

In [1]:

```
pip install pyodbc
```

Requirement already satisfied: pyodbc in c:\users\marky\anaconda3\envs\geo_env\lib\site-packages (4.0.30)
Note: you may need to restart the kernel to use updated packages.

In [2]:

```
pip install openpyxl
```

Requirement already satisfied: openpyxl in c:\users\marky\anaconda3\envs\geo_env\lib\site-packages (3.0.3)
Requirement already satisfied: et-xmlfile in c:\users\marky\anaconda3\envs\geo_env\lib\site-packages (from openpyxl) (1.0.1)
Requirement already satisfied: jdcal in c:\users\marky\anaconda3\envs\geo_env\lib\site-packages (from openpyxl) (1.4.1)
Note: you may need to restart the kernel to use updated packages.

In [3]:

```
pip install sqlalchemy
```

Requirement already satisfied: sqlalchemy in c:\users\marky\anaconda3\envs\geo_env\lib\site-packages (1.3.16)
Note: you may need to restart the kernel to use updated packages.

In [4]:

```
from bs4 import BeautifulSoup

import pandas as pd

import pyodbc

import requests

import sqlalchemy
from sqlalchemy import create_engine

import time
```

In [5]:

```
# Read website into 'BeautifulSoup' to help parse the website
website_soup = BeautifulSoup(requests.get('https://vcnewsdaily.com/').text)
```

In [6]:

```
# Read funding activity and store into list
funding_activity = website_soup.find_all('div',class_='col-md-8')
```

In [7]:

```
# For loop to read in 'funding_activity', get information(date, headline, website link), and store into list
funding_activity_list = [[funding_activity[i].small.text,funding_activity[i].h5.text,funding_activity[i].a.get('href')] for i in range(len(funding_activity))]
df_funding_activity_list = pd.DataFrame(funding_activity_list,columns=['Funding Date','Funding Title','Website Article Link'])
df_funding_activity_list.head()
```

Out[7]:

	Funding Date	Funding Title	Website Article Link
0	08 May 2020	Jellyfish Announces \$12M Series A Funding	https://vcnewsdaily.com/jellyfish/venture-capi...
1	08 May 2020	Finite State Snares \$12.5M in Series A	https://vcnewsdaily.com/finite-state/venture-c...
2	08 May 2020	Limbox Lands \$9M Series A	https://vcnewsdaily.com/limbox/venture-capital...
3	08 May 2020	White Ops Receives Investment	https://vcnewsdaily.com/white-ops/venture-capi...
4	07 May 2020	Burst Oral Care Grabs \$20M in Funding	https://vcnewsdaily.com/burst-oral-care/ventur...

In [8]:

```
df_funding_activity_list.shape
```

Out[8]:

(30, 3)

In [9]:

```
# Articles from VC News Daily put into series
website_article_links = df_funding_activity_list['Website Article Link']
```

In [10]:

```
# Articles from our CSV download put into a series
website_article_links_total = pd.read_csv('VC_News_Daily.csv').drop(columns='Unnamed: 0')['Website Article Link']
```

In [11]:

```
# Checking for nnumber of Links that aren't already saved
website_article_links_length = len(website_article_links[~website_article_links_total.isin(website_article_links)])
website_article_links_length
```

Out[11]:

0

In [12]:

```
df_funding_activity_list = df_funding_activity_list[0:website_article_links_length]
df_funding_activity_list.head()
```

Out[12]:

Funding Date	Funding Title	Website Article Link
--------------	---------------	----------------------

In [13]:

```
funding_activity_funding=[]
if website_article_links_length > 0:
    for i in range(len(df_funding_activity_list)):
        company_text = BeautifulSoup(requests.get(website_article_links[i]).text)
        funding_headline = company_text.find_all(id='summary')[0].text.strip('\n').strip('\t')
        company_city = funding_headline.title().split(', ')[0]
        company_state = funding_headline.upper().split(', ')[1]
        website_company_link = company_text.find_all(id='fullArticle')[0].a.get('href')
        funding_article = company_text.find_all(id='fullArticle')[1].text
        funding_activity_funding.append([funding_headline,company_city,company_state,website_company_link,funding_article])

df_funding_activity_funding = pd.DataFrame(funding_activity_funding,columns=['Funding Headline','City','State','Website Company Link','Funding Article'])
df_funding_activity_funding.head()
```

Out[13]:

Funding Headline	City	State	Website Company Link	Funding Article
------------------	------	-------	----------------------	-----------------

In [14]:

```
df_funding_activity_funding.shape
```

Out[14]:

(0, 5)

In [15]:

```
website_company_links = df_funding_activity_funding['Website Company Link']
```

In [16]:

```
funding_company_list = []
if website_article_links_length > 0:
    for i in range(len(df_funding_activity_list)):
        company_info = BeautifulSoup(requests.get(website_company_links[i]).text)
        company_name = company_info.find(class_='col-md-7 d-flex align-items-center').h5.text
        company_desc = company_info.find(class_='mt-5').p.text
        funding_type = company_info.find_all('td')[-2].text.strip('\n')
        funding_amount = company_info.find_all('td')[-3].text.strip('\n')
        funding_company_list.append([company_name,company_desc,funding_type,funding_amount])

df_funding_company_list = pd.DataFrame(funding_company_list,columns=['Company','Company Description','Latest Funding Round','Latest Funding Amount'])
df_funding_company_list.head()
```

Out[16]:

Company	Company Description	Latest Funding Round	Latest Funding Amount
---------	---------------------	----------------------	-----------------------

In [17]:

```
df_funding_company_list.shape
```

Out[17]:

(0, 4)

In [18]:

```
df_funding_data = pd.concat([df_funding_company_list,df_funding_activity_list,df_funding_activity_funding], axis=1)
df_funding_data.head()
```

Out[18]:

Company	Company Description	Latest Funding Round	Latest Funding Amount	Funding Date	Funding Title	Website Article Link	Funding Headline	City	State	Website Company Link	Funding Article
---------	---------------------	----------------------	-----------------------	--------------	---------------	----------------------	------------------	------	-------	----------------------	-----------------

In [19]:

```
df_funding_data.shape
```

Out[19]:

(0, 12)

In [20]:

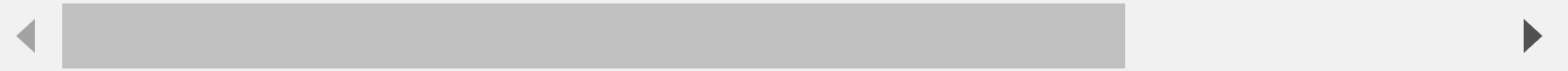
```
total_funding_data = pd.read_csv('VC_News_Daily.csv').drop(columns='Unnamed: 0')
```

In [21]:

```
total_funding_data.head()
```

Out[21]:

	Company	Company Description	Latest Funding Round	Latest Funding Amount	Funding Date	Funding Title	Website Article Link	Funding Headline	City	State	
0	Jellyfish	Jellyfish is the leading Engineering Management...	Series A	\$12,000,000	08 May 2020	Jellyfish Announces \$12M Series A Funding	https://vcnewsdaily.com/jellyfish/venture-capi...	BOSTON, MA, Jellyfish has raised over \$12 mill...	Boston	MA	http
1	Finite State	The mission of Finite State is to protect the ...	Series A	\$12,500,000	08 May 2020	Finite State Snares \$12.5M in Series A	https://vcnewsdaily.com/finite-state/venture-c...	COLUMBUS, OH, Finite State has raised \$12.5 Mi...	Columbus	OH	https:
2	Limbix	Limbix is a prescription digital therapeutic c...	Series A	\$9,000,000	08 May 2020	Limbix Lands \$9M Series A	https://vcnewsdaily.com/limbix/venture-capital...	PALO ALTO, CA, Limbix has successfully raised ...	Palo Alto	CA	https
3	White Ops	White Ops is a pioneer in the detection of and...	NaN	undisclosed	08 May 2020	White Ops Receives Investment	https://vcnewsdaily.com/white-ops/venture-capi...	NEW YORK, NY, White Ops today announced the co...	New York	NY	https://v
4	Burst Oral Care	BURST has one simple aim: to make sonic oral c...	Series C	\$20,000,000	07 May 2020	Burst Oral Care Grabs \$20M in Funding	https://vcnewsdaily.com/burst-oral-care/ventur...	WALNUT, CA, Burst Oral Care has received \$20 m...	Walnut	CA	https://vc



In [22]:

```
total_funding_data = pd.concat([df_funding_data,total_funding_data],axis=0)#.reset_index(drop=True)
```

In [23]:

```
total_funding_data.head()
```

Out[23]:

	Company	Company Description	Latest Funding Round	Latest Funding Amount	Funding Date	Funding Title	Website Article Link	Funding Headline	City	State	
0	Jellyfish	Jellyfish is the leading Engineering Management...	Series A	\$12,000,000	08 May 2020	Jellyfish Announces \$12M Series A Funding	https://vcnewsdaily.com/jellyfish/venture-capi...	BOSTON, MA, Jellyfish has raised over \$12 mill...	Boston	MA	http
1	Finite State	The mission of Finite State is to protect the ...	Series A	\$12,500,000	08 May 2020	Finite State Snares \$12.5M in Series A	https://vcnewsdaily.com/finite-state/venture-c...	COLUMBUS, OH, Finite State has raised \$12.5 Mi...	Columbus	OH	https:
2	Limbix	Limbix is a prescription digital therapeutic c...	Series A	\$9,000,000	08 May 2020	Limbix Lands \$9M Series A	https://vcnewsdaily.com/limbix/venture-capital...	PALO ALTO, CA, Limbix has successfully raised ...	Palo Alto	CA	https
3	White Ops	White Ops is a pioneer in the detection of and...	NaN	undisclosed	08 May 2020	White Ops Receives Investment	https://vcnewsdaily.com/white-ops/venture-capi...	NEW YORK, NY, White Ops today announced the co...	New York	NY	https://\
4	Burst Oral Care	BURST has one simple aim: to make sonic oral c...	Series C	\$20,000,000	07 May 2020	Burst Oral Care Grabs \$20M in Funding	https://vcnewsdaily.com/burst-oral-care/ventur...	WALNUT, CA, Burst Oral Care has received \$20 m...	Walnut	CA	https://vc

In [24]:

```
total_funding_data.shape
```

Out[24]:

(49, 12)

In [25]:

```
total_funding_data.to_csv('VC_News_Daily.csv')
```

In [26]:

```
# [Only once]  
# Create excel file in order to upload into Microsoft SQL in order to upload dataset  
# total_funding_data.to_excel('VC_News_Daily.xlsx')
```

In [27]:

```
# Sensitive information withheld  
driver_name = '{SENSITIVE_INFO_WITHHELD}'  
server_name = 'SENSITIVE_INFO_WITHHELD'  
database_name = 'SENSITIVE_INFO_WITHHELD'  
trusted_connection = 'yes'
```

In [28]:

```
# Connect to SQL Server  
conn = pyodbc.connect(Driver=driver_name,  
                      Server=server_name,  
                      Database=database_name,  
                      Trusted_Connection=trusted_connection)  
cursor = conn.cursor()
```


In [29]:

```
# Before adding new information - SQL Table as is, without the reordering
sql_raw_total_funding_data = pd.read_sql('SELECT * FROM VC_News_Daily',conn)
sql_raw_total_funding_data.head()
```

Out[29]:

	index	Company	Company Description	Latest Funding Round	Latest Funding Amount	Funding Date	Funding Title	Website Article Link	Funding Headline	City	State
0	0	Jellyfish	Jellyfish is the leading Engineering Managemen...	Series A	\$12,000,000	08 May 2020	Jellyfish Announces \$12M Series A Funding	https://vcnewsdaily.com/jellyfish/venture-capi...	BOSTON, MA, Jellyfish has raised over \$12 mill...	Boston	MA
1	4	Burst Oral Care	BURST has one simple aim: to make sonic oral c...	Series C	\$20,000,000	07 May 2020	Burst Oral Care Grabs \$20M in Funding	https://vcnewsdaily.com/burst-oral-care/ventur...	WALNUT, CA, Burst Oral Care has received \$20 m...	Walnut	CA I
2	11	Every Mother	Every Mother is designed for women at all stag...	Seed	\$1,500,000	07 May 2020	Every Mother Collects \$1.5M Seed Round	https://vcnewsdaily.com/every-mother/venture-c...	BROOKLYN, NY, Every Mother has raised a \$1.5 m...	Brooklyn	NY
3	1	Finite State	The mission of Finite State is to protect the ...	Series A	\$12,500,000	08 May 2020	Finite State Snares \$12.5M in Series A	https://vcnewsdaily.com/finite-state/venture-c...	COLUMBUS, OH, Finite State has raised \$12.5 Mi...	Columbus	OH
4	2	Limbox	Limbox is a prescription digital therapeutic c...	Series A	\$9,000,000	08 May 2020	Limbox Lands \$9M Series A	https://vcnewsdaily.com/limbox/venture-capital...	PALO ALTO, CA, Limbox has successfully raised ...	Palo Alto	CA

In [30]:

```
# Sensitive information withheld  
engine = sqlalchemy.create_engine('mssql+pyodbc://SENSITIVE_INFO_WITHHELD')
```

In [31]:

```
total_funding_data.to_sql('VC_News_Daily',con=engine,if_exists='replace',index=True)
```

In [32]:

```
sql_total_funding_data = pd.read_sql('SELECT * FROM VC_News_Daily',conn).sort_values('index').reset_index().drop(columns=['level_0','index'])
sql_total_funding_data.head()
```

Out[32]:

	Company	Company Description	Latest Funding Round	Latest Funding Amount	Funding Date	Funding Title	Website Article Link	Funding Headline	City	State	
0	Jellyfish	Jellyfish is the leading Engineering Managemen...	Series A	\$12,000,000	08 May 2020	Jellyfish Announces \$12M Series A Funding	https://vcnewsdaily.com/jellyfish/venture-capi...	BOSTON, MA, Jellyfish has raised over \$12 mill...	Boston	MA	http
1	Finite State	The mission of Finite State is to protect the ...	Series A	\$12,500,000	08 May 2020	Finite State Snares \$12.5M in Series A	https://vcnewsdaily.com/finite-state/venture-c...	COLUMBUS, OH, Finite State has raised \$12.5 Mi...	Columbus	OH	https:.
2	Limbix	Limbix is a prescription digital therapeutic C...	Series A	\$9,000,000	08 May 2020	Limbix Lands \$9M Series A	https://vcnewsdaily.com/limbix/venture-capital...	PALO ALTO, CA, Limbix has successfully raised ...	Palo Alto	CA	https
3	White Ops	White Ops is a pioneer in the detection of and...	None	undisclosed	08 May 2020	White Ops Receives Investment	https://vcnewsdaily.com/white-ops/venture-capi...	NEW YORK, NY, White Ops today announced the co...	New York	NY	https://\
4	Burst Oral Care	BURST has one simple aim: to make sonic oral c...	Series C	\$20,000,000	07 May 2020	Burst Oral Care Grabs \$20M in Funding	https://vcnewsdaily.com/burst-oral-care/ventur...	WALNUT, CA, Burst Oral Care has received \$20 m...	Walnut	CA	https://vc

In [33]:

```
sql_total_funding_data.shape
```

Out[33]:

(49, 12)

In []: