

Data Description

The data you are provided with are tweets scraped directly from the public Twitter API with a query that collected all tweets that mention the twitter handle of the following airlines or are from/to the account ID of that particular airline:

• KLM	56377143
• AirFrance	106062176
• British_Airways	18332190
• AmericanAir	22536055
• Lufthansa	124476322
• AirBerlin	26223583
• AirBerlin assist	2182373406
• easyJet	38676903
• RyanAir	1542862735
• SingaporeAir	253340062
• Qantas	218730857
• EtihadAirways	45621423
• VirginAtlantic	20626359

The data has been collected between May 22, 2019 and March 30, 2020. Some interruptions due to technical issues may have occurred.

The Twitter API returns plain-text JavaScript Object Notation (JSON). These are grouped together into plain-text files of 64MB each. Each of these files contains between 9000 and 15000 lines of text, where each line is one tweet, for a total of approximately 6.5 million tweets in 35.2GB.

For a full overview of the output provided by the Twitter API and the exact definition of all the information see:

<https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object>

JSON Example: Tweet Object

```
{ "created_at": "Wed May 22 12:20:49 +0000 2019", "id": 1131173065122078720,
  "id_str": "1131173065122078720", "text": "So @AmericanAir @EWRairport lied to an old lady causing her
  to miss her flight #934 to LHR. She arrived at 6:05am f\u0026 https://t.co/Q9Ae8JW0tP",
  "source": "\u003ca href=\"http://twitter.com/download/iphone\" rel=\"nofollow\"\u003eTwitter for
  iPhone\u003c/a\u003e", "truncated": true, "in_reply_to_status_id": null, "in_reply_to_status_id_str": null,
  "in_reply_to_user_id": null, "in_reply_to_user_id_str": null, "in_reply_to_screen_name": null }
```

Example of a Tweet in a Database

```
_id: ObjectId("5e94446754effd0af8cdcc1a")
created_at: "Wed May 22 12:20:49 +0000 2019"
id: 1131173065122078720
id_str: "1131173065122078720"
text: "So @AmericanAir @EWRairport lied to an old lady causing her to miss he..."
source: "<a href='\"http://twitter.com/download/iphone\"' rel='\"nofollow\">Twitter fo..."
truncated: true
in_reply_to_status_id: null
in_reply_to_status_id_str: null
in_reply_to_user_id: null
in_reply_to_user_id_str: null
in_reply_to_screen_name: null
> user: Object
  geo: null
  coordinates: null
  place: null
  contributors: null
  is_quote_status: false
> extended_tweet: Object
  quote_count: 0
  reply_count: 0
  retweet_count: 0
  favorite_count: 0
> entities: Object
  favorited: false
  retweeted: false
  filter_level: "low"
  lang: "en"
  timestamp_ms: "1558527649561"
```

Next to the information about the tweet itself, each line has information about the user sending the tweet, extended metadata on the tweet, and if available the geolocation data. These are larger blocks of information associated with every tweet in so-called child objects. For example, if a user is mentioned in the tweet text this creates additional variables in the `extended_tweet` object:

```
  > extended_tweet: Object
    full_text: "So @AmericanAir @EWRairport lied to an old lady causing her to miss he..."
    > display_text_range: Array
      0: 0
      1: 279
    > entities: Object
      > hashtags: Array
      > urls: Array
      > user_mentions: Array
        > 0: Object
          screen_name: "AmericanAir"
          name: "American Airlines"
          id: 22536055
          id_str: "22536055"
          > indices: Array
        > 1: Object
          screen_name: "EWRairport"
          name: "Newark Liberty Airport"
          id: 234423098
          id_str: "234423098"
          > indices: Array
      > symbols: Array
```

Every tweet also contains the information about the user profile of the poster:

```
_id: ObjectId("5e94446754effd0af8cdcc1a")
created_at: "Wed May 22 12:20:49 +0000 2019"
id: 1131173065122078720
id_str: "1131173065122078720"
text: "So @AmericanAir @EWRAirport lied to an old lady causing her to miss he..."
source: "<a href='\"http://twitter.com/download/iphone\"' rel='\"nofollow\"'>Twitter fo..."
truncated: true
in_reply_to_status_id: null
in_reply_to_status_id_str: null
in_reply_to_user_id: null
in_reply_to_user_id_str: null
in_reply_to_screen_name: null
user: Object
  id: 2730172562
  id_str: "2730172562"
  name: "Daman"
  screen_name: "DamanMalone"
  location: "A rock in space"
  url: null
  description: "Business owner, Animal advocate, horror movie fan and collector of jea..."
  translator_type: "none"
  protected: false
  verified: false
  followers_count: 163
  friends_count: 746
  listed_count: 6
  favourites_count: 5403
  statuses_count: 1256
  created_at: "Wed Aug 13 19:45:35 +0000 2014"
  utc_offset: null
  time_zone: null
  geo_enabled: true
  lang: null
  contributors_enabled: false
  is_translator: false
  profile_background_color: "000000"
  profile_background_image_url: "http://abs.twimg.com/images/themes/theme1/bg.png"
  profile_background_image_url_https: "https://abs.twimg.com/images/themes/theme1/bg.png"
  profile_background_tile: false
  profile_link_color: "FF691F"
  profile_sidebar_border_color: "000000"
  profile_sidebar_fill_color: "000000"
  profile_text_color: "000000"
  profile_use_background_image: false
  profile_image_url: "http://pbs.twimg.com/profile_images/1082748978763059201/EvXohsLa_norma..."
  profile_image_url_https: "https://pbs.twimg.com/profile_images/1082748978763059201/EvXohsLa_norm..."
  profile_banner_url: "https://pbs.twimg.com/profile_banners/2730172562/1542426215"
  default_profile: false
  default_profile_image: false
```