In [ ]:
```
# Installing covidcast
!pip install covidcast
```

In [ ]:
```
from datetime import date
import covidcast
```

For the ground-truth number of daily Covid cases (labels), use the confirmed incidence num signal from the Indicator Combination source:

In [88]:
```
#By CA county, daily freq
sDate = date(2020, 5, 1)
eDate = date(2021, 10, 31)
counties = ['0'+str(x) for x in range(6001,6120)]
labels = covidcast.signal('indicator-combination', 'confirmed_incidence_num',
sDate, eDate, geo_type = 'county', geo_values=counties)
```

In [89]:
```
labels.set_index(['geo_value','time_value'], inplace = True)
labels['label'] = labels['value']
```

In [ ]:
```
#By CA county, daily freq

features = [('google-symptoms', 'anosmia_raw_search'),('google-symptoms', 'age
usia_raw_search'),
            ('hospital-admissions', 'smoothed_covid19_from_claims'),('doctor-v
isits','smoothed_cli'),
            ('chng','smoothed_outpatient_covid')]

for i,f in enumerate(features):
  print(f)
  globals()[f'feature{i}'] = covidcast.signal(f[0], f[1], sDate, eDate, geo_ty
pe = 'county', geo_values=counties).set_index(['geo_value','time_value'])[['si
gnal','value']]
```

In [103]:
```
import pandas as pd
featureset = [globals()[f'feature{i}'] for i in range(5)]
featureset += [pd.DataFrame(labels['label'])]
df = pd.concat(featureset, axis=1, join="outer")

a = list(df.columns)
for i,x in enumerate(df.iloc[1,]):
  if '_' in str(x):
    a[i+1] = x
df.columns = a

df.drop('signal',axis=1, inplace=True)
df.fillna(method="ffill",inplace=True)

# data not available for all counties and times, so the NAs are replaced using
the forward-fill methodology (default to day before)
df.to_csv('TrainingData.csv')
```