

```
In [144... import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [146... import glob
```

```
In [148... glob.glob(r'C:\Users\lebon\OneDrive\Desktop\LEANDRINHO\Udemy\DataAnalytic'
```

[illegible]

```
In [12]: len(glob.glob(r'C:\Users\lebon\OneDrive\Desktop\LEANDRINH0\Udemy\DataAnal\
```

```
Out[12]: 505
```

```
In [14]: company_list = [  
    r'C:\\Users\\lebon\\OneDrive\\Desktop\\LEANDRINH0\\Udemy\\DataAnalyti  
    r'C:\\Users\\lebon\\OneDrive\\Desktop\\LEANDRINH0\\Udemy\\DataAnalyt:  
    r'C:\\Users\\lebon\\OneDrive\\Desktop\\LEANDRINH0\\Udemy\\DataAnalyt:  
    r'C:\\Users\\lebon\\OneDrive\\Desktop\\LEANDRINH0\\Udemy\\DataAnalyt:  
    ]
```

```
In [150... import warnings  
from warnings import filterwarnings  
filterwarnings('ignore')
```

```
In [154... import glob  
  
company_list = glob.glob(r'C:\Users\lebon\OneDrive\Desktop\LEANDRINH0\Uder  
  
# Now your loop will work  
all_data = pd.DataFrame()  
  
for file in company_list:  
    current_df = pd.read_csv(file)  
    all_data = pd.concat([all_data, current_df], ignore_index=True)
```

```
In [35]: all_data.shape
```

```
Out[35]: (4752, 7)
```

```
In [37]: all_data.head(6)
```

```
Out[37]:
```

	date	open	high	low	close	volume	Name
0	2013-02-08	67.7142	68.4014	66.8928	67.8542	158168416	AAPL
1	2013-02-11	68.0714	69.2771	67.6071	68.5614	129029425	AAPL
2	2013-02-12	68.5014	68.9114	66.8205	66.8428	151829363	AAPL
3	2013-02-13	66.7442	67.6628	66.1742	66.7156	118721995	AAPL
4	2013-02-14	66.3599	67.3771	66.2885	66.6556	88809154	AAPL
5	2013-02-15	66.9785	67.1656	65.7028	65.7371	97924631	AAPL

```
In [41]: all_data['Name'].unique()
```

```
Out[41]: array(['AAPL', 'AMZN', 'GOOG', 'MSFT'], dtype=object)
```

```
In [45]: all_data.isnull().sum()
```

```
Out[45]: date      0  
open      0  
high      0  
low       0  
close     0  
volume    0  
Name      0  
dtype: int64
```

```
In [47]: all_data.dtypes
```

```
Out[47]: date      object
open    float64
high    float64
low      float64
close    float64
volume   int64
Name     object
dtype: object
```

```
In [ ]:
```

```
In [57]: all_data['date'] = pd.to_datetime(all_data['date'])
```

```
In [59]: all_data['date']
```

```
Out[59]: 0      2013-02-08
1      2013-02-11
2      2013-02-12
3      2013-02-13
4      2013-02-14
...
4747   2018-02-01
4748   2018-02-02
4749   2018-02-05
4750   2018-02-06
4751   2018-02-07
Name: date, Length: 4752, dtype: datetime64[ns]
```

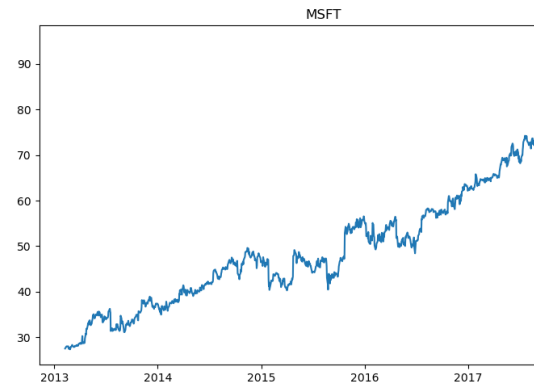
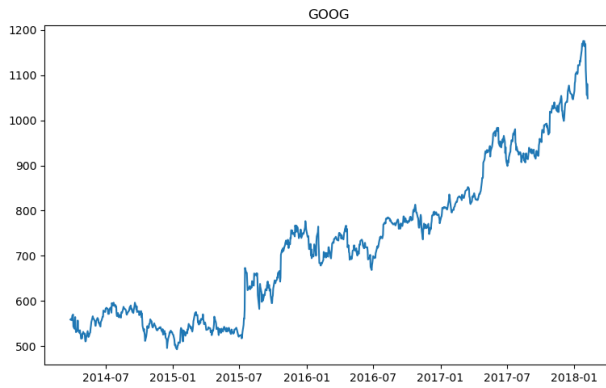
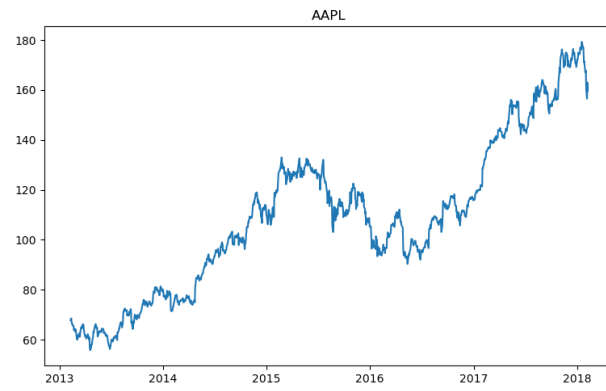
```
In [61]: tech_list = all_data['Name'].unique()
```

```
In [63]: tech_list
```

```
Out[63]: array(['AAPL', 'AMZN', 'GOOG', 'MSFT'], dtype=object)
```

```
In [67]: plt.figure(figsize=(20,12))
```

```
for index, company in enumerate(tech_list, 1):
    plt.subplot(2,2, index)
    filter1 = all_data['Name']==company
    df = all_data[filter1]
    plt.plot(df['date'], df['close'])
    plt.title(company)
```



```
In [69]: all_data['close'].rolling(window=10).mean().head(14)
```

```
Out[69]: 0      NaN
1      NaN
2      NaN
3      NaN
4      NaN
5      NaN
6      NaN
7      NaN
8      NaN
9      66.03251
10     65.57280
11     65.13051
12     64.79722
13     64.43137
Name: close, dtype: float64
```

```
In [71]: new_data = all_data.copy()
```

```
In [77]: ma_day = [10, 20, 50]
```

```
for ma in ma_day:
    new_data['close_'+str(ma)] = new_data['close'].rolling(ma).mean()
```

```
In [79]: new_data.tail(7)
```

```
Out[79]:
```

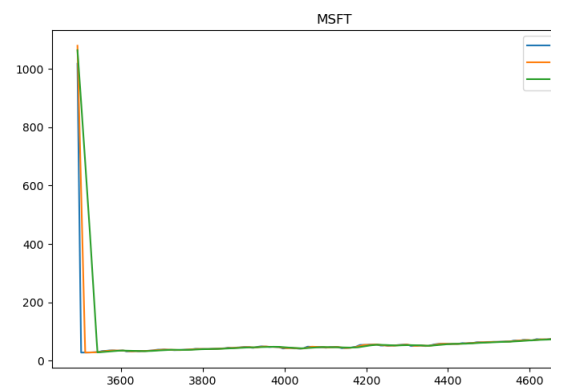
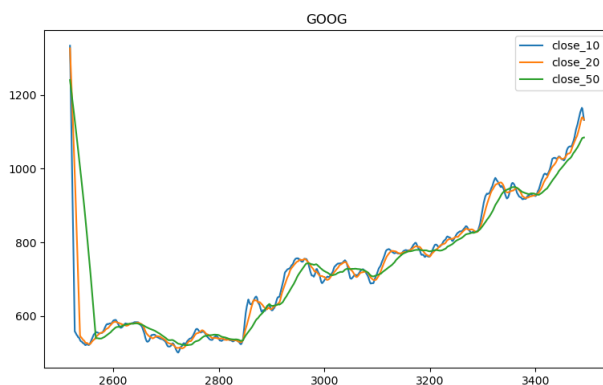
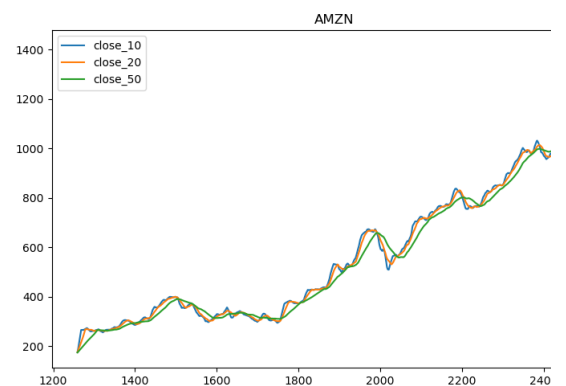
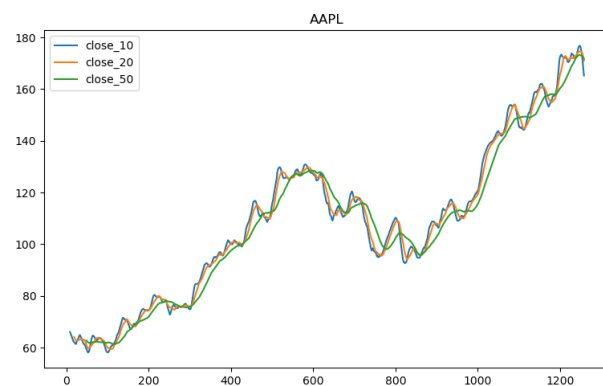
	date	open	high	low	close	volume	Name	close_10	close_20	cl
<b>4745</b>	2018-01-30	93.30	93.660	92.1000	92.74	38635053	MSFT	91.862	89.8285	8
<b>4746</b>	2018-01-31	93.75	95.400	93.5100	95.01	48756338	MSFT	92.349	90.2815	8
<b>4747</b>	2018-02-01	94.79	96.070	93.5813	94.26	47227882	MSFT	92.765	90.6770	8
<b>4748</b>	2018-02-02	93.64	93.970	91.5000	91.78	47867753	MSFT	92.943	90.9105	8
<b>4749</b>	2018-02-05	90.56	93.240	88.0000	88.00	51031465	MSFT	92.582	90.9010	8
<b>4750</b>	2018-02-06	86.89	91.475	85.2500	91.33	67998564	MSFT	92.525	91.0535	8
<b>4751</b>	2018-02-07	90.49	91.770	89.2000	89.61	41107592	MSFT	92.304	91.1230	8

```
In [81]: new_data.columns
```

```
Out[81]: Index(['date', 'open', 'high', 'low', 'close', 'volume', 'Name', 'close_10',
               'close_20', 'close_50'],
              dtype='object')
```

```
In [87]: plt.figure(figsize=(20,12))
```

```
for index, company in enumerate(tech_list, 1):
    plt.subplot(2,2, index)
    filter1 = new_data['Name']==company
    df = new_data[filter1]
    df[['close_10', 'close_20', 'close_50']].plot(ax=plt.gca())
    plt.title(company)
```



```
In [ ]:
```

```
In [12]: apple = pd.read_csv(r'C:\\Users\\lebon\\OneDrive\\Desktop\\LEANDRINHO\\Ud
```

```
In [14]: apple.head(4)
```

```
Out[14]:
```

	date	open	high	low	close	volume	Name
0	2013-02-08	67.7142	68.4014	66.8928	67.8542	158168416	AAPL
1	2013-02-11	68.0714	69.2771	67.6071	68.5614	129029425	AAPL
2	2013-02-12	68.5014	68.9114	66.8205	66.8428	151829363	AAPL
3	2013-02-13	66.7442	67.6628	66.1742	66.7156	118721995	AAPL

```
In [100... apple['Daily Return (in %)'] = apple['close'].pct_change()*100
```

```
In [102... apple.head(4)
```

```
Out[102...
```

	open	high	low	close	volume	name	daily return (in %)	Daily
date								
2013-02-08	67.7142	68.4014	66.8928	67.8542	158168416	AAPL	NaN	
2013-02-11	68.0714	69.2771	67.6071	68.5614	129029425	AAPL	1.042235	1.
2013-02-12	68.5014	68.9114	66.8205	66.8428	151829363	AAPL	-2.506658	-2.
2013-02-13	66.7442	67.6628	66.1742	66.7156	118721995	AAPL	-0.190297	-0.

```
In [156... import plotly.express as px
```

```
In [ ]:
```

```
In [123... apple.dtypes
```

```
Out[123...
```

open	float64
high	float64
low	float64
close	float64
volume	int64
name	object
daily return (in %)	float64
daily return (in %)	float64
dtype:	object

```
In [173... apple.rename(columns={'date': 'date'}, inplace=True)
```

```
In [175... apple.head(4)
```

```
Out[175...
```

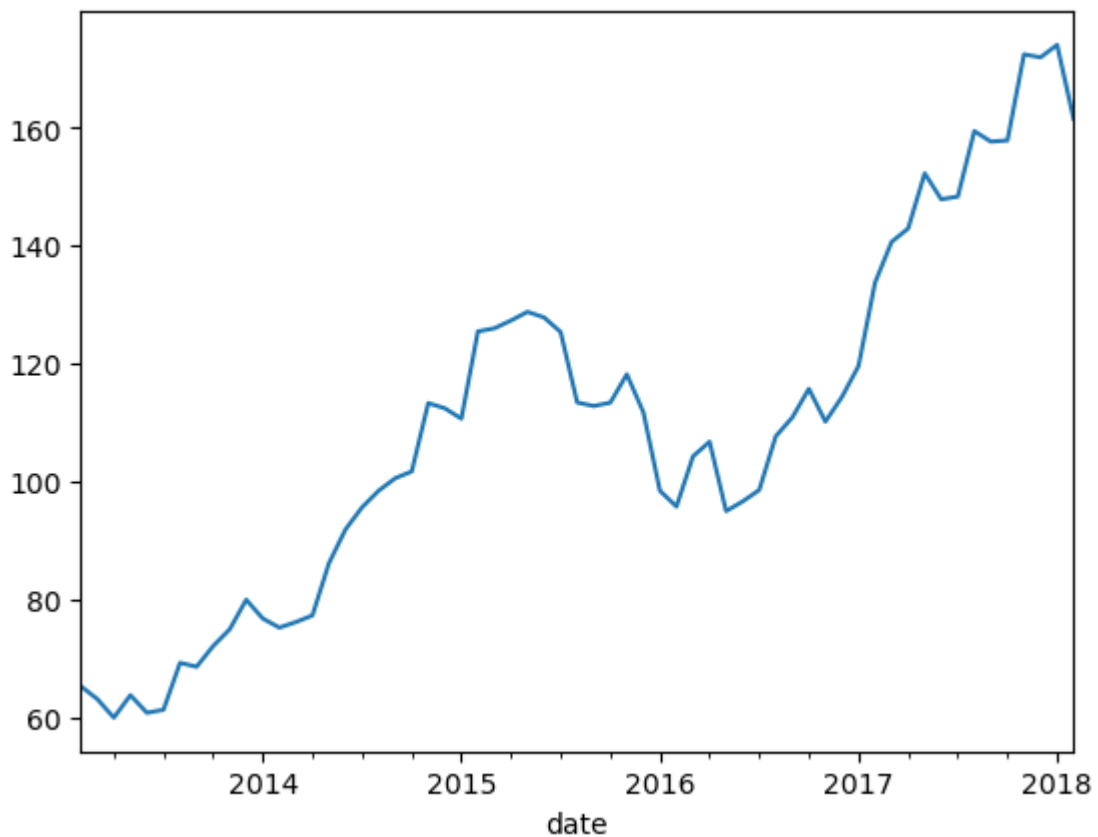
	open	high	low	close	volume	name	daily return (in %)	daily re
date								
2013-02-08	67.7142	68.4014	66.8928	67.8542	158168416	AAPL	NaN	
2013-02-11	68.0714	69.2771	67.6071	68.5614	129029425	AAPL	1.042235	1.
2013-02-12	68.5014	68.9114	66.8205	66.8428	151829363	AAPL	-2.506658	-2.
2013-02-13	66.7442	67.6628	66.1742	66.7156	118721995	AAPL	-0.190297	-0.

```
In [180... apple['close'].resample('M').mean()
```

```
Out[180... date
2013-02-28      65.306264
2013-03-31      63.120110
2013-04-30      59.966432
2013-05-31      63.778927
2013-06-30      60.791120
...
2017-10-31     157.817273
2017-11-30     172.406190
2017-12-31     171.891500
2018-01-31     174.005238
2018-02-28     161.468000
Freq: ME, Name: close, Length: 61, dtype: float64
```

```
In [182... apple['close'].resample('M').mean().plot()
```

```
Out[182... <Axes: xlabel='date'>
```



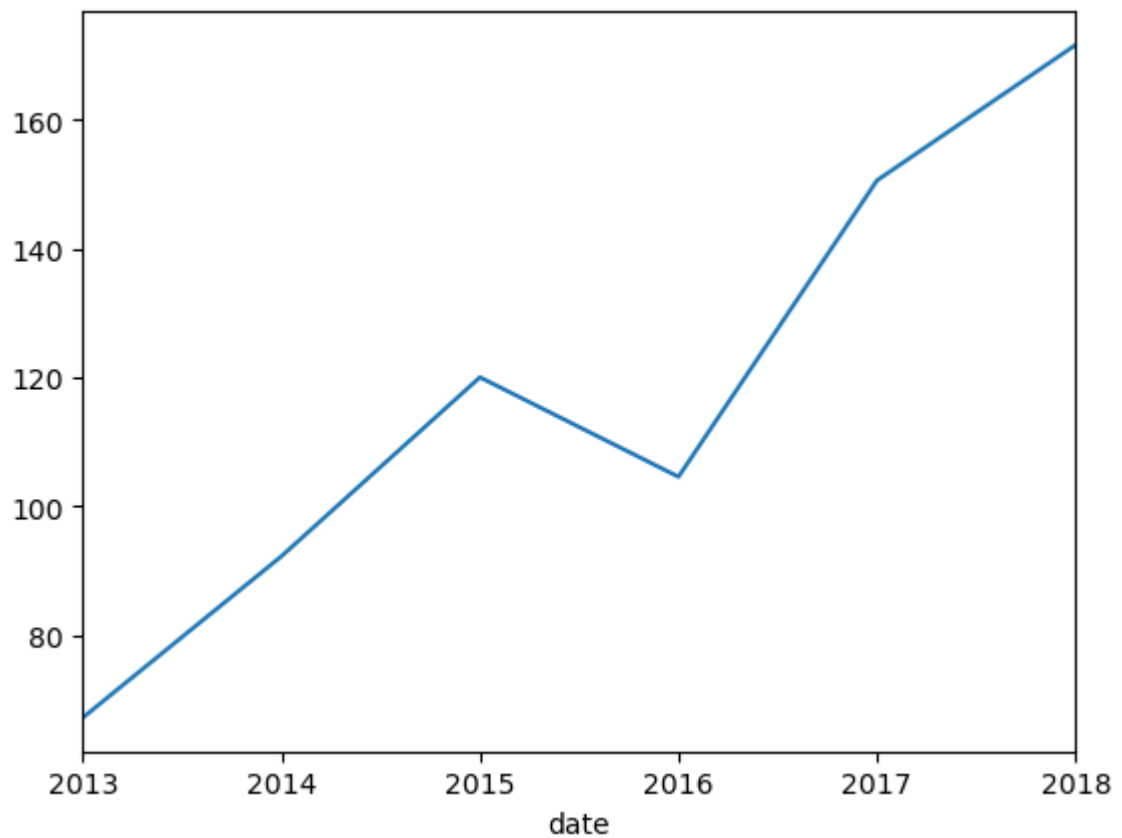
```
In [184... apple['close'].resample('Y').mean()
```

```
Out[184... date
2013-12-31      67.237839
2014-12-31      92.264531
2015-12-31     120.039861
2016-12-31     104.604008
2017-12-31     150.585080
2018-12-31     171.594231
Freq: YE-DEC, Name: close, dtype: float64
```

```
In [186... apple['close'].resample('Y').mean().plot()
```

```
Out[186... <Axes: xlabel='date'>
```





```
In [188... apple['close'].resample('Q').mean()
```

```
Out[188... date
2013-03-31    64.020291
2013-06-30    61.534692
2013-09-30    66.320670
2013-12-31    75.567478
2014-03-31    76.086293
2014-06-30    85.117475
2014-09-30    98.163311
2014-12-31   108.821016
2015-03-31   120.776721
2015-06-30   127.937937
2015-09-30   117.303438
2015-12-31   114.299297
2016-03-31    99.655082
2016-06-30    99.401250
2016-09-30   105.866094
2016-12-31   113.399048
2017-03-31   131.712500
2017-06-30   147.875397
2017-09-30   155.304603
2017-12-31   167.148254
2018-03-31   171.594231
Freq: QE-DEC, Name: close, dtype: float64
```

```
In [298... company_list[320]
```

```
Out[298... 'C:\\Users\\lebon\\OneDrive\\Desktop\\LEANDRINHO\\Udemy\\DataAnalyticsReal-
WorldProjectsPythonR\\resources\\individual_stocks_5yr\\MSFT_data.csv'
```

```
In [300... app = pd.read_csv(company_list[1])
amzn = pd.read_csv(company_list[37])
google = pd.read_csv(company_list[207])
msft = pd.read_csv(company_list[320])
```

```
In [302... closing_price = pd.DataFrame()
```

```
In [304... closing_price['apple_close'] = app['close']
closing_price['amzn_close'] = amzn['close']
closing_price['goog_close'] = google['close']
closing_price['msft_close'] = msft['close']
```

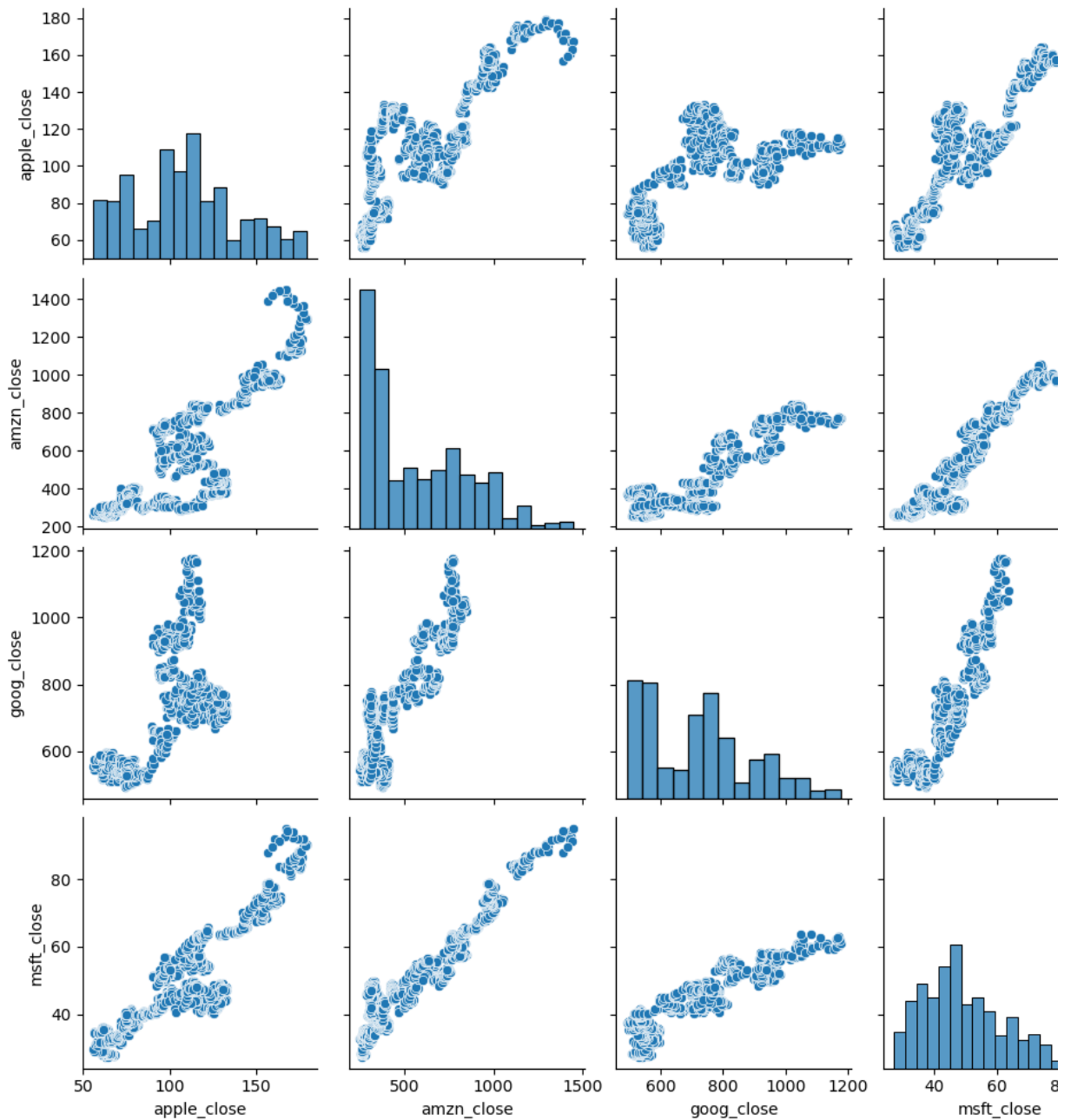
```
In [306... closing_price
```

```
Out[306...      apple_close  amzn_close  goog_close  msft_close
0      67.8542      261.95      558.46      27.55
1      68.5614      257.21      559.99      27.86
2      66.8428      258.70      556.97      27.88
3      66.7156      269.47      567.16      28.03
4      66.6556      269.24      567.00      28.04
...      ...      ...      ...      ...
1254    167.7800    1390.00      NaN      94.26
1255    160.5000    1429.95      NaN      91.78
1256    156.4900    1390.00      NaN      88.00
1257    163.0300    1442.84      NaN      91.33
1258    159.5400    1416.78      NaN      89.61
```

1259 rows × 4 columns

```
In [308... sns.pairplot(closing_price)
```

```
Out[308... <seaborn.axisgrid.PairGrid at 0x16c2b5faae0>
```



In [ ]:

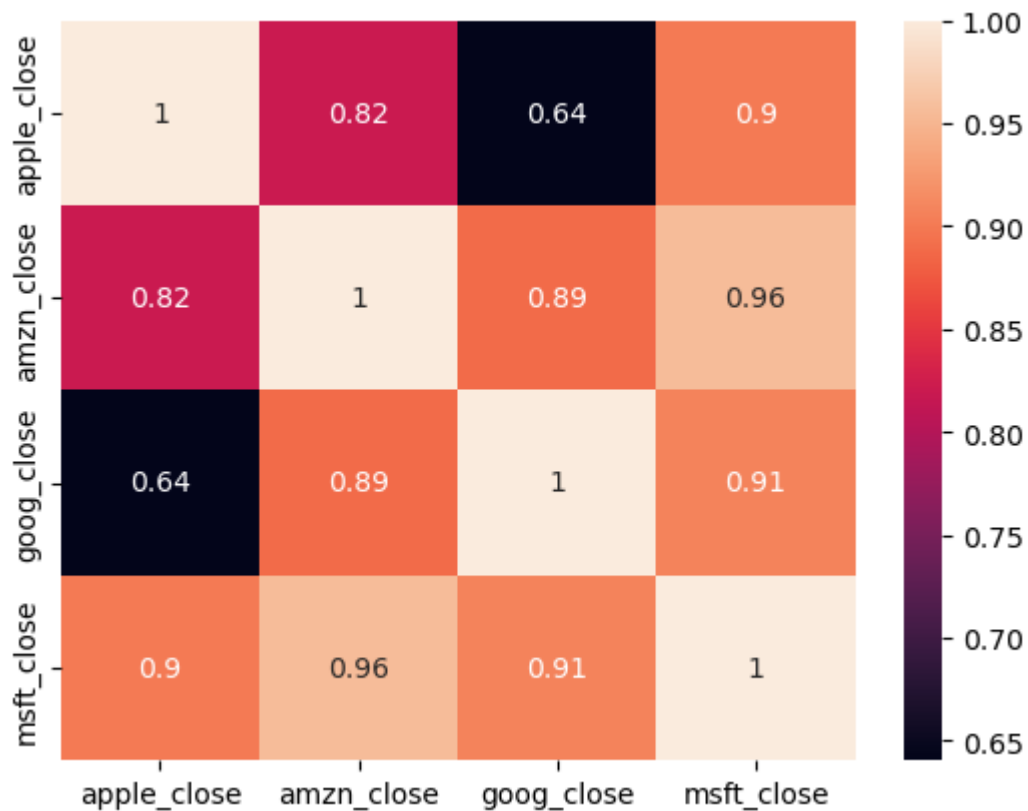
```
In [312...] closing_price.corr()
```

```
Out[312...]
           apple_close  amzn_close  goog_close  msft_close
apple_close      1.000000      0.819078      0.640522      0.899689
amzn_close       0.819078      1.000000      0.888456      0.955977
goog_close       0.640522      0.888456      1.000000      0.907011
msft_close       0.899689      0.955977      0.907011      1.000000
```

In [ ]:

```
In [319...] sns.heatmap(closing_price.corr(), annot=True)
```

```
Out[319...] <Axes: >
```



In [ ]:

```
In [322... for col in closing_price.columns:
            closing_price[col + '_pct_change'] = (closing_price[col] - closing_pr:
```

In [324... closing\_price

```
Out[324...
   apple_close  amzn_close  goog_close  msft_close  apple_close_pct_change  amzn_c
0      67.8542      261.95      558.46      27.55                      NaN
1      68.5614      257.21      559.99      27.86          1.042235
2      66.8428      258.70      556.97      27.88         -2.506658
3      66.7156      269.47      567.16      28.03         -0.190297
4      66.6556      269.24      567.00      28.04         -0.089934
...         ...         ...         ...         ...         ...
1254     167.7800     1390.00         NaN      94.26          0.209043
1255     160.5000     1429.95         NaN      91.78         -4.339015
1256     156.4900     1390.00         NaN      88.00         -2.498442
1257     163.0300     1442.84         NaN      91.33          4.179181
1258     159.5400     1416.78         NaN      89.61         -2.140710
```

1259 rows × 8 columns

In [326... closing\_price.columns

```
Out[326...] Index(['apple_close', 'amzn_close', 'goog_close', 'msft_close',  
                  'apple_close_pct_change', 'amzn_close_pct_change',  
                  'goog_close_pct_change', 'msft_close_pct_change'],  
                  dtype='object')
```

```
In [328...] clsing_p = closing_price[['apple_close_pct_change', 'amzn_close_pct_change',  
                                     'goog_close_pct_change', 'msft_close_pct_change']]
```

```
In [330...] clsing_p
```

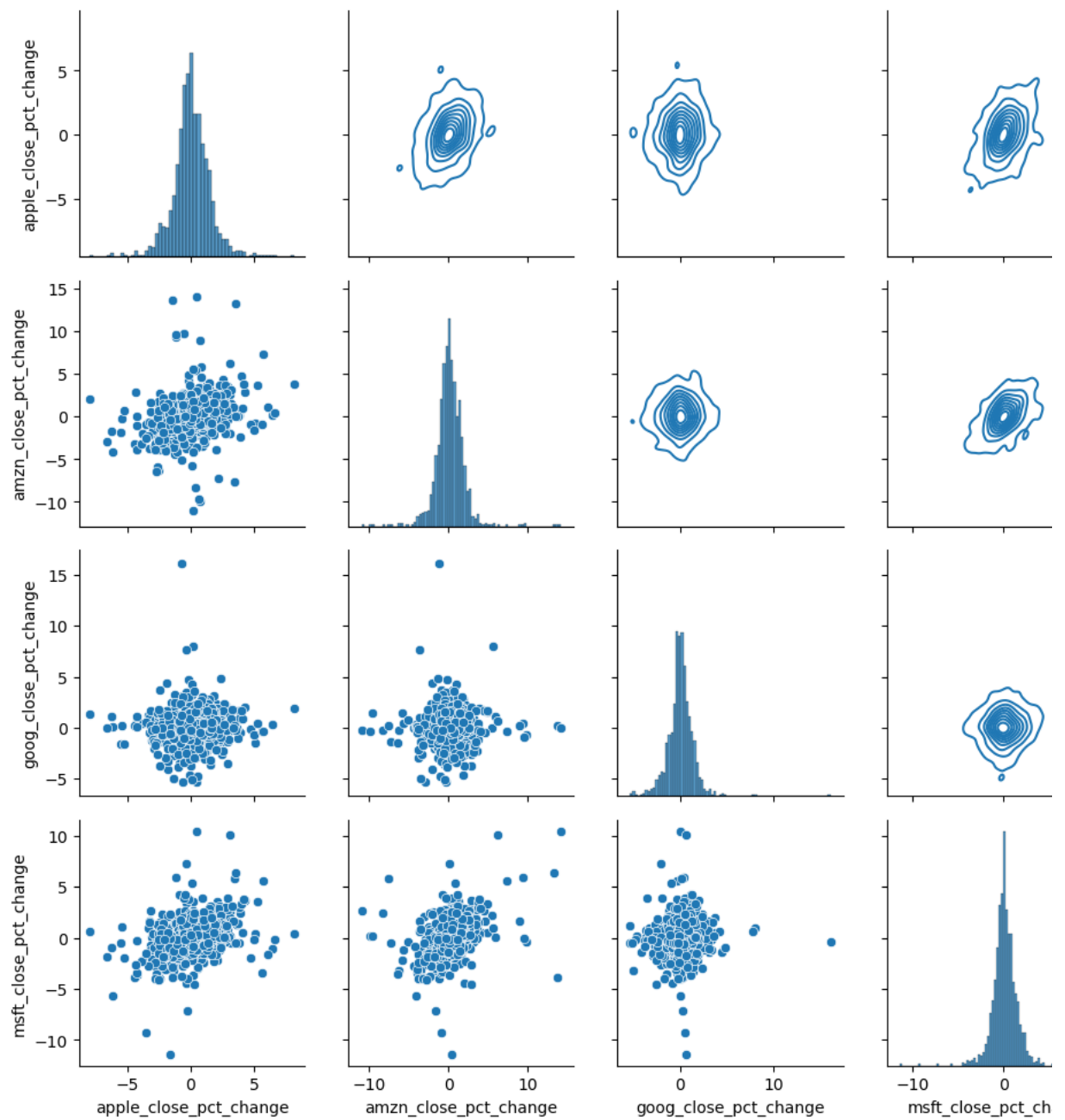
```
Out[330...]      apple_close_pct_change  amzn_close_pct_change  goog_close_pct_change  msft_close_pct_change  
0                NaN                NaN                NaN  
1             1.042235             -1.809506             0.273968  
2             -2.506658              0.579293            -0.539295  
3             -0.190297              4.163123             1.829542  
4             -0.089934             -0.085353            -0.028211  
...                ...                ...                ...  
1254            0.209043            -4.196734             NaN  
1255           -4.339015              2.874101             NaN  
1256           -2.498442            -2.793804             NaN  
1257            4.179181              3.801439             NaN  
1258           -2.140710            -1.806160             NaN
```

1259 rows × 4 columns

```
In [ ]:
```

```
In [335...] g = sns.PairGrid(data = clsing_p)  
g.map_diag(sns.histplot)  
g.map_lower(sns.scatterplot)  
g.map_upper(sns.kdeplot)
```

```
Out[335...] <seaborn.axisgrid.PairGrid at 0x16c31879670>
```



In [ ]:

In [338... clsing\_p.corr()

Out[338...

	apple_close_pct_change	amzn_close_pct_change	goog_close_pc
apple_close_pct_change	1.000000	0.287659	
amzn_close_pct_change	0.287659	1.000000	
goog_close_pct_change	0.036202	0.027698	
msft_close_pct_change	0.366598	0.402678	

In [ ]: