

Ben Kahan
DS210
Homework 5
Collaborators: None

Dataset 1

For my first data set, I chose a set called `us-counties` provided by The New York Times. This dataset contains US COVID-19 cases by county throughout the country.

One of the important features I was looking for was a larger set with time-series data. In addition, this dataset is relevant and it would be interesting to look at COVID trends paired with another dataset, `president_county_candidate`. I would like to see if there is any correlation between counties that voted REP versus DEM versus OTHER and their COVID cases.

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

```
In [10]: covid = pd.read_csv('us-counties.csv')  
pres = pd.read_csv('president_county_candidate.csv')
```

```
In [8]: covid.head()
```

Out [8]:

	date	county	state	fips	cases	deaths
0	2022-10-21	Carroll	New Hampshire	33003.0	11568	68.0
1	2022-10-21	St. Joseph	Michigan	26149.0	15421	223.0
2	2022-10-21	Cheshire	New Hampshire	33005.0	17887	125.0
3	2022-10-21	Coos	New Hampshire	33007.0	8007	124.0
4	2022-10-21	Grafton	New Hampshire	33009.0	22040	82.0

In [9]: covid.describe()

Out [9]:

	fips	cases	deaths
count	3217.000000	3.257000e+03	3179.000000
mean	31501.370532	2.975667e+04	334.327147
std	16368.431244	1.118776e+05	1259.674228
min	1001.000000	0.000000e+00	0.000000
25%	19039.000000	3.002000e+03	44.000000
50%	30029.000000	7.553000e+03	105.000000
75%	46121.000000	1.961800e+04	249.000000
max	78030.000000	3.476634e+06	42238.000000

In [11]: pres.head()

Out [11]:

	state	county	candidate	party	total_votes	won
0	Delaware	Kent County	Joe Biden	DEM	44552	True
1	Delaware	Kent County	Donald Trump	REP	41009	False
2	Delaware	Kent County	Jo Jorgensen	LIB	1044	False
3	Delaware	Kent County	Howie Hawkins	GRN	420	False
4	Delaware	New Castle County	Joe Biden	DEM	195034	True

In [12]: `pres.describe()`

Out [12]:

	total_votes
count	3.217700e+04
mean	4.960321e+03
std	3.581577e+04
min	0.000000e+00
25%	3.000000e+00
50%	3.400000e+01
75%	7.450000e+02
max	3.028885e+06

Dataset 2

For my second dataset, I chose a set called `Top 100 Cryptocurrencies Historical Dataset`. I began

crypto trading in 2014 and this dataset contains the top altcoins. While it would be a challenge, I would like to use this dataset to identity the most viable altcoin(s) for the future.

In []: