

Problem 1.

<https://github.com/meesuny/lab2>

Problem 2.

```
In [1]: !pip install openpyxl # for Excel
```

Collecting openpyxl

Downloading openpyxl-3.0.10-py2.py3-none-any.whl (242 kB)

242.1/242.1 kB 4.3 MB/s eta 0:00:00a 0:00:01

Collecting et-xmlfile

Downloading et_xmlfile-1.1.0-py3-none-any.whl (4.7 kB)

Installing collected packages: et-xmlfile, openpyxl

Successfully installed et-xmlfile-1.1.0 openpyxl-3.0.10

WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager. It is recommended to use a virtual environment instead: <https://pip.pypa.io/warnings/venv>

```
In [2]: !pip install pyreadstat # for SPSS
```

Collecting pyreadstat

Downloading pyreadstat-1.1.9-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (2.7 MB)

2.7/2.7 MB 10.1 MB/s eta 0:00:0000:0100:01

Requirement already satisfied: pandas>=1.2.0 in /usr/local/lib/python3.10/site-packages (from pyreadstat) (1.4.4)

Requirement already satisfied: numpy>=1.21.0 in /usr/local/lib/python3.10/site-packages (from pandas>=1.2.0->pyreadstat) (1.23.3)

Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/site-packages (from pandas>=1.2.0->pyreadstat) (2022.4)

Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/site-packages (from pandas>=1.2.0->pyreadstat) (2.8.2)

Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/site-packages (from python-dateutil>=2.8.1->pandas>=1.2.0->pyreadstat) (1.16.0)

Installing collected packages: pyreadstat

Successfully installed pyreadstat-1.1.9

WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager. It is recommended to use a virtual environment instead: <https://pip.pypa.io/warnings/venv>

```
In [3]: !pip install pystata
```

Collecting pystata

Downloading pystata-0.0.1-py3-none-any.whl (21 kB)

Requirement already satisfied: pandas in /usr/local/lib/python3.10/site-packages (from pystata) (1.4.4)

Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/site-packages (from pandas->pystata) (2022.4)

Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/site-packages (from pandas->pystata) (2.8.2)

Requirement already satisfied: numpy>=1.21.0 in /usr/local/lib/python3.10/site-packages (from pandas->pystata) (1.23.3)

Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/site-packages (from python-dateutil>=2.8.1->pandas->pystata) (1.16.0)

Installing collected packages: pystata

Successfully installed pystata-0.0.1

WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager. It is recommended to use a virtual environment instead: <https://pip.pypa.io/warnings/venv>

```
In [4]: import pandas as pd
import numpy as np
import requests
import json
import os
import openpyxl
from bs4 import BeautifulSoup
import collections
collections.Callable = collections.abc.Callable
```

```
In [5]: data1 = pd.read_csv('data/data1.csv', header=2)
data1
```

Out[5]:

	Country	Happiness score	Whisker-high	Whisker-low	Dystopia (1.92) + residual	Explained by: GDP per capita	Explained by: Social support	Explained by: Healthy life expectancy	Explained by: Freedom to make life choices	Explained by: Generosity	Explained by: Perceptions of corruption
0	Finland	7.632	7.695	7.569	2.595	1.305	1.592	0.874	0.681	0.192	0.393
1	Norway	7.594	7.657	7.530	2.383	1.456	1.582	0.861	0.686	0.286	0.340
2	Denmark	7.555	7.623	7.487	2.370	1.351	1.590	0.868	0.683	0.284	0.408
3	Iceland	7.495	7.593	7.398	2.426	1.343	1.644	0.914	0.677	0.353	0.138
4	Switzerland	7.487	7.570	7.405	2.320	1.420	1.549	0.927	0.660	0.256	0.357
...
151	Yemen	3.355	3.448	3.262	1.106	0.442	1.073	0.343	0.244	0.083	0.064
152	Tanzania	3.303	3.414	3.193	0.628	0.455	0.991	0.381	0.481	0.270	0.097
153	South Sudan	3.254	3.385	3.123	1.691	0.337	0.608	0.177	0.112	0.224	0.106
154	Central African Republic	3.083	3.227	2.939	2.487	0.024	0.000	0.010	0.305	0.218	0.038
155	Burundi	2.905	3.074	2.735	1.752	0.091	0.627	0.145	0.065	0.149	0.076

156 rows × 11 columns

In [6]:

```
data2 = pd.read_csv('data/data2.txt', header=2, comment = '/')
data2
```

Out[6]:

	Country	Happiness score	Whisker-high	Whisker-low	Dystopia (1.92) + residual	Explained by: GDP per capita	Explained by: Social support	Explained by: Healthy life expectancy	Explained by: Freedom to make life choices	Explained by: Generosity	Explained by: Perceptions of corruption
0	Finland	7.632	7.695	7.569	2.595	1.305	1.592	0.874	0.681	0.192	0.393
1	Norway	7.594	7.657	7.530	2.383	1.456	1.582	0.861	0.686	0.286	0.340
2	Denmark	7.555	7.623	7.487	2.370	1.351	1.590	0.868	0.683	0.284	0.408
3	Iceland	7.495	7.593	7.398	2.426	1.343	1.644	0.914	0.677	0.353	0.138
4	Switzerland	7.487	7.570	7.405	2.320	1.420	1.549	0.927	0.660	0.256	0.357
...
151	Yemen	3.355	3.448	3.262	1.106	0.442	1.073	0.343	0.244	0.083	0.064
152	Tanzania	3.303	3.414	3.193	0.628	0.455	0.991	0.381	0.481	0.270	0.097
153	South Sudan	3.254	3.385	3.123	1.691	0.337	0.608	0.177	0.112	0.224	0.106
154	Central African Republic	3.083	3.227	2.939	2.487	0.024	0.000	0.010	0.305	0.218	0.038
155	Burundi	2.905	3.074	2.735	1.752	0.091	0.627	0.145	0.065	0.149	0.076

156 rows × 11 columns

In [7]:

```
data3 = pd.read_csv('data/data3.txt', header=2, sep='\t')
data3
```

Out[7]:

	Country	Happiness score	Whisker-high	Whisker-low	Dystopia (1.92) + residual	Explained by: GDP per capita	Explained by: Social support	Explained by: Healthy life expectancy	Explained by: Freedom to make life choices	Explained by: Generosity	Explained by: Perceptions of corruption
0	Finland	7.632	7.695	7.569	2.595	1.305	1.592	0.874	0.681	0.192	0.393
1	Norway	7.594	7.657	7.530	2.383	1.456	1.582	0.861	0.686	0.286	0.340
2	Denmark	7.555	7.623	7.487	2.370	1.351	1.590	0.868	0.683	0.284	0.408
3	Iceland	7.495	7.593	7.398	2.426	1.343	1.644	0.914	0.677	0.353	0.138
4	Switzerland	7.487	7.570	7.405	2.320	1.420	1.549	0.927	0.660	0.256	0.357
...
151	Yemen	3.355	3.448	3.262	1.106	0.442	1.073	0.343	0.244	0.083	0.064
152	Tanzania	3.303	3.414	3.193	0.628	0.455	0.991	0.381	0.481	0.270	0.097
153	South Sudan	3.254	3.385	3.123	1.691	0.337	0.608	0.177	0.112	0.224	0.106
154	Central African Republic	3.083	3.227	2.939	2.487	0.024	0.000	0.010	0.305	0.218	0.038
155	Burundi	2.905	3.074	2.735	1.752	0.091	0.627	0.145	0.065	0.149	0.076

156 rows × 11 columns

In [8]:

```
data4 = pd.read_csv('data/data4.txt', header=None, sep='$')
data4

# Add headers on this one?
```

Out[8]:

		0	1	2	3	4	5	6	7	8	9	10
0	Finland	7.632	7.695	7.569	2.595	1.305	1.592	0.874	0.681	0.192	0.393	
1	Norway	7.594	7.657	7.530	2.383	1.456	1.582	0.861	0.686	0.286	0.340	
2	Denmark	7.555	7.623	7.487	2.370	1.351	1.590	0.868	0.683	0.284	0.408	
3	Iceland	7.495	7.593	7.398	2.426	1.343	1.644	0.914	0.677	0.353	0.138	
4	Switzerland	7.487	7.570	7.405	2.320	1.420	1.549	0.927	0.660	0.256	0.357	
...
151	Yemen	3.355	3.448	3.262	1.106	0.442	1.073	0.343	0.244	0.083	0.064	
152	Tanzania	3.303	3.414	3.193	0.628	0.455	0.991	0.381	0.481	0.270	0.097	
153	South Sudan	3.254	3.385	3.123	1.691	0.337	0.608	0.177	0.112	0.224	0.106	
154	Central African Republic	3.083	3.227	2.939	2.487	0.024	0.000	0.010	0.305	0.218	0.038	
155	Burundi	2.905	3.074	2.735	1.752	0.091	0.627	0.145	0.065	0.149	0.076	

156 rows × 11 columns

In [9]:

```
data5 = pd.read_csv('data/data5.csv', nrows=156)
data5
```

Out[9]:

	Country	Happiness score	Whisker-high	Whisker-low	Dystopia (1.92) + residual	Explained by: GDP per capita	Explained by: Social support	Explained by: Healthy life expectancy	Explained by: Freedom to make life choices	Explained by: Generosity	Explained by: Perceptions of corruption
0	Finland	7.632	7.695	7.569	2.595	1.305	1.592	0.874	0.681	0.192	0.393
1	Norway	7.594	7.657	7.530	2.383	1.456	1.582	0.861	0.686	0.286	0.340
2	Denmark	7.555	7.623	7.487	2.370	1.351	1.590	0.868	0.683	0.284	0.408
3	Iceland	7.495	7.593	7.398	2.426	1.343	1.644	0.914	0.677	0.353	0.138
4	Switzerland	7.487	7.570	7.405	2.320	1.420	1.549	0.927	0.660	0.256	0.357
...
151	Yemen	3.355	3.448	3.262	1.106	0.442	1.073	0.343	0.244	0.083	0.064
152	Tanzania	3.303	3.414	3.193	0.628	0.455	0.991	0.381	0.481	0.270	0.097
153	South Sudan	3.254	3.385	3.123	1.691	0.337	0.608	0.177	0.112	0.224	0.106
154	Central African Republic	3.083	3.227	2.939	2.487	0.024	0.000	0.010	0.305	0.218	0.038
155	Burundi	2.905	3.074	2.735	1.752	0.091	0.627	0.145	0.065	0.149	0.076

156 rows × 11 columns

In [10]:

```
data6 = pd.read_csv('data/data6.dat', na_values = '999.000')
data6
```

Out[10]:

	Country	Happiness score	Whisker-high	Whisker-low	Dystopia (1.92) + residual	Explained by: GDP per capita	Explained by: Social support	Explained by: Healthy life expectancy	Explained by: Freedom to make life choices	Explained by: Generosity	Explained by: Perceptions of corruption
0	Finland	7.632	7.695	7.569	2.595	NaN	NaN	NaN	0.681	0.192	0.393
1	Norway	7.594	7.657	7.530	NaN	NaN	1.582	NaN	0.686	0.286	0.340
2	Denmark	7.555	7.623	7.487	2.370	1.351	1.590	NaN	0.683	0.284	0.408
3	Iceland	7.495	7.593	NaN	2.426	1.343	1.644	0.914	0.677	0.353	NaN
4	Switzerland	7.487	7.570	7.405	2.320	1.420	1.549	0.927	0.660	0.256	0.357
...
151	NaN	3.355	3.448	3.262	1.106	0.442	1.073	0.343	0.244	NaN	0.064
152	Tanzania	NaN	NaN	3.193	0.628	NaN	0.991	0.381	0.481	0.270	0.097
153	South Sudan	3.254	NaN	3.123	1.691	0.337	NaN	0.177	0.112	0.224	0.106
154	Central African Republic	3.083	3.227	2.939	2.487	0.024	0.000	0.010	0.305	NaN	0.038
155	Burundi	2.905	3.074	NaN	1.752	0.091	NaN	0.145	0.065	0.149	0.076

156 rows × 11 columns

In [11]:

```
data7 = pd.read_excel('data/data7.xlsx', sheet_name = "Data")
data7

# Had to pip install openpyxl in order for this to run.
# Collecting openpyxl
# Downloading openpyxl-3.0.10-py2.py3-none-any.whl (242 kB)
# _____ 242.1/242.1 kB 367.7 kB/s eta 0:00:00a 0:00:01
# Collecting et-xmlfile
# Downloading et_xmlfile-1.1.0-py3-none-any.whl (4.7 kB)
# Installing collected packages: et-xmlfile, openpyxl
# Successfully installed et-xmlfile-1.1.0 openpyxl-3.0.10
# WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with the system package manager. It is recommended to use c
```


Out[11]:

	Country	Happiness score	Whisker-high	Whisker-low	Dystopia (1.92) + residual	Explained by: GDP per capita	Explained by: Social support	Explained by: Healthy life expectancy	Explained by: Freedom to make life choices	Explained by: Generosity	Explained by: Perceptions of corruption
0	Finland	7.632	7.695	7.569	2.595	1.305	1.592	0.874	0.681	0.192	0.393
1	Norway	7.594	7.657	7.530	2.383	1.456	1.582	0.861	0.686	0.286	0.340
2	Denmark	7.555	7.623	7.487	2.370	1.351	1.590	0.868	0.683	0.284	0.408
3	Iceland	7.495	7.593	7.398	2.426	1.343	1.644	0.914	0.677	0.353	0.138
4	Switzerland	7.487	7.570	7.405	2.320	1.420	1.549	0.927	0.660	0.256	0.357
...
151	Yemen	3.355	3.448	3.262	1.106	0.442	1.073	0.343	0.244	0.083	0.064
152	Tanzania	3.303	3.414	3.193	0.628	0.455	0.991	0.381	0.481	0.270	0.097
153	South Sudan	3.254	3.385	3.123	1.691	0.337	0.608	0.177	0.112	0.224	0.106
154	Central African Republic	3.083	3.227	2.939	2.487	0.024	0.000	0.010	0.305	0.218	0.038
155	Burundi	2.905	3.074	2.735	1.752	0.091	0.627	0.145	0.065	0.149	0.076

156 rows × 11 columns

In [12]:

```
data8 = pd.read_stata('data/data8.dta')
data8

# STATA pip install
```

Out[12]:

	country	happinessscore	whiskerhigh	whiskerlow	dystopia192residual	explainedbygdppercapita	explainedbysocialsupport	explainedbyhealthylifeexpectancy	explainedbyfreedomtor
0	Finland	7.632	7.695	7.569	2.595	1.305	1.592	0.874	
1	Norway	7.594	7.657	7.530	2.383	1.456	1.582	0.861	
2	Denmark	7.555	7.623	7.487	2.370	1.351	1.590	0.868	
3	Iceland	7.495	7.593	7.398	2.426	1.343	1.644	0.914	
4	Switzerland	7.487	7.570	7.405	2.320	1.420	1.549	0.927	
...	
151	Yemen	3.355	3.448	3.262	1.106	0.442	1.073	0.343	
152	Tanzania	3.303	3.414	3.193	0.628	0.455	0.991	0.381	
153	South Sudan	3.254	3.385	3.123	1.691	0.337	0.608	0.177	
154	Central African Republic	3.083	3.227	2.939	2.487	0.024	0.000	0.010	
155	Burundi	2.905	3.074	2.735	1.752	0.091	0.627	0.145	

156 rows × 11 columns



In [13]:

data9 = pd.read_spss('data/data9.sav')
data9

Out[13]:

	country	happiness	whiskerhigh	whiskerlow	dystopia	gdpPC	socsupport	lifeexp	lifechoice	generous	corrupt
0	Finland	7.632	7.695	7.569	2.595	1.305	1.592	0.874	0.681	0.192	0.393
1	Norway	7.594	7.657	7.530	2.383	1.456	1.582	0.861	0.686	0.286	0.340
2	Denmark	7.555	7.623	7.487	2.370	1.351	1.590	0.868	0.683	0.284	0.408
3	Iceland	7.495	7.593	7.398	2.426	1.343	1.644	0.914	0.677	0.353	0.138
4	Switzerland	7.487	7.570	7.405	2.320	1.420	1.549	0.927	0.660	0.256	0.357
...
151	Yemen	3.355	3.448	3.262	1.106	0.442	1.073	0.343	0.244	0.083	0.064
152	Tanzania	3.303	3.414	3.193	0.628	0.455	0.991	0.381	0.481	0.270	0.097
153	South Sudan	3.254	3.385	3.123	1.691	0.337	0.608	0.177	0.112	0.224	0.106
154	Central African Republic	3.083	3.227	2.939	2.487	0.024	0.000	0.010	0.305	0.218	0.038
155	Burundi	2.905	3.074	2.735	1.752	0.091	0.627	0.145	0.065	0.149	0.076

156 rows × 11 columns

In [14]:

```
data10 = pd.read_sas('data/data10.xpt')
data10
```

Out[14]:

	COUNTRY	HAPPINES	WHISKERH	WHISKERL	DYSTOPIA	EXPLAINE	EXPLAIN2	EXPLAIN3	EXPLAIN4	EXPLAIN5	EXPLAIN6
0	b'Finland'	7.632	7.695	7.569	2.595	1.305	1.592000e+00	0.874	0.681	0.192	0.393
1	b'Norway'	7.594	7.657	7.530	2.383	1.456	1.582000e+00	0.861	0.686	0.286	0.340
2	b'Denmark'	7.555	7.623	7.487	2.370	1.351	1.590000e+00	0.868	0.683	0.284	0.408
3	b'Iceland'	7.495	7.593	7.398	2.426	1.343	1.644000e+00	0.914	0.677	0.353	0.138
4	b'Switzerland'	7.487	7.570	7.405	2.320	1.420	1.549000e+00	0.927	0.660	0.256	0.357
...
151	b'Yemen'	3.355	3.448	3.262	1.106	0.442	1.073000e+00	0.343	0.244	0.083	0.064
152	b'Tanzania'	3.303	3.414	3.193	0.628	0.455	9.910000e-01	0.381	0.481	0.270	0.097
153	b'South Sudan'	3.254	3.385	3.123	1.691	0.337	6.080000e-01	0.177	0.112	0.224	0.106
154	b'Central African Republic'	3.083	3.227	2.939	2.487	0.024	5.397605e-79	0.010	0.305	0.218	0.038
155	b'Burundi'	2.905	3.074	2.735	1.752	0.091	6.270000e-01	0.145	0.065	0.149	0.076

156 rows × 11 columns

Problem 3.

```
In [15]: nasatoken = os.environ['nasatoken']

In [16]: useragent_url = 'https://httpbin.org/user-agent'
r = requests.get(useragent_url)
useragent = json.loads(r.text)['user-agent']

In [17]: headers = {'X-API-Key': nasatoken,
                    'User-Agent': useragent,
                    'From': 'tby8aj@virginia.edu'}

root = 'https://api.nasa.gov'

endpoint = '/neo/rest/v1/feed?start_date=2022-10-01&end_date=2022-10-08&api_key=nasatoken'

r = requests.get(root + endpoint,
```

```
headers = headers)
r
```

Out[17]: <Response [200]>

```
In [18]: myjson = json.loads(r.text)
near_earth_objects1 = pd.json_normalize(myjson, record_path = ['near_earth_objects', '2022-10-01'])
near_earth_objects2 = pd.json_normalize(myjson, record_path = ['near_earth_objects', '2022-10-02'])
near_earth_objects3 = pd.json_normalize(myjson, record_path = ['near_earth_objects', '2022-10-03'])
near_earth_objects4 = pd.json_normalize(myjson, record_path = ['near_earth_objects', '2022-10-04'])
near_earth_objects5 = pd.json_normalize(myjson, record_path = ['near_earth_objects', '2022-10-05'])
near_earth_objects6 = pd.json_normalize(myjson, record_path = ['near_earth_objects', '2022-10-06'])
near_earth_objects7 = pd.json_normalize(myjson, record_path = ['near_earth_objects', '2022-10-07'])
near_earth_objects_all = pd.concat([near_earth_objects1, near_earth_objects2, near_earth_objects3, near_earth_objects4,
                                   near_earth_objects5, near_earth_objects6, near_earth_objects7], ignore_index=True)

#near_earth_objects_all

# This adds "close_approach_date" as the very last column of the resulting data frame.
df = pd.DataFrame(data=near_earth_objects_all)
df['close_approach_date'] = [d[0]['close_approach_date'] for d in df['close_approach_data']]

near_earth_objects_all
```

Out[18]:

	id	neo_reference_id	name	nasa_jpl_url	absolute_magnitude_h	is_potentially_hazardous_asteroid	close_approach_data	is_sentry_object	
0	3429684	3429684	(2008 TZ)	http://ssd.jpl.nasa.gov/sbdb.cgi?sstr=3429684	25.40	False	[{'close_approach_date': '2022-10-01', 'close_...	False	http://api.nasa.gov/neo
1	3548667	3548667	(2010 TX54)	http://ssd.jpl.nasa.gov/sbdb.cgi?sstr=3548667	20.98	False	[{'close_approach_date': '2022-10-01', 'close_...	False	http://api.nasa.gov/neo
2	3650847	3650847	(2013 TN127)	http://ssd.jpl.nasa.gov/sbdb.cgi?sstr=3650847	26.20	False	[{'close_approach_date': '2022-10-01', 'close_...	False	http://api.nasa.gov/neo
3	3772755	3772755	(2017 FT90)	http://ssd.jpl.nasa.gov/sbdb.cgi?sstr=3772755	26.20	False	[{'close_approach_date': '2022-10-01', 'close_...	False	http://api.nasa.gov/neo
4	3782015	3782015	(2017 SZ20)	http://ssd.jpl.nasa.gov/sbdb.cgi?sstr=3782015	24.50	False	[{'close_approach_date': '2022-10-01', 'close_...	False	http://api.nasa.gov/neo
...	
75	3342645	3342645	(2006 SG7)	http://ssd.jpl.nasa.gov/sbdb.cgi?sstr=3342645	22.90	False	[{'close_approach_date': '2022-10-07', 'close_...	False	http://api.nasa.gov/neo
76	3384488	3384488	(2007 RD20)	http://ssd.jpl.nasa.gov/sbdb.cgi?sstr=3384488	21.00	False	[{'close_approach_date': '2022-10-07', 'close_...	False	http://api.nasa.gov/neo
77	3625706	3625706	(2013 BD74)	http://ssd.jpl.nasa.gov/sbdb.cgi?sstr=3625706	24.20	False	[{'close_approach_date': '2022-10-07', 'close_...	False	http://api.nasa.gov/neo
78	3648879	3648879	(2013 TJ6)	http://ssd.jpl.nasa.gov/sbdb.cgi?sstr=3648879	25.20	False	[{'close_approach_date': '2022-10-07', 'close_...	False	http://api.nasa.gov/neo
79	54055076	54055076	(2020 TE2)	http://ssd.jpl.nasa.gov/sbdb.cgi?sstr=54055076	26.30	False	[{'close_approach_date': '2022-10-07', 'close_...	False	http://api.nasa.gov/neo/

80 rows × 18 columns

◀

▶

In []:

Problem 4.

a) Zillow.com disallows the following pages and subpages to be crawled, which suggests that they do not allow pulling the current listed prices for houses for sale in Charlottesville from zillow.com.

- Disallow: /homes/comps_for_sale/
- Disallow: /homes/comps/
- Disallow: /homes/for_sale/*_agent_list/
- Disallow: /homes/for_sale/1_favorite_list/1_rs/1_fr/

b) Google.com allows both /m/finance and /finance to be crawled, so scraping the current stock prices off of google.com/finance is permitted.

c) Twitter disallows */followers to be crawled, so we wouldn't be able to copy the list of twitter accounts that each NBA player follows.

d) Genius.com disallows crawling /songs*/metadata, and metadata presumably includes the lyrics, so it doesn't allow getting lyrics from their site. This means that we can't get the lyrics to Lizzo's Good as Hell from <https://genius.com/Lizzo-good-as-hell-lyrics>.

In []:

Problem 5.

```
In [1]: import pandas as pd
import numpy as np
import requests
import json
import os
from bs4 import BeautifulSoup
import collections
collections.Callable = collections.abc.Callable
import time
```

```
In [2]: geniustoken = os.environ['geniustoken']
useragent_url = 'https://httpbin.org/user-agent'
r = requests.get(useragent_url)
useragent = json.loads(r.text)['user-agent']
```

```
In [3]: urltosrape = 'https://genius.com/Lizzo-good-as-hell-lyrics'

r = requests.get(urltosrape,
                 headers = {'User-Agent': useragent,
                             'From': 'tby8aj@virginia.edu'})
myhtml = BeautifulSoup(r.text, 'html.parser')
r
```

```
Out[3]: <Response [200]>
```

```
In [4]: # (a) Find the lyrics to "Good as Hell"

myhtml.find_all('div', attrs={'data-lyrics-container': True})[0].text
```

```
Out[4]: "[Chorus]I do my hair toss, check my nailsBaby, how you feelin'? (Feelin' good as hell)Hair toss, check my nailsBaby, how you feelin'? (Feelin' good as hell)[Verse 1]Woo, child, tired of the bullshitGo on, dust your shoulders off, keep it movingYes, Lord, tryna get some new shitIn there, swimwear, going-to-the-pool s hitCome now, come dry your eyesYou know you a star, you can touch the skyI know that it's hard, but you have to tryIf you need advice, let me simplify[Pre-Chorus]If he don't love you anymoreJust walk your fine ass out the door[Chorus]I do my hair toss, check my nailsBaby, how you feelin'? (Feelin' good as hell)Hair toss, check my nailsBaby, how you feelin'? (Feelin' good as hell)(Feeling good as hell)Baby, how you feelin'? (Feelin' good as hell)[Verse 2]Woo, girl, need to kick off your shoesGotta take a deep breath, time to focus on youAll the big fights, long nights that you been throughI got a bottle of tequila I been saving for youBoss up and change your lifeYou can have it all, no sacrificeI know he did you wrong, we can make it rightSo go and let it all hang out tonight[Pre-Chorus]'Cause he don't love you anymoreSo walk your fine ass out the door"
```

```
In [5]: # (b) Find links to the other songs and append to song_list
```



```

song_list=[]

for x in range(4,19):
    song_list.append(myhtml.find_all('a', attrs={'class':"Link__StyledLink-h3isu4-0 fUgcxf"})[x])

song_list

```

```

Out[5]: [<a class="Link__StyledLink-h3isu4-0 fUgcxf" font-weight="light" href="https://genius.com/Lizzo-cuz-i-love-you-lyrics">Cuz I Love You</a>,
<a class="Link__StyledLink-h3isu4-0 fUgcxf" font-weight="light" href="https://genius.com/Lizzo-like-a-girl-lyrics">Like a Girl</a>,
<a class="Link__StyledLink-h3isu4-0 fUgcxf" font-weight="light" href="https://genius.com/Lizzo-juice-lyrics">Juice</a>,
<a class="Link__StyledLink-h3isu4-0 fUgcxf" font-weight="light" href="https://genius.com/Lizzo-soulmate-lyrics">Soulmate</a>,
<a class="Link__StyledLink-h3isu4-0 fUgcxf" font-weight="light" href="https://genius.com/Lizzo-jerome-lyrics">Jerome</a>,
<a class="Link__StyledLink-h3isu4-0 fUgcxf" font-weight="light" href="https://genius.com/Lizzo-cry-baby-lyrics">Cry Baby</a>,
<a class="Link__StyledLink-h3isu4-0 fUgcxf" font-weight="light" href="https://genius.com/Lizzo-tempo-lyrics">Tempo</a>,
<a class="Link__StyledLink-h3isu4-0 fUgcxf" font-weight="light" href="https://genius.com/Lizzo-exactly-how-i-feel-lyrics">Exactly How I Feel</a>,
<a class="Link__StyledLink-h3isu4-0 fUgcxf" font-weight="light" href="https://genius.com/Lizzo-better-in-color-lyrics">Better in Color</a>,
<a class="Link__StyledLink-h3isu4-0 fUgcxf" font-weight="light" href="https://genius.com/Lizzo-heaven-help-me-lyrics">Heaven Help Me</a>,
<a class="Link__StyledLink-h3isu4-0 fUgcxf" font-weight="light" href="https://genius.com/Lizzo-lingerie-lyrics">Lingerie</a>,
<a class="Link__StyledLink-h3isu4-0 fUgcxf" font-weight="light" href="https://genius.com/Lizzo-boys-lyrics">Boys</a>,
<a class="Link__StyledLink-h3isu4-0 fUgcxf" font-weight="light" href="https://genius.com/Lizzo-truth-hurts-lyrics">Truth Hurts</a>,
<a class="Link__StyledLink-h3isu4-0 fUgcxf" font-weight="light" href="https://genius.com/Lizzo-water-me-lyrics">Water Me</a>,
<a class="Link__StyledLink-h3isu4-0 fUgcxf" font-weight="light" href="https://genius.com/Lizzo-good-as-hell-remix-lyrics">Good as Hell (Remix)</a>]

```

```

In [6]: song_links=[]

for x in song_list:
    song_links.append(x.get('href'))

song_links

```

```

Out[6]: ['https://genius.com/Lizzo-cuz-i-love-you-lyrics',
'https://genius.com/Lizzo-like-a-girl-lyrics',
'https://genius.com/Lizzo-juice-lyrics',
'https://genius.com/Lizzo-soulmate-lyrics',
'https://genius.com/Lizzo-jerome-lyrics',
'https://genius.com/Lizzo-cry-baby-lyrics',
'https://genius.com/Lizzo-tempo-lyrics',
'https://genius.com/Lizzo-exactly-how-i-feel-lyrics',
'https://genius.com/Lizzo-better-in-color-lyrics',
'https://genius.com/Lizzo-heaven-help-me-lyrics',
'https://genius.com/Lizzo-lingerie-lyrics',
'https://genius.com/Lizzo-boys-lyrics',
'https://genius.com/Lizzo-truth-hurts-lyrics',
'https://genius.com/Lizzo-water-me-lyrics',
'https://genius.com/Lizzo-good-as-hell-remix-lyrics']

```

```
In [7]: def scrape_song(url):
        time.sleep(5)
        print('Now getting the lyrics from ' + url)
        r = requests.get(url, headers = {'User-Agent': useragent, 'From': 'tby8aj@virginia.edu'})
        myhtml = BeautifulSoup(r.text, 'html.parser')
        try:
            lyrictext = myhtml.find_all('div', attrs={'data-lyrics-container': True})
            return lyrictext
        except: pass
```

```
In [8]: lyrics = [scrape_song(u) for u in song_links]
```

```
Now getting the lyrics from https://genius.com/Lizzo-cuz-i-love-you-lyrics
Now getting the lyrics from https://genius.com/Lizzo-like-a-girl-lyrics
Now getting the lyrics from https://genius.com/Lizzo-juice-lyrics
Now getting the lyrics from https://genius.com/Lizzo-soulmate-lyrics
Now getting the lyrics from https://genius.com/Lizzo-jerome-lyrics
Now getting the lyrics from https://genius.com/Lizzo-cry-baby-lyrics
Now getting the lyrics from https://genius.com/Lizzo-tempo-lyrics
Now getting the lyrics from https://genius.com/Lizzo-exactly-how-i-feel-lyrics
Now getting the lyrics from https://genius.com/Lizzo-better-in-color-lyrics
Now getting the lyrics from https://genius.com/Lizzo-heaven-help-me-lyrics
Now getting the lyrics from https://genius.com/Lizzo-lingerie-lyrics
Now getting the lyrics from https://genius.com/Lizzo-boys-lyrics
Now getting the lyrics from https://genius.com/Lizzo-truth-hurts-lyrics
Now getting the lyrics from https://genius.com/Lizzo-water-me-lyrics
Now getting the lyrics from https://genius.com/Lizzo-good-as-hell-remix-lyrics
```

```
In [9]: print(lyrics[3])
```

```
[<div class="Lyrics__Container-sc-1ynbvzw-6 YYrds" data-lyrics-container="true">[Intro]<br/>True story<br/>No glory, let's go<br/><br/>[Verse 1]<br/><a class="ReferentFragmentdesktop__ClickTarget-sc-110r0d9-0 hYSBRZ" href="/16984532/Lizzo-soulmate/Yeah-the-old-me-used-to-love-a-gemini-like-a-threesome-fuckin-with-him-every-night-a-lotta-two-faced-people-show-me-both-sides-so-i-figured-out-i-gotta-be-my-own-type"><span class="ReferentFragmentdesktop__Highlight-sc-110r0d9-1 jAzSMw">Yeah, the old me used to love a Gemini<br/>Like a threesome, fuckin' with him every night<br/>A lotta two-faced people show me both sides<br/>So I figured out I gotta be my own type</span></a><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span><span><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span></span><br/><br/>[Pre-Chorus 1]<br/><a class="ReferentFragmentdesktop__ClickTarget-sc-110r0d9-0 hYSBRZ" href="/25844455/Lizzo-soulmate/They-used-to-say-to-get-a-man-you-had-to-know-how-to-look-they-used-to-say-to-keep-a-man-you-had-to-know-how-to-cook"><span class="ReferentFragmentdesktop__Highlight-sc-110r0d9-1 jAzSMw">They used to say to get a man, you had to know how to look<br/>They used to say to keep a man, you had to know how to cook</span></a><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span><span><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span></span><br/><a class="ReferentFragmentdesktop__ClickTarget-sc-110r0d9-0 hYSBRZ" href="/17467866/Lizzo-soulmate/But-im-solo-in-soho-sippin-soju-in-malibu"><span class="ReferentFragmentdesktop__Highlight-sc-110r0d9-1 jAzSMw">But I'm solo in Soho, sippin' Soju in Malibu</span></a><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span><span><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span></span><br/>It's a me, myself kinda attitude<br/><div class="InreadContainer__Container-sc-19040w5-0 cuJbPY PrimisPlayer__InreadContainer-sc-1tvdtf7-0 ju0VWZ" data-exclude-from-selection="true"><div class="PrimisPlayer__Container-sc-1tvdtf7-1 csMTdh"></div></div>[Chorus]<br/>'Cause I'm my own soulmate (Yeah, yeah)<br/>I know how to love me (Love me)<br/>I know that I'm always gonna hold me down<br/>Yeah, I'm my own soulmate (Yeah, yeah)<br/>No, I'm never lonely (Lonely)<br/><a class="ReferentFragmentdesktop__ClickTarget-sc-110r0d9-0 hYSBRZ" href="/17124343/Lizzo-soulmate/I-know-im-a-queen-but-i-dont-need-no-crown"><span class="ReferentFragmentdesktop__Highlight-sc-110r0d9-1 jAzSMw">I know I'm a queen, but I don't need no crown</span></a><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span><span><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span></span><br/>Look up in the mirror like "Damn, she the one"<br/><br/>[Post-Chorus]<br/>One, one, one, one, one, one (The one)<br/>Like "Damn, she the one"<br/>One, one, one (The one)<br/>Bad bitch in the mirror like "Yeah, I'm in love"<br/>Love, love, love, love, love, love (In love)<br/>Like "Yeah, I'm in love"<br/>Love, love, love (In love)<br/>Look up in the mirror like "Damn, she the one"<br/><br/>[Verse 2]<br/>And she never tell me to exercise<br/>We always get extra fries<br/>And you know the sex is fire<br/>And I gotta testify<br/>I get flowers every Sunday<br/><a class="ReferentFragmentdesktop__ClickTarget-sc-110r0d9-0 hYSBRZ" href="/16979790/Lizzo-soulmate/Ima-marry-me-one-day"><span class="ReferentFragmentdesktop__Highlight-sc-110r0d9-1 jAzSMw">I'ma marry me one day</span></a><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span><span><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span></span><br/></div>, <div class="Lyrics__Container-sc-1ynbvzw-6 YYrds" data-lyrics-container="true">[Pre-Chorus 2]<br/>True love ain't something you can buy yourself<br/>True love finally happens when you by yourself<br/>So if you by yourself, then go and buy yourself<br/><a class="ReferentFragmentdesktop__ClickTarget-sc-110r0d9-0 hYSBRZ" href="/17158285/Lizzo-soulmate/Another-round-from-the-bottle-on-the-higher-shelf"><span class="ReferentFragmentdesktop__Highlight-sc-110r0d9-1 jAzSMw">Another round from the bottle on the higher shelf</span></a><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span><span><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span></span><br/><br/>[Chorus]<br/>'Cause I'm my own soulmate (Yeah, yeah)<br/>I know how to love me (Love me)<br/>I know that I'm always gonna hold me down<br/>(Always gonna hold me down)<br/>Yeah, I'm my own soulmate (Yeah, yeah)<br/>No, I'm never lonely (Lonely)<br/>(No, I'm never lonely, no)<br/><a class="ReferentFragmentdesktop__ClickTarget-sc-110r0d9-0 hYSBRZ" href="/17124343/Lizzo-soulmate/I-know-im-a-queen-but-i-dont-need-no-crown"><span class="ReferentFragmentdesktop__Highlight-sc-110r0d9-1 jAzSMw">I know I'm a queen, but I don't need no crown</span></a><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span><span><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span></span><br/>Look up in the mirror like "Damn, she the one"<br/><br/>[Post-Chorus]<br/>One, one, one, one, one, one (The one)<br/>(She the one, she the one)<br/>Like, "Damn, she the one"<br/>(She the one, she the one, she the one)<br/>One, one, one (The one)<br/>Bad bitch in the mirror like, "Yeah, I'm in love"<br/>(I'm in love, I'm in love, I'm in love)<br/>Love, love, love, love, love, love (In love)<br/>(I'm in love, I'm in love)<br/>Like "Yeah, I'm in love"<br/>(I'm in love, yeah, I'm in love)<br/>Love, love, love (In love)<br/>Look up in the mirror like "Damn, she the one"<br/></div>, <div class="Lyrics__Container-sc-1ynbvzw-6 YYrds" data-lyrics-container="true">[Bridge]<br/>I'm the one (Ooh)<br/>I'm the one (You're th
```

```
e one)<br/>I'm the one (Oh, you're the one)<br/>Look up in the mirror like "Damn, she the one"<br/>One, one, one, one, one, one<br/>Like "Damn, she the one"<br/>One, one, one, one<br/>Bad bitch in the mirror like "Yeah, I'm in love"<br/>Love, love, love, love, love, love (In love)<br/>(I'm in love, yeah, I'm in love)<br/>Love, love, love (In love)<br/>Look up in the mirror (Ow)<br/><br/>[Chorus]<br/>'Cause I'm my own soulmate (Yeah, yeah)<br/>I know how to love me (Love me)<br/>I know that I'm always gonna hold me down<br/>Look up in the mirror like "Damn, she the one"<br/>Yeah, I'm my own soulmate (Yeah, yeah)<br/>No, I'm never lonely (Lonely)<br/><a class="ReferentFragmentdesktop__ClickTarget-sc-110r0d9-0 hYSBRZ" href="/17124343/Lizzo-soulmate/I-know-im-a-queen-but-i-dont-need-no-crown"><span class="ReferentFragmentdesktop__Highlight-sc-110r0d9-1 jAzSMw">I know I'm a queen, but I don't need no crown</span></a><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span><span><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span><span style="position:absolute;opacity:0;width:0;height:0;pointer-events:none;z-index:-1" tabindex="0"></span></span><br/>Look up in the mirror like "Damn, she the one"</div>]
```

In []:

Problem 6.

In [10]:

```
useragent_url = 'https://httpbin.org/user-agent'
r = requests.get(useragent_url)
useragent = json.loads(r.text)['user-agent']
```

In [17]:

```
headers = {'User-Agent': useragent,
           'From': 'tby8aj@virginia.edu'}

root = 'https://maps.spirithalloween.com'

endpoint = '/api/getAsyncLocations'

myparams = {'template':'domain','level':'domain','lat':40.380028, 'lng':-97.910156, 'radius':1800, 'limit':1500}

r = requests.get(root + endpoint,
                 headers = headers,
                 params = myparams)

myjson = json.loads(r.text)
spirit = pd.json_normalize(myjson, record_path = ['markers'])
spirit
```

Out[17]:

	lat	lng	tooltip	info	iconURL	clickable	locationId	specialties
0	40.931555	-98.380914		<div class="tlsmap_popup"> <div class="map-popu...	//assets.spirithalloween.com/images/mapIconSm.gif	True	1323	None
1	40.785642	-96.653696		<div class="tlsmap_popup"> <div class="map-popu...	//assets.spirithalloween.com/images/mapIconSm.gif	True	2125	None
2	40.862625	-96.681946		<div class="tlsmap_popup"> <div class="map-popu...	//assets.spirithalloween.com/images/mapIconSm.gif	True	2837	None
3	39.180132	-96.556783		<div class="tlsmap_popup"> <div class="map-popu...	//assets.spirithalloween.com/images/mapIconSm.gif	True	957	None
4	41.229689	-96.132165		<div class="tlsmap_popup"> <div class="map-popu...	//assets.spirithalloween.com/images/mapIconSm.gif	True	925	None
...
1443	44.311523	-69.785648		<div class="tlsmap_popup"> <div class="map-popu...	//assets.spirithalloween.com/images/mapIconSm.gif	True	2537	None
1444	44.835581	-68.743224		<div class="tlsmap_popup"> <div class="map-popu...	//assets.spirithalloween.com/images/mapIconSm.gif	True	2181	None
1445	45.940145	-66.663654		<div class="tlsmap_popup"> <div class="map-popu...	//assets.spirithalloween.com/images/mapIconSm.gif	True	2333	None
1446	46.114718	-64.83408		<div class="tlsmap_popup"> <div class="map-popu...	//assets.spirithalloween.com/images/mapIconSm.gif	True	2323	None
1447	44.654522	-63.60234		<div class="tlsmap_popup"> <div class="map-popu...	//assets.spirithalloween.com/images/mapIconSm.gif	True	3205	None

1448 rows × 8 columns

In [18]:

```
mylist = [BeautifulSoup(x).find_all('div', attrs={'class': 'address-two'})[0].text for x in spirit['info']]
mylist
```

```
Out[18]: ['Near Dillards',  
          'Former Hobby Lobby',  
          'Next to Famous Footwear',  
          'Former GAP',  
          'Former Dress Barn',  
          'Next to Hibbet Sports',  
          'Former American Eagle',  
          'Next to Ulta',  
          'next to Ross',  
          'Former Burlington',  
          'Former Goodys',  
          'Next to B B theater',  
          'Next to Younkers',  
          'Next to Spectrum Paint',  
          'Former Colemans',  
          'Former Best Buy',  
          'Next to Dicks Sporting Goods',  
          'Former Express',  
          'Next to JC Penney',  
          'Former Pier 1',  
          'next to Childrens Mercy North',  
          'Former Office Depot',  
          'Next to TJ Maxx',  
          'Former Hometown Flooring',  
          'Next to Dollar Tree',  
          'Former Gordmans',  
          'Next to Five Guys',  
          'next to T Mobile',  
          'Former Gordans',  
          'Across from Target',  
          'Former Charming Charlies',  
          'Former Younkers',  
          'Next to Younkers',  
          'Former JC Penneys',  
          'Behind Popeyes Chicken',  
          'Former Forever 21',  
          'Next to Subway',  
          'Former Natural Grocer',  
          'Former Sears',  
          'Next to Dollar General',  
          'Next to Mathis Brothers',  
          'Former Dress Barn',  
          'Across from Reasors',  
          'Former Old Navy',  
          'Former Party City',
```

'Next to Natural Grocer',
'Next to Lowes',
'Next to JC Penney',
'Next to PhoEver',
'Former Sears',
'Next to JoAnn Fabrics',
'Former Pier 1',
'Next to Goodwill',
'Across from Michael Kors',
'Next to Ollies',
'Former Rack Room Shoes',
'Next to Taco Bueno',
'Former Younkers',
'Former Petco',
'Next to Post Office',
'Next to Urban Air',
'Next to Journeys',
'Northwest Arkansas Mall - former Abercrombie',
'Former Gym',
'Former Office Max',
'Cheyenne Plaza',
'Next to Sierra',
'Former Charming Charlies',
'Former Schnidermans',
'Former Aspire Fitness',
'Near Sportsmans Warehouse',
'Former Bed Bath Beyond',
'Former Pier One',
'Former Herbergers',
'Former H M',
'Next to Michael Kors',
'Former Bed Bath Beyond',
'Former Dress Barn',
'Former Office Depot',
'Former Pier One',
'Next to Big Lots',
'Former Harbor Freight',
'Former Performance Bicycle Shop',
'Former Fitness 19',
'Former Gordmans',
'Next to Fast Signs',
'Former Pier 1',
'Former Planet Fitness',
'Former Charming Charlies',
'Discounted Home',

'Old Navy',
'Next to Jeweler Design Studio',
'Next to Bath Body Works',
'Former Halloween City',
'Next to Napa Auto Parts',
'Former Big Thrill Factory',
'Former K-Mart',
'Former Total Wine',
'Former Golfsmith',
'Former Gordmans',
'Next to Petco',
'Near LL Bean and Forever 21',
'Next to Banfield Pet Hospital',
'Former Toby Keiths Restaurant',
'Next to Davids Bridal former Rogans Shoes',
'Next to BC Restaurant',
'Next to Moline High School behind Subway',
'Former Dress Barn',
'Near Friday's and Hampton Inn',
'Former TJ Maxx',
'Next to Sheridan Inn',
'Next to Thunderbird Lanes',
'Former Shop N Save by Lowes',
'Former Pier 1',
'Former H M',
'Former Office Depot',
'Next to JoAnn Fabric',
'Former Party City',
'Next to Salvation Army',
'Former OReilly Auto Parts',
'Former Pier One',
'Former Pier 1',
'Hobby Lobby Center',
'Ashley Furniture',
'Next to Burlington',
'Next to Ollies Bargain Outlet',
'Amarillo Westgate Mall former Bealls',
'Former Restoration Hardware',
'Kohls Center',
'Former Shop N Save',
'Shop N Save',
'Next to Ollies Save-A-Lot',
'Former Nike Outlet',
'Former Dress Barn',
'Former Sears',

'Former Sports Authority',
'Next to Mattress Outlet',
'Next to UPS',
'Former Gordmans',
'Next to Jenny Craig',
'Former Gordmans',
'Former Old Navy',
'Former JC Penney',
'Former Macys',
'former Herbergers',
'Former Wallys',
'Former Macys',
'Conway Towne Center - former Sears',
'Former NY Co',
'Former Pier 1',
'Former Furniture Row',
'Former Choice furniture',
'Next to Dicks Sporting Goods',
'Former Sears Hometown',
'Former Sears',
'Former TJMaxx',
'Lakewood Village Center - former Dressbarn',
'Former Sears',
'Former WG R Express',
'Across from Bed Bath Beyond',
'Former Macys',
'Former Macys',
'Clovis',
'Glenwood Springs Mall',
'Former-Pier 1-Highlands Caraway',
'Next to Jennys Donuts',
'Former Mattress Firm',
'Next to Bed Bath Beyond',
'Next to Shoal Creek Tavern',
'Former Party City',
'Former DSW',
'Former TJMaxx',
'Dress Barn',
'Former Shopko',
'Former Office Depot',
'Next to Conns',
'Next to Target',
'Next to Goodwill',
'Former Pet Sense',
'Former Party City',

'Former Charter Fitness',
'Former Old Navy',
'Former Toys R Us',
'Former Famous Footwear',
'Former Aarons',
'Former Stein Mart',
'Former Office Max',
'Former Rockler',
'Former Pier One',
'Former Charming Charlies',
'Former CVS',
'Former Beauty Supply',
'Former Party City',
'Former Covid Center',
'Former Ben Franklin Crafts',
'Former HH Gregg',
'Former Dollar Western Wear',
'Former Bealls',
'Former Tuesday Morning',
'Former Dress Barn',
'Former Lifeway',
'Former Olsons Floor Covering',
'Next to Macys',
'Next to Ulta',
'Former Marshalls',
'Former Carsons Furniture',
'Next to Sports Clips',
'Former Off 5th',
'Former Fresh Market',
'Former Hobby Lobby',
'Next to Pet Smart',
'Target Center',
'Former Sports Authority',
'Former GamerTagz',
'Former Joanns',
'Former Dress Barn',
'Former Bed Bath Beyond',
'Across from Express Kay Jewelers',
'Former Hallmark',
'Former Old Navy',
'Former Charming Charlies',
'Former Paws Claws',
'Former SteinMart',
'Former World Market',
'Former 5 Below',

'Former Stein Mart',
'Former Dollar Store',
'Former Saks',
'Former Sears Appliance',
'Former Fitness Center',
'Former Bass Shoe Factory Outlet',
'Former HH Gregg',
'Former Bed Bath and Beyond',
'Former TJMaxx',
'Next to Binnys',
'Former Party City',
'Next to Walmart',
'Former Tuesday Morning',
'Next to Whole Foods',
'Former Kohls',
'Former Pier One',
'Former Party City',
'Former Office Depot',
'Former Bed Bath Beyond',
'Former OM',
'Former Foot locker',
'Former Hobby Lobby',
'Former Forever 21',
'Former Dress Barn',
'Former DSW Shoes',
'Fomer Dollar General',
'Former Nike Outlet',
'Former Discovery',
'Former Sears Appliance',
'Former The Room Place',
'Abilene - Brookhollow',
'Former Macys',
'Bosque River Centre next to Maurices',
'Commons at WolfCreek - next to Dave Busters',
'Ridgeway Trace Center by Best Buy',
'Rite Aid',
'Southaven Towne Center - former Lifeway Store',
'Near Super 1',
'Former Sutherlands Lumber',
'Former Dress Barn',
'Former Pier 1',
'Next to Burkes',
'Former JCPenney',
'Next to Home Depot',
'Former Dress Barn',

'Carriage Crossing Marketplace- next to Old Navy',
'Former Sears',
'Next to Dicks Sporting Goods',
'Former Sleep Outfitters',
'Pier 1',
'Next to Dirt Cheap',
'Old Hickory Mall - former Macys',
'Former Home Zone',
'Near Burlington',
'Former Burkes Outlet',
'Coors Bypass and Ellison in Cottonwood corners',
'Next to Ollies',
'Corner of 4th st NW and Osuna NW',
'Former 511 Tactical',
'Across from Walmart On Carlisle Ave',
'Former Staples',
'Roswell Mall next to Galaxy 8 - Allen Theatres',
'Former Advantage Furniture',
'Former Builders First Source',
'Former Office Depot',
'Former Forever 21',
'Former Citi Trends',
'Next to Edward Jones',
'Former The Avenue',
'Former Sears',
'former Dress Barn',
'Former Marsh',
'Former Fashion Bug',
'Kings Center Mall',
'Next to Home Depot',
'Former HH Gregg',
'Behind CVS',
'Governors Square Mall TENT',
'Northview Plaza Shopping Center',
'Former Toys R Us',
'Former Fashion Bug',
'Former Stein Mart',
'Odessa Chimney Rock',
'Sunset Mall - Former Bealls',
'Next to Clairs',
'Next to Planet Fitness',
'Former Clothing Store',
'Former Violets',
'Former JcPenney',
'Next to Harbor Freight',

'Next to Planet Fitness',
'Lufkin Mall - former Sears',
'Former Sears',
'Behind Shopko',
'Former Thrifty Jeans',
'Next to Chuck E Cheese',
'Located next to Hobby Lobby',
'Barnes Crossing former Sears',
'Former EQ3 Furniture store',
'Former XS Cargo',
'Former Carsons',
'Former Lifeway Christian',
'Next to Burlington',
'Former Gordmans',
'Steinmart',
'Former Tuesday Morning',
'Pier 1',
'Next to Off Broadway shoes',
'Former Sears Location',
'Next to Dave and Busters',
'Former Hibbett',
'Former Annies Attic',
'Former Gordmans',
'former Burkes Outlet',
'Tejas Center next to JoAnn Fabrics',
'Former CVS',
'former Steinmart',
'Former Ultimattress',
'Next to American Eagle Outfitters',
'Former Performance Bicycle',
'Next to TJ Maxx',
'Former Mattress Firm',
'Former Forever 21',
'Next to Michaels',
'Former Marks',
'Former Neiman Marcus Last Call',
'Former Klingmans',
'Former Orange Coworking',
'Near Maurices',
'Next to Dollar Tree',
'Former Palais Royal',
'Fashion Bug',
'Former Tuesday Morning',
'Located next to TJ Max',
'Former Dollar General',

'Next to DSW',
'Located next to Schuler Books',
'Former Pier 1',
'Forever 21',
'Old Davids Bridal',
'Former Bealls',
'Former Randalls',
'Former Petsmart',
'Next to DSW',
'Formerr Elder Beerman',
'Former Elder-Beerman',
'Former Sears',
'Former Rite Aid',
'Behind University Mall',
'Former JC Penney',
'Former Petco',
'Near DSW Shoes',
'Next to Big Lots',
'Former Famous Footwear',
'next to Sears',
'Former Family Christian',
'Salvation Army',
'Former Sears',
'Near Rue 21 Mall Entrance',
'Former Stein Mart',
'Former Shoe Carnival',
'Bealls Outlet Store',
'Former Dress Barn',
'Next to Sears',
'Former Bed Bath Beyond',
'Former Staples',
'Former Pottery Barn',
'Beavercreek',
'Next to Petsmart',
'Next to Barnes Noble',
'Former Aarons',
'Macys',
'Across from H M',
'Former Lifeway',
'Former Bennion Crafts',
'Former Goodwill',
'Next to Ulta',
'Former Bed Bath Beyond',
'Gen X Clothing',
'Former Rue 21',

'Former Bed Bath Beyond',
'Near Starbucks',
'Former Art Van Furniture',
'Next to Zumiez',
'Next to Alter States',
'next to ROSS on Dowlen Road',
'Lane Bryant',
'Former Fashion Bug',
'Former Rue 21',
'Former Tuesday Morning',
'Telshore Blvd By Outback Steakhouse',
'Former Charming Charlies',
'Former Dicks Sporting Goods',
'Former Charlottes Saddlery',
'Former Sam Moon',
'Former Forever 21',
'Located next to Summit Sports',
'Former Party City',
'Former Sears',
'Former Sears Space',
'Former Sears',
'Former Bealls',
'Former Mens Warehouse',
'Next to Burlington former MJM Shoes',
'Former World Market',
'Located next to CVS',
'Next to Maurices',
'Former Elder Beerman',
'Former JoAnn',
'Former Office Depot',
'Former Michaels',
'Next to Hobby Lobby',
'Park North',
'Former Best Buy',
'Near JC Penney',
'Former Dicks',
'Former Family Christian Book Store',
'corner of McNeese and Ryan',
'San Antonio Quarry market',
'Former Loves Furniture',
'Furniture World',
'Former Palais Royal',
'Former Annas Linens I-10 and 410 Wonderland Mall',
'Located across from Old Navy',
'Former Stein Mart',

'Central Mall - former Stage',
'Former Goodwill',
'Former Mattress Firm',
'Bonita Lakes Crossing next to JoAnn Fabrics',
'Former Albertsons',
'At the intersection of McFarland Blvd N and Waterm',
'Former Salvation Army',
'Corner of Joe Battle and Montwood Dr.',
'Former Staples',
'Former E G Amish Furniture',
'Former TJ Maxx',
'Former Party City',
'Former Bealls',
'Former Mazda Dealership',
'Former Bed Bath and Beyond',
'Former North Face in Outlet Shoppes of El Paso',
'Next to Five Below',
'Former Sears',
'Former Jo Anns Fabrics',
'Former Salvation Army',
'Former Stein Mart - North Mesa St. Remcon',
'Former Melrose',
'Former Payless',
'Next to Goodwill',
'Former Sears',
'Former Pier One',
'Home Depot Center',
'Former Buybuy Baby',
'Former Dollar General',
'Down East',
'Former Furniture Liquidators',
'Former Sears Home Store',
'Next to Bed Bath and Beyond',
'Next to Five Below',
'Located next to JoAnn Fabrics',
'Next to Dollar General',
'Former Kirklands',
'former Mattress Firm',
'Former Buckle',
'Next to Big Lots',
'Former Sears',
'Next to Mall Entrance',
'Former CVS',
'Former Ulta',
'Near Burlington ULTA',

'Next to Ollies',
'Roadhouse Centre',
'Next to Cattlemans',
'Pennant Sports',
'Former Giant Eagle',
'Former Family Dollar',
'Next to Big Sandys',
'Next to American Home Fitness',
'next to Davids Bridal',
'Next to Marshalls',
'Former LifeWay',
'Former Staples',
'Located next to TJ Max',
'Village Plaza- Zehrs Centre',
'Former Aldi',
'Former Payless Shoes',
'Former Ashley Furniture',
'Next to Aldi',
'Victoria Mall former Bealls',
'Next to JCPenney',
'Former Babies R Us',
'Former Hibbett Sports',
'Former Drug Mart',
'near Hammond Square across from Jimmy Johns',
'Former Chuck E Cheese',
'Next to Tuesday Morning',
'Former Elder Beerman',
'Next to Dollar Tree',
'Former TRU',
'Former Dress Barn',
'Former Kroger',
'Located behind Arbys',
'Mandeville - Premier Center',
'Mall de las Aguilas',
'Near AMC Theater',
'Former Teppermans Furniture',
'Former Trampoline Place',
'TENT - Next to Mall',
'Former Shopko next to Jo Anns Fabrics',
'Former Best Buy',
'Former Shopko',
'Midtown Square - next to Office Depot',
'former Party City in Kenner',
'Next to Home Depot',
'Former MC Sports',

'Winn Dixie Shopping Center next to Roses',
'Luling - former Sears Hometown',
'Former Half Price Books',
'Former Deb',
'Big Lots Center',
'Former Sears',
'Next to Dunhams',
'Former Toys R Us',
'former Kirklands',
'Former Spring',
'Former CVS',
'Next to Ollies',
'Former Best Buy',
'Former Ethan Allen',
'Former Charming Charlie',
'Next to Ace Hardware',
'Next to Petco',
'Former Tuesday Morning',
'Next to Milestone Athletic Club',
'Former HH Gregg',
'Former Earth Fare',
'Former Bed Bath Beyond',
'Former CVS',
'Tuesday Morning Center',
'Former Ollies',
'Former Elder Beerman',
'Former Toys R Us',
'Former Bed Bath and Beyond',
'Former Office Depot',
'Former Saks 5th Ave',
'Former Steinmart',
'Former Shopko',
'Next to Fitness Center',
'Next to Walmart',
'Former Goodwill',
'Value City Furniture Center',
'Former Kroger',
'Unknown',
'Near Dollar Tree',
'Former Dollar Tree',
'by Ace Hardware',
'Former Dicks Sporting Goods',
'Next to Planet Fitness',
'SPID and Weber Road',
'Hyland Centre Near Best Buy',

'Former JC Penney',
'Near Staples',
'Former Bealls',
'Former Rooms to Go',
'Former Rite Aid',
'American Freight',
'Former DSW',
'Next to Post Office',
'Former Barnes Noble',
'former Factory to U',
'Former Cato',
'Next to Pitt Boss BBQ',
'Former Dillards',
'Former Sears',
'Next to Shoe Carnival',
'Former YouFit',
'Next to Floor Decor',
'Former BBQ Island',
'Inline Location',
'Next to T-Mobile',
'Former Walgreens',
'Former Dress Barn',
'Former Hi-Health',
'Former Dillards 735',
'Next to Bed Bath and Beyond',
'Former YouFit',
'Former Stein Mart',
'Former Pet Club',
'Former Pier 1 Imports',
'Former Z Gallery',
'Medicine Hat Mall',
'Former Planet Fitness',
'Former Bealls',
'Between Sunnys and Lotus Market',
'Next to Red Robin',
'Former Big Lots',
'Former Lane Bryant',
'Former Hi-Health',
'Former Sports Authority',
'Next to Mall Entrance',
'Former Office Max',
'Next to Santisi Brothers',
'Former Kirklands',
'Midtown Plaza',
'Former Kmart',

'Former Stein Mart',
'Former At Home',
'former SLFGA Liquor Store',
'Former Staples',
'Former Sears',
'Former Stein Mart',
'Next to The Dump',
'Next to Jerseys',
'Former Bealls',
'Former Pure Fitness',
'Former Forever 21',
'Between Whole Foods and Bookmans',
'Former Joanns',
'Former Dress Barn',
'Former Bealls Outlet',
'Former Dress Barn',
'Longstaff st. Next to Diamond Bar Meats',
'Next to Kohls',
'Former Sears',
'',
'Lowes Plaza',
'Next to Ross',
'Former Gabes',
'Former Shoe Warehouse',
'Former Jo Ann Fabrics',
'Pet Value',
'Former Party City',
'Former AC Moore',
'Next to Dollarama',
'Former Sears',
'Former Sears',
'Former American Eagle',
'Former Rite Aid',
'Former Elder Beerman',
'Next to Papa Johns',
'Former Charming Charlie',
'Next to Stride Rite',
'Former Old Navy',
'Former Tri-State Surplus',
'Next to Ollies Bargain Outlet',
'Next to Massage Envy',
'Former Victorias Secret',
'Near Boston Pizza',
'Outside entrance near Bed Bath and Beyond',
'Next to fitness 19',

'Former Sport Chek',
'Former No Frills',
'Former Dunhams Sports',
'Former JC Penney',
'Former Fallas',
'Next to Staples',
'Former Fitness Equipment',
'Former Fabricland',
'Right of LL Flooring',
'Next to Canadian Appliance Source',
'Near Shoe Dept Encore Justice',
'Next to Kroger',
'Former Sears',
'Next to Starbucks',
'Former Hastings',
'Next to Ice Arena',
'Former Chapters',
'Former HCD Furniture',
'Former Aarons',
'IN PARKING LOT',
'behind A W and Arbys',
'Former Lazboy',
'Former Sears',
'Former National Sports',
'Former Gordmans',
'Race Track exit on Boulder Hwy Former KMart',
'Former Pier 1',
'Former Sears',
'Next to the Movie Theater',
'Former Fallas Paredes',
'Former Pier One Mountain View Plaza',
'Former Lil Niblets',
'Gates of York Plaza',
'Former LCBO',
'Former Big Lots - Corner of Craig Revere',
'Warm Springs rd and Green Valley Pky',
'Former DOT Furniture',
'Former Lacks Furniture',
'Corner of Ann and Decatur former CVS',
'Former TJ Max Maryland Pky and E Flamingo',
'Former Sears at Meadows Mall',
'Next to Walmart',
'Former Big Lots on 10th St.',
'Between PetSmart and Hobby Lobby',
'Former Bed Bath Beyond',

'Montecito Crossing Durango and the 215',
'Former Home Town Sears',
'By Black Bear Diner',
'Former Kmart',
'X Gordmans',
'Plaza DeColores Former Furniture Royal',
'Next to Dollar Tree',
'Former Penningtons',
'Next to Piggly Wiggly',
'Former Smiths Supermarket',
'Former RV Center',
'Charleston Ave next to Home Depot',
'Former Rite-Aid',
'Shoe Carnival',
'Old Country Plaza',
'Former Jainlee',
'Former Pier 1',
'Pickering Ridge',
'Former Charming Charlies',
'Next to Lowes Food',
'Former Sears',
'Back of Mall on Walmart side',
'Former Habitat for Humanity',
'Former Dollar General',
'Next to Walmart',
'Former Bon Ton',
'Former Michaels',
'Former Pottery Barn',
'Next to Hamricks',
'Former Wolf Furniture',
'Former Bed Bath Beyond',
'Former Sears at the Oshawa Centre',
'Next to Food Lion',
'Former Sears',
'AC Moore',
'Former Marshalls',
'Former Gap',
'Former Sears',
'Former Bonton next to Kohls',
'Former Save A Lot',
'Next to Little Gym',
'Former Home Furniture',
'Harrisonburg',
'Former Sears',
'Beside Burlington',

'Former Dillards',
'Former Advanced Auto Parts',
'Near Food Lion',
'Next to Ashley Furniture',
'Next to Planet Fitness',
'Former Sears',
'Former Kmart',
'Formerly JC Penney',
'Next to Post Office',
'Former Pier 1',
'Former KMart',
'Former Sears',
'Former JC Penney',
'Former Sears',
'Former Sears',
'Former JCPenney',
'Next to Giant',
'Next to Harbor Freight',
'Next to Brick Furniture',
'Former Bed Bath Beyond',
'Burlington Coat Factory',
'Former Dollar General',
'Former BMC Lumber Yard',
'Former Bed Bath Beyond',
'Former Fashion Bug',
'Former Bed Bath and Beyond',
'Next to HomeSense',
'Next to Once Upon A Child',
'Next to HomeSense',
'Former Reebok',
'Former CVS',
'Former Bowling Alley',
'Former Olympia Sports',
'Former Optigolf',
'Former Tuesday Morning',
'Former Reno Depot',
'Former JCPenney',
'Former Peebles',
'Former Pier 1',
'Former Toys R Us',
'Dress Barn',
'Former Petco',
'Former Pier 1',
'Former Stein Mart',
'Former Bed Bath Beyond',

'Former Sears',
'Former Babies R Us',
'Former Sears',
'Former Pier One',
'Former Albertsons',
'Near Bealls Outlet',
'Next to Planet Fitness',
'Former Old Navy',
'Former Tuesday Morning',
'Between Pier 1 Imports Skyzone',
'Stein Mart',
'Former Gap',
'Left of Abba Worship',
'Former Saks OFF 5th',
'Former Party Co',
'Fayetteville - Cross Creek Plaza',
'Former Partytime',
'Next to Once Upon a Child',
'Former Lee Furniture',
'Next to Kohls',
'Former Marshalls',
'Former Dollar General',
'Former Charming Charlies',
'Former Bon Ton',
'Former Dress Barn',
'Bon Ton',
'Former Lord and Taylor',
'Next to Boscovs',
'Former Office Depot',
'Next to Petland Carolinas',
'Next to Juice Bar',
'Former Pier One',
'Former Golf Galaxy',
'Former Pier One',
'Former Desert Warehouse Furniture',
'Former Sears',
'Former Sears',
'Former Pier 1',
'Former Party City',
'Former Sears',
'Former CVS',
'Former A.C. Moore',
'Former Dress Barn',
'Former Office Depot',
'Former Old Navy',

'Former Mattress Firm',
'Former Great Outdoors',
'Former Pier 1',
'Former Peavy Mart',
'Former Gander Outdoors',
'Former Sears',
'Former ACMoore',
'Former Sears',
'Former Forever 21',
'Former DSW',
'Former Monkey Joes',
'Former Annas Linens',
'Beside Staples',
'Former Planet Fitness',
'Former Pier One',
'Former PetSmart',
'Former Office Depot',
'Former Total Wine',
'Former Pier One',
'Former AC Moore',
'Former BiLo',
'Former Sears',
'Former Rite Aid',
'Former Bed Bath Beyond',
'Next to GameStop',
'Former DSW Shoes',
'Former Dicks Sporting Goods',
'Next to Conns',
'Former Home Goods',
'Former Michaels next to Game Stop',
'Former Bon Ton',
'Former Party City',
'Former Pier One',
'Former K Mart',
'Former Pet Supplies Plus',
'Former WSS by Home Depot',
'Former Office Max',
'Former SteinMart',
'Former Charlotte Russe',
'Former Tuesday Morning',
'Toys R Us',
'Former Rite Aid',
'Next to Aldi',
'Former Rack Room Shoes',
'Former Office Depot',

'Former K-Mart',
'Former Fashion Joy',
'Former Bed Bath Beyond',
'Bon Ton',
'Former Sears',
'Behind PetSmart',
'Former Bealls',
'Former Rite Aid',
'Former CVS',
'Former Big Lots',
'Former Gap',
'Former Sears',
'Former Rack Room Shoes',
'Former Fallas',
'Former Fine Wine Spirits',
'Next to Office Depot',
'Former Office Depot',
'Former Stein Mart',
'Next to Swoozies',
'Former Kmart',
'Former Party City',
'Next to Big Lots',
'Former HH Gregg',
'Former Sports authority Across from Sams Club',
'Former Saks Fifth Avenue',
'Former Bouclair Home',
'Former Party City',
'Former Dicks Sporting Goods',
'Next to DSW',
'Kantana Centrium - Across from Hockey Life on Corn',
'Former Steinmart',
'On the corner next to Petco',
'Former Just Cabinets',
'Next to Bed Bath Beyond',
'Former Pier One',
'Former Vons',
'Riverside',
'Former Steinmart',
'Former Fallas',
'Former Nike',
'Former Circuit City',
'next to Indigo Books-Gold Gym',
'Former Fashion Bug',
'Next to Hide N Seek Playground',
'Near Starlight Dos Lagos Luxury 15 Theaters',

'Former Tools R Us',
'Former Kahoots Pet Store',
'Next to Ross',
'Former Shoe City',
'Next to DSW',
'Former Famous Footwear',
'Former Book Store',
'Escondido Mall Tent',
'Next to Sally Beauty',
'Next to Produce Junction',
'Former H M',
'South Keys Mall',
'Former Kmart',
'Near Sakura Buffet',
'Beside 99 Store',
'Located next to Ross',
'Former Michaels',
'Former Pier 1',
'Rite Aid',
'Former Steinmart',
'Former Chuck E Cheese',
'Former Macys',
'Former Rock Thrift',
'Former H M next to spencers',
'Former 24 HR Fitness',
'Next to Hibachi Grill',
'Former JoAnns',
'Former Aarons',
'Former Family Dollar',
'Former Planet Fitness',
'Former Pier 1',
'Former Performance Bicycle',
'Next to Ollies',
'Former Staples',
'Former SteinMart',
'Former Blue Ridge Mt Sports',
'Former Thomasville Furniture',
'Former Dollar City Plus',
'Former Petco',
'Former Sears',
'By Dollar Tree',
'Former AAA Insurance',
'Former Pier One',
'Former Sitn Sleep',
'by Ollies and Redners',

```
'Next to Art Gallery',  
'Former Big Lots',  
'TENT - Next to Target',  
'Next to Express',  
'Next to Fitness 19',  
'TENT - Former Halloween Express',  
'Former Lowes',  
'Former KMart by LA Fitness',  
'Across from Burlington',  
'Former Hollywood Video',  
...]
```

```
In [19]: spirit_store_list_DF = pd.DataFrame()  
spirit_store_list_DF['mylist'] = mylist  
spirit_store_list_DF['mylist'].value_counts()
```

```
Out[19]: Former Sears                63  
Former Pier 1                      32  
Former Dress Barn                  25  
Former Office Depot                23  
Former Party City                  20  
..  
Former Buckle                      1  
Near Burlington  ULTA              1  
Roadhouse Centre                   1  
Next to Cattlemans                 1  
Halifax Forum - Maritime Hall      1  
Name: mylist, Length: 859, dtype: int64
```

```
In [ ]:
```