: struct (nullable = true)
|    |    |-- coordinates: array (nullable = true)
|    |    |    |-- element: double (containsNull = true)
|    |    |-- type: string (nullable = true)
|    |-- created_at: string (nullable = true)
|    |-- display_text_range: array (nullable = true)
```

```
|    |    |-- element: long (containsNull = true)
|    |-- entities: struct (nullable = true)
|    |    |-- hashtags: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- text: string (nullable = true)
|    |    |-- media: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |-- url: string (nullable = true)
|    |    |-- symbols: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
```

```
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- text: string (nullable = true)
|    |    |-- urls: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- url: string (nullable = true)
|    |    |-- user_mentions: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- name: string (nullable = true)
|    |    |    |    |-- screen_name: string (nullable = true)
|    |-- extended_entities: struct (nullable = true)
|    |    |-- media: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
```

```
|    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |-- video_info: struct (nullable = true)
|    |    |    |    |    |-- aspect_ratio: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- duration_millis: long (nullable = true)
|    |    |    |    |    |-- variants: array (nullable = true)
|    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
|    |    |    |    |    |    |    |-- content_type: string (nullable = true)
|    |    |    |    |    |    |    |-- url: string (nullable = true)
|    |-- extended_tweet: struct (nullable = true)
|    |    |-- display_text_range: array (nullable = true)
|    |    |    |-- element: long (containsNull = true)
|    |    |-- entities: struct (nullable = true)
|    |    |    |-- hashtags: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- text: string (nullable = true)
|    |    |    |-- media: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
```

```
 |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |-- medium: struct (nullable = true)
 |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |-- small: struct (nullable = true)
 |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |-- thumb: struct (nullable = true)
 |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |-- source_status_id: long (nullable = true)
 |    |    |    |    |    |-- source_status_id_str: string (nullable = true)
 |    |    |    |    |    |-- source_user_id: long (nullable = true)
 |    |    |    |    |    |-- source_user_id_str: string (nullable = true)
 |    |    |    |    |    |-- type: string (nullable = true)
 |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |    |    |    |-- video_info: struct (nullable = true)
 |    |    |    |    |    |    |-- aspect_ratio: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |    |-- duration_millis: long (nullable = true)
 |    |    |    |    |    |    |-- variants: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- content_type: string (nullable =
true)
 |    |    |    |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |    |-- symbols: array (nullable = true)
 |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |-- text: string (nullable = true)
 |    |    |    |-- urls: array (nullable = true)
 |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |-- display_url: string (nullable = true)
 |    |    |    |    |    |-- expanded_url: string (nullable = true)
 |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |    |-- user_mentions: array (nullable = true)
 |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |-- id: long (nullable = true)
 |    |    |    |    |    |-- id_str: string (nullable = true)
 |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |-- name: string (nullable = true)
```

```
|    |    |    |    |    |-- screen_name: string (nullable = true)
|    |    |-- extended_entities: struct (nullable = true)
|    |    |    |-- media: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |    |-- video_info: struct (nullable = true)
|    |    |    |    |    |    |-- aspect_ratio: array (nullable = true)
|    |    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |    |-- duration_millis: long (nullable = true)
|    |    |    |    |    |    |-- variants: array (nullable = true)
|    |    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
```

```
 |    |    |    |    |    |    |    |    |-- content_type: string (nullable =
true)
 |    |    |    |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |    |-- full_text: string (nullable = true)
 |    |-- favorite_count: long (nullable = true)
 |    |-- favorited: boolean (nullable = true)
 |    |-- filter_level: string (nullable = true)
 |    |-- geo: struct (nullable = true)
 |    |    |-- coordinates: array (nullable = true)
 |    |    |    |-- element: double (containsNull = true)
 |    |    |-- type: string (nullable = true)
 |    |-- id: long (nullable = true)
 |    |-- id_str: string (nullable = true)
 |    |-- in_reply_to_screen_name: string (nullable = true)
 |    |-- in_reply_to_status_id: long (nullable = true)
 |    |-- in_reply_to_status_id_str: string (nullable = true)
 |    |-- in_reply_to_user_id: long (nullable = true)
 |    |-- in_reply_to_user_id_str: string (nullable = true)
 |    |-- is_quote_status: boolean (nullable = true)
 |    |-- lang: string (nullable = true)
 |    |-- place: struct (nullable = true)
 |    |    |-- bounding_box: struct (nullable = true)
 |    |    |    |-- coordinates: array (nullable = true)
 |    |    |    |    |-- element: array (containsNull = true)
 |    |    |    |    |    |-- element: array (containsNull = true)
 |    |    |    |    |    |    |-- element: double (containsNull = true)
 |    |    |    |-- type: string (nullable = true)
 |    |    |-- country: string (nullable = true)
 |    |    |-- country_code: string (nullable = true)
 |    |    |-- full_name: string (nullable = true)
 |    |    |-- id: string (nullable = true)
 |    |    |-- name: string (nullable = true)
 |    |    |-- place_type: string (nullable = true)
 |    |    |-- url: string (nullable = true)
 |    |-- possibly_sensitive: boolean (nullable = true)
 |    |-- quote_count: long (nullable = true)
 |    |-- quoted_status_id: long (nullable = true)
 |    |-- quoted_status_id_str: string (nullable = true)
 |    |-- reply_count: long (nullable = true)
 |    |-- retweet_count: long (nullable = true)
 |    |-- retweeted: boolean (nullable = true)
 |    |-- scopes: struct (nullable = true)
 |    |    |-- followers: boolean (nullable = true)
 |    |-- source: string (nullable = true)
 |    |-- text: string (nullable = true)
 |    |-- truncated: boolean (nullable = true)
 |    |-- user: struct (nullable = true)
 |    |    |-- contributors_enabled: boolean (nullable = true)
```

```
|    |    |-- created_at: string (nullable = true)
|    |    |-- default_profile: boolean (nullable = true)
|    |    |-- default_profile_image: boolean (nullable = true)
|    |    |-- description: string (nullable = true)
|    |    |-- favourites_count: long (nullable = true)
|    |    |-- followers_count: long (nullable = true)
|    |    |-- friends_count: long (nullable = true)
|    |    |-- geo_enabled: boolean (nullable = true)
|    |    |-- id: long (nullable = true)
|    |    |-- id_str: string (nullable = true)
|    |    |-- is_translator: boolean (nullable = true)
|    |    |-- listed_count: long (nullable = true)
|    |    |-- location: string (nullable = true)
|    |    |-- name: string (nullable = true)
|    |    |-- profile_background_color: string (nullable = true)
|    |    |-- profile_background_image_url: string (nullable = true)
|    |    |-- profile_background_image_url_https: string (nullable = true)
|    |    |-- profile_background_tile: boolean (nullable = true)
|    |    |-- profile_banner_url: string (nullable = true)
|    |    |-- profile_image_url: string (nullable = true)
|    |    |-- profile_image_url_https: string (nullable = true)
|    |    |-- profile_link_color: string (nullable = true)
|    |    |-- profile_sidebar_border_color: string (nullable = true)
|    |    |-- profile_sidebar_fill_color: string (nullable = true)
|    |    |-- profile_text_color: string (nullable = true)
|    |    |-- profile_use_background_image: boolean (nullable = true)
|    |    |-- protected: boolean (nullable = true)
|    |    |-- screen_name: string (nullable = true)
|    |    |-- statuses_count: long (nullable = true)
|    |    |-- translator_type: string (nullable = true)
|    |    |-- url: string (nullable = true)
|    |    |-- verified: boolean (nullable = true)
|    |    |-- verified_type: string (nullable = true)
|    |    |-- withheld_in_countries: array (nullable = true)
|    |    |    |-- element: string (containsNull = true)
|    |-- withheld_copyright: boolean (nullable = true)
|    |-- withheld_in_countries: array (nullable = true)
|    |    |-- element: string (containsNull = true)
|-- quoted_status_id: long (nullable = true)
|-- quoted_status_id_str: string (nullable = true)
|-- quoted_status_permalink: struct (nullable = true)
|    |-- display: string (nullable = true)
|    |-- expanded: string (nullable = true)
|    |-- url: string (nullable = true)
|-- quoted_text: string (nullable = true)
|-- reply_count: long (nullable = true)
|-- retweet_count: long (nullable = true)
|-- retweeted: string (nullable = true)
```

```
|-- retweeted_from: string (nullable = true)
|-- retweeted_status: struct (nullable = true)
|    |-- coordinates: struct (nullable = true)
|    |    |-- coordinates: array (nullable = true)
|    |    |    |-- element: double (containsNull = true)
|    |    |-- type: string (nullable = true)
|    |-- created_at: string (nullable = true)
|    |-- display_text_range: array (nullable = true)
|    |    |-- element: long (containsNull = true)
|    |-- entities: struct (nullable = true)
|    |    |-- hashtags: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- text: string (nullable = true)
|    |    |-- media: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
```

```
|    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |-- url: string (nullable = true)
|    |    |    |-- symbols: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- text: string (nullable = true)
|    |    |    |-- urls: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |-- user_mentions: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- name: string (nullable = true)
|    |    |    |    |    |-- screen_name: string (nullable = true)
|    |-- extended_entities: struct (nullable = true)
|    |    |-- media: array (nullable = true)
|    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
```

```
|    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |-- type: string (nullable = true)
|    |    |    |    |-- url: string (nullable = true)
|    |    |    |    |-- video_info: struct (nullable = true)
|    |    |    |    |    |-- aspect_ratio: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- duration_millis: long (nullable = true)
|    |    |    |    |    |-- variants: array (nullable = true)
|    |    |    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
|    |    |    |    |    |    |    |-- content_type: string (nullable = true)
|    |    |    |    |    |    |    |-- url: string (nullable = true)
|    |-- extended_tweet: struct (nullable = true)
|    |    |-- display_text_range: array (nullable = true)
|    |    |    |-- element: long (containsNull = true)
|    |    |-- entities: struct (nullable = true)
|    |    |    |-- hashtags: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- text: string (nullable = true)
|    |    |    |-- media: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |-- id_str: string (nullable = true)
```

```
 |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |-- media_url: string (nullable = true)
 |    |    |    |    |    |-- media_url_https: string (nullable = true)
 |    |    |    |    |    |-- sizes: struct (nullable = true)
 |    |    |    |    |    |    |-- large: struct (nullable = true)
 |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |-- medium: struct (nullable = true)
 |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |-- small: struct (nullable = true)
 |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |-- thumb: struct (nullable = true)
 |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |-- source_status_id: long (nullable = true)
 |    |    |    |    |    |-- source_status_id_str: string (nullable = true)
 |    |    |    |    |    |-- source_user_id: long (nullable = true)
 |    |    |    |    |    |-- source_user_id_str: string (nullable = true)
 |    |    |    |    |    |-- type: string (nullable = true)
 |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |    |    |    |-- video_info: struct (nullable = true)
 |    |    |    |    |    |    |-- aspect_ratio: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |    |-- duration_millis: long (nullable = true)
 |    |    |    |    |    |    |-- variants: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- content_type: string (nullable =
true)
 |    |    |    |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |    |-- symbols: array (nullable = true)
 |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |-- text: string (nullable = true)
 |    |    |    |-- urls: array (nullable = true)
 |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |-- display_url: string (nullable = true)
 |    |    |    |    |    |-- expanded_url: string (nullable = true)
 |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |-- element: long (containsNull = true)
```

```
|    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |-- user_mentions: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- name: string (nullable = true)
|    |    |    |    |    |-- screen_name: string (nullable = true)
|    |    |-- extended_entities: struct (nullable = true)
|    |    |    |-- media: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |    |-- type: string (nullable = true)
```

```
 |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |    |    |    |-- video_info: struct (nullable = true)
 |    |    |    |    |    |    |-- aspect_ratio: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |    |-- duration_millis: long (nullable = true)
 |    |    |    |    |    |    |-- variants: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- content_type: string (nullable =
true)
 |    |    |    |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |-- full_text: string (nullable = true)
 |    |-- favorite_count: long (nullable = true)
 |    |-- favorited: boolean (nullable = true)
 |    |-- filter_level: string (nullable = true)
 |    |-- geo: struct (nullable = true)
 |    |    |-- coordinates: array (nullable = true)
 |    |    |    |-- element: double (containsNull = true)
 |    |    |-- type: string (nullable = true)
 |    |-- id: long (nullable = true)
 |    |-- id_str: string (nullable = true)
 |    |-- in_reply_to_screen_name: string (nullable = true)
 |    |-- in_reply_to_status_id: long (nullable = true)
 |    |-- in_reply_to_status_id_str: string (nullable = true)
 |    |-- in_reply_to_user_id: long (nullable = true)
 |    |-- in_reply_to_user_id_str: string (nullable = true)
 |    |-- is_quote_status: boolean (nullable = true)
 |    |-- lang: string (nullable = true)
 |    |-- place: struct (nullable = true)
 |    |    |-- bounding_box: struct (nullable = true)
 |    |    |    |-- coordinates: array (nullable = true)
 |    |    |    |    |-- element: array (containsNull = true)
 |    |    |    |    |    |-- element: array (containsNull = true)
 |    |    |    |    |    |    |-- element: double (containsNull = true)
 |    |    |    |-- type: string (nullable = true)
 |    |    |-- country: string (nullable = true)
 |    |    |-- country_code: string (nullable = true)
 |    |    |-- full_name: string (nullable = true)
 |    |    |-- id: string (nullable = true)
 |    |    |-- name: string (nullable = true)
 |    |    |-- place_type: string (nullable = true)
 |    |    |-- url: string (nullable = true)
 |    |-- possibly_sensitive: boolean (nullable = true)
 |    |-- quote_count: long (nullable = true)
 |    |-- quoted_status: struct (nullable = true)
 |    |    |-- coordinates: struct (nullable = true)
 |    |    |    |-- coordinates: array (nullable = true)
 |    |    |    |    |-- element: double (containsNull = true)
```

```
|    |    |    |-- type: string (nullable = true)
|    |    |-- created_at: string (nullable = true)
|    |    |-- display_text_range: array (nullable = true)
|    |    |    |-- element: long (containsNull = true)
|    |    |-- entities: struct (nullable = true)
|    |    |    |-- hashtags: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- text: string (nullable = true)
|    |    |    |-- media: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- small: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- thumb: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |-- source_status_id: long (nullable = true)
|    |    |    |    |    |-- source_status_id_str: string (nullable = true)
|    |    |    |    |    |-- source_user_id: long (nullable = true)
|    |    |    |    |    |-- source_user_id_str: string (nullable = true)
|    |    |    |    |    |-- type: string (nullable = true)
```

```
|    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |-- symbols: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- text: string (nullable = true)
|    |    |    |-- urls: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- url: string (nullable = true)
|    |    |    |-- user_mentions: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- name: string (nullable = true)
|    |    |    |    |    |-- screen_name: string (nullable = true)
|    |    |-- extended_entities: struct (nullable = true)
|    |    |    |-- media: array (nullable = true)
|    |    |    |    |-- element: struct (containsNull = true)
|    |    |    |    |    |-- additional_media_info: struct (nullable = true)
|    |    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |    |-- embeddable: boolean (nullable = true)
|    |    |    |    |    |    |-- monetizable: boolean (nullable = true)
|    |    |    |    |    |    |-- title: string (nullable = true)
|    |    |    |    |    |-- description: string (nullable = true)
|    |    |    |    |    |-- display_url: string (nullable = true)
|    |    |    |    |    |-- expanded_url: string (nullable = true)
|    |    |    |    |    |-- id: long (nullable = true)
|    |    |    |    |    |-- id_str: string (nullable = true)
|    |    |    |    |    |-- indices: array (nullable = true)
|    |    |    |    |    |    |-- element: long (containsNull = true)
|    |    |    |    |    |-- media_url: string (nullable = true)
|    |    |    |    |    |-- media_url_https: string (nullable = true)
|    |    |    |    |    |-- sizes: struct (nullable = true)
|    |    |    |    |    |    |-- large: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- medium: struct (nullable = true)
|    |    |    |    |    |    |    |-- h: long (nullable = true)
|    |    |    |    |    |    |    |-- resize: string (nullable = true)
|    |    |    |    |    |    |    |-- w: long (nullable = true)
|    |    |    |    |    |    |-- small: struct (nullable = true)
```

```
 |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |-- thumb: struct (nullable = true)
 |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |-- source_status_id: long (nullable = true)
 |    |    |    |    |    |-- source_status_id_str: string (nullable = true)
 |    |    |    |    |    |-- source_user_id: long (nullable = true)
 |    |    |    |    |    |-- source_user_id_str: string (nullable = true)
 |    |    |    |    |    |-- type: string (nullable = true)
 |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |    |    |    |-- video_info: struct (nullable = true)
 |    |    |    |    |    |    |-- aspect_ratio: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |    |-- duration_millis: long (nullable = true)
 |    |    |    |    |    |    |-- variants: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |    |    |    |-- bitrate: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- content_type: string (nullable =
true)
 |    |    |    |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |-- extended_tweet: struct (nullable = true)
 |    |    |    |-- display_text_range: array (nullable = true)
 |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |-- entities: struct (nullable = true)
 |    |    |    |    |-- hashtags: array (nullable = true)
 |    |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |    |-- text: string (nullable = true)
 |    |    |    |    |-- media: array (nullable = true)
 |    |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |    |-- additional_media_info: struct (nullable =
true)
 |    |    |    |    |    |    |    |-- description: string (nullable = true)
 |    |    |    |    |    |    |    |-- embeddable: boolean (nullable = true)
 |    |    |    |    |    |    |    |-- monetizable: boolean (nullable = true)
 |    |    |    |    |    |    |    |-- title: string (nullable = true)
 |    |    |    |    |    |    |-- description: string (nullable = true)
 |    |    |    |    |    |    |-- display_url: string (nullable = true)
 |    |    |    |    |    |    |-- expanded_url: string (nullable = true)
 |    |    |    |    |    |    |-- id: long (nullable = true)
 |    |    |    |    |    |    |-- id_str: string (nullable = true)
 |    |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |    |-- media_url: string (nullable = true)
```

```
 |    |    |    |    |    |    |-- media_url_https: string (nullable = true)
 |    |    |    |    |    |    |-- sizes: struct (nullable = true)
 |    |    |    |    |    |    |    |-- large: struct (nullable = true)
 |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |    |-- medium: struct (nullable = true)
 |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |    |-- small: struct (nullable = true)
 |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |    |-- thumb: struct (nullable = true)
 |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |-- source_status_id: long (nullable = true)
 |    |    |    |    |    |    |-- source_status_id_str: string (nullable =
true)
 |    |    |    |    |    |    |-- source_user_id: long (nullable = true)
 |    |    |    |    |    |    |-- source_user_id_str: string (nullable = true)
 |    |    |    |    |    |    |-- type: string (nullable = true)
 |    |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |    |    |    |    |-- video_info: struct (nullable = true)
 |    |    |    |    |    |    |    |-- aspect_ratio: array (nullable = true)
 |    |    |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |    |    |-- duration_millis: long (nullable = true)
 |    |    |    |    |    |    |    |-- variants: array (nullable = true)
 |    |    |    |    |    |    |    |    |-- element: struct (containsNull =
true)
 |    |    |    |    |    |    |    |    |    |-- bitrate: long (nullable =
true)
 |    |    |    |    |    |    |    |    |    |-- content_type: string (nullable
= true)
 |    |    |    |    |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |    |    |-- symbols: array (nullable = true)
 |    |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |    |-- text: string (nullable = true)
 |    |    |    |    |-- urls: array (nullable = true)
 |    |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |    |-- display_url: string (nullable = true)
 |    |    |    |    |    |    |-- expanded_url: string (nullable = true)
 |    |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: long (containsNull = true)
```

```
 |    |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |    |    |-- user_mentions: array (nullable = true)
 |    |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |    |-- id: long (nullable = true)
 |    |    |    |    |    |    |-- id_str: string (nullable = true)
 |    |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |    |-- name: string (nullable = true)
 |    |    |    |    |    |    |-- screen_name: string (nullable = true)
 |    |    |    |-- extended_entities: struct (nullable = true)
 |    |    |    |    |-- media: array (nullable = true)
 |    |    |    |    |    |-- element: struct (containsNull = true)
 |    |    |    |    |    |    |-- additional_media_info: struct (nullable =
true)
 |    |    |    |    |    |    |    |-- description: string (nullable = true)
 |    |    |    |    |    |    |    |-- embeddable: boolean (nullable = true)
 |    |    |    |    |    |    |    |-- monetizable: boolean (nullable = true)
 |    |    |    |    |    |    |    |-- title: string (nullable = true)
 |    |    |    |    |    |    |-- description: string (nullable = true)
 |    |    |    |    |    |    |-- display_url: string (nullable = true)
 |    |    |    |    |    |    |-- expanded_url: string (nullable = true)
 |    |    |    |    |    |    |-- id: long (nullable = true)
 |    |    |    |    |    |    |-- id_str: string (nullable = true)
 |    |    |    |    |    |    |-- indices: array (nullable = true)
 |    |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |    |-- media_url: string (nullable = true)
 |    |    |    |    |    |    |-- media_url_https: string (nullable = true)
 |    |    |    |    |    |    |-- sizes: struct (nullable = true)
 |    |    |    |    |    |    |    |-- large: struct (nullable = true)
 |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |    |-- medium: struct (nullable = true)
 |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |    |-- small: struct (nullable = true)
 |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |    |-- thumb: struct (nullable = true)
 |    |    |    |    |    |    |    |    |-- h: long (nullable = true)
 |    |    |    |    |    |    |    |    |-- resize: string (nullable = true)
 |    |    |    |    |    |    |    |    |-- w: long (nullable = true)
 |    |    |    |    |    |    |-- source_status_id: long (nullable = true)
 |    |    |    |    |    |    |-- source_status_id_str: string (nullable =
true)
 |    |    |    |    |    |    |-- source_user_id: long (nullable = true)
```

```
 |    |    |    |    |    |    |-- source_user_id_str: string (nullable = true)
 |    |    |    |    |    |    |-- type: string (nullable = true)
 |    |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |    |    |    |    |-- video_info: struct (nullable = true)
 |    |    |    |    |    |    |    |-- aspect_ratio: array (nullable = true)
 |    |    |    |    |    |    |    |    |-- element: long (containsNull = true)
 |    |    |    |    |    |    |    |-- duration_millis: long (nullable = true)
 |    |    |    |    |    |    |    |-- variants: array (nullable = true)
 |    |    |    |    |    |    |    |    |-- element: struct (containsNull =
true)
 |    |    |    |    |    |    |    |    |    |-- bitrate: long (nullable =
true)
 |    |    |    |    |    |    |    |    |    |-- content_type: string (nullable
= true)
 |    |    |    |    |    |    |    |    |    |-- url: string (nullable = true)
 |    |    |    |-- full_text: string (nullable = true)
 |    |    |-- favorite_count: long (nullable = true)
 |    |    |-- favorited: boolean (nullable = true)
 |    |    |-- filter_level: string (nullable = true)
 |    |    |-- geo: struct (nullable = true)
 |    |    |    |-- coordinates: array (nullable = true)
 |    |    |    |    |-- element: double (containsNull = true)
 |    |    |    |-- type: string (nullable = true)
 |    |    |-- id: long (nullable = true)
 |    |    |-- id_str: string (nullable = true)
 |    |    |-- in_reply_to_screen_name: string (nullable = true)
 |    |    |-- in_reply_to_status_id: long (nullable = true)
 |    |    |-- in_reply_to_status_id_str: string (nullable = true)
 |    |    |-- in_reply_to_user_id: long (nullable = true)
 |    |    |-- in_reply_to_user_id_str: string (nullable = true)
 |    |    |-- is_quote_status: boolean (nullable = true)
 |    |    |-- lang: string (nullable = true)
 |    |    |-- place: struct (nullable = true)
 |    |    |    |-- bounding_box: struct (nullable = true)
 |    |    |    |    |-- coordinates: array (nullable = true)
 |    |    |    |    |    |-- element: array (containsNull = true)
 |    |    |    |    |    |    |-- element: array (containsNull = true)
 |    |    |    |    |    |    |    |-- element: double (containsNull = true)
 |    |    |    |    |-- type: string (nullable = true)
 |    |    |    |-- country: string (nullable = true)
 |    |    |    |-- country_code: string (nullable = true)
 |    |    |    |-- full_name: string (nullable = true)
 |    |    |    |-- id: string (nullable = true)
 |    |    |    |-- name: string (nullable = true)
 |    |    |    |-- place_type: string (nullable = true)
 |    |    |    |-- url: string (nullable = true)
 |    |    |-- possibly_sensitive: boolean (nullable = true)
 |    |    |-- quote_count: long (nullable = true)
```

```
|    |    |-- quoted_status_id: long (nullable = true)
|    |    |-- quoted_status_id_str: string (nullable = true)
|    |    |-- reply_count: long (nullable = true)
|    |    |-- retweet_count: long (nullable = true)
|    |    |-- retweeted: boolean (nullable = true)
|    |    |-- scopes: struct (nullable = true)
|    |    |    |-- followers: boolean (nullable = true)
|    |    |-- source: string (nullable = true)
|    |    |-- text: string (nullable = true)
|    |    |-- truncated: boolean (nullable = true)
|    |    |-- user: struct (nullable = true)
|    |    |    |-- contributors_enabled: boolean (nullable = true)
|    |    |    |-- created_at: string (nullable = true)
|    |    |    |-- default_profile: boolean (nullable = true)
|    |    |    |-- default_profile_image: boolean (nullable = true)
|    |    |    |-- description: string (nullable = true)
|    |    |    |-- favourites_count: long (nullable = true)
|    |    |    |-- followers_count: long (nullable = true)
|    |    |    |-- friends_count: long (nullable = true)
|    |    |    |-- geo_enabled: boolean (nullable = true)
|    |    |    |-- id: long (nullable = true)
|    |    |    |-- id_str: string (nullable = true)
|    |    |    |-- is_translator: boolean (nullable = true)
|    |    |    |-- listed_count: long (nullable = true)
|    |    |    |-- location: string (nullable = true)
|    |    |    |-- name: string (nullable = true)
|    |    |    |-- profile_background_color: string (nullable = true)
|    |    |    |-- profile_background_image_url: string (nullable = true)
|    |    |    |-- profile_background_image_url_https: string (nullable = true)
|    |    |    |-- profile_background_tile: boolean (nullable = true)
|    |    |    |-- profile_banner_url: string (nullable = true)
|    |    |    |-- profile_image_url: string (nullable = true)
|    |    |    |-- profile_image_url_https: string (nullable = true)
|    |    |    |-- profile_link_color: string (nullable = true)
|    |    |    |-- profile_sidebar_border_color: string (nullable = true)
|    |    |    |-- profile_sidebar_fill_color: string (nullable = true)
|    |    |    |-- profile_text_color: string (nullable = true)
|    |    |    |-- profile_use_background_image: boolean (nullable = true)
|    |    |    |-- protected: boolean (nullable = true)
|    |    |    |-- screen_name: string (nullable = true)
|    |    |    |-- statuses_count: long (nullable = true)
|    |    |    |-- translator_type: string (nullable = true)
|    |    |    |-- url: string (nullable = true)
|    |    |    |-- verified: boolean (nullable = true)
|    |    |    |-- verified_type: string (nullable = true)
|    |    |    |-- withheld_in_countries: array (nullable = true)
|    |    |    |    |-- element: string (containsNull = true)
|    |    |-- withheld_in_countries: array (nullable = true)
```

```
|    |    |    |-- element: string (containsNull = true)
|    |-- quoted_status_id: long (nullable = true)
|    |-- quoted_status_id_str: string (nullable = true)
|    |-- quoted_status_permalink: struct (nullable = true)
|    |    |-- display: string (nullable = true)
|    |    |-- expanded: string (nullable = true)
|    |    |-- url: string (nullable = true)
|    |-- reply_count: long (nullable = true)
|    |-- retweet_count: long (nullable = true)
|    |-- retweeted: boolean (nullable = true)
|    |-- scopes: struct (nullable = true)
|    |    |-- followers: boolean (nullable = true)
|    |    |-- place_ids: array (nullable = true)
|    |    |    |-- element: string (containsNull = true)
|    |-- source: string (nullable = true)
|    |-- text: string (nullable = true)
|    |-- truncated: boolean (nullable = true)
|    |-- user: struct (nullable = true)
|    |    |-- contributors_enabled: boolean (nullable = true)
|    |    |-- created_at: string (nullable = true)
|    |    |-- default_profile: boolean (nullable = true)
|    |    |-- default_profile_image: boolean (nullable = true)
|    |    |-- description: string (nullable = true)
|    |    |-- favourites_count: long (nullable = true)
|    |    |-- followers_count: long (nullable = true)
|    |    |-- friends_count: long (nullable = true)
|    |    |-- geo_enabled: boolean (nullable = true)
|    |    |-- id: long (nullable = true)
|    |    |-- id_str: string (nullable = true)
|    |    |-- is_translator: boolean (nullable = true)
|    |    |-- listed_count: long (nullable = true)
|    |    |-- location: string (nullable = true)
|    |    |-- name: string (nullable = true)
|    |    |-- profile_background_color: string (nullable = true)
|    |    |-- profile_background_image_url: string (nullable = true)
|    |    |-- profile_background_image_url_https: string (nullable = true)
|    |    |-- profile_background_tile: boolean (nullable = true)
|    |    |-- profile_banner_url: string (nullable = true)
|    |    |-- profile_image_url: string (nullable = true)
|    |    |-- profile_image_url_https: string (nullable = true)
|    |    |-- profile_link_color: string (nullable = true)
|    |    |-- profile_sidebar_border_color: string (nullable = true)
|    |    |-- profile_sidebar_fill_color: string (nullable = true)
|    |    |-- profile_text_color: string (nullable = true)
|    |    |-- profile_use_background_image: boolean (nullable = true)
|    |    |-- protected: boolean (nullable = true)
|    |    |-- screen_name: string (nullable = true)
|    |    |-- statuses_count: long (nullable = true)
```

```
|    |    |-- translator_type: string (nullable = true)
|    |    |-- url: string (nullable = true)
|    |    |-- verified: boolean (nullable = true)
|    |    |-- verified_type: string (nullable = true)
|    |    |-- withheld_in_countries: array (nullable = true)
|    |    |    |-- element: string (containsNull = true)
|    |-- withheld_in_countries: array (nullable = true)
|    |    |-- element: string (containsNull = true)
|-- source: string (nullable = true)
|-- text: string (nullable = true)
|-- timestamp_ms: string (nullable = true)
|-- truncated: boolean (nullable = true)
|-- tweet_text: string (nullable = true)
|-- user: struct (nullable = true)
|    |-- contributors_enabled: boolean (nullable = true)
|    |-- created_at: string (nullable = true)
|    |-- default_profile: boolean (nullable = true)
|    |-- default_profile_image: boolean (nullable = true)
|    |-- description: string (nullable = true)
|    |-- favourites_count: long (nullable = true)
|    |-- followers_count: long (nullable = true)
|    |-- friends_count: long (nullable = true)
|    |-- geo_enabled: boolean (nullable = true)
|    |-- id: long (nullable = true)
|    |-- id_str: string (nullable = true)
|    |-- is_translator: boolean (nullable = true)
|    |-- listed_count: long (nullable = true)
|    |-- location: string (nullable = true)
|    |-- name: string (nullable = true)
|    |-- profile_background_color: string (nullable = true)
|    |-- profile_background_image_url: string (nullable = true)
|    |-- profile_background_image_url_https: string (nullable = true)
|    |-- profile_background_tile: boolean (nullable = true)
|    |-- profile_banner_url: string (nullable = true)
|    |-- profile_image_url: string (nullable = true)
|    |-- profile_image_url_https: string (nullable = true)
|    |-- profile_link_color: string (nullable = true)
|    |-- profile_sidebar_border_color: string (nullable = true)
|    |-- profile_sidebar_fill_color: string (nullable = true)
|    |-- profile_text_color: string (nullable = true)
|    |-- profile_use_background_image: boolean (nullable = true)
|    |-- protected: boolean (nullable = true)
|    |-- screen_name: string (nullable = true)
|    |-- statuses_count: long (nullable = true)
|    |-- translator_type: string (nullable = true)
|    |-- url: string (nullable = true)
|    |-- verified: boolean (nullable = true)
|    |-- verified_type: string (nullable = true)
```

```
|    |-- withheld_in_countries: array (nullable = true)
|    |    |-- element: string (containsNull = true)
|-- withheld_in_countries: array (nullable = true)
|    |-- element: string (containsNull = true)
```

Due to the huge amount of data in the full dataset, I will try taking one file as an example to understand the data structures and conduct EDA

0.2.3 Sample Data

[14]:
```
# Load one json file as a sample to test
twitter_sample = spark.read.json('gs://msca-bdp-tweets/final_project/
 ↪part-50692-aa6d3cb4-7022-4df2-9921-218307589ce2-c000.json')
twitter_sample.limit(5).toPandas()
```

[14]:
```
  coordinates                      created_at display_text_range  \
0        None  Sun Jan 01 05:05:44 +0000 2023               None
1        None  Sun Jan 01 05:05:44 +0000 2023           [17, 38]
2        None  Sun Jan 01 05:05:45 +0000 2023               None
3        None  Sun Jan 01 05:05:45 +0000 2023           [0, 140]
4        None  Sun Jan 01 05:05:46 +0000 2023               None

           entities  \
0
([([19, 43],   _ _  _   _ )], None, [], [], [(1393144504924839937,
1393144504924839937, [3, 17], Abhisek Jantar Mantar, AbhisekJantar), (91851084,
91851084, [44, 59], Sant Rampal Ji Maharaj, SaintRampalJiM)])
1  ([], [(None, None, pic.twitter.com/jlKoje7cFG,
https://twitter.com/ko_kayi/status/1609415556373041152/photo/1,
1609415533123997696, 1609415533123997696, [39, 62],
http://pbs.twimg.com/tweet_video_thumb/FlXK5nvaEAANufR.jpg,
https://pbs.twimg.com/tweet_video_thumb/FlXK5nvaEAANufR.jpg,
Row(large=Row(h=288, resize='fit', w=500), medium=Row(h=288, resize='fit',
w=500), small=Row(h=288, resize='fit', w=500), thumb=Row(h=150, resize='crop',
w=150)), None, None, None, None, photo, https://t.co/jlKoje7cFG)], [], [],
[(3247181346, 3247181346, [0, 16], SoFloMan, Iamtherealpiman)])
2
([], None, [], [], [])
3
([], None, [], [(twitter.com/i/web/status/1…,
https://twitter.com/i/web/status/1609415558050775040, [116, 139],
https://t.co/GVXLlpDqw0)], [])
4
([], None, [], [], [])

                     extended_entities  \
0
```

```
None
1  ([(None, None, pic.twitter.com/jlKoje7cFG,
https://twitter.com/ko_kayi/status/1609415556373041152/photo/1,
1609415533123997696, 1609415533123997696, [39, 62],
http://pbs.twimg.com/tweet_video_thumb/FlXK5nvaEAANufR.jpg,
https://pbs.twimg.com/tweet_video_thumb/FlXK5nvaEAANufR.jpg,
Row(large=Row(h=288, resize='fit', w=500), medium=Row(h=288, resize='fit',
w=500), small=Row(h=288, resize='fit', w=500), thumb=Row(h=150, resize='crop',
w=150)), None, None, None, None, animated_gif, https://t.co/jlKoje7cFG,
Row(aspect_ratio=[125, 72], duration_millis=None, variants=[Row(bitrate=0,
content_type='video/mp4',
url='https://video.twimg.com/tweet_video/FlXK5nvaEAANufR.mp4')]))],)
2
None
3
None
4
None

                          extended_tweet  \
0
None
1
None
2
None
3  ([0, 163], ([], [Row(additional_media_info=None, description=None,
display_url='pic.twitter.com/65L9ODITNy', expanded_url='https://twitter.com/Geof
fRhymer/status/1609415558050775040/photo/1', id=1609415553042505728,
id_str='1609415553042505728', indices=[164, 187],
media_url='http://pbs.twimg.com/media/FlXK6x8WYAACOgk.jpg',
media_url_https='https://pbs.twimg.com/media/FlXK6x8WYAACOgk.jpg',
sizes=Row(large=Row(h=900, resize='fit', w=1200), medium=Row(h=900,
resize='fit', w=1200), small=Row(h=510, resize='fit', w=680), thumb=Row(h=150,
resize='crop', w=150)), source_status_id=None, source_status_id_str=None,
source_user_id=None, source_user_id_str=None, type='photo',
url='https://t.co/65L9ODITNy', video_info=None)], [], [], []),
([Row(additional_media_info=None, description=None,
display_url='pic.twitter.com/65L9ODITNy', expanded_url='https://twitter.com/Geof
fRhymer/status/1609415558050775040/photo/1', id=1609415553042505728,
id_str='1609415553042505728', indices=[164, 187],
media_url='http://pbs.twimg.com/media/FlXK6x8WYAACOgk.jpg',
media_url_https='https://pbs.twimg.com/media/FlXK6x8WYAACOgk.jpg',
sizes=Row(large=Row(h=900, resize='fit', w=1200), medium=Row(h=900,
resize='fit', w=1200), small=Row(h=510, resize='fit', w=680), thumb=Row(h=150,
resize='crop', w=150)), source_status_id=None, source_status_id_str=None,
source_user_id=None, source_user_id_str=None, type='photo',
```

```
url='https://t.co/65L9ODITNy', video_info=None)],), Happy New Year 2023 YALL  I
Hope Everything You Inspire To Be Will Be FullFilled . Also This Is The Year I
Will Be Graduating College!! \n\nI'm Claiming It!!!
https://t.co/65L9ODITNy)
4
None

   favorite_count  favorited filter_level   geo  …  retweet_count retweeted  \
0               0      False          low  None  …              0        RT
1               0      False          low  None  …              0
2               0      False          low  None  …              0
3               0      False          low  None  …              0
4               0      False          low  None  …              0

  retweeted_from  \
0  AbhisekJantar
1           None
2           None
3           None
4           None

                      retweeted_status  \
0  (Sun Jan 01 01:53:30 +0000 2023, [0, 140], ([Row(indices=[0, 24],
text='  _ _  _   _ ')], None, [],
[Row(display_url='twitter.com/i/web/status/1…',
expanded_url='https://twitter.com/i/web/status/1609367179950174208',
indices=[116, 139], url='https://t.co/mlNRpTNypS')], [Row(id=91851084,
id_str='91851084', indices=[25, 40], name='Sant Rampal Ji Maharaj',
screen_name='SaintRampalJiM')]), None, ([0, 262], ([Row(indices=[0, 24],
text='  _ _  _   _ ')], [Row(additional_media_info=None,
description=None, display_url='pic.twitter.com/dgY54f1Y6s', expanded_url='https:
//twitter.com/AbhisekJantar/status/1609367179950174208/photo/1',
id=1609367170705940482, id_str='1609367170705940482', indices=[263, 286],
media_url='http://pbs.twimg.com/media/FlWe6jsaYAIlV-r.jpg',
media_url_https='https://pbs.twimg.com/media/FlWe6jsaYAIlV-r.jpg',
sizes=Row(large=Row(h=1250, resize='fit', w=1000), medium=Row(h=1200,
resize='fit', w=960), small=Row(h=680, resize='fit', w=544), thumb=Row(h=150,
resize='crop', w=150)), source_status_id=None, source_status_id_str=None,
source_user_id=None, source_user_id_str=None, type='photo',
url='https://t.co/dgY54f1Y6s', video_info=None)], [], [], [Row(id=91851084,
id_str='91851084', indices=[25, 40], name='Sant Rampal Ji Maharaj',
screen_name='SaintRampalJiM')]), ([Row(additional_media_info=None,
description=None, display_url='pic.twitter.com/dgY54f1Y6s', expanded_url='https:
//twitter.com/AbhisekJantar/status/1609367179950174208/photo/1',
id=1609367170705940482, id_str='1609367170705940482', indices=[263, 286],
media_url='http://pbs.twimg.com/media/FlWe6jsaYAIlV-r.jpg',
media_url_https='https://pbs.twimg.com/media/FlWe6jsaYAIlV-r.jpg',
```

```
sizes=Row(large=Row(h=1250, resize='fit', w=1000), medium=Row(h=1200,
resize='fit', w=960), small=Row(h=680, resize='fit', w=544), thumb=Row(h=150,
resize='crop', w=150)), source_status_id=None, source_status_id_str=None,
source_user_id=None, source_user_id_str=None, type='photo',
url='https://t.co/dgY54f1Y6s', video_info=None)],),
#  _ _  _   _ \n@SaintRampalJiM says, I can guarantee that if the
school and college going youth are sent to my ashram for satsang even for one
Sunday each month, then in a year's time they will give up on the western
influences and dressing, and follow https://t.co/dgY54f1Y6s), 178, False, low,
1609367179950174208, 1609367179950174208, None, None, None, None, None, False,
en, ((([[72.436739, 22.923256], [72.436739, 23.104662], [72.703725, 23.104662],
[72.703725, 22.923256]]], Polygon), India, IN, Ahmadabad City, India,
272983f6b52c196e, Ahmadabad City, city,
https://api.twitter.com/1.1/geo/id/272983f6b52c196e.json), False, 0, None, None,
None, None, 7, 181, False, None, <a href="http://twitter.com/download/android"
rel="nofollow">Twitter for Android</a>,
#  _ _  _   _ \n@SaintRampalJiM says, I can guarantee that if the
school and college going youth are sent… https://t.co/mlNRpTNypS, True, (False,
Fri May 14 10:02:20 +0000 2021, True, False,             ,
      \n             ,            , 6635, 13236,
2565, True, 1393144504924839937, 1393144504924839937, False, 4, India, Abhisek
Jantar Mantar, F5F8FA, , , False,
https://pbs.twimg.com/profile_banners/1393144504924839937/1620987851,
http://pbs.twimg.com/profile_images/1393549114269306884/UYPyPNVl_normal.jpg,
https://pbs.twimg.com/profile_images/1393549114269306884/UYPyPNVl_normal.jpg,
1DA1F2, C0DEED, DDEEF6, 333333, True, False, AbhisekJantar, 5677, none,
http://supremegod.org, False, []))
1
None
2
None
3
None
4
None

 source  \
0  <a href="http://twitter.com/download/android" rel="nofollow">Twitter for
Android</a>
1    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
2   <a href="https://cheapbotsdonequick.com" rel="nofollow">Cheap Bots, Done
Quick!</a>
3    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
4    <a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for
iPhone</a>
```

```
                                                                text  \
0  RT @AbhisekJantar: #  _ _  _   _ \n@SaintRampalJiM says, I can
guarantee that if the school and college going youth are sent to my…
1
@Iamtherealpiman Let’s go old school  https://t.co/jlKoje7cFG
2               What shall we talk about, professor? I've no opinion on tea so
long as it's potable, but expensive leaves do pique my interest.
3    Happy New Year 2023 YALL  I Hope Everything You Inspire To Be Will Be
FullFilled . Also This Is The Year I Will Be… https://t.co/GVXLlpDqw0
4                                                                    how do you
even show up to school after missing such a huge kick that badly

    timestamp_ms  truncated  \
0  1672549544870      False
1  1672549544791      False
2  1672549545261      False
3  1672549545191       True
4  1672549546335      False

                                       tweet_text  \
0  #  _ _  _   _ \n@SaintRampalJiM says, I can guarantee that if the
school and college going youth are sent to my ashram for satsang even for one
Sunday each month, then in a year's time they will give up on the western
influences and dressing, and follow https://t.co/dgY54f1Y6s
1
@Iamtherealpiman Let’s go old school  https://t.co/jlKoje7cFG
2
What shall we talk about, professor? I've no opinion on tea so long as it's
potable, but expensive leaves do pique my interest.
3
Happy New Year 2023 YALL  I Hope Everything You Inspire To Be Will Be FullFilled
. Also This Is The Year I Will Be Graduating College!! \n\nI’m Claiming It!!!
    https://t.co/65L9ODITNy
4
how do you even show up to school after missing such a huge kick that badly

                                user
0                                          (False, Thu Oct 21 03:26:33 +0000
2021, True, False,   ,          ,
   ,  -        \n'LORD SANT RAMPAL JI MAHARAJ', 178045, 2671, 575,
False, 1451026985963048966, 1451026985963048966, False, 0,  Sonipat, Haryana,
MANJEET ROHILLA //@ Follow Me, F5F8FA, , , False,
https://pbs.twimg.com/profile_banners/1451026985963048966/1662986586,
http://pbs.twimg.com/profile_images/1564178018897371136/vXOgHXrw_normal.jpg,
https://pbs.twimg.com/profile_images/1564178018897371136/vXOgHXrw_normal.jpg,
1DA1F2, C0DEED, DDEEF6, 333333, True, False, MANJEET95715329, 184004, none,
```

```
None, False, [])
1
(False, Tue May 03 13:22:16 +0000 2022, True, False, None, 117, 262, 114, False,
1521478992506761216, 1521478992506761216, False, 0, None, Kokai, F5F8FA, , ,
False, https://pbs.twimg.com/profile_banners/1521478992506761216/1670420415,
http://pbs.twimg.com/profile_images/1600485333073727488/IzYIUnH9_normal.jpg,
https://pbs.twimg.com/profile_images/1600485333073727488/IzYIUnH9_normal.jpg,
1DA1F2, C0DEED, DDEEF6, 333333, True, False, ko_kayi, 209, none, None, False,
[])
2  (False, Sun Feb 06 07:41:06 +0000 2022, True, False, A Linhardt quote bot,
tweets every 30 minutes. Not spoiler-free. Does not respond., 0, 74, 6, False,
1490228974793871361, 1490228974793871361, False, 2, lovingly made by Rex  more
→, Linhardt von Hevring, F5F8FA, , , False,
https://pbs.twimg.com/profile_banners/1490228974793871361/1644133387,
http://pbs.twimg.com/profile_images/1542691909725958144/AQc1_70v_normal.jpg,
https://pbs.twimg.com/profile_images/1542691909725958144/AQc1_70v_normal.jpg,
1DA1F2, C0DEED, DDEEF6, 333333, True, False, Linhardtbot_, 15637, none, https://
twitter.com/BylethBot/status/1542687049257365504?s=20&t=Ce9-SQoF3WJUq4bZv-8Zmw,
False, [])
3              (False, Fri Jul 20 16:06:47 +0000 2012, False, False, #VI   #BVI
  #VirginIslands #BritishVirginIslands #WWE #Scorpio #BlackLivesMatter  ,
206967, 7280, 8008, True, 707389129, 707389129, False, 87, British Virgin
Islands | ATL,      , 131516,
http://abs.twimg.com/images/themes/theme14/bg.gif,
https://abs.twimg.com/images/themes/theme14/bg.gif, False,
https://pbs.twimg.com/profile_banners/707389129/1577747732,
http://pbs.twimg.com/profile_images/1588035133449314305/GtFCBkyo_normal.jpg,
https://pbs.twimg.com/profile_images/1588035133449314305/GtFCBkyo_normal.jpg,
2196F3, EEEEEE, EFEFEF, 333333, True, False, GeoffRhymer, 222200, none, None,
False, [])
4
(False, Wed Mar 11 17:06:45 +0000 2020, True, False, #ElTri | #LevelUp |
#WeAreTexans | #Rockets | i love the houston astros, 59998, 454, 530, False,
1237787021000802305, 1237787021000802305, False, 1, Houston, TX, ral , F5F8FA, ,
, False, https://pbs.twimg.com/profile_banners/1237787021000802305/1646957116,
http://pbs.twimg.com/profile_images/1539843884443901952/Tl520Dta_normal.jpg,
https://pbs.twimg.com/profile_images/1539843884443901952/Tl520Dta_normal.jpg,
1DA1F2, C0DEED, DDEEF6, 333333, True, False, MillsMissed, 9116, none, None,
False, [])

[5 rows x 38 columns]
```

[ ]:
```
twitter_sample.count()
```

[ ]:
```
3891
```

[ ]:
```
# Reference: https://developer.twitter.com/en/docs/twitter-api/v1/
 ↪data-dictionary/object-model/tweet
#twitter_sample.columns
```

[ ]:
```
twitter_sample.describe().show()
```

```
+-------+------------------+-------------+-----------+--------------------+-
------------------+----------------------+--------------------+-------------
----------+--------------------+----------------------+----+----------+------
-------------+--------------------+--------------------+----------+----------
--+---------+--------------------+--------------------+--------------------+----
---------------+--------------------+
|summary|          created_at|favorite_count|filter_level|                  id|
id_str|in_reply_to_screen_name|in_reply_to_status_id|in_reply_to_status_id_str|
in_reply_to_user_id|in_reply_to_user_id_str|lang|quote_count|
quoted_status_id|quoted_status_id_str|
quoted_text|reply_count|retweet_count|retweeted|      retweeted_from|
source|                text|        timestamp_ms|          tweet_text|
+-------+------------------+-------------+-----------+--------------------+-
------------------+----------------------+--------------------+-------------
----------+--------------------+----------------------+----+----------+------
-------------+--------------------+--------------------+----------+----------
--+---------+--------------------+--------------------+--------------------+----
---------------+--------------------+
|  count|                3891|         3891|       3891|                3891|
3891|                   509|                 491|                    491|
509|                   509|3891|       3891|                 222|
222|                222|       3891|         3891|     3891|
1880|                3891|               3891|                3891|
3891|
|   mean|                null|          0.0|
null|1.586465943139159…|1.586465943139159…|                   null|
1.592608443771351…|     1.592608443771351…|5.366150258415684…|
5.366150258415684…|null|        0.0|1.585821218656793…|1.585821218656793…|
null|        0.0|          0.0|     null|                null|
null|                null|1.667077930612886…|               null|
| stddev|                null|          0.0|
null|2.134056251690485…|2.134056251690485…|                   null|
2.145571707658750…|     2.145571707658750…|6.457986503456906…|
6.457986503456906…|null|        0.0|5.225229950850277…|5.225229950850277…|
null|        0.0|          0.0|     null|                null|
null|                null| 5.087986592496483E9|                null|
|    min|Mon Sep 05 03:19:…|            0|         low| 1566627112303206400|
1566627112303206400|                0ldjma|  1484610686164652035|
1484610686164652035|             890891|    1002989942350450688|  en|
0|  961320185180377093| 1315679765374947328|"In 2023, let’s p…|          0|
```

```
0|         | Bloomsburg Unive…|<a href="http://i…|"The Transfer Por…|
1662347984754|" I think it's fa…|
|    max|Sun Jan 01 05:13:…|             0|         low| 1609417525757870080|
1609417525757870080|            zachmarsh22|  1609417227979329537|
1609417227979329537| 1605749420141813760|     998939688261079040|  en|
0| 1609417465527951360|  961320185180377093|  Opening Day is…|          0|
0|       RT|              zzntwt|<a href="https://…|   wow   Co…|
1672550014329|  A new quarter …|
+-------+--------------------+-------------+-----------+--------------------+-
------------------+----------------------+--------------------+-------------
----------+--------------------+----------------------+----+----------+------
--------------+--------------------+--------------------+-----------+-------
--+---------+--------------------+--------------------+--------------------+----
---------------+--------------------+
```

Filtering Sample

[39]:
```
# Filter the sample data by selecting only relevant columns for analysis
## Focus on user and place columns since they have a number of child objects,␣
 ↪so here I will only choose important related features

filtered_sample = twitter_sample.
 ↪select(['coordinates','favorite_count','filter_level','in_reply_to_screen_name',\
                                          ␣
 ↪'retweeted','retweeted_from','retweeted_status','retweeted_status.
 ↪retweet_count',\
                                           'text',\
                                           'place.country','place.
 ↪full_name','place.place_type',\
                                           'timestamp_ms',\
                                           'user.id_str','user.name','user.
 ↪screen_name','user.location','user.description','user.followers_count','user.
 ↪statuses_count','user.created_at','user.verified',\
                                           'lang'\
                                          ])
filtered_sample.limit(5)
```

[39]:
```
+----------+-------------+-----------+----------------------+---------+-----
--------+--------------------+-------------+--------------------+-------+------
---+---------+------------+------------------+--------------------+----------
-----+--------------------+--------------------+--------------+--------------+-
------------------+--------+----+
|coordinates|favorite_count|filter_level|in_reply_to_screen_name|retweeted|retwe
eted_from|    retweeted_status|retweet_count|
text|country|full_name|place_type| timestamp_ms|             id_str|
name|    screen_name|            location|
description|followers_count|statuses_count|          created_at|verified|lang|
```

```
+----------+-------------+-----------+--------------------+---------+-----
--------+-------------------+-------------+-------------------+-------+------
---+----------+-------------+-------------------+-------------------+----------
-----+-------------------+-------------------+--------------+--------------+-
------------------+--------+----+
|       null|             0|        low|                null|       RT|
AbhisekJantar|{Sun Jan 01 01:53…|          181|RT @AbhisekJantar…|   null|
null|      null|1672549544870|1451026985963048966|MANJEET ROHILLA
/…|MANJEET95715329|    Sonipat, Haryana|   ,       …|           2671|
184004|Thu Oct 21 03:26:…|   false|  en|
|       null|             0|        low|       Iamtherealpiman|         |
null|                null|         null|@Iamtherealpiman …|   null|     null|
null|1672549544791|1521478992506761216|               Kokai|        ko_kayi|
null|                null|            262|           209|Tue May 03 13:22:…|
false|  en|
|       null|             0|        low|                null|         |
null|                null|         null|What shall we tal…|   null|     null|
null|1672549545261|1490228974793871361|Linhardt von Hevring|
Linhardtbot_|lovingly made by …|A Linhardt quote …|             74|
15637|Sun Feb 06 07:41:…|   false|  en|
|       null|             0|        low|                null|         |
null|                null|         null|Happy New Year 20…|   null|     null|
null|1672549545191|          707389129|           |    GeoffRhymer|British
Virgin Is…|#VI   #BVI  ?…|           7280|        222200|Fri Jul 20
16:06:…|   false|  en|
|       null|             0|        low|                null|         |
null|                null|         null|how do you even s…|   null|     null|
null|1672549546335|1237787021000802305|               ral |    MillsMissed|
Houston, TX|#ElTri | #LevelUp…|            454|          9116|Wed Mar 11
17:06:…|   false|  en|
+----------+-------------+-----------+--------------------+---------+-----
--------+-------------------+-------------+-------------------+-------+------
---+----------+-------------+-------------------+-------------------+----------
-----+-------------------+-------------------+--------------+--------------+-
------------------+--------+----+
```

[ ]:

[9]:
```
# create a list of words related to education
edu_words =␣
↪['education','K-12','teachers','professors','students','university','universities','college
```

[33]:
```
# From the filtered sample above, do another filter to identify␣
↪education-related tweets only

filtered_sample2 = filtered_sample.filter((filtered_sample.text.rlike('|'.
↪join(edu_words))) & (filtered_sample.lang == 'en'))
```

```
filtered_sample2.limit(10)
```

[33]:
```
+-----------+-------------+-----------+----------------------+---------+-----
--------------+--------------------+-------------+--------------------+-------+
---------+----------+-------------+------------------+------------------+-----
----------+--------------------+--------------------+---------------+----------
---+--------------------+--------+----+
|coordinates|favorite_count|filter_level|in_reply_to_screen_name|retweeted|
retweeted_from|    retweeted_status|retweet_count|
text|country|full_name|place_type| timestamp_ms|             id_str|
name|    screen_name|            location|
description|followers_count|statuses_count|          created_at|verified|lang|
+-----------+-------------+-----------+----------------------+---------+-----
--------------+--------------------+-------------+--------------------+-------+
---------+----------+-------------+------------------+------------------+-----
----------+--------------------+--------------------+---------------+----------
---+--------------------+--------+----+
|       null|            0|         low|         DougLesmerises|         |
null|                null|         null|@DougLesmerises F…|   null|     null|
null|1672549795589|1442296833548566532|The Ultimate Critic|critic_ultimate|
null|                 Nah|            0|           511|Mon Sep 27 01:16:…|
false|  en|
|       null|            0|         low|              TexansCap|         |
null|                null|         null|@TexansCap Man… u…|   null|     null|
null|1672549796062|1065369515129749504|    KrazyKarl   |         kp8912|
Houston, TX|Houston Sports Ps…|            261|          5545|Wed Nov 21
22:21:…|   false|  en|
|       null|            0|         low|                  null|       RT|
mmpadellan|{Sat Dec 31 23:00…|          707|RT @mmpadellan: I…|   null|
null|      null|1672549796249|1503902422598516739|  Sheryl
Bridgeford|SherylBridgefo1|Washington, the S…|I bailed on Tweet…|
55|          1594|Wed Mar 16 01:14:…|   false|  en|
|       null|            0|         low|                  null|         |
null|                null|         null|I love college fo…|   null|     null|
null|1672549796276|          857052152|         Terry Teel|       teeltron|
Springfield, MO|USAF VET. PBR. FO…|            49|          4331|Mon Oct 01
21:30:…|   false|  en|
|       null|            0|         low|                KevanGP|         |
null|                null|         null|@KevanGP You're w…|   null|     null|
null|1672549798147|           55918804|        Kent Ahrens|    kent_ahrens|
Columbia, MO|               null|            550|         28833|Sat Jul 11
20:14:…|   false|  en|
|       null|            0|         low|                  null|         |
null|                null|         null|Feliz not Navidad…|   null|     null|
null|1672549798514|1279413188765810688|          Claes  |     ClaesAzure|
```

```
Ontario, Canada|21       …|             56|         21863|Sat Jul 04
13:54:…|   false|  en|
|       null|             0|         low|             alabama313|
|AggieFootball @TJ…|               null|         null|@alabama313 @Aggi…|
null|     null|      null|1672549798875|1519536023759425536|          EIon
Musk|EIonMus52986652|               null|                null|             0|
133|Thu Apr 28 04:36:…|   false|  en|
|       null|             0|         low|               echenze|
|xysist Funny enou…|               null|         null|@echenze @xysist …|
null|     null|      null|1672549800743| 959573829537460224|          P. Otieno|
DataNomadKE|               Kenya|.Anything Data. D…|           2830|
4516|Fri Feb 02 23:46:…|   false|  en|
|       null|             0|         low|                  null|       RT|
mattufford|{Sun Jan 01 05:05…|           11|RT @mattufford: t…|   null|
null|      null|1672549801209|          326412173|      Chris Jackson|
ChrisCJackson|         Houston, TX|Lawyer but don't …|          13822|
131814|Wed Jun 29 21:43:…|   false|  en|
|       null|             0|         low|                  null|       RT|
PalanShail|{Thu Dec 29 14:54…|           10|RT @PalanShail: T…|   null|
null|      null|1672549802352| 820598109990309888|           Ahir Het|
Ahirhet623|       Kukma,Gujarat| astronomy lover…|             74|
134|Sun Jan 15 11:46:…|   false|  en|
+-----------+-------------+-----------+----------------------+---------+-----
--------------+--------------------+------------+--------------------+-------+
---------+----------+-------------+------------------+------------------+-----
----------+--------------------+--------------------+---------------+----------
---+--------------------+--------+----+
```

[28]:
```
# Count the number of tweets in the sample after filtering
filtered_sample2.count()
```

[28]:
```
1533
```

[17]:
```
filtered_sample2.describe().show()
```

```
[Stage 33:=============================================>              (3 + 1) / 4]

+-------+-------------+-----------+----------------------+---------+---------
----+--------------------+--------------------------+------------+-----------
-+----------+--------------------+-----------+-----------------+--------------
-----+-----------------+-----------------+--------------------+----+
|summary|favorite_count|filter_level|in_reply_to_screen_name|retweeted|retweet_c
ount|      retweeted_from|                      text|      country|
name|place_type|              id_str| screen_name|         location|
description|  followers_count|   statuses_count|          created_at|lang|
+-------+-------------+-----------+----------------------+---------+---------
----+--------------------+--------------------------+------------+-----------
-+----------+--------------------+-----------+-----------------+--------------
```

```
-----+----------------+----------------+--------------------+----+
|  count|          1533|        1533|                   230|     1533|
1533|                628|                    1533|           56|
56|        56|                1533|        1533|             1012|
1330|             1533|             1533|                1533|1533|
|   mean|           0.0|        null|                  null|     null|
0.0|                null|                      null|         null|
null|      null|5.245674498878506…|        null|616.3333333333334|
3.66670000745516E14|2861.569471624266|37609.95629484671|
null|null|
| stddev|           0.0|        null|                  null|     null|
0.0|                null|                      null|         null|
null|      null|6.396983794871625E17|
null|794.3648196305439|6.350910698326311E14|20662.10159194182|89716.64440352617|
null|null|
|    min|             0|         low|              AKMartich|         |
0|1053SS I'm pretty…|#BOB   #     #…|       Canada|      Arizona|
admin| 10037142179989790 72|10HernandezM|   Houston Texas |  ! ggs + bangtxteez|
0|                3|Fri Apr 01 02:18:…|  en|
|    max|             0|         low|               wasssam_|       RT|
0|              zzntwt|         my college f…|United States|West Hartford|
city|           998788866|     zoosey_|                | astronomy lover…|
580079|          1342059|Wed Sep 29 18:54:…|  en|
+-------+--------------+------------+----------------------+---------+---------
----+--------------------+---------------------------+-------------+-----------
-+----------+--------------------+------------+-----------------+--------------
-----+----------------+----------------+--------------------+----+
```

0.2.4 Filtering the Original Data and save it

[10]:
```
%%time
filtered_df = twitter_df.filter((twitter_df.text.rlike('|'.join(edu_words))) &␣
↪(twitter_df.lang == 'en'))\
                        .
↪select(['coordinates','favorite_count','filter_level','in_reply_to_screen_name',\
                                  ␣
↪'retweeted','retweet_count','retweeted_from','retweeted_status','text',\
                                    'place.country','place.country_code','place.
↪full_name','place.place_type','place.bounding_box',\
                                    'timestamp_ms',\
                                    'user.id_str','user.name','user.
↪screen_name','user.location','user.description','user.followers_count','user.
↪statuses_count','user.created_at','user.verified',\
                                    'lang'])
```

```
filtered_df.limit(5)
```

```
CPU times: user 18.2 ms, sys: 2.82 ms, total: 21 ms
Wall time: 382 ms
```

[10]:

```
+-----------+-------------+-----------+----------------------+---------+-----
-------+---------------+-------------------+-------------------+-------+-----
-------+---------+----------+-----------+-------------+-------------------+----
---------------+-------------+-------------------+-------------------+------
---------+-------------+-------------------+--------+----+
|coordinates|favorite_count|filter_level|in_reply_to_screen_name|retweeted|retwe
et_count| retweeted_from|    retweeted_status|
text|country|country_code|full_name|place_type|bounding_box| timestamp_ms|
id_str|                name|   screen_name|            location|
description|followers_count|statuses_count|          created_at|verified|lang|
+-----------+-------------+-----------+----------------------+---------+-----
-------+---------------+-------------------+-------------------+-------+-----
-------+---------+----------+-----------+-------------+-------------------+----
---------------+-------------+-------------------+-------------------+------
---------+-------------+-------------------+--------+----+
|       null|             0|         low|                   null|       RT|
0|       egggpoke|{null, Sun Sep 04…|RT @egggpoke: WTF…|   null|
null|     null|      null|        null|1662263541651|1433212678537891842|eddy !
thankyoujerma|    _eddywave_|             they/he|hi im ed :)
hobby…|            119|          1104|Wed Sep 01 23:38:…|   false|  en|
|       null|             0|         low|                   null|       RT|
0|    klllngsleyy|{null, Sat Sep 03…|RT @klllngsleyy: …|   null|
null|     null|      null|        null|1662263542731|
821489779845931009|  pitobi YJAS|   zhaneennovy|Western Visayas,
…|°Everything happe…|            112|          4245|Tue Jan 17 22:50:…|
false|  en|
|       null|             0|         low|                   null|       RT|
0|     tooopscoop|{null, Sat Sep 03…|RT @tooopscoop: S…|   null|
null|     null|      null|        null|1662263543586| 784543632242122752|
hot diggity dog|tgeonemisfit35|                null|     not new to this|
172|         45168|Fri Oct 07 23:59:…|   false|  en|
|       null|             0|         low|                   null|       RT|
0|     miera_rrip|{null, Thu Aug 25…|RT @miera_rrip: F…|   null|
null|     null|      null|        null|1662263544459|1259673673587404802|amway |
kaftannnn…|sincerelybyfaa|                null|BAJU KAFTAN | TEL…|
1797|         56617|Mon May 11 02:36:…|   false|  en|
|       null|             0|         low|                   null|       RT|
0|ScooterMagruder|{null, Sun Sep 04…|RT @ScooterMagrud…|   null|
null|     null|      null|        null|1662263545337| 919407662252593153|
Ashley Bueno|   heather_b27|            Florida |                null|
176|          4577|Sun Oct 15 03:40:…|   false|  en|
```

```
+-----------+-------------+-----------+--------------------+---------+-----
--------+---------------+-------------------+-------------------+-------+-----
-------+---------+----------+-----------+------------+-------------------+----
---------------+-------------+-------------------+-------------------+------
---------+-------------+-------------------+--------+----+
```

[ ]:
```
# Count the data volume in the original df after applying filters

filtered_df.count()
```

[ ]:
```
24721729
```

[ ]:
```
# Save the filtered original dataset for future use

save_path = 'gs://msca-bdp-students-bucket/shared_data/mdvo/BDP-Final-Project'
filtered_df.write.format('parquet').\
            mode('overwrite').\
            save(save_path)
```

[ ]:
```
!hadoop fs -ls "gs://msca-bdp-students-bucket/shared_data/mdvo/
↪BDP-Final-Project" | head -7
```

```
Found 5742 items
-rwx------   3 root root          0 2023-03-09 07:30 gs://msca-bdp-students-
bucket/shared_data/mdvo/BDP-Final-Project/_SUCCESS
-rwx------   3 root root    2605946 2023-03-09 07:14 gs://msca-bdp-students-
bucket/shared_data/mdvo/BDP-Final-
Project/part-00000-59df6c97-c835-4c6d-8159-1dfe42cf74f5-c000.snappy.parquet
-rwx------   3 root root    2855860 2023-03-09 07:14 gs://msca-bdp-students-
bucket/shared_data/mdvo/BDP-Final-
Project/part-00001-59df6c97-c835-4c6d-8159-1dfe42cf74f5-c000.snappy.parquet
-rwx------   3 root root    2960462 2023-03-09 07:14 gs://msca-bdp-students-
bucket/shared_data/mdvo/BDP-Final-
Project/part-00002-59df6c97-c835-4c6d-8159-1dfe42cf74f5-c000.snappy.parquet
-rwx------   3 root root    2707638 2023-03-09 07:14 gs://msca-bdp-students-
bucket/shared_data/mdvo/BDP-Final-
Project/part-00003-59df6c97-c835-4c6d-8159-1dfe42cf74f5-c000.snappy.parquet
-rwx------   3 root root    3077704 2023-03-09 07:14 gs://msca-bdp-students-
bucket/shared_data/mdvo/BDP-Final-
Project/part-00004-59df6c97-c835-4c6d-8159-1dfe42cf74f5-c000.snappy.parquet
```