Project Part 2

Names: Annaliese Adams, Anum Damani, Andrew Bissell, Eric Cha

In [1]:
```
# Installing covidcast
!pip install covidcast
```

```
Collecting covidcast
  Downloading covidcast-0.1.5-py3-none-any.whl (12.3 MB)
     |████████████████████████████████| 12.3 MB 20.9 MB/s
Collecting epiweeks
  Downloading epiweeks-2.1.3-py3-none-any.whl (5.9 kB)
Requirement already satisfied: pandas in /shared-libs/python3.7/py/lib/python3
.7/site-packages (from covidcast) (1.2.5)
Collecting descartes
  Downloading descartes-1.1.0-py3-none-any.whl (5.8 kB)
Requirement already satisfied: tqdm in /shared-libs/python3.7/py/lib/python3.7
/site-packages (from covidcast) (4.62.3)
Collecting delphi-epidata>=0.0.11
  Downloading delphi_epidata-0.3.1-py3-none-any.whl (6.8 kB)
Collecting imageio
  Downloading imageio-2.13.0-py3-none-any.whl (3.3 MB)
     |████████████████████████████████| 3.3 MB 37.3 MB/s
Requirement already satisfied: numpy in /shared-libs/python3.7/py/lib/python3.
7/site-packages (from covidcast) (1.19.5)
Requirement already satisfied: matplotlib in /shared-libs/python3.7/py/lib/pyt
hon3.7/site-packages (from covidcast) (3.4.3)
Requirement already satisfied: requests in /shared-libs/python3.7/py/lib/pytho
n3.7/site-packages (from covidcast) (2.26.0)
Collecting geopandas
  Downloading geopandas-0.10.2-py2.py3-none-any.whl (1.0 MB)
     |████████████████████████████████| 1.0 MB 38.8 MB/s
Collecting imageio-ffmpeg
  Downloading imageio_ffmpeg-0.4.5-py3-none-manylinux2010_x86_64.whl (26.9 MB)
     |████████████████████████████████| 26.9 MB 42.0 MB/s
Requirement already satisfied: python-dateutil>=2.7.3 in /shared-libs/python3.
7/py-core/lib/python3.7/site-packages (from pandas->covidcast) (2.8.2)
Requirement already satisfied: pytz>=2017.3 in /shared-libs/python3.7/py/lib/p
ython3.7/site-packages (from pandas->covidcast) (2021.3)
Requirement already satisfied: tenacity in /shared-libs/python3.7/py/lib/pytho
n3.7/site-packages (from delphi-epidata>=0.0.11->covidcast) (8.0.1)
Requirement already satisfied: aiohttp in /shared-libs/python3.7/py-core/lib/p
ython3.7/site-packages (from delphi-epidata>=0.0.11->covidcast) (3.8.0)
Requirement already satisfied: pillow>=8.3.2 in /shared-libs/python3.7/py/lib/
python3.7/site-packages (from imageio->covidcast) (8.4.0)
Requirement already satisfied: kiwisolver>=1.0.1 in /shared-libs/python3.7/py/
lib/python3.7/site-packages (from matplotlib->covidcast) (1.3.2)
Requirement already satisfied: cycler>=0.10 in /shared-libs/python3.7/py/lib/p
ython3.7/site-packages (from matplotlib->covidcast) (0.11.0)
Requirement already satisfied: pyparsing>=2.2.1 in /shared-libs/python3.7/py-c
ore/lib/python3.7/site-packages (from matplotlib->covidcast) (2.4.7)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /shared-libs/python3.7
/py/lib/python3.7/site-packages (from requests->covidcast) (1.26.7)
Requirement already satisfied: idna<4,>=2.5; python_version >= "3" in /shared-
```

```
libs/python3.7/py-core/lib/python3.7/site-packages (from requests->covidcast)
(3.3)
Requirement already satisfied: charset-normalizer~=2.0.0; python_version >= "3
" in /shared-libs/python3.7/py-core/lib/python3.7/site-packages (from requests
->covidcast) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /shared-libs/python3.7/py
/lib/python3.7/site-packages (from requests->covidcast) (2021.10.8)
Collecting shapely>=1.6
  Downloading Shapely-1.8.0-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86_64.
whl (1.1 MB)
     |████████████████████████████████| 1.1 MB 31.5 MB/s
Collecting pyproj>=2.2.0
  Downloading pyproj-3.2.1-cp37-cp37m-manylinux2010_x86_64.whl (6.3 MB)
     |████████████████████████████████| 6.3 MB 35.3 MB/s
Requirement already satisfied: fiona>=1.8 in /shared-libs/python3.7/py/lib/pyt
hon3.7/site-packages (from geopandas->covidcast) (1.8.20)
Requirement already satisfied: six>=1.5 in /shared-libs/python3.7/py-core/lib/
python3.7/site-packages (from python-dateutil>=2.7.3->pandas->covidcast) (1.16
.0)
Requirement already satisfied: aiosignal>=1.1.2 in /shared-libs/python3.7/py-c
ore/lib/python3.7/site-packages (from aiohttp->delphi-epidata>=0.0.11->covidca
st) (1.2.0)
Requirement already satisfied: yarl<2.0,>=1.0 in /shared-libs/python3.7/py-cor
e/lib/python3.7/site-packages (from aiohttp->delphi-epidata>=0.0.11->covidcast
) (1.7.2)
Requirement already satisfied: frozenlist>=1.1.1 in /shared-libs/python3.7/py-
core/lib/python3.7/site-packages (from aiohttp->delphi-epidata>=0.0.11->covidc
ast) (1.2.0)
Requirement already satisfied: typing-extensions>=3.7.4; python_version < "3.8
" in /shared-libs/python3.7/py-core/lib/python3.7/site-packages (from aiohttp-
>delphi-epidata>=0.0.11->covidcast) (3.10.0.2)
Requirement already satisfied: multidict<7.0,>=4.5 in /shared-libs/python3.7/p
y-core/lib/python3.7/site-packages (from aiohttp->delphi-epidata>=0.0.11->covi
dcast) (5.2.0)
Requirement already satisfied: asynctest==0.13.0; python_version < "3.8" in /s
hared-libs/python3.7/py-core/lib/python3.7/site-packages (from aiohttp->delphi
-epidata>=0.0.11->covidcast) (0.13.0)
Requirement already satisfied: attrs>=17.3.0 in /shared-libs/python3.7/py-core
/lib/python3.7/site-packages (from aiohttp->delphi-epidata>=0.0.11->covidcast)
(21.2.0)
Requirement already satisfied: async-timeout<5.0,>=4.0.0a3 in /shared-libs/pyt
hon3.7/py-core/lib/python3.7/site-packages (from aiohttp->delphi-epidata>=0.0.
11->covidcast) (4.0.1)
Requirement already satisfied: setuptools in /root/venv/lib/python3.7/site-pac
kages (from fiona>=1.8->geopandas->covidcast) (47.1.0)
Requirement already satisfied: cligj>=0.5 in /shared-libs/python3.7/py/lib/pyt
hon3.7/site-packages (from fiona>=1.8->geopandas->covidcast) (0.7.2)
Requirement already satisfied: click>=4.0 in /shared-libs/python3.7/py/lib/pyt
hon3.7/site-packages (from fiona>=1.8->geopandas->covidcast) (8.0.3)
Requirement already satisfied: munch in /shared-libs/python3.7/py/lib/python3.
7/site-packages (from fiona>=1.8->geopandas->covidcast) (2.5.0)
Requirement already satisfied: click-plugins>=1.0 in /shared-libs/python3.7/py
/lib/python3.7/site-packages (from fiona>=1.8->geopandas->covidcast) (1.1.1)
Requirement already satisfied: importlib-metadata; python_version < "3.8" in /
shared-libs/python3.7/py-core/lib/python3.7/site-packages (from click>=4.0->fi
ona>=1.8->geopandas->covidcast) (4.8.2)
Requirement already satisfied: zipp>=0.5 in /shared-libs/python3.7/py-core/lib
```

```
/python3.7/site-packages (from importlib-metadata; python_version < "3.8"->cli
ck>=4.0->fiona>=1.8->geopandas->covidcast) (3.6.0)
Installing collected packages: epiweeks, descartes, delphi-epidata, imageio, s
hapely, pyproj, geopandas, imageio-ffmpeg, covidcast
Successfully installed covidcast-0.1.5 delphi-epidata-0.3.1 descartes-1.1.0 ep
iweeks-2.1.3 geopandas-0.10.2 imageio-2.13.0 imageio-ffmpeg-0.4.5 pyproj-3.2.1
shapely-1.8.0
WARNING: You are using pip version 20.1.1; however, version 21.3.1 is availabl
e.
You should consider upgrading via the '/root/venv/bin/python -m pip install --
upgrade pip' command.
```

In [2]:
```python
from datetime import date
import covidcast
```

In [3]:
```python
[ca_counties] = covidcast.fips_to_name("^06.*", ties_method="all")
ca_counties
```

Out[3]:
```
{'06000': ['California'],
 '06001': ['Alameda County'],
 '06003': ['Alpine County'],
 '06005': ['Amador County'],
 '06007': ['Butte County'],
 '06009': ['Calaveras County'],
 '06011': ['Colusa County'],
 '06013': ['Contra Costa County'],
 '06015': ['Del Norte County'],
 '06017': ['El Dorado County'],
 '06019': ['Fresno County'],
 '06021': ['Glenn County'],
 '06023': ['Humboldt County'],
 '06025': ['Imperial County'],
 '06027': ['Inyo County'],
 '06029': ['Kern County'],
 '06031': ['Kings County'],
 '06033': ['Lake County'],
 '06035': ['Lassen County'],
 '06037': ['Los Angeles County'],
 '06039': ['Madera County'],
 '06041': ['Marin County'],
 '06043': ['Mariposa County'],
 '06045': ['Mendocino County'],
 '06047': ['Merced County'],
 '06049': ['Modoc County'],
 '06051': ['Mono County'],
 '06053': ['Monterey County'],
 '06055': ['Napa County'],
 '06057': ['Nevada County'],
 '06059': ['Orange County'],
 '06061': ['Placer County'],
 '06063': ['Plumas County'],
 '06065': ['Riverside County'],
 '06067': ['Sacramento County'],
 '06069': ['San Benito County'],
 '06071': ['San Bernardino County'],
 '06073': ['San Diego County'],
```

```
             '06075': ['San Francisco County'],
             '06077': ['San Joaquin County'],
             '06079': ['San Luis Obispo County'],
             '06081': ['San Mateo County'],
             '06083': ['Santa Barbara County'],
             '06085': ['Santa Clara County'],
             '06087': ['Santa Cruz County'],
             '06089': ['Shasta County'],
             '06091': ['Sierra County'],
             '06093': ['Siskiyou County'],
             '06095': ['Solano County'],
             '06097': ['Sonoma County'],
             '06099': ['Stanislaus County'],
             '06101': ['Sutter County'],
             '06103': ['Tehama County'],
             '06105': ['Trinity County'],
             '06107': ['Tulare County'],
             '06109': ['Tuolumne County'],
             '06111': ['Ventura County'],
             '06113': ['Yolo County'],
             '06115': ['Yuba County']}
```

In [4]:
```python
len(ca_counties)
```

Out[4]: 59

In [5]:
```python
#fetch the "US Facts Cases and Deaths" data source
#Ground Truth

#2/13 - 8/13
covidCases = covidcast.signal("usa-facts", "confirmed_incidence_num", date(20
covidCases
```

Out[5]:

	geo_value	signal	time_value	issue	lag	missing_value	missing_stder
0	06037	confirmed_incidence_num	2020-02-13	2021-09-16	581	0	!
1	06059	confirmed_incidence_num	2020-02-13	2021-09-16	581	0	!
2	06073	confirmed_incidence_num	2020-02-13	2021-09-16	581	0	!
3	06077	confirmed_incidence_num	2020-02-13	2021-09-16	581	0	!
4	06085	confirmed_incidence_num	2020-02-13	2021-09-16	581	0	!
...	...	...	...	...	...	...	..
54	06107	confirmed_incidence_num	2020-08-13	2021-09-16	399	0	!
55	06109	confirmed_incidence_num	2020-08-13	2021-09-16	399	0	!
56	06111	confirmed_incidence_num	2020-08-13	2021-09-16	399	0	!
57	06113	confirmed_incidence_num	2020-08-13	2021-09-16	399	0	!
58	06115	confirmed_incidence_num	2020-08-13	2021-09-16	399	0	!

10421 rows × 13 columns

In [6]:
```python
#Source: Change Healthcare
covidOutpatient = covidcast.signal("chng", "smoothed_outpatient_covid",date(2
covidOutpatient
```

Out[6]:

	geo_value	signal	time_value	issue	lag	missing_value	missing_stde
0	06000	smoothed_outpatient_covid	2020-02-13	2021-02-21	374	0	
1	06001	smoothed_outpatient_covid	2020-02-13	2021-02-21	374	0	
2	06005	smoothed_outpatient_covid	2020-02-13	2021-02-21	374	0	
3	06007	smoothed_outpatient_covid	2020-02-13	2021-02-21	374	0	
4	06009	smoothed_outpatient_covid	2020-02-13	2021-02-21	374	0	
...	...	...	...	...	...	...	
53	06107	smoothed_outpatient_covid	2020-08-13	2021-02-21	192	0	
54	06109	smoothed_outpatient_covid	2020-08-13	2021-02-21	192	0	
55	06111	smoothed_outpatient_covid	2020-08-13	2021-02-21	192	0	
56	06113	smoothed_outpatient_covid	2020-08-13	2021-02-21	192	0	
57	06115	smoothed_outpatient_covid	2020-08-13	2021-02-21	192	0	

10592 rows × 13 columns

In [7]:
```python
symptomOutpatient = covidcast.signal("chng", "smoothed_outpatient_cli",date(2
symptomOutpatient
```

Out[7]:

	geo_value	signal	time_value	issue	lag	missing_value	missing_stderr
0	06000	smoothed_outpatient_cli	2020-02-13	2021-02-21	374	0	5
1	06001	smoothed_outpatient_cli	2020-02-13	2021-02-21	374	0	5
2	06005	smoothed_outpatient_cli	2020-02-13	2021-02-21	374	0	5
3	06007	smoothed_outpatient_cli	2020-02-13	2021-02-21	374	0	5
4	06009	smoothed_outpatient_cli	2020-02-13	2021-02-21	374	0	5
...	...	...	...	...	...	...	...
53	06107	smoothed_outpatient_cli	2020-08-13	2021-02-21	192	0	5
54	06109	smoothed_outpatient_cli	2020-08-13	2021-02-21	192	0	5
55	06111	smoothed_outpatient_cli	2020-08-13	2021-02-21	192	0	5
56	06113	smoothed_outpatient_cli	2020-08-13	2021-02-21	192	0	5
57	06115	smoothed_outpatient_cli	2020-08-13	2021-02-21	192	0	5

10592 rows × 13 columns

In [8]:
```python
adjustedOutpatient = covidcast.signal("chng", "smoothed_adj_outpatient_covid"
adjustedOutpatient
```

Out[8]:

	geo_value	signal	time_value	issue	lag	missing_value	missing_
0	06000	smoothed_adj_outpatient_covid	2020-02-13	2021-02-21	374	0	
1	06001	smoothed_adj_outpatient_covid	2020-02-13	2021-02-21	374	0	
2	06005	smoothed_adj_outpatient_covid	2020-02-13	2021-02-21	374	0	
3	06007	smoothed_adj_outpatient_covid	2020-02-13	2021-02-21	374	0	
4	06009	smoothed_adj_outpatient_covid	2020-02-13	2021-02-21	374	0	
...	...	...	...	...	...	...	
53	06107	smoothed_adj_outpatient_covid	2020-08-13	2021-02-21	192	0	
54	06109	smoothed_adj_outpatient_covid	2020-08-13	2021-02-21	192	0	
55	06111	smoothed_adj_outpatient_covid	2020-08-13	2021-02-21	192	0	
56	06113	smoothed_adj_outpatient_covid	2020-08-13	2021-02-21	192	0	
57	06115	smoothed_adj_outpatient_covid	2020-08-13	2021-02-21	192	0	

10592 rows × 13 columns

In [9]:
```python
#Doctor Visits
#Signals: smoothed_cli

#February 1st
covidSymptomDoctorVisits = covidcast.signal("doctor-visits", "smoothed_cli",d
covidSymptomDoctorVisits
```

Out[9]:

	geo_value	signal	time_value	issue	lag	missing_value	missing_stderr	missing_s
0	06000	smoothed_cli	2020-02-13	2020-06-09	117	0	5	
1	06001	smoothed_cli	2020-02-13	2020-06-09	117	0	5	
2	06005	smoothed_cli	2020-02-13	2020-06-09	117	0	5	
3	06007	smoothed_cli	2020-02-13	2020-06-09	117	0	5	
4	06009	smoothed_cli	2020-02-13	2020-06-09	117	0	5	
...	...	...	...	...	...	...	...	
42	06107	smoothed_cli	2020-08-13	2020-10-15	63	0	5	
43	06109	smoothed_cli	2020-08-13	2020-10-15	63	0	5	
44	06111	smoothed_cli	2020-08-13	2020-10-15	63	0	5	
45	06113	smoothed_cli	2020-08-13	2020-10-15	63	0	5	
46	06115	smoothed_cli	2020-08-13	2020-10-15	63	0	5	

8177 rows × 13 columns

In [10]:
```python
#Hospital Admissions
#Signals: smoothed_covid19

#Febuary 1st
hospitalAdmissions = covidcast.signal("hospital-admissions", "smoothed_covid1
hospitalAdmissions
```

Out[10]:

	geo_value	signal	time_value	issue	lag	missing_value	missing_stderr	miss
0	06000	smoothed_covid19	2020-02-13	2020-06-30	138	0	5	
1	06001	smoothed_covid19	2020-02-13	2020-06-30	138	0	5	
2	06013	smoothed_covid19	2020-02-13	2020-06-30	138	0	5	
3	06019	smoothed_covid19	2020-02-13	2020-06-30	138	0	5	
4	06029	smoothed_covid19	2020-02-13	2020-06-30	138	0	5	
...	...	...	...	...	...	...	...	
20	06095	smoothed_covid19	2020-08-13	2020-09-30	48	0	5	
21	06097	smoothed_covid19	2020-08-13	2020-09-30	48	0	5	
22	06099	smoothed_covid19	2020-08-13	2020-09-30	48	0	5	
23	06107	smoothed_covid19	2020-08-13	2020-09-30	48	0	5	
24	06111	smoothed_covid19	2020-08-13	2020-09-30	48	0	5	

4464 rows × 13 columns

In [11]:
```python
#Merge Data

merged_covid_data = covidcast.aggregate_signals([covidCases, covidOutpatient,
merged_covid_data
```

Out[11]:

	geo_value	time_value	usa-facts_confirmed_incidence_num_0_issue	usa-facts_confirmed_incide...
0	06000	2020-02-13	NaT	
1	06001	2020-02-13	NaT	
2	06005	2020-02-13	NaT	
3	06007	2020-02-13	NaT	
4	06009	2020-02-13	NaT	
...	...	...	...	
10785	06107	2020-08-13	2021-09-16	
10786	06109	2020-08-13	2021-09-16	
10787	06111	2020-08-13	2021-09-16	
10788	06113	2020-08-13	2021-09-16	
10789	06115	2020-08-13	2021-09-16	

10790 rows × 51 columns

In [12]:
```python
#Determine Null Columns
merged_covid_data[merged_covid_data.isnull().any(axis = 1)]

#Columns with All Missing: stderr and sample_size (all null values):

#usa-facts_confirmed_incidence_num_0_stderr
#usa-facts_confirmed_incidence_num_0_sample_size
#chng_smoothed_outpatient_covid_1_stderr
#chng_smoothed_outpatient_covid_1_sample_size
#chng_smoothed_outpatient_cli_2_stderr
#chng_smoothed_outpatient_cli_2_sample_size
#chng_smoothed_adj_outpatient_covid_3_stderr
#chng_smoothed_adj_outpatient_covid_3_sample_size
#doctor-visits_smoothed_cli_4_stderr
#doctor-visits_smoothed_cli_4_sample_size
#hospital-admissions_smoothed_covid19_5_stderr
#hospital-admissions_smoothed_covid19_5_sample_size


#issue columns (when data was uploaded/updated on the GitHub site):

#usa-facts_confirmed_incidence_num_0_issue
#chng_smoothed_outpatient_covid_1_issue
#chng_smoothed_outpatient_cli_2_issue
#chng_smoothed_adj_outpatient_covid_3_issue
#doctor-visits_smoothed_cli_4_issue
#hospital-admissions_smoothed_covid19_5_issue


#Columns to Keep:

#Make new df with desired columns
new_merged = merged_covid_data[["geo_value", "time_value", "usa-facts_confirm
"chng_smoothed_adj_outpatient_covid_3_value", "doctor-visits_smoothed_cli_4_v
#rename
new_merged = new_merged.rename(columns = {"usa-facts_confirmed_incidence_num_
 "chng_smoothed_adj_outpatient_covid_3_value": "(ADJ) % Confirmed COVID Docto

#print
new_merged
```

Out[12]:

	geo_value	time_value	Ground Truth Cases	% Confirmed COVID Doctor Visits	(1) % Visits w/ Symptoms	(ADJ) % Confirmed COVID Doctor Visits	(2) % Visits w/ Symptoms	% N Vis COV associa
0	06000	2020-02-13	NaN	0.203214	0.791640	0.203214	0.000000	0.087∠
1	06001	2020-02-13	NaN	0.014638	0.057907	0.015211	0.000000	0.269€
2	06005	2020-02-13	NaN	0.167716	0.167716	0.167716	0.000000	N
3	06007	2020-02-13	NaN	0.050255	0.071938	0.050255	0.000000	N
4	06009	2020-02-13	NaN	0.330854	2.822597	0.330854	0.000000	N
...	...	...	...	...	...	...	...	
10785	06107	2020-08-13	126.0	1.021952	3.792120	1.164212	8.723067	3.351
10786	06109	2020-08-13	4.0	0.274820	1.157530	0.302357	1.003629	N
10787	06111	2020-08-13	68.0	0.195529	3.534633	0.192925	2.613487	6.753ٍ
10788	06113	2020-08-13	22.0	0.343868	3.893518	0.378412	2.185835	N
10789	06115	2020-08-13	11.0	0.449591	9.168948	0.533066	2.207103	N

10790 rows × 8 columns

In [25]:

```
new_merged
```

Out[25]:

	geo_value	time_value	Ground Truth Cases	% Confirmed COVID Doctor Visits	(1) % Visits w/ Symptoms	(ADJ) % Confirmed COVID Doctor Visits	(2) % Visits w/ Symptoms	% N Vis COV associa
0	06000	2020-02-13	NaN	0.203214	0.791640	0.203214	0.000000	0.087
1	06001	2020-02-13	NaN	0.014638	0.057907	0.015211	0.000000	0.2696
2	06005	2020-02-13	NaN	0.167716	0.167716	0.167716	0.000000	N
3	06007	2020-02-13	NaN	0.050255	0.071938	0.050255	0.000000	N
4	06009	2020-02-13	NaN	0.330854	2.822597	0.330854	0.000000	N
...	...	...	...	...	...	...	...	...
10785	06107	2020-08-13	126.0	1.021952	3.792120	1.164212	8.723067	3.351
10786	06109	2020-08-13	4.0	0.274820	1.157530	0.302357	1.003629	N
10787	06111	2020-08-13	68.0	0.195529	3.534633	0.192925	2.613487	6.753
10788	06113	2020-08-13	22.0	0.343868	3.893518	0.378412	2.185835	N
10789	06115	2020-08-13	11.0	0.449591	9.168948	0.533066	2.207103	N

10790 rows × 8 columns

In [40]:

```python
#Impute Missing Data
import numpy as np
from sklearn.impute import SimpleImputer
new_merged2=new_merged.iloc[:, 3:8]
new_merged2
```

Out[40]:

	% Confirmed COVID Doctor Visits	(1) % Visits w/ Symptoms	(ADJ) % Confirmed COVID Doctor Visits	(2) % Visits w/ Symptoms	% NEW Visits COVID-associated
0	0.203214	0.791640	0.203214	0.000000	0.087496
1	0.014638	0.057907	0.015211	0.000000	0.269634
2	0.167716	0.167716	0.167716	0.000000	NaN
3	0.050255	0.071938	0.050255	0.000000	NaN
4	0.330854	2.822597	0.330854	0.000000	NaN
...	...	...	...	...	...
10785	1.021952	3.792120	1.164212	8.723067	3.351107
10786	0.274820	1.157530	0.302357	1.003629	NaN
10787	0.195529	3.534633	0.192925	2.613487	6.753834
10788	0.343868	3.893518	0.378412	2.185835	NaN
10789	0.449591	9.168948	0.533066	2.207103	NaN

10790 rows × 5 columns

In [42]:
```python
import pandas as pd
imp = SimpleImputer(missing_values=np.nan, strategy='mean')
imp=imp.fit(new_merged2)
imputed_df=imp.transform(new_merged2.values)
imputed_df = pd.DataFrame(imputed_df)
print(imputed_df)
```

```
              0         1         2         3         4
0      0.203214  0.791640  0.203214  0.000000  0.087496
1      0.014638  0.057907  0.015211  0.000000  0.269634
2      0.167716  0.167716  0.167716  0.000000  2.923763
3      0.050255  0.071938  0.050255  0.000000  2.923763
4      0.330854  2.822597  0.330854  0.000000  2.923763
...         ...       ...       ...       ...       ...
10785  1.021952  3.792120  1.164212  8.723067  3.351107
10786  0.274820  1.157530  0.302357  1.003629  2.923763
10787  0.195529  3.534633  0.192925  2.613487  6.753834
10788  0.343868  3.893518  0.378412  2.185835  2.923763
10789  0.449591  9.168948  0.533066  2.207103  2.923763

[10790 rows x 5 columns]
/shared-libs/python3.7/py/lib/python3.7/site-packages/sklearn/base.py:446: Use
rWarning: X does not have valid feature names, but SimpleImputer was fitted wi
th feature names
  "X does not have valid feature names, but"
```

In [43]:
```python
new_merged3=new_merged.iloc[:,0:2]
new_merged3
```

Out[43]:

	geo_value	time_value
0	06000	2020-02-13
1	06001	2020-02-13
2	06005	2020-02-13
3	06007	2020-02-13
4	06009	2020-02-13
...	...	...
10785	06107	2020-08-13
10786	06109	2020-08-13
10787	06111	2020-08-13
10788	06113	2020-08-13
10789	06115	2020-08-13

10790 rows × 2 columns

In [45]:
```python
final_merge=[imputed_df, new_merged3]
covid_data=pd.concat(final_merge, axis = 1)
covid_data
```

Out[45]:

	0	1	2	3	4	geo_value	time_value
0	0.203214	0.791640	0.203214	0.000000	0.087496	06000	2020-02-13
1	0.014638	0.057907	0.015211	0.000000	0.269634	06001	2020-02-13
2	0.167716	0.167716	0.167716	0.000000	2.923763	06005	2020-02-13
3	0.050255	0.071938	0.050255	0.000000	2.923763	06007	2020-02-13
4	0.330854	2.822597	0.330854	0.000000	2.923763	06009	2020-02-13
...	...	...	...	...	...	...	...
10785	1.021952	3.792120	1.164212	8.723067	3.351107	06107	2020-08-13
10786	0.274820	1.157530	0.302357	1.003629	2.923763	06109	2020-08-13
10787	0.195529	3.534633	0.192925	2.613487	6.753834	06111	2020-08-13
10788	0.343868	3.893518	0.378412	2.185835	2.923763	06113	2020-08-13
10789	0.449591	9.168948	0.533066	2.207103	2.923763	06115	2020-08-13

10790 rows × 7 columns

In [49]:
```python
covid_data.isnull().sum()
```

Out[49]:

```
0               0
1               0
2               0
3               0
4               0
geo_value       0
time_value      0
dtype: int64
```

Created in **Deepnote**