In [1]:
```
#dependencies and setup
import pandas as pd
# import os
# from openpyxl import Workbook
# import numpy as np
# import re
# import datetime as dt
import requests
# from pprint import pprint
from bs4 import BeautifulSoup
from splinter import Browser
from webdriver_manager.chrome import ChromeDriverManager
import time
```

In [2]:
```
# SQLite dependencies
import sqlite3
from sqlalchemy import create_engine
from sqlalchemy import Column, Integer, String, Float
```

In [37]:
```
# create a browser instance using splinter
executable_path = {'executable_path': ChromeDriverManager().install()}
browser = Browser('chrome', **executable_path, headless=False)
time.sleep(1)
```

```
====== WebDriver manager ======
Current google-chrome version is 95.0.4638
Get LATEST chromedriver version for 95.0.4638 google-chrome
Driver [C:\Users\mosab\.wdm\drivers\chromedriver\win32\95.0.4638.69\chromedrive
r.exe] found in cache
```

In [43]:
```
# Empty lists
names = []
ms_cat = []
ytd_daily = []
yr1 = []
yr3= []
yr5 = []
yr10 = []
life_of_fund = []
net_expense_ratio = []
gross_expense_ratio = []
ms_rating_overall = []

for i in range(1,98):
    # visit fidelity URL
    fidelity_url = f"https://fundresearch.fidelity.com/fund-screener/results/tab
    browser.visit(fidelity_url)
    time.sleep(4)

    # create HTML object
    html = browser.html

    # parse HTML with BeautifulSoup
    soup = BeautifulSoup(html, 'html.parser')
```

```
    div = soup.find('div', id ='static-table-container')
    table = div.find('table', id = 'static-table')
    tbody = table.find('tbody', id = 'static-tbody')
    for listing in tbody.find_all('td', class_ = 'name left'):
        for name in listing.find_all('a'):
            names.append(name.text)
    div2 = soup.find('div', id ='scrollable-results-table-wrapper')
    table2 = div2.find('table', id = 'scrollable-results-table')
    tbody2 = table2.find('tbody', id = 'results-tbody')
    for listing in tbody2.find_all('td', class_ = 'morningstarCategory left'):
            ms_cat.append(listing.text)
    for listing in tbody2.find_all('td', class_ = 'ytdDaily right'):
            ytd_daily.append(listing.text)
    for listing in tbody2.find_all('td', class_ = 'yr1 left'):
            yr1.append(listing.text)
    for listing in tbody2.find_all('td', class_ = 'yr3 left sorted-column-cell')
            yr3.append(listing.text)
    for listing in tbody2.find_all('td', class_ = 'yr5 left'):
            yr5.append(listing.text)
    for listing in tbody2.find_all('td', class_ = 'yr10 left'):
            yr10.append(listing.text)
    for listing in tbody2.find_all('td', class_ = 'lifeOfFund left'):
            life_of_fund.append(listing.text)
    for listing in tbody2.find_all('td', class_ = 'netExpenseRatio right'):
            net_expense_ratio.append(listing.text)
    for listing in tbody2.find_all('td', class_ = 'grossExpenseRatio right'):
            gross_expense_ratio.append(listing.text)
    for listing in tbody2.find_all('td', class_ = 'morningstarRatingOverall cent
            if type(listing.find('span')) == type(None):
                    ms_rating_overall.append("")
            else:
                    ms_rating_overall.append(listing.find('span').text)
    print(f"{len(names)} funds scraped until page {i}")
```

```
100 funds scraped on page 1
200 funds scraped on page 2
300 funds scraped on page 3
400 funds scraped on page 4
500 funds scraped on page 5
600 funds scraped on page 6
700 funds scraped on page 7
800 funds scraped on page 8
900 funds scraped on page 9
1000 funds scraped on page 10
1100 funds scraped on page 11
1200 funds scraped on page 12
1300 funds scraped on page 13
1400 funds scraped on page 14
1500 funds scraped on page 15
1600 funds scraped on page 16
1700 funds scraped on page 17
1800 funds scraped on page 18
1900 funds scraped on page 19
2000 funds scraped on page 20
2100 funds scraped on page 21
2200 funds scraped on page 22
2300 funds scraped on page 23
2400 funds scraped on page 24
2500 funds scraped on page 25
2600 funds scraped on page 26
```

```
2700 funds scraped on page 27
2800 funds scraped on page 28
2900 funds scraped on page 29
3000 funds scraped on page 30
3100 funds scraped on page 31
3200 funds scraped on page 32
3300 funds scraped on page 33
3400 funds scraped on page 34
3500 funds scraped on page 35
3600 funds scraped on page 36
3700 funds scraped on page 37
3800 funds scraped on page 38
3900 funds scraped on page 39
4000 funds scraped on page 40
4100 funds scraped on page 41
4200 funds scraped on page 42
4300 funds scraped on page 43
4400 funds scraped on page 44
4500 funds scraped on page 45
4600 funds scraped on page 46
4700 funds scraped on page 47
4800 funds scraped on page 48
4900 funds scraped on page 49
5000 funds scraped on page 50
5100 funds scraped on page 51
5200 funds scraped on page 52
5300 funds scraped on page 53
5400 funds scraped on page 54
5500 funds scraped on page 55
5600 funds scraped on page 56
5700 funds scraped on page 57
5800 funds scraped on page 58
5900 funds scraped on page 59
6000 funds scraped on page 60
6100 funds scraped on page 61
6200 funds scraped on page 62
6300 funds scraped on page 63
6400 funds scraped on page 64
6500 funds scraped on page 65
6600 funds scraped on page 66
6700 funds scraped on page 67
6800 funds scraped on page 68
6900 funds scraped on page 69
7000 funds scraped on page 70
7100 funds scraped on page 71
7200 funds scraped on page 72
7300 funds scraped on page 73
7400 funds scraped on page 74
7500 funds scraped on page 75
7600 funds scraped on page 76
7700 funds scraped on page 77
7800 funds scraped on page 78
7900 funds scraped on page 79
8000 funds scraped on page 80
8100 funds scraped on page 81
8200 funds scraped on page 82
8300 funds scraped on page 83
8400 funds scraped on page 84
8500 funds scraped on page 85
8600 funds scraped on page 86
8700 funds scraped on page 87
```

```
8800 funds scraped on page 88
8900 funds scraped on page 89
9000 funds scraped on page 90
9100 funds scraped on page 91
9200 funds scraped on page 92
9300 funds scraped on page 93
9400 funds scraped on page 94
9500 funds scraped on page 95
9600 funds scraped on page 96
```

In [53]:

```
# close the browser session
browser.quit()
```

In [49]:

```
df_overview = pd.DataFrame({'name' : names, 'morningstar_category': ms_cat, 'ytd
                            'yr3': yr3, 'yr5': yr5, 'yr10': yr10, 'life_of_fund'
                            'net_expense_ratio': net_expense_ratio, 'gross_expen
                            'morningstar_rating_overall': ms_rating_overall})
df_overview.head()
```

Out[49]:

	name	morningstar_category	ytdDaily	yr1	yr3	yr5	yr10	life_of_fu
0	Baron Partners Fund Institutional Shares (BPTIX)	Large Growth	+44.60%	+110.27%	+65.56%	+47.54%	+29.02%	+27.0
1	Baron Partners Fund Retail Shares (BPTRX)	Large Growth	+44.28%	+109.72%	+65.13%	+47.15%	+28.68%	+20.4
2	Morgan Stanley Institutional Fund, Inc. Incept...	Small Growth	+20.05%	+81.15%	+55.11%	+38.37%	+22.37%	+14.1
3	Morgan Stanley Institutional Fund, Inc. Incept...	Small Growth	+19.75%	+80.67%	+54.70%	+37.98%	+22.01%	+13.8
4	Morgan Stanley Institutional Fund, Inc. Incept...	Small Growth	--	+79.39%	+53.46%	+36.90%	+21.11%	+13.0

In [51]:

```
df_overview.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9626 entries, 0 to 9625
Data columns (total 11 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
```

```
 0   name                        9626 non-null   object
 1   morningstar_category        9626 non-null   object
 2   ytdDaily                    9626 non-null   object
 3   yr1                         9626 non-null   object
 4   yr3                         9626 non-null   object
 5   yr5                         9626 non-null   object
 6   yr10                        9626 non-null   object
 7   life_of_fund                9626 non-null   object
 8   net_expense_ratio           9626 non-null   object
 9   gross_expense_ratio         9626 non-null   object
 10  morningstar_rating_overall  9626 non-null   object
dtypes: object(11)
memory usage: 827.4+ KB
```

In [52]:

```
df_overview.to_csv('fidelity_mutual_fund_overview.csv', index = False)
```