DATA ASSIGNMENT

SUBMITTED BY: AAKASH PATEL

1. DATA IMPORT

In [1]:
```
# Installing required libs and importing them
!pip install pandas-profiling
!pip install ipywidgets
!pip install plotly

import requests
import wget
import pandas as pd
import json
from datetime import date, timedelta
from pandas_profiling import ProfileReport
import plotly.express as px
```

```
Requirement already satisfied: pandas-profiling in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (3.2.0)
Requirement already satisfied: requests>=2.24.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pandas-profiling) (2.27.1)
Requirement already satisfied: joblib~=1.1.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pandas-profiling) (1.1.0)
Requirement already satisfied: missingno>=0.4.2 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pandas-profiling) (0.5.1)
Requirement already satisfied: multimethod>=1.4 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pandas-profiling) (1.8)
Requirement already satisfied: jinja2>=2.11.1 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pandas-profiling) (3.0.3)
Requirement already satisfied: tangled-up-in-unicode==0.2.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pandas-profiling) (0.2.0)
Requirement already satisfied: htmlmin>=0.1.12 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pandas-profiling) (0.1.12)
Requirement already satisfied: seaborn>=0.10.1 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pandas-profiling) (0.11.2)
Requirement already satisfied: matplotlib>=3.2.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pandas-profiling) (3.5.2)
Requirement already satisfied: PyYAML>=5.0.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pandas-profiling) (6.0)
Requirement already satisfied: markupsafe~=2.1.1 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pandas-profiling) (2.1.1)
Requirement already satisfied: numpy>=1.16.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pandas-profiling) (1.22.3)
```

```
WARNING: There was an error checking the latest version of pip.
Requirement already satisfied: scipy>=1.4.1 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pandas-profiling) (1.8.1)
Requirement already satisfied: phik>=0.11.1 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pandas-profiling) (0.12.2)
Requirement already satisfied: pydantic>=1.8.1 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pandas-profiling) (1.9.1)
Requirement already satisfied: pandas!=1.0.0,!=1.0.1,!=1.0.2,!=1.1.0,>=0.25.3 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pandas-profiling) (1.4.1)
Requirement already satisfied: visions[type_image_path]==0.7.4 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pandas-profiling) (0.7.4)
Requirement already satisfied: tqdm>=4.48.2 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pandas-profiling) (4.64.0)
Requirement already satisfied: attrs>=19.3.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from visions[type_image_path]==0.7.4->pandas-profiling) (21.4.0)
Requirement already satisfied: networkx>=2.4 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from visions[type_image_path]==0.7.4->pandas-profiling) (2.8.3)
Requirement already satisfied: Pillow in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from visions[type_image_path]==0.7.4->pandas-profiling) (9.1.1)
Requirement already satisfied: imagehash in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from visions[type_image_path]==0.7.4->pandas-profiling) (4.2.1)
Requirement already satisfied: packaging>=20.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from matplotlib>=3.2.0->pandas-profiling) (21.3)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from matplotlib>=3.2.0->pandas-profiling) (3.0.7)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from matplotlib>=3.2.0->pandas-profiling) (2.8.2)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from matplotlib>=3.2.0->pandas-profiling) (1.4.2)
Requirement already satisfied: cycler>=0.10 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from matplotlib>=3.2.0->pandas-profiling) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from matplotlib>=3.2.0->pandas-profiling) (4.33.3)
Requirement already satisfied: pytz>=2020.1 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pandas!=1.0.0,!=1.0.1,!=1.0.2,!=1.1.0,>=0.25.3->pandas-profiling) (202
2.1)
Requirement already satisfied: typing-extensions>=3.7.4.3 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from pydantic>=1.8.1->pandas-profiling) (4.1.1)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from requests>=2.24.0->pandas-profiling) (2021.10.8)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from requests>=2.24.0->pandas-profiling) (1.26.9)
Requirement already satisfied: idna<4,>=2.5 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from requests>=2.24.0->pandas-profiling) (3.3)
Requirement already satisfied: charset-normalizer~=2.0.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from requests>=2.24.0->pandas-profiling) (2.0.12)
Requirement already satisfied: colorama in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from tqdm>=4.48.2->pandas-profiling) (0.4.4)
Requirement already satisfied: six>=1.5 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from python-dateutil>=2.7->matplotlib>=3.2.0->pandas-profiling) (1.16.0)
Requirement already satisfied: PyWavelets in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from imagehash->visions[type_image_path]==0.7.4->pandas-profiling) (1.3.0)
Requirement already satisfied: ipywidgets in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (7.7.0)
Requirement already satisfied: nbformat>=4.2.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipywidgets) (5.2.0)
```

```
WARNING: There was an error checking the latest version of pip.
Requirement already satisfied: widgetsnbextension~=3.6.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipywidgets) (3.6.0)
Requirement already satisfied: ipython>=4.0.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipywidgets) (8.1.1)
Requirement already satisfied: traitlets>=4.3.1 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipywidgets) (5.1.1)
Requirement already satisfied: ipython-genutils~=0.2.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipywidgets) (0.2.0)
Requirement already satisfied: ipykernel>=4.5.1 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipywidgets) (6.9.2)
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipywidgets) (1.1.0)
Requirement already satisfied: tornado<7.0,>=4.2 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipykernel>=4.5.1->ipywidgets) (6.1)
Requirement already satisfied: jupyter-client<8.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipykernel>=4.5.1->ipywidgets) (7.1.2)
Requirement already satisfied: matplotlib-inline<0.2.0,>=0.1.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipykernel>=4.5.1->ipywidgets) (0.1.3)
Requirement already satisfied: nest-asyncio in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipykernel>=4.5.1->ipywidgets) (1.5.4)
Requirement already satisfied: psutil in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipykernel>=4.5.1->ipywidgets) (5.9.0)
Requirement already satisfied: debugpy<2.0,>=1.0.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipykernel>=4.5.1->ipywidgets) (1.5.1)
Requirement already satisfied: colorama in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipython>=4.0.0->ipywidgets) (0.4.4)
Requirement already satisfied: jedi>=0.16 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipython>=4.0.0->ipywidgets) (0.18.1)
Requirement already satisfied: stack-data in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipython>=4.0.0->ipywidgets) (0.2.0)
Requirement already satisfied: backcall in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipython>=4.0.0->ipywidgets) (0.2.0)
Requirement already satisfied: decorator in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipython>=4.0.0->ipywidgets) (5.1.1)
Requirement already satisfied: pygments>=2.4.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipython>=4.0.0->ipywidgets) (2.11.2)
Requirement already satisfied: prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipython>=4.0.0->ipywidgets) (3.0.28)
Requirement already satisfied: pickleshare in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipython>=4.0.0->ipywidgets) (0.7.5)
Requirement already satisfied: setuptools>=18.5 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from ipython>=4.0.0->ipywidgets) (49.2.1)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from nbformat>=4.2.0->ipywidgets) (4.4.0)
Requirement already satisfied: jupyter-core in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from nbformat>=4.2.0->ipywidgets) (4.9.2)
Requirement already satisfied: notebook>=4.4.1 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from widgetsnbextension~=3.6.0->ipywidgets) (6.4.10)
Requirement already satisfied: parso<0.9.0,>=0.8.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from jedi>=0.16->ipython>=4.0.0->ipywidgets) (0.8.3)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.
2.0->ipywidgets) (0.18.1)
Requirement already satisfied: attrs>=17.4.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.2.0->ipywidgets) (21.4.0)
Requirement already satisfied: pyzmq>=13 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from jupyter-client<8.0->ipykernel>=4.5.1->ipywidgets) (22.3.0)
Requirement already satisfied: entrypoints in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from jupyter-client<8.0->ipykernel>=4.5.1->ipywidgets) (0.4)
Requirement already satisfied: python-dateutil>=2.1 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from jupyter-client<8.0->ipykernel>=4.5.1->ipywidgets) (2.8.2)
Requirement already satisfied: pywin32>=1.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from jupyter-core->nbformat>=4.2.0->ipywidgets) (303)
Requirement already satisfied: argon2-cffi in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets) (21.3.0)
Requirement already satisfied: nbconvert>=5 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets) (6.4.4)
Requirement already satisfied: prometheus-client in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets) (0.13.1)
Requirement already satisfied: jinja2 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets) (3.0.3)
Requirement already satisfied: Send2Trash>=1.8.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets) (1.8.0)
Requirement already satisfied: terminado>=0.8.3 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets) (0.13.3)
Requirement already satisfied: wcwidth in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0->ipython>=4.0.0->ipywidgets)
(0.2.5)
Requirement already satisfied: executing in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from stack-data->ipython>=4.0.0->ipywidgets) (0.8.3)
Requirement already satisfied: asttokens in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from stack-data->ipython>=4.0.0->ipywidgets) (2.0.5)
Requirement already satisfied: pure-eval in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from stack-data->ipython>=4.0.0->ipywidgets) (0.2.2)
Requirement already satisfied: pandocfilters>=1.4.1 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipyw
idgets) (1.5.0)
Requirement already satisfied: bleach in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets) (4.1.
0)
Requirement already satisfied: beautifulsoup4 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidget
s) (4.10.0)
Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ip
ywidgets) (0.5.13)
Requirement already satisfied: mistune<2,>=0.8.1 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidg
ets) (0.8.4)
Requirement already satisfied: defusedxml in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets)
(0.7.1)
Requirement already satisfied: testpath in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets) (0.
6.0)
Requirement already satisfied: jupyterlab-pygments in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywi
dgets) (0.1.2)
Requirement already satisfied: MarkupSafe>=2.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from jinja2->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets) (2.
1.1)
Requirement already satisfied: six>=1.5 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from python-dateutil>=2.1->jupyter-client<8.0->ipykernel>=4.5.1->ipywidgets)
(1.16.0)
Requirement already satisfied: pywinpty>=1.1.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from terminado>=0.8.3->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywi
dgets) (2.0.5)
Requirement already satisfied: argon2-cffi-bindings in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from argon2-cffi->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywi
dgets) (21.2.0)
Requirement already satisfied: cffi>=1.0.1 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from argon2-cffi-bindings->argon2-cffi->notebook>=4.4.1->widgetsnbextension~
=3.6.0->ipywidgets) (1.15.0)
Requirement already satisfied: soupsieve>1.2 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from beautifulsoup4->nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.
6.0->ipywidgets) (2.3.1)
Requirement already satisfied: packaging in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from bleach->nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidg
ets) (21.3)
Requirement already satisfied: webencodings in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from bleach->nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipyw
idgets) (0.5.1)
Requirement already satisfied: pycparser in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi->notebook>=4.4.1->widgetsn
bextension~=3.6.0->ipywidgets) (2.21)
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from packaging->bleach->nbconvert>=5->notebook>=4.4.1->widgetsn
bextension~=3.6.0->ipywidgets) (3.0.7)
```

```
WARNING: There was an error checking the latest version of pip.
Requirement already satisfied: plotly in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (5.8.1)
Requirement already satisfied: tenacity>=6.2.0 in c:\users\aakash.patel\appdata\local\programs\python\python39\lib\site-packages (from plotly) (8.0.1)
```

In [2]:
```
data_source_url = "https://healthdata.gov/resource/j8mb-icvb.json"
x = requests.get(data_source_url)
json_res = x.json()
data_source = pd.DataFrame.from_dict(json_res)
data_source.head(5)
```

Out[2]:

	state	state_name	state_fips	fema_region	overall_outcome	date	new_results_reported	total_results_reported
0	AL	Alabama	01	Region 4	Negative	2020-03-01T00:00:00.000	96	96
1	AL	Alabama	01	Region 4	Positive	2020-03-01T00:00:00.000	16	16
2	AL	Alabama	01	Region 4	Negative	2020-03-02T00:00:00.000	72	168
3	AL	Alabama	01	Region 4	Positive	2020-03-02T00:00:00.000	6	22
4	AL	Alabama	01	Region 4	Negative	2020-03-03T00:00:00.000	94	262

In [3]:
```
print("Total Row count: ", len(data_source))
```

```
Total Row count:  1000
```

In [4]:
```
# Download latest dataset csv
download_url = "https://healthdata.gov/api/views/j8mb-icvb/rows.csv?accessType=DOWNLOAD"
res = wget.download(download_url)
df_main  = pd.read_csv(res)
df_main.head()
```

```
-1 / unknown
```

Out[4]:

	state	state_name	state_fips	fema_region	overall_outcome	date	new_results_reported	total_results_reported	geocoded_state
0	AL	Alabama	1	Region 4	Negative	2020/03/01	96	96	NaN
1	AL	Alabama	1	Region 4	Positive	2020/03/01	16	16	NaN
2	AL	Alabama	1	Region 4	Negative	2020/03/02	72	168	NaN
3	AL	Alabama	1	Region 4	Positive	2020/03/02	6	22	NaN
4	AL	Alabama	1	Region 4	Negative	2020/03/03	94	262	NaN

2. EXPLORATORY DATA ANALYSIS

In [5]:
```
# Profiling the dataset

df_profile = ProfileReport(df_main)
df_profile
```

```
Summarize dataset:   0%|          | 0/5 [00:00<?, ?it/s]
Generate report structure:   0%|          | 0/1 [00:00<?, ?it/s]
Render HTML:   0%|          | 0/1 [00:00<?, ?it/s]
```

Pandas Profiling Report

Overview Variables Interactions Correlations Missing values Sample

Overview

Overview Alerts 21 Reproduction

Dataset statistics

Number of variables	9
Number of observations	134773
Missing cells	134773
Missing cells (%)	11.1%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	9.3 MiB
Average record size in memory	72.0 B

Variable types

Categorical	5
Numeric	3
Unsupported	1

Variables

state
Categorical

Distinct	56
Distinct (%)	< 0.1%

WA 2490
NE 2487
FL 2487

Out[5]:

In [6]:
```
print("Row count: ", len(df_main))
print("Latest available date: ", max(df_main["date"]))
```

```
Row count:  134773
Latest available date:  2022/06/09
```

SOLUTIONS:

METRIC 1: Total number of PCR Tests as of yesterday in USA

In [7]:
```
# Determining Yesterday's date
yesterday = date.today() - timedelta(days=1)
considered_date = yesterday

# If yesterday's data isn't available in data most recent date available would be considered
if len(df_main[df_main["date"] == yesterday]) == 0:
    print("Yesterday's data not available!")
    print("Considering available data for most recent date: "+ max(df_main["date"]))
    considered_date = max(df_main["date"])
    d1 = df_main[df_main["date"] == considered_date]
else:
    d1 = df_main[df_main["date"] == yesterday]
```

```
Yesterday's data not available!
Considering available data for most recent date: 2022/06/09
```

In [8]:
```
d1["state_region"] = d1.apply(lambda r: r["state"] +"-" +r["fema_region"].split(" ")[1:][0], axis=1)
d1.head()
```

```
C:\Users\aakash.patel\AppData\Local\Temp\ipykernel_43628\2056466593.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  d1["state_region"] = d1.apply(lambda r: r["state"] +"-" +r["fema_region"].split(" ")[1:][0], axis=1)
```

Out[8]:

	state	state_name	state_fips	fema_region	overall_outcome	date	new_results_reported	total_results_reported	geocoded_state	state_region
9837	AR	Arkansas	5	Region 6	Inconclusive	2022/06/09	3	9820	NaN	AR-6
9838	AR	Arkansas	5	Region 6	Negative	2022/06/09	2213	4304407	NaN	AR-6
9839	AR	Arkansas	5	Region 6	Positive	2022/06/09	273	503714	NaN	AR-6
14767	CO	Colorado	8	Region 8	Inconclusive	2022/06/09	58	55179	NaN	CO-8
14768	CO	Colorado	8	Region 8	Negative	2022/06/09	14739	14894597	NaN	CO-8

In [9]:
```
d2 = d1[["state_region","overall_outcome","total_results_reported"]]
d2.head()
```

Out[9]:

	state_region	overall_outcome	total_results_reported
9837	AR-6	Inconclusive	9820
9838	AR-6	Negative	4304407
9839	AR-6	Positive	503714
14767	CO-8	Inconclusive	55179
14768	CO-8	Negative	14894597

In [10]:
```
d3 = pd.DataFrame()
state_region = []
positives=[]
negatives=[]
inconclusives=[]
for index, row in d2.iterrows():

    if row["state_region"] not in state_region:
        state_region.append(row["state_region"])

    if row['overall_outcome']=="Inconclusive":
        inconclusives.append(row["total_results_reported"])

    elif row["overall_outcome"]=="Negative":
        negatives.append(row["total_results_reported"])

    else:
        positives.append(row["total_results_reported"])

d3["state_region"] =state_region
d3["Inconclusive"] = inconclusives
d3["Negative"] = negatives
d3["Positive"] = positives

d3.head()
```

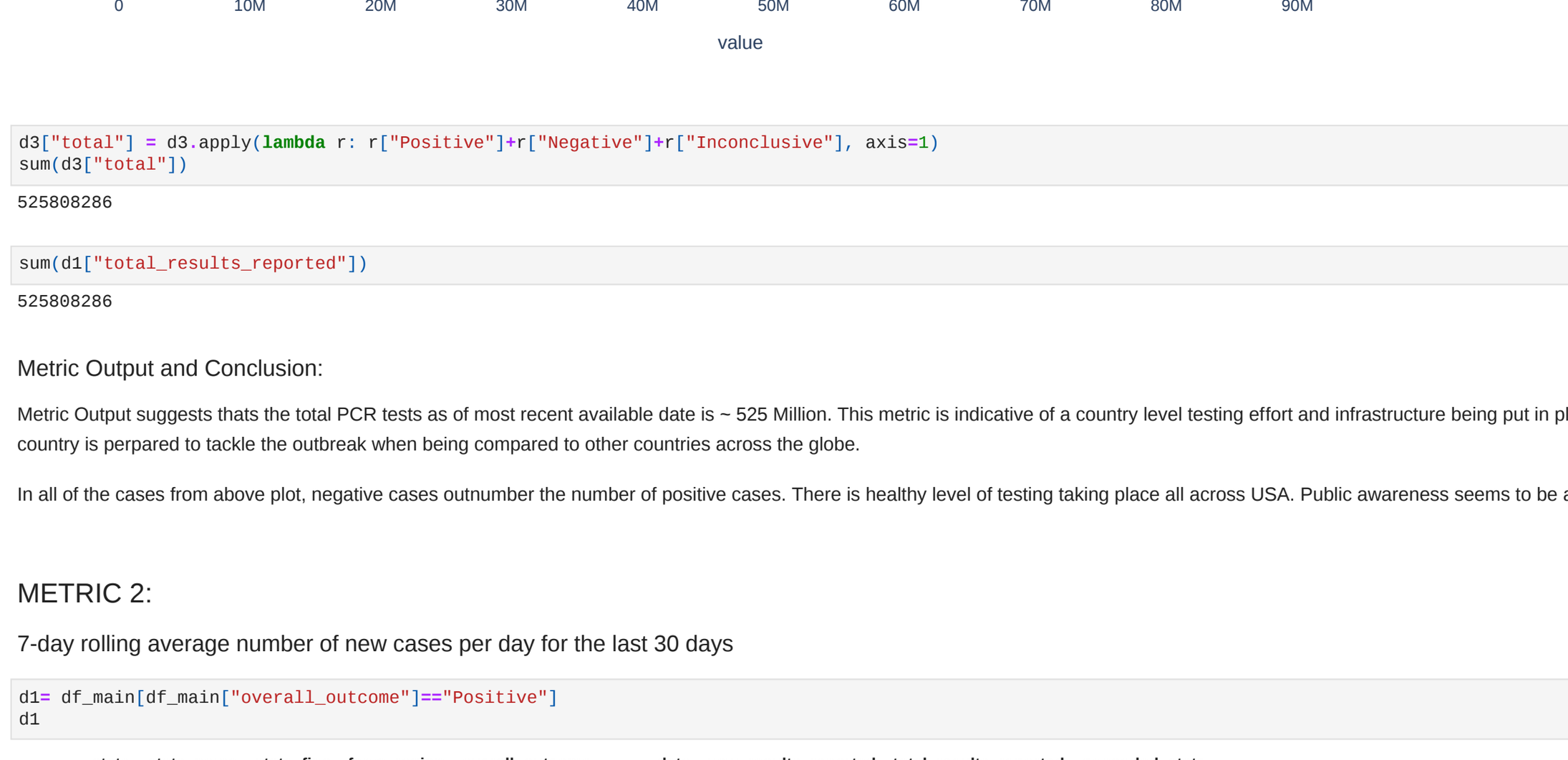
Out[10]:

	state_region	Inconclusive	Negative	Positive
0	AR-6	9820	4304407	503714
1	CO-8	55179	14894597	1510727
2	DC-3	12066	5023709	239337
3	FL-4	235694	51785873	7140548
4	GU-9	32	281230	38738

In [11]:
```
fig = px.bar(d3, y="state_region", x=["Inconclusive", "Positive", "Negative"], title="State-Region Wise Total PCR Tests as of "+considered_date, orientation='h')

# To resize figure vertically and verify for all available state_regions.
# fig.update_layout(
#     autosize=False,
#     width=800,
#     height=1000,)

fig.show()
```



In [12]:
```
d3["total"] = d3.apply(lambda r: r["Positive"]+r["Negative"]+r["Inconclusive"], axis=1)
sum(d3["total"])
```

Out[12]: 525808286

In [13]:
```
sum(d1["total_results_reported"])
```

Out[13]: 525808286

Metric Output and Conclusion:

Metric Output suggests that the total PCR tests as of most recent available date is ~ 525 Million. This metric is indicative of a country level testing effort and infrastructure being put in place to counter the pandemic. It gives the sense of scale at which a country is perpared to tackle the outbreak when being compared to other countries across the globe.

In all of the cases from above plot, negative cases outnumber the number of positive cases. There is healthy level of testing taking place all across USA. Public awareness seems to be at the maximum about COVID.

METRIC 2:

7-day rolling average number of new cases per day for the last 30 days

In [14]:
```
d1= df_main[df_main["overall_outcome"]=="Positive"]
d1
```

Out[14]:

	state	state_name	state_fips	fema_region	overall_outcome	date	new_results_reported	total_results_reported	geocoded_state
1	AL	Alabama	1	Region 4	Positive	2020/03/01	16	16	NaN
3	AL	Alabama	1	Region 4	Positive	2020/03/02	6	22	NaN
5	AL	Alabama	1	Region 4	Positive	2020/03/03	9	31	NaN
7	AL	Alabama	1	Region 4	Positive	2020/03/04	2	33	NaN
9	AL	Alabama	1	Region 4	Positive	2020/03/05	6	39	NaN
...	...	...	...	...	...	...	...	...	...
134760	WY	Wyoming	56	Region 8	Positive	2022/06/04	88	122009	NaN
134763	WY	Wyoming	56	Region 8	Positive	2022/06/05	56	122065	NaN
134766	WY	Wyoming	56	Region 8	Positive	2022/06/06	113	122178	NaN
134769	WY	Wyoming	56	Region 8	Positive	2022/06/07	122	122300	NaN
134772	WY	Wyoming	56	Region 8	Positive	2022/06/08	96	122396	NaN

45174 rows × 9 columns

In [15]:
```
d2 = d1.groupby(["date"])["new_results_reported"].sum()
d2 = d2.to_frame()
d2
```

Out[15]:

date	new_results_reported
2020/03/01	63
2020/03/02	85
2020/03/03	35
2020/03/04	117
2020/03/05	116
...	...
2022/06/05	51466
2022/06/06	71147
2022/06/07	75604
2022/06/08	48509
2022/06/09	11395

831 rows × 1 columns

In [16]:
```
# Verifying above output
print("Total new Positive Cases for ", considered_date, ":")
sum(d1[d1["date"]==considered_date].new_results_reported)
```

```
Total new Positive Cases for  2022/06/09 :
```

Out[16]: 11395

In [17]:
```
d2['rolling_avg'] = d2['new_results_reported'].rolling(7).mean()
d2.reset_index(inplace=True)
```

In [18]:
```
d2.head(10)
```

Out[18]:

	date	new_results_reported	rolling_avg
0	2020/03/01	63	NaN
1	2020/03/02	85	NaN
2	2020/03/03	35	NaN
3	2020/03/04	117	NaN
4	2020/03/05	116	NaN
5	2020/03/06	171	NaN
6	2020/03/07	149	105.142857
7	2020/03/08	234	129.571429
8	2020/03/09	273	156.428571
9	2020/03/10	397	208.142857

In [19]:
```
# Determining Prior date
from datetime import datetime
dt_object = datetime.strptime(
  considered_date, '%Y/%m/%d').date()
from_dt = dt_object - timedelta(days=30)
from_dt = from_dt.strftime("%Y/%m/%d")
print(from_dt)
```

```
2022/05/10
```

In [20]:
```
# Result output for last 30 days
d2[d2["date"] > str(from_dt)]
```

Out[20]:

	date	new_results_reported	rolling_avg
801	2022/05/11	103501	77166.714286
802	2022/05/12	95128	80068.571429
803	2022/05/13	88831	82043.428571
804	2022/05/14	74859	84074.714286
805	2022/05/15	59590	85505.142857
806	2022/05/16	96071	88098.285714
807	2022/05/17	116278	90608.285714
808	2022/05/18	119447	92886.285714
809	2022/05/19	107343	94631.285714
810	2022/05/20	94318	95415.142857
811	2022/05/21	71752	94971.285714
812	2022/05/22	57979	94741.142857
813	2022/05/23	84411	93075.428571
814	2022/05/24	97570	90402.857143
815	2022/05/25	98284	87379.571429
816	2022/05/26	83560	83982.000000
817	2022/05/27	70127	80526.142857
818	2022/05/28	54653	78083.428571
819	2022/05/29	47676	76611.571429
820	2022/05/30	37492	69908.857143
821	2022/05/31	77620	67058.857143
822	2022/06/01	89112	65748.571429
823	2022/06/02	92320	67000.000000
824	2022/06/03	78097	68138.571429
825	2022/06/04	57048	68480.714286
826	2022/06/05	51466	69022.142857
827	2022/06/06	71147	73830.000000
828	2022/06/07	75604	73542.000000
829	2022/06/08	48509	67741.571429
830	2022/06/09	11395	56180.857143

Metric Output and Conclusion:

Metric Output suggests the 7-day rolling average of COVID cases to be in range 56000-95000 approx. Rolling Average is a handy metric to get an idea about how number of cases are within last one week. It gives a good sense of the latest figures along with indications of where the pandemic may be heading. But even with a noticable sharp drop (~37k) in Positive cases in the last days, the rolling average tends to be requiring some days to ease off and represent the change. Also during the first 7 days of computing the metric, when previous days data isn't available, metric would be showing off unreliable data. However since above results just cover past 30 days of data, Day 1-7 of the data isn't impactful. Above results indicate positive signs of easing COVID cases across the states.

METRIC 3:

10 states with the highest test positivity rate (positive tests / tests performed) for tests performed in the last 30 days

In [21]:
```
d1=df_main
d1["state_region"] = d1.apply(lambda r: r["state"] +"-" +r["fema_region"].split(" ")[1:][0], axis=1)
d1.head()
```

Out[21]:

	state	state_name	state_fips	fema_region	overall_outcome	date	new_results_reported	total_results_reported	geocoded_state	state_region
0	AL	Alabama	1	Region 4	Negative	2020/03/01	96	96	NaN	AL-4
1	AL	Alabama	1	Region 4	Positive	2020/03/01	16	16	NaN	AL-4
2	AL	Alabama	1	Region 4	Negative	2020/03/02	72	168	NaN	AL-4
3	AL	Alabama	1	Region 4	Positive	2020/03/02	6	22	NaN	AL-4
4	AL	Alabama	1	Region 4	Negative	2020/03/03	94	262	NaN	AL-4

In [22]:
```
# Filtering data from last 30 days
d2 = d1[d1["date"]> str(from_dt)]
d2
```

Out[22]:

	state	state_name	state_fips	fema_region	overall_outcome	date	new_results_reported	total_results_reported	geocoded_state	state_region
2391	AL	Alabama	1	Region 4	Inconclusive	2022/05/11	9	16624	NaN	AL-4
2392	AL	Alabama	1	Region 4	Negative	2022/05/11	4797	6568776	NaN	AL-4
2393	AL	Alabama	1	Region 4	Positive	2022/05/11	322	1078969	NaN	AL-4
2394	AL	Alabama	1	Region 4	Inconclusive	2022/05/12	4	16628	NaN	AL-4
2395	AL	Alabama	1	Region 4	Negative	2022/05/12	3779	6572555	NaN	AL-4
...	...	...	...	...	...	...	...	...	...	...
134768	WY	Wyoming	56	Region 8	Negative	2022/06/07	787	1313490	NaN	WY-8
134769	WY	Wyoming	56	Region 8	Positive	2022/06/07	122	122300	NaN	WY-8
134770	WY	Wyoming	56	Region 8	Inconclusive	2022/06/08	5	3379	NaN	WY-8
134771	WY	Wyoming	56	Region 8	Negative	2022/06/08	635	1314125	NaN	WY-8
134772	WY	Wyoming	56	Region 8	Positive	2022/06/08	96	122396	NaN	WY-8

4725 rows × 10 columns

In [23]:
```
# Preparing state wise total positive cases data
d3= d2[d2["overall_outcome"]=="Positive"]
d4 = d3.groupby(["state","state_name"])["new_results_reported"].sum().to_frame()
d4.reset_index(inplace=True)
d4.rename(columns={'new_results_reported': 'total_positive_cases_reported'}, inplace=True)
d4.head()
```

Out[23]:

	state	state_name	total_positive_cases_reported
0	AK	Alaska	1949
1	AL	Alabama	12475
2	AR	Arkansas	6257
3	AZ	Arizona	38933
4	CA	California	155553

In [24]:
```
# Preparing state wise total tests results data
d5= d2.groupby(["state"])["new_results_reported"].sum().to_frame()
d5.reset_index(inplace=True)
d5.rename(columns={'new_results_reported': 'total_tests_reported'}, inplace=True)
d5.head()
```

Out[24]:

	state	total_tests_reported
0	AK	14551
1	AL	112591
2	AR	68402
3	AZ	208204
4	CA	2583202

In [25]:
```
merged_df = pd.merge(d4,d5,how='outer',on=['state'])
merged_df["positivity_rate"] = merged_df.apply(lambda r: r["total_positive_cases_reported"]/r["total_tests_reported"], axis=1)
sorted_df = merged_df.sort_values(by=['positivity_rate'], ascending=False)
```

In [26]:
```
# Final results
sorted_df[["state","state_name", "positivity_rate"]].head(10)
```

Out[26]:

	state	state_name	positivity_rate
13	IA	Iowa	1.000000
48	VI	U.S. Virgin Islands	0.290981
40	PR	Puerto Rico	0.287522
34	NV	Nevada	0.251092
46	UT	Utah	0.221546
12	HI	Hawaii	0.205764
9	FL	Florida	0.190072
3	AZ	Arizona	0.186994
8	DE	Delaware	0.183878
47	VA	Virginia	0.179586

In [27]:
```
# Validating positivity rate for IA
t1 = df_main[df_main["state"]=="IA"]
t1= t1[t1["date"]>str(from_dt)]
t2 = t1[t1["overall_outcome"]=="Positive"]

print("Positive cases: ", t2.new_results_reported.sum())
print("Total cases : ", t1.new_results_reported.sum())
print("Positivity rate: ", t2.new_results_reported.sum()/t1.new_results_reported.sum())
```

```
Positive cases:  15117
Total cases :  15117
Positivity rate:  1.0
```

In [28]:
```
t1 = df_main[df_main["state"]=="IA"]
t1= t1[t1["date"]>str(from_dt)]
t2 = t1[t1["overall_outcome"]=="Negative"]

print("Negative cases: ", t2.new_results_reported.sum())
print("Total cases : ", t1.new_results_reported.sum())
```

```
Negative cases:  0
Total cases :  15117
```

Metric Output and Conclusion:

Posivity metric gives a sense of how the ratio of positive to total cases is performing for the states. Higher the positivity rate for a state, higher the rates of COVID cases among those who are getting tested. Lower positive rate indicates less ratio of infections to total tests being performed in the state.

Metric Output suggests the total positivity rates to be ~32% or less in most of the states. Iowa with the highest positivity rate could be at the centre for a potential new wave. Positivity rate of 1 suggests that only positive cases were being reported by labs of the state. With the absense of Negative and Inconclusive data from the state, positivity rate is becoming unreasonably higher. So, more data for the state should be collected and scrutinized. With Virginia having 10th lowest positivity rate, it can be due to the that people of Virginia are comparatively immune to infection than those from Iowa.