

```
In [1]: import pandas as pd
import altair as alt
```

```
In [28]: raw_data = pd.read_csv('clean_tweet.csv')
raw_data
```

Out[28]:

|     | _id                      | data.created_at                | data.fav_count | data.id      | data.lang | data.possibly_sensitive | data.retweet_count | data.source                                       | data.text                                         |
|-----|--------------------------|--------------------------------|----------------|--------------|-----------|-------------------------|--------------------|---------------------------------------------------|---------------------------------------------------|
| 0   | 638625d1073ea59a7af54da7 | Tue Nov 29 15:31:21 +0000 2022 | 0              | 1.597610e+18 | en        | NaN                     | 0                  | href="http://twitter.com/download/iphone" r...    | I take 0 action without strategy                  |
| 1   | 638625d2073ea59a7af54da8 | Tue Nov 29 15:31:20 +0000 2022 | 0              | 1.597610e+18 | en        | NaN                     | 0                  | <a href="https://mobile.twitter.com" rel="nofo... | @GuidoDisalle gm Guido!                           |
| 2   | 638625d2073ea59a7af54da9 | Tue Nov 29 15:31:18 +0000 2022 | 0              | 1.597610e+18 | en        | NaN                     | 0                  | href="http://twitter.com/download/iphone" r...    | @SleuthieGoosie Absolutely agree with you. Not... |
| 3   | 638625d2073ea59a7af54daa | Tue Nov 29 15:31:16 +0000 2022 | 0              | 1.597610e+18 | en        | NaN                     | 0                  | href="http://twitter.com/download/iphone" r...    | @ThisIsKyleR @elonmusk Yes, life must be soooo... |
| 4   | 638625d2073ea59a7af54dab | Tue Nov 29 15:31:13 +0000 2022 | 0              | 1.597610e+18 | en        | False                   | 0                  | href="http://twitter.com/download/android" ...    | @funstonpaleo So looking forward to seeing the... |
| ... | ...                      | ...                            | ...            | ...          | ...       | ...                     | ...                | ...                                               | ...                                               |
| 936 | 638afc07e9ecfc21816ed142 | Sat Dec 03 07:15:26 +0000 2022 | 1              | 1.598940e+18 | und       | False                   | 0                  | href="http://twitter.com/download/android" ...    | Cry https://t.co/i0yoZW9CUp                       |
| 937 | 638afc07e9ecfc21816ed143 | Sat Dec 03 07:15:20 +0000 2022 | 0              | 1.598940e+18 | en        | NaN                     | 0                  | href="http://twitter.com/download/android" ...    | @TheRealKean Who's the journalist?                |
| 938 | 638afc09e9ecfc21816ed144 | Sat Dec 03 07:15:17 +0000 2022 | 1              | 1.598940e+18 | en        | False                   | 0                  | href="http://twitter.com/download/android" ...    | Let's! https://t.co/LLkOotHZPq                    |
| 939 | 638afc0ae9ecfc21816ed145 | Sat Dec 03 07:15:14 +0000 2022 | 0              | 1.598940e+18 | en        | NaN                     | 0                  | href="http://twitter.com/download/android" ...    | @MsPrinBoo @Justin_Ling I second this, please 🍌   |
| 940 | 638afc0be9ecfc21816ed146 | Sat Dec 03 07:14:54 +0000 2022 | 0              | 1.598940e+18 | en        | NaN                     | 0                  | href="http://twitter.com/download/android" ...    | @s_thompsett @Apollol69 @patdubois @TheoFleury... |

941 rows × 13 columns

## Dataframe tweets only containing:

- data.text
- data.lang

```
In [29]: tweets = raw_data[['data.text', 'data.lang']].copy()
tweets.rename(columns={'data.text': 'text', 'data.lang': 'language'}, inplace=True)
tweets
```

Out[29]:

|     | text                                              | language |
|-----|---------------------------------------------------|----------|
| 0   | I take 0 action without strategy                  | en       |
| 1   | @GuidoDisalle gm Guido!                           | en       |
| 2   | @SleuthieGoosie Absolutely agree with you. Not... | en       |
| 3   | @ThisIsKyleR @elonmusk Yes, life must be soooo... | en       |
| 4   | @funstonpaleo So looking forward to seeing the... | en       |
| ... | ...                                               | ...      |
| 936 | Cry https://t.co/i0yoZW9CUp                       | und      |
| 937 | @TheRealKean Who's the journalist?                | en       |
| 938 | Let's! https://t.co/LLkOotHZPq                    | en       |
| 939 | @MsPrinBoo @Justin_Ling I second this, please 🍌   | en       |
| 940 | @s_thompsett @Apollol69 @patdubois @TheoFleury... | en       |

941 rows × 2 columns

```
In [36]: c1 = tweets['text'].str.contains('@elonmusk')
c2 = tweets['text'].str.contains(' elon ')
c3 = tweets['text'].str.contains(' musk ')
elon = tweets[c1|c2|c3].copy().reset_index()
elon
```

Out[36]:

|    | index | text                                               | language |
|----|-------|----------------------------------------------------|----------|
| 0  | 3     | @ThisIsKyleR @elonmusk Yes, life must be soooo...  | en       |
| 1  | 52    | @Liz_Wheeler @elonmusk Everyone talking about ...  | en       |
| 2  | 152   | @FermisSocks @DJSnM @ctilley79 @elonmusk Umm.....  | en       |
| 3  | 165   | @WallStreetSilv @tim_cook @Apple @elonmusk Or ...  | en       |
| 4  | 318   | When the government tries to tell you what is o... | en       |
| 5  | 339   | @elonmusk So does Google unfortunately althoug...  | en       |
| 6  | 392   | @MovesTurtle @FlatEarthZone Or musk said he'd ...  | en       |
| 7  | 431   | @FLVoiceNews @elonmusk Nobody listens to you,W...  | en       |
| 8  | 461   | @DeificDragon @elonmusk My shocked face:           | en       |
| 9  | 600   | @elonmusk Whatever people may say good or bad ...  | en       |
| 10 | 654   | @ColeRWeav @G2StanAccount @realzedekiah04 @Cas...  | en       |
| 11 | 771   | @MasonTron9000 elon took away my little text s...  | en       |
| 12 | 838   | @farzyness @elonmusk Lol, :).\n\nhttps://t.co/...  | und      |
| 13 | 841   | @willcain @mtaibbi @elonmusk Lol, you mean thi...  | en       |
| 14 | 846   | @ClayTravis @elonmusk @Twitter Lol, :).\n\nhtt...  | und      |
| 15 | 847   | @elonmusk @hodgetwins @micsolana It wasn't und...  | en       |

```
In [37]: len(elon)/len(tweets)
```

Out[37]: 0.01700318809776833

```
In [43]: cond1 = tweets['text'].str.contains('covid')
cond2 = tweets['text'].str.contains('mask')

covid = tweets[cond1|cond2].copy().reset_index()

covid
```

Out[43]:

|   | index | text                                              | language |
|---|-------|---------------------------------------------------|----------|
| 0 | 190   | @aaronhoyland The Chinese get scooped up in ma... | en       |

# Test of zero-shot classification algorithm

## Preparation of model

```
In [44]: from transformers import pipeline
classifier = pipeline("zero-shot-classification", model="MoritzLaurer/mDeBERTa-v3-base-mnli-xnli")
```

C:\Users\Noura\anaconda3\envs\SDSPProject\lib\site-packages\transformers\convert\_slow\_tokenizer.py:446: UserWarning: The sentencepiece tokenizer that you are converting to a fast tokenizer uses the byte fallback option which is not implemented in the fast tokenizers. In practice this means that the fast version of the tokenizer can produce unknown tokens whereas the sentencepiece version would have converted these unknown tokens into a sequence of byte tokens matching the original piece of text. warnings.warn(

```
In [40]: tags = ["elon musk", "illness", "neutral"]
test = tweets["text"].copy().head(5).tolist()
output = classifier(test, tags, multi_label=False)
print(output)
```

[{'sequence': 'I take 0 action without strategy', 'labels': ['neutral', 'elon musk', 'illness'], 'scores': [0.8391062021255493, 0.10489579290151596, 0.055997978895902634]}, {'sequence': '@GuidoDisalle gm Guido!', 'labels': ['neutral', 'illness', 'elon musk'], 'scores': [0.4901840388774872, 0.31519070267677307, 0.19462524354457855]}, {'sequence': '@SleuthieGoosie Absolutely agree with you. Not normal.', 'labels': ['neutral', 'illness', 'elon musk'], 'scores': [0.7111625075340271, 0.20479916036128998, 0.08403828740119934]}, {'sequence': '@ThisIsKyleR @elonmusk Yes, life must be soooooo hard for you. Being able to murder whoever you want isn't enough,... https://t.co/Mzp3Li38s0', 'labels': ['elon musk', 'illness', 'neutral'], 'scores': [0.935019850730896, 0.032652948051691055, 0.032327212393283844]}, {'sequence': '@funstonpaleo So looking forward to seeing the results that come out of this!! https://t.co/o9Esvh5PXw', 'labels': ['illness', 'neutral', 'elon musk'], 'scores': [0.5565128326416016, 0.3380279541015625, 0.10545921325683594]}]

```
In [41]: raw_result = pd.DataFrame.from_dict(output)
raw_result
```

Out[41]:

|   | sequence                                          | labels                        | scores                                            |
|---|---------------------------------------------------|-------------------------------|---------------------------------------------------|
| 0 | I take 0 action without strategy                  | [neutral, elon musk, illness] | [0.8391062021255493, 0.10489579290151596, 0.05... |
| 1 | @GuidoDisalle gm Guido!                           | [neutral, illness, elon musk] | [0.4901840388774872, 0.31519070267677307, 0.19... |
| 2 | @SleuthieGoosie Absolutely agree with you. Not... | [neutral, illness, elon musk] | [0.7111625075340271, 0.20479916036128998, 0.08... |
| 3 | @ThisIsKyleR @elonmusk Yes, life must be soooo... | [elon musk, illness, neutral] | [0.935019850730896, 0.032652948051691055, 0.03... |
| 4 | @funstonpaleo So looking forward to seeing the... | [illness, neutral, elon musk] | [0.5565128326416016, 0.3380279541015625, 0.105... |