In [2]:

```
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
from selenium.webdriver.common.action_chains import ActionChains
import time
from bs4 import BeautifulSoup
import pandas as pd
from selenium.webdriver.chrome.service import Service
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
```

In [3]:

```
# Initialize the WebDriver
driver = webdriver.Chrome()

# Open Yahoo Finance
driver.get('https://finance.yahoo.com/screener/predefined/sec-ind_sec-t

# Define a function to fetch data from the current page
def fetch_data():
    try:
        # Get the page source
        page_source = driver.page_source

        # Parse the HTML using BeautifulSoup
        soup = BeautifulSoup(page_source, 'html.parser')

        # Find the table element
        table = soup.find('table')

        # Check if the table exists
        if table:
            # Find the table body
            tbody = table.find('tbody')

            # Check if the table body exists
            if tbody:
                # Extract the table data into a list of lists
                data = []
                # Find all rows in the table body
                for row in tbody.find_all('tr'):
                    # Extract text from all cells in the row
                    row_data = [cell.get_text(strip=True) for cell in r
                    # Append row data to the main data list
                    data.append(row_data)

                # Convert the data into a DataFrame
                df = pd.DataFrame(data, columns=["Symbol", "Name", "Pri

                return df
            else:
                print("Table body not found.")
                return None
        else:
            print("Table not found.")
            return None
    except Exception as e:
        print("Error occurred while fetching data:", e)
        return None

# Fetch data from the current page
df_list = [fetch_data()]

# Define the number of times to click the "Next" button
num_clicks = 19

# Click the "Next" button and fetch data from subsequent pages
for _ in range(num_clicks):
    try:
        # Find the "Next" button
        next_button = WebDriverWait(driver, 20).until(
            EC.visibility_of_element_located((By.XPATH, '//*[@id="scr-r
        )
        # Scroll the page to bring the button into view
        driver.execute_script("arguments[0].scrollIntoView();", next_bu

        # Click the "Next" button using JavaScript
        driver.execute_script("arguments[0].click();", next_button)

        # Wait for the page to load
        time.sleep(3)  # Adjust the sleep time if needed

        # Fetch data from the current page and add it to the list
        df = fetch_data()
        if df is not None:
            df_list.append(df)
    except Exception as e:
        print("Error occurred while clicking Next button:", e)

# Concatenate all the DataFrames into a single DataFrame
final_df = pd.concat(df_list, ignore_index=True)

# Close the browser
driver.quit()

# Print the final DataFrame
final_df
```

```
Table not found.
Error occurred while clicking Next button: Message:
Stacktrace:
        GetHandleVerifier [0x00007FF704EA1502+60802]
        (No symbol) [0x00007FF704E1AC02]
        (No symbol) [0x00007FF704CD7CE4]
        (No symbol) [0x00007FF704D26D4D]
        (No symbol) [0x00007FF704D26E1C]
        (No symbol) [0x00007FF704D6CE37]
        (No symbol) [0x00007FF704D4ABBF]
        (No symbol) [0x00007FF704D6A224]
        (No symbol) [0x00007FF704D4A923]
        (No symbol) [0x00007FF704D18FEC]
        (No symbol) [0x00007FF704D19C21]
        GetHandleVerifier [0x00007FF7051A411D+3217821]
        GetHandleVerifier [0x00007FF7051E60B7+3488055]
        GetHandleVerifier [0x00007FF7051DF03F+3459263]
        GetHandleVerifier [0x00007FF704F5B846+823494]
        (No symbol) [0x00007FF704E25F9F]
        (No symbol) [0x00007FF704E20EC4]
        (No symbol) [0x00007FF704E21052]
        (No symbol) [0x00007FF704E118A4]
        BaseThreadInitThunk [0x00007FFF975A257D+29]
        RtlUserThreadStart [0x00007FFF97FEAA48+40]
```

Out[3]:

	Symbol	Name	Price (Intraday)	Change	% Change	Volume	Avg Vol (3 month)	Market Cap
0	BFAFX	American Funds Bond Fund of Amer F1	-0.05	-0.45%	11.01	11.17	11.14	-0.84%
1	ABNFX	American Funds Bond Fund of Amer F2	-0.05	-0.45%	11.01	11.17	11.14	-0.76%
2	CFAEX	American Funds Bond Fund of Amer 529E	-0.05	-0.45%	11.01	11.17	11.14	-0.88%
3	CFAFX	American Funds Bond Fund of Amer 529F	-0.05	-0.45%	11.01	11.17	11.14	-0.79%
4	RBFEX	American Funds Bond Fund of Amer R4	-0.05	-0.45%	11.01	11.17	11.14	-0.82%
...	...	...	...	...	...	...	...	...
445	0P0001M384	Polar Capital Fut Healthcare A acc USD	+0.83	+0.85%	96.93	100.68	93.09	9.32%
446	0P0001LUQZ	Invesco Funds - Invesco China Health Care Equi...	-0.0100	-0.23%	4.2800	4.22	4.52	-11.42%
447	0P00000BON	Candriam Equities L Biotechnology	+5.16	+0.71%	728.37	765.35	718.69	2.78%
448	0P0001HJ7Z	Bellevue (Lux) Bellevue Dgtl Hthl AI2USD	+1.53	+1.18%	129.78	135.55	133.36	1.49%
449	0P0001BM5E	Sector Healthcare Value B USD	+1.47	+0.91%	162.39	167.81	165.17	3.27%

450 rows × 10 columns

In [4]:

```
final_df.to_csv('healthcare_finance.csv')
```

In [5]:

```
pd.read_csv("healthcare_finance.csv")
```

Out[5]:

	Unnamed: 0	Symbol	Name	Price (Intraday)	Change	% Change	Volume	Av Vol mont
0	0	BFAFX	American Funds Bond Fund of Amer F1	-0.05	-0.45%	11.01	11.17	11.
1	1	ABNFX	American Funds Bond Fund of Amer F2	-0.05	-0.45%	11.01	11.17	11.
2	2	CFAEX	American Funds Bond Fund of Amer 529E	-0.05	-0.45%	11.01	11.17	11.
3	3	CFAFX	American Funds Bond Fund of Amer 529F	-0.05	-0.45%	11.01	11.17	11.
4	4	RBFEX	American Funds Bond Fund of Amer R4	-0.05	-0.45%	11.01	11.17	11.
...	...	...	...	...	...	...	...	
445	445	0P0001M384	Polar Capital Fut Healthcare A acc USD	0.83	+0.85%	96.93	100.68	93.
446	446	0P0001LUQZ	Invesco Funds - Invesco China Health Care Equi...	-0.01	-0.23%	4.2800	4.22	4.
447	447	0P00000BON	Candriam Equities L Biotechnology	5.16	+0.71%	728.37	765.35	718.
448	448	0P0001HJ7Z	Bellevue (Lux) Bellevue Dgtl Hthl AI2USD	1.53	+1.18%	129.78	135.55	133.
449	449	0P0001BM5E	Sector Healthcare Value B USD	1.47	+0.91%	162.39	167.81	165.

450 rows × 11 columns

In [ ]:

```

```