In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

PART 1

1.

In [19]:
```
dataframe = pd.read_csv('CCI.csv')

if (df['Open'] == df['Close']).all():
    print("CCI has no difference ")
else:
    print("CCI has differences between Open and Close prices .")
```

```
CCI has differences between Open and Close prices .
```

In [23]:
```
dataframe = pd.read_csv('KHC.csv')

if (df['Open'] == df['Close']).all():
    print("KHC has no difference ")
else:
    print("KHC has differences between Open and Close prices .")
```

```
KHC has differences between Open and Close prices .
```

In [25]:
```
dataframe = pd.read_csv('VRSK.csv')

if (df['Open'] == df['Close']).all():
    print("VRSK has no difference ")
else:
    print("VRSK has differences between Open and Close prices .")
```

```
VRSK has differences between Open and Close prices .
```

In [27]:
```
dataframe = pd.read_csv('WFC.csv')

if (df['Open'] == df['Close']).all():
    print("WFC has no difference ")
else:
    print("WFC has differences between Open and Close prices .")
```

```
WFC has differences between Open and Close prices .
```

In [29]:
```
dataframe = pd.read_csv('WRB.csv')

if (df['Open'] == df['Close']).all():
    print("WRB has no difference ")
else:
    print("WRB has differences between Open and Close prices .")
```

```
WRB has differences between Open and Close prices .
```

In [33]:
```
# ALL the companies have difference between open and close price,  which mean th
#or we can say risk and uncertainty
```

2

In [63]:
```
DOC = ['CCI.csv', 'KHC.csv', 'VRSK.csv', 'WFC.csv', 'WRB.csv']
stocks = ['CCI', 'KHC', 'VRSK', 'WFC', 'WRB']
prices = {}
for csv, stock in zip(DOC, stocks):
    df = pd.read_csv(file)
    highestprice = df['Close'].max()
    lowestprice = df['Close'].min()
    prices[stock] = (highestprice, lowestprice)

for stock, prices in prices.items():
    print(f"{stock}: Highest Close = {prices[0]}, Lowest Close = {prices[1]}")
```

```
CCI: Highest Close = 48.64, Lowest Close = 15.877
KHC: Highest Close = 48.64, Lowest Close = 15.877
VRSK: Highest Close = 48.64, Lowest Close = 15.877
WFC: Highest Close = 48.64, Lowest Close = 15.877
WRB: Highest Close = 48.64, Lowest Close = 15.877
```

3

In [49]:
```
for csv, stock in zip(DOC, stocks):
    df = pd.read_csv(csv, parse_dates=True, index_col='Date')

    df['Log_Returns'] = np.log(df['Close'] / df['Close'].shift(1))

    print(f"{stock} - Min: {df['Log_Returns'].min()}, Max: {df['Log_Returns'].ma
```

```
CCI - Min: -0.132673239113567, Max: 0.10740862959777037, Mean: 0.0004347240878579
206, Median: 0.0012675420438606549
KHC - Min: -0.3210442740312378, Max: 0.17807957517444345, Mean: -0.00033414413286
79492, Median: 0.0003135410568503709
VRSK - Min: -0.1129339603752813, Max: 0.12474656931109922, Mean: 0.00053066926993
50444, Median: 0.0013574614352954886
WFC - Min: -0.17278806543107672, Max: 0.13570561636402745, Mean: -8.7771795151241
42e-05, Median: 0.0
WRB - Min: -0.16763138866956478, Max: 0.11296805669969835, Mean: 0.00067462139076
74952, Median: 0.0011591866832374544
```

4

In [61]:
```
for csv, stock in zip(DOC, stocks):
    df = pd.read_csv(csv, parse_dates=True, index_col='Date')
    df['Log_Returns'] = np.log(df['Close'] / df['Close'].shift(1))
    MAXgaindate = df['Log_Returns'].idxmax()
    MAXlossdate = df['Log_Returns'].idxmin()
    print(f"{stock} - Highest Gain on: {MAXgaindate}, Highest Loss on: {MAXlossd
```

```
CCI - Highest Gain on: 2020-03-26 00:00:00, Highest Loss on: 2020-03-16 00:00:00
KHC - Highest Gain on: 2020-03-13 00:00:00, Highest Loss on: 2019-02-22 00:00:00
VRSK - Highest Gain on: 2020-03-24 00:00:00, Highest Loss on: 2020-03-16 00:00:00
WFC - Highest Gain on: 2020-03-24 00:00:00, Highest Loss on: 2020-03-12 00:00:00
WRB - Highest Gain on: 2020-03-17 00:00:00, Highest Loss on: 2020-03-16 00:00:00
```

5

In [76]:
```
for csv, stock in zip(DOC, stocks):
    df = pd.read_csv(csv, parse_dates=True, index_col='Date')
    weeklyvolume = df['Volume'].resample('W').mean()
    print(f"{stock} - Average Weekly Volume: {weeklyvolume.mean()}")
```

```
CCI - Average Weekly Volume: 2083065.8785942493
KHC - Average Weekly Volume: 6521755.670926518
VRSK - Average Weekly Volume: 859686.805111821
WFC - Average Weekly Volume: 25499706.42172524
WRB - Average Weekly Volume: 1655727.634025559
```

6

In [81]:
```
totalreturns = {}
for csv, stock in zip(DOC, stocks):
    df = pd.read_csv(csv, parse_dates=True, index_col='Date')
    totalreturn = ((df['Close'].iloc[-1] / df['Close'].iloc[0]) - 1) * 100
    totalreturns[stock] = totalreturn

max_return_company = max(totalreturns, key=totalreturns.get)
max_return = totalreturns[max_return_company]

print(f"The company with the highest total return is {max_return_company} with a
```

```
The company with the highest total return is WRB with a return of 176.77%.
```

part2

2.1

In [85]:
```
url = 'https://ies-fsv.s3.eu-central-1.amazonaws.com/companies/companies_no_subi
```

In [87]:
```
df = pd.read_csv(url)
```

In [95]:
```
missingdates = df['included'].isna().sum()
print({missing_inclusion_dates}, f" companies do not filled-in the date of inclu
```

```
{45}  companies do not filled-in the date of inclusion
```

2.2

In [129…
```
df = df.dropna(subset=['included'])
df['included'] = pd.to_datetime(df['included'], errors='coerce')
oldest_constituent = df[df['included'] == df['included'].min()]
youngest_constituent = df[df['included'] == df['included'].max()]
average_age = (pd.Timestamp.now() - df['included']).mean()
print("Oldest Constituent:")
print(oldest_constituent)
```

```
print("\nYoungest Constituent:")
print(youngest_constituent)
print(f"\nAverage Age of a Constituent: {average_age}")
```

Oldest Constituent:

	Symbol	Security	GICS_Sector	hq	included	Founded
25	MO	Altria	Consumer Staples	Richmond, Virginia	1957-04-03	1985
31	AEP	American Electric Power	Utilities	Columbus, Ohio	1957-04-03	1906
72	BA	Boeing	Industrials	Chicago, Illinois	1957-04-03	1916
77	BMY	Bristol Myers Squibb	Health Care	New York City, New York	1957-04-03	1989
86	CPB	Campbell Soup Company	Consumer Staples	Camden, New Jersey	1957-04-03	1869
93	CAT	Caterpillar Inc.	Industrials	Irving, Texas	1957-04-03	1925
105	CVX	Chevron Corporation	Energy	San Ramon, California	1957-04-03	1879
118	KO	The Coca-Cola Company	Consumer Staples	Atlanta, Georgia	1957-04-03	1886
120	CL	Colgate-Palmolive	Consumer Staples	New York City, New York	1957-04-03	1806
124	COP	ConocoPhillips	Energy	Houston, Texas	1957-04-03	2002
138	CVS	CVS Health	Health Care	Woonsocket, Rhode Island	1957-04-03	1996
143	DE	John Deere	Industrials	Moline, Illinois	1957-04-03	1837
159	DTE	DTE Energy	Utilities	Detroit, Michigan	1957-04-03	1995
167	EIX	Edison International	Utilities	Rosemead, California	1957-04-03	1886
174	ETR	Entergy	Utilities	New Orleans, Louisiana	1957-04-03	1913
187	EXC	Exelon	Utilities	Chicago, Illinois	1957-04-03	2000
191	XOM	ExxonMobil	Energy	Irving, Texas	1957-04-03	1999
204	F	Ford Motor Company	Consumer Discretionary	Dearborn, Michigan	1957-04-03	1903
216	GD	General Dynamics	Industrials	Falls Church, Virginia	1957-04-03	1899
225	HAL	Halliburton	Energy			
226	HIG	Hartford (The)	Financials			
231	HSY	Hershey's	Consumer Staples			
245	IBM	IBM	Information Technology			
254	IP	International Paper	Materials			
274	KMB	Kimberly-Clark	Consumer Staples			
279	KR	Kroger	Consumer Staples			
310	MRK	Merck & Co.	Health Care			
342	NSC	Norfolk Southern Railway	Industrials			
368	PEP	PepsiCo	Consumer Staples			
370	PFE	Pfizer	Health Care			
378	PPG	PPG Industries	Materials			
381	PG	Procter & Gamble	Consumer Staples			
385	PEG	Public Service Enterprise Group	Utilities			
413	SEE	Sealed Air	Materials			
423	SO	Southern Company	Utilities			
460	UNP	Union Pacific Corporation	Industrials			
496	XEL	Xcel Energy	Utilities			

```
225             Houston, Texas 1957-04-03              1919
226      Hartford, Connecticut 1957-04-03              1810
231      Hershey, Pennsylvania 1957-04-03              1894
245           Armonk, New York 1957-04-03              1911
254         Memphis, Tennessee 1957-04-03              1898
274              Irving, Texas 1957-04-03              1872
279           Cincinnati, Ohio 1957-04-03              1883
310     Kenilworth, New Jersey 1957-04-03              1891
342          Norfolk, Virginia 1957-04-03  1881/1894 (1980)
368         Purchase, New York 1957-04-03              1898
370    New York City, New York 1957-04-03              1849
378   Pittsburgh, Pennsylvania 1957-04-03              1883
381           Cincinnati, Ohio 1957-04-03              1837
385         Newark, New Jersey 1957-04-03              1903
413  Charlotte, North Carolina 1957-04-03              1960
423           Atlanta, Georgia 1957-04-03              1945
460            Omaha, Nebraska 1957-04-03              1862
496     Minneapolis, Minnesota 1957-04-03              1909

Youngest Constituent:
    Symbol         Security GICS_Sector              hq   included Founded
438   TRGP  Targa Resources      Energy  Houston, Texas 2022-12-10    2005

Average Age of a Constituent: 9065 days 14:08:42.890651392
```

In [ ]: