# 01_AM_Transform_Covid_Tweets

July 13, 2021

```python
[1]: import pandas as pd
     import numpy as np
     import reverse_geocode
```

## 1 Transform Covid tweets

- add country based on coordinates
- store as csv

```python
[2]: hydrated_json = pd.read_json("../data/raw/outdated/Hydrated_Tweets.jsonl",␣
     ↪lines=True)
```

Drop faulty rows (i.e. rows without coordinates):

```python
[3]: hydrated_json = hydrated_json.drop(hydrated_json[hydrated_json['coordinates'].
     ↪isnull()].index).reset_index(drop=True)
```

Add country name based on coordinates:

```python
[4]: hydrated_json["country"] = ""
     for i in range(len(hydrated_json)):
         hydrated_json["country"][i] = reverse_geocode.
     ↪search(tuple([hydrated_json["geo"][i]["coordinates"],(1,1)]))[0]['country']␣
     ↪#second tuple is to avoid error
```

```
<ipython-input-4-8763949fbb50>:5: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  hydrated_json["country"][i] = reverse_geocode.search(tuple([hydrated_json["geo
"][i]["coordinates"],(1,1)]))[0]['country'] #second tuple is to avoid error
```

```python
[5]: len(hydrated_json["country"].unique())
```

```
[5]: 204
```

Export dataset as CSV:

```python
[6]: hydrated_json.to_csv("../data/interim/Hydrated_Tweets_with_countries.csv")
```