

Data Analysis Stock price

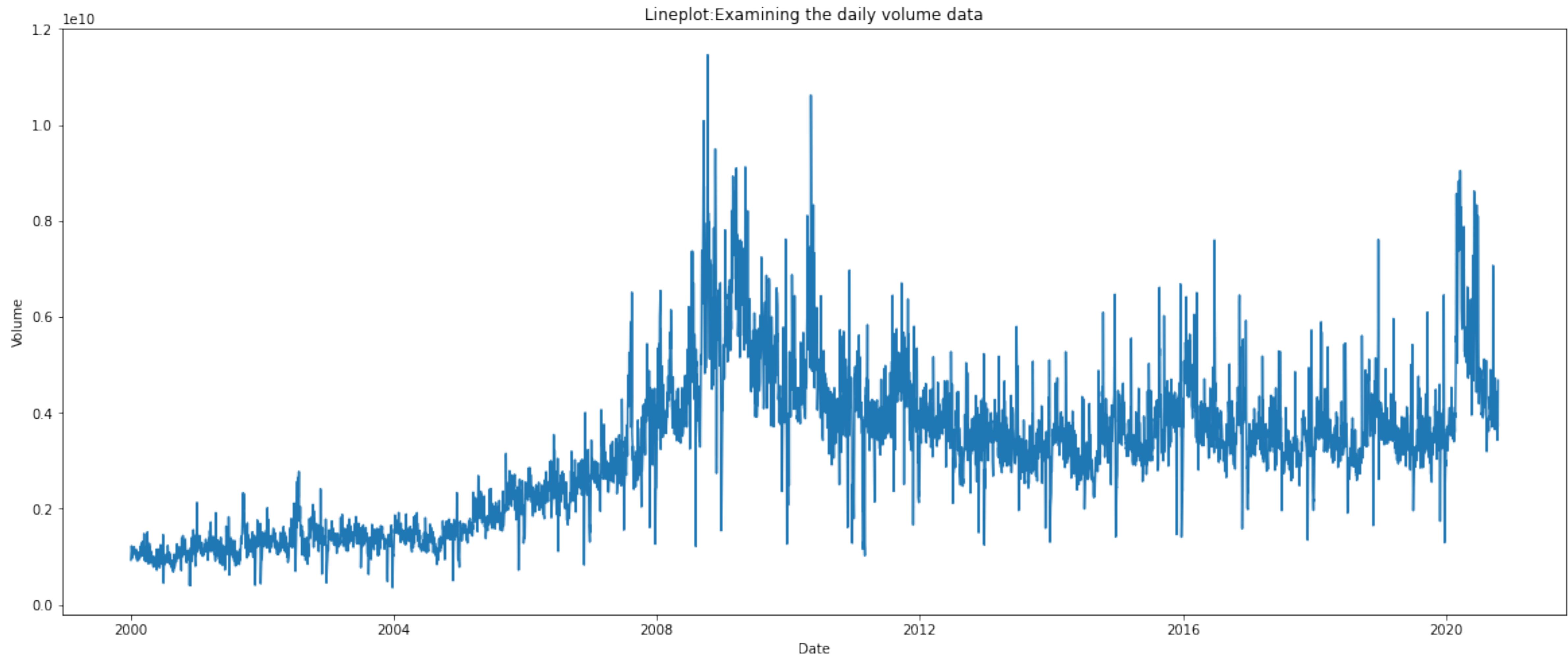
Examine the daily open, close, volume data

There is a saying on Wall Street, “Santa brings the Christmas rally”.

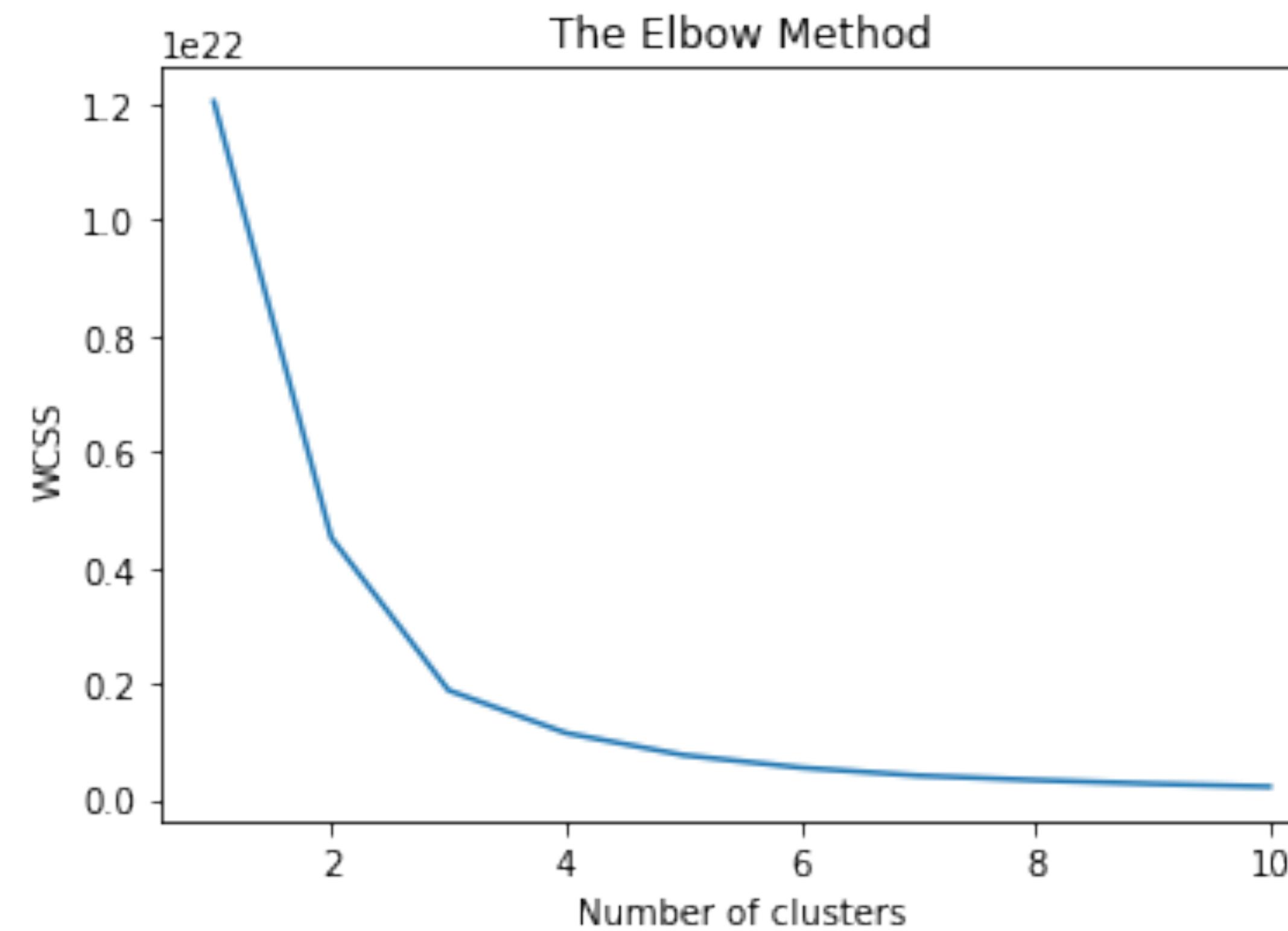
Our Data

	Date	Open	High	Low	Close	Adj Close	Volume
0	2000-01-03	1469.250000	1478.000000	1438.359985	1455.219971	1455.219971	931800000
1	2000-01-04	1455.219971	1455.219971	1397.430054	1399.420044	1399.420044	1009000000
2	2000-01-05	1399.420044	1413.270020	1377.680054	1402.109985	1402.109985	1085500000
3	2000-01-06	1402.109985	1411.900024	1392.099976	1403.449951	1403.449951	1092300000
4	2000-01-07	1403.449951	1441.469971	1400.729980	1441.469971	1441.469971	1225200000

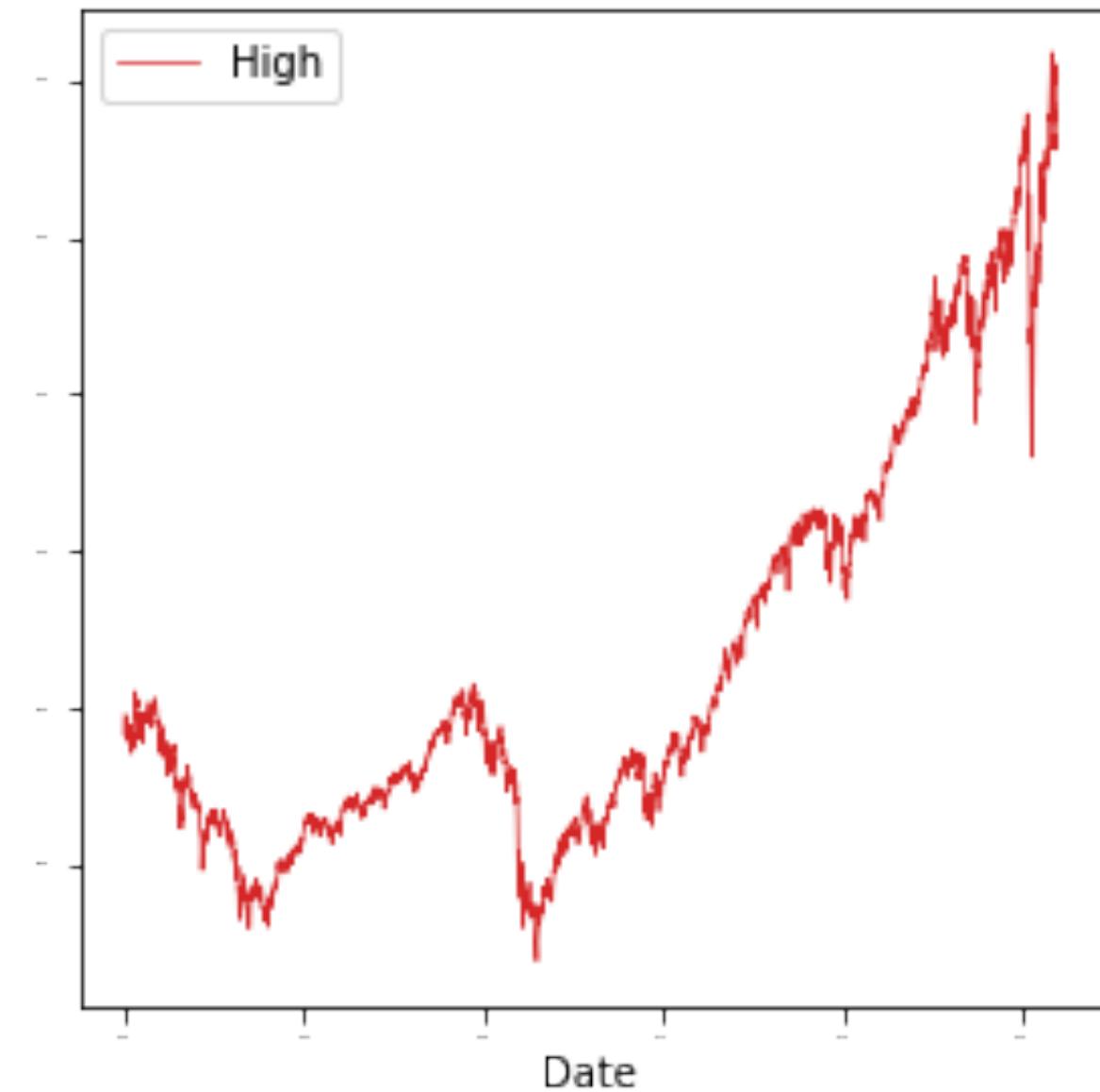
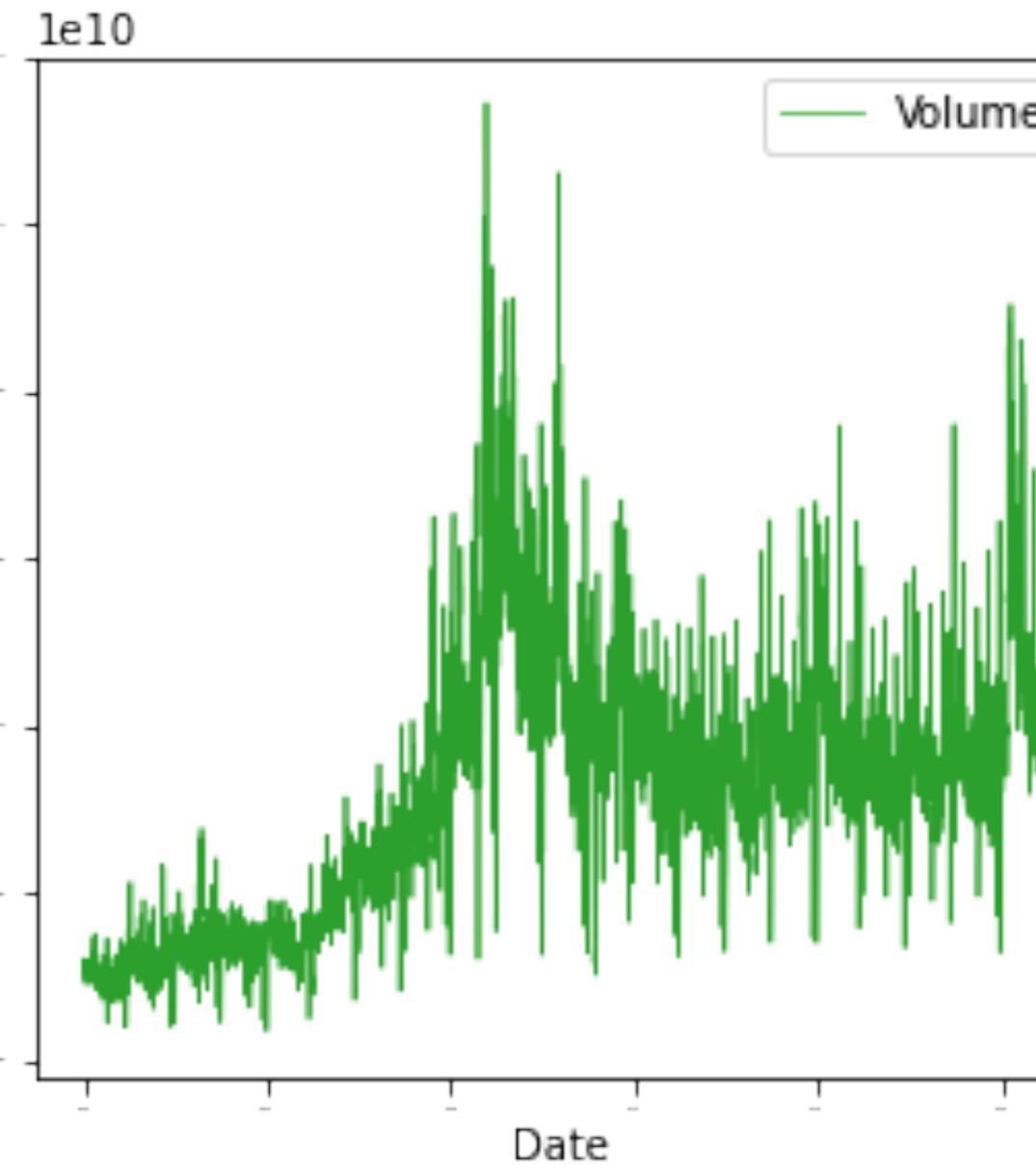
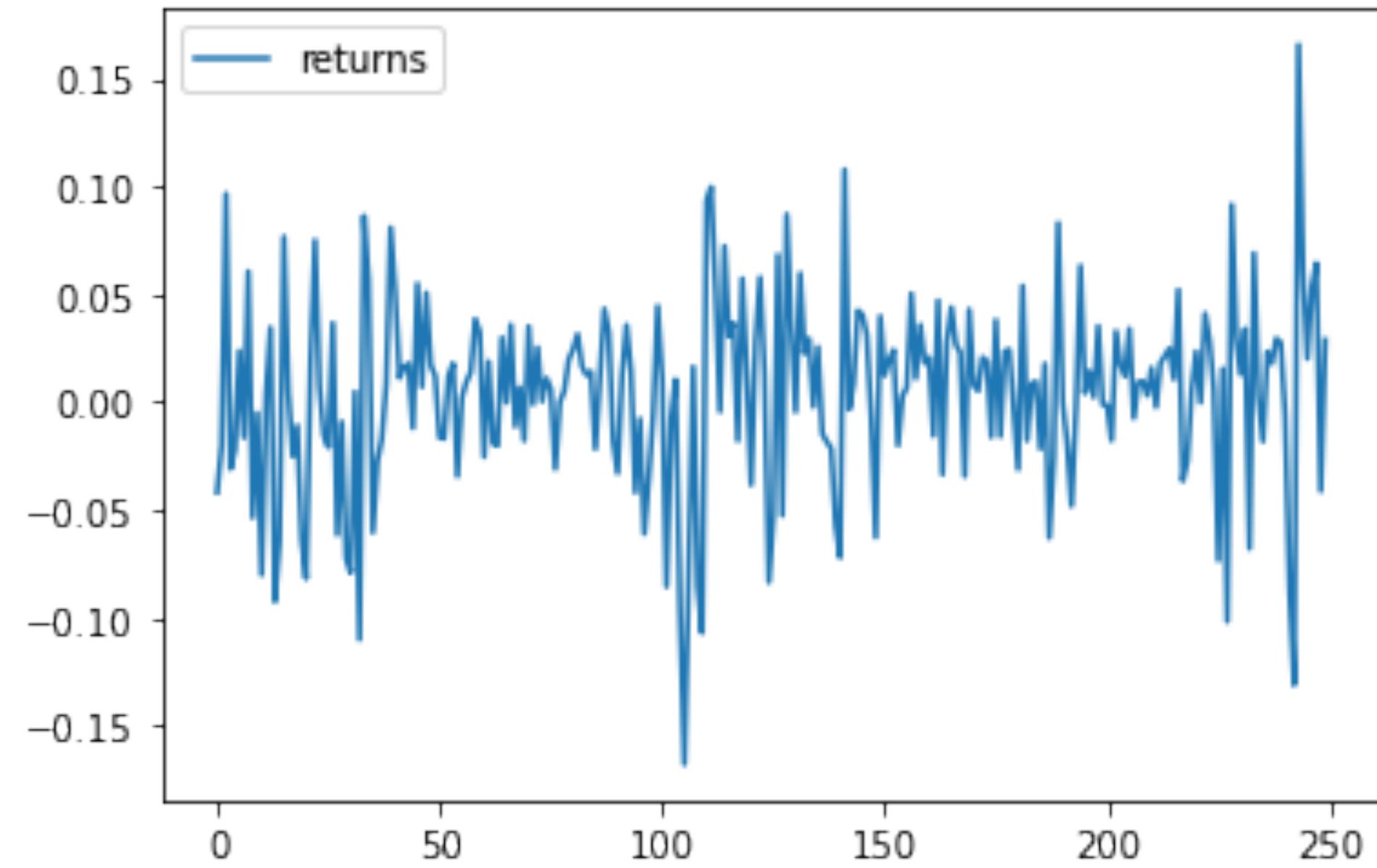
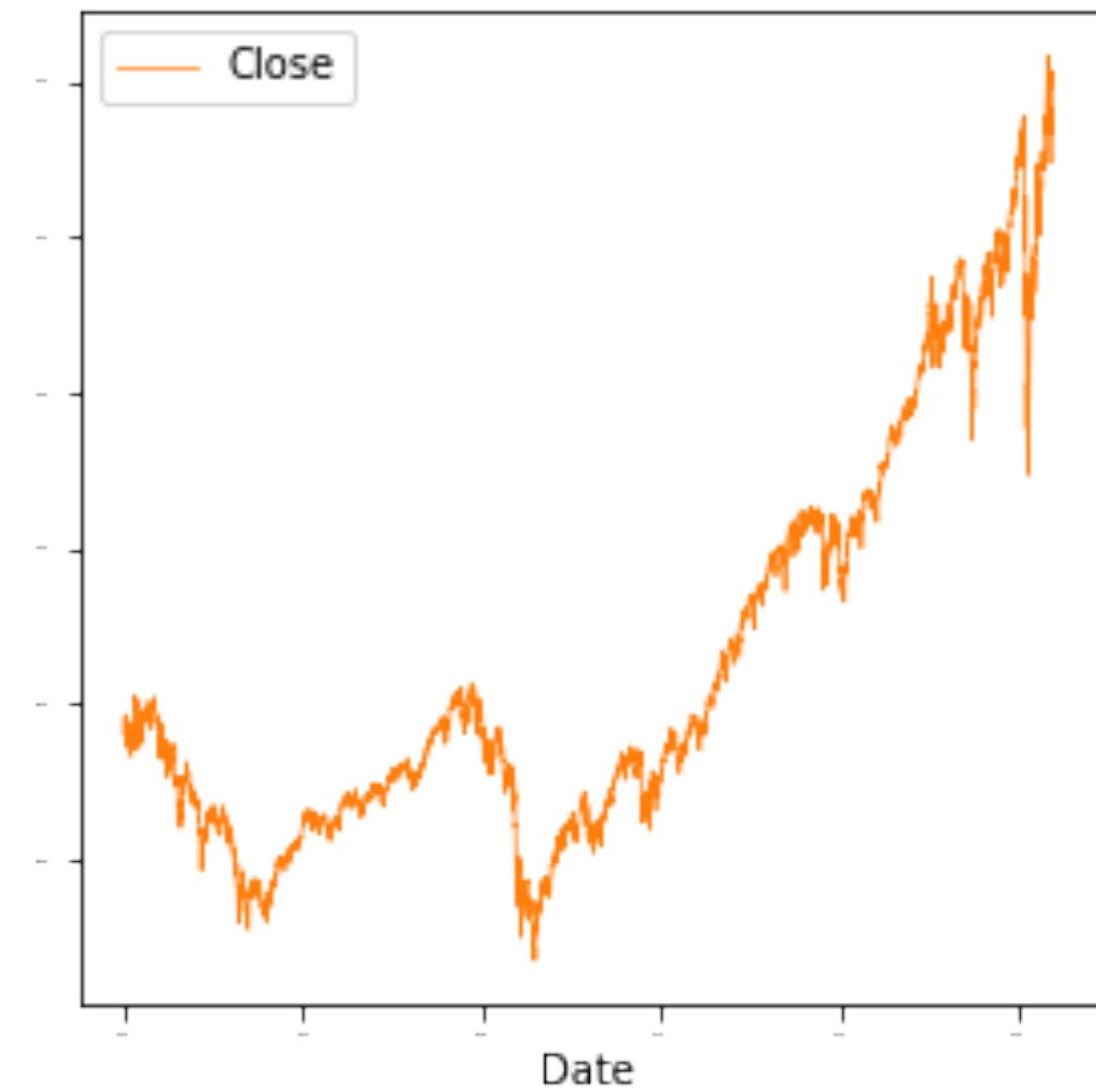
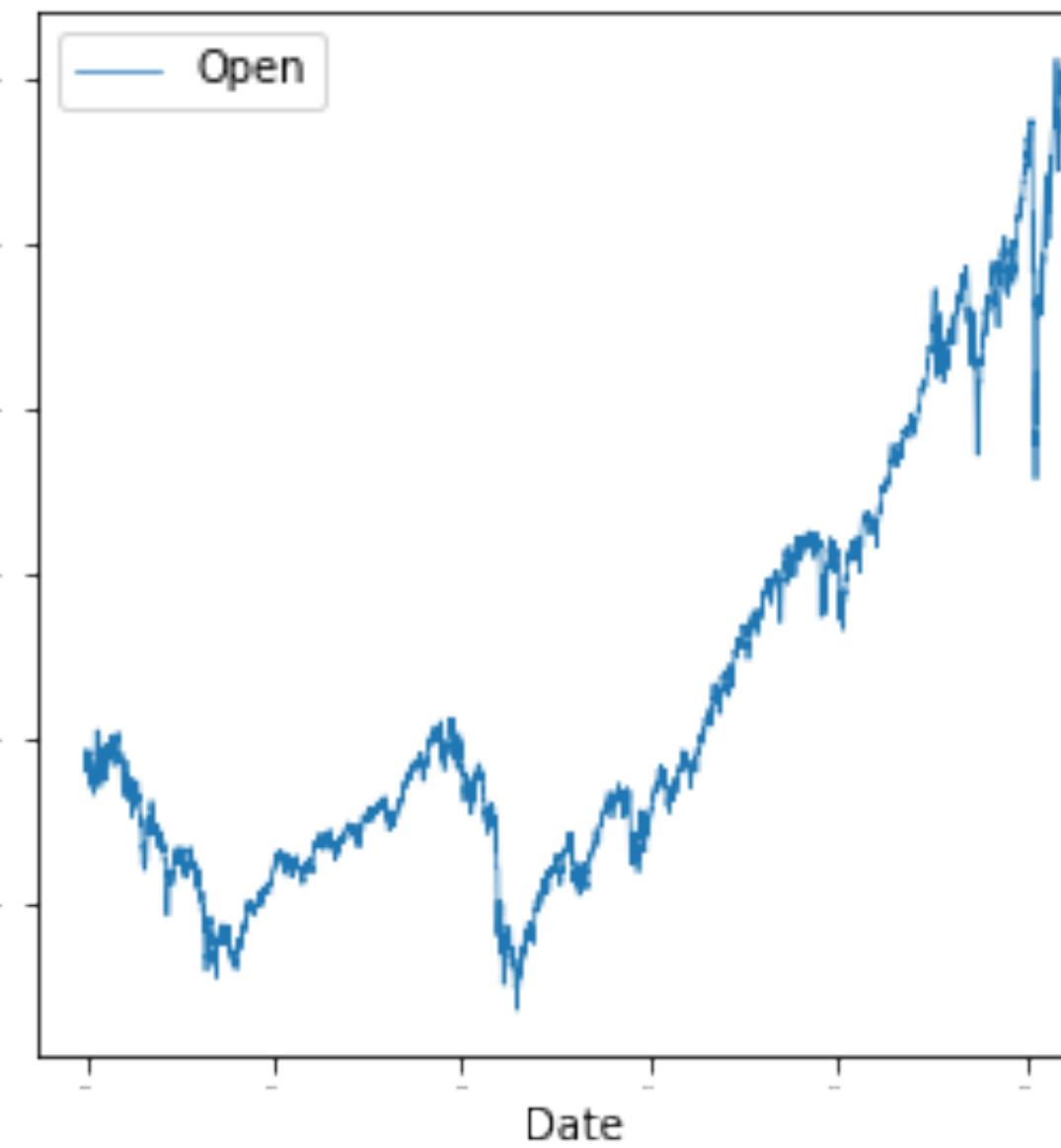
Daily Volume data



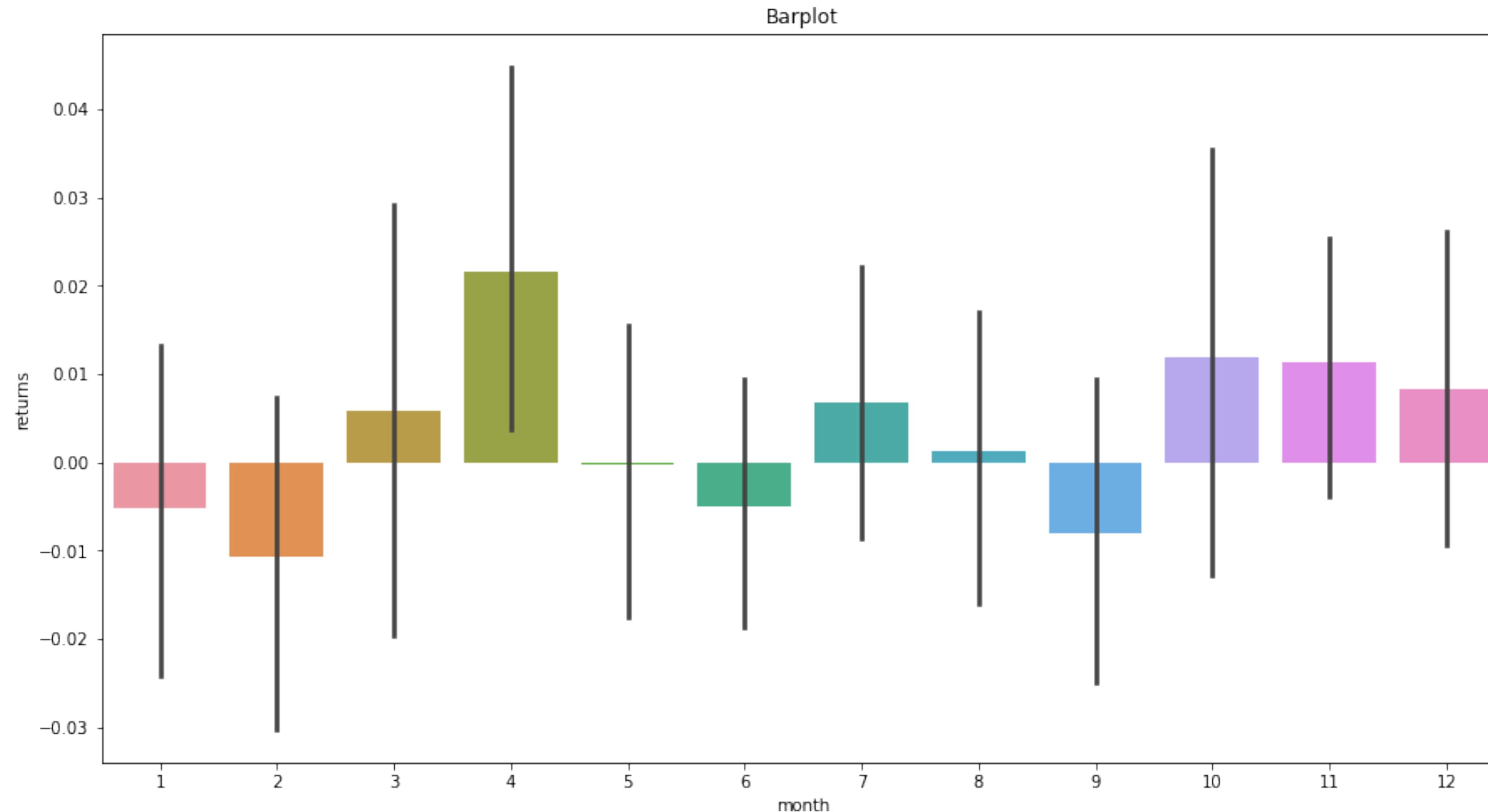
Volume data cluster



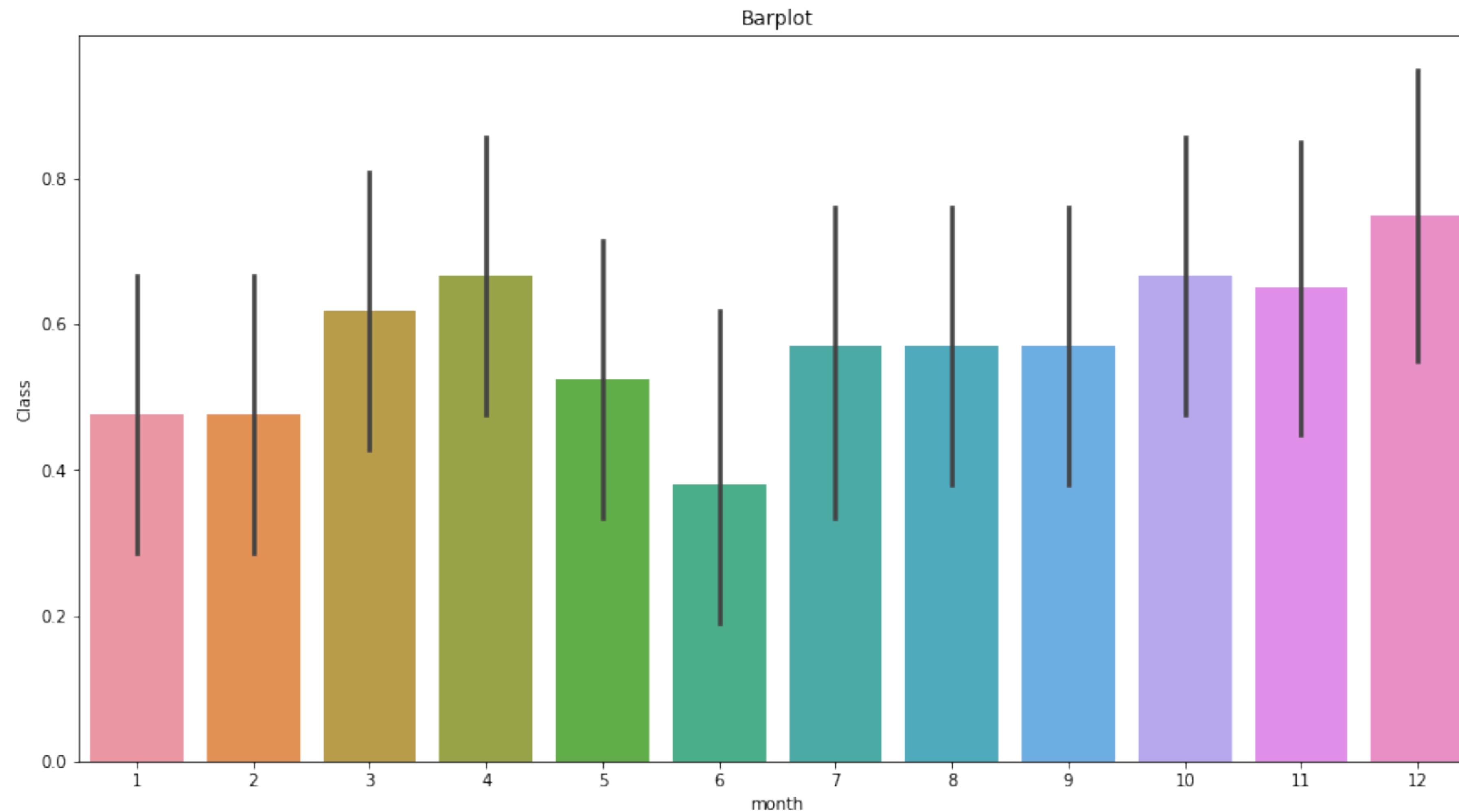
Daily Open, Close, Volume, high, Returns data



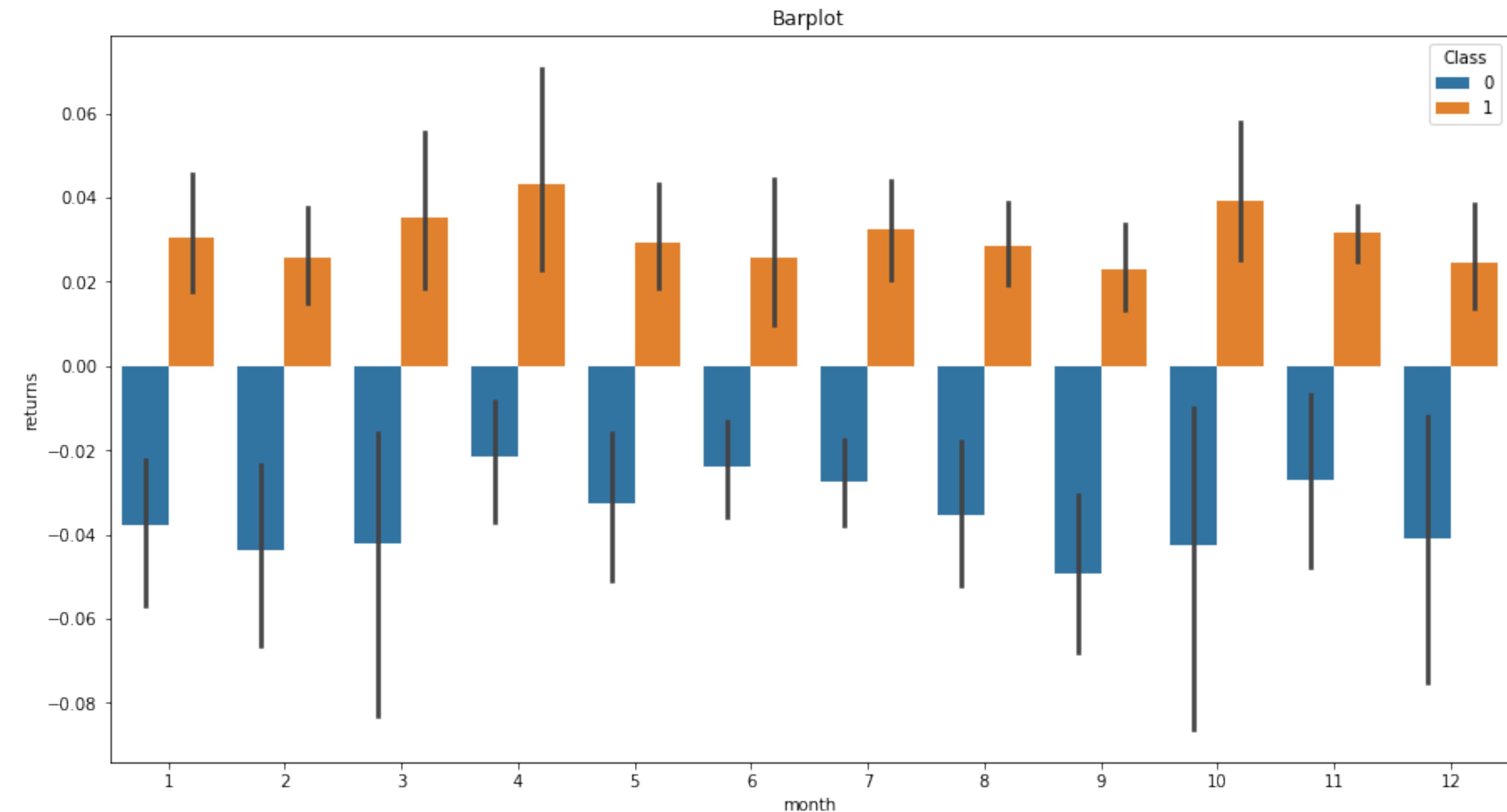
Returns for each month



Class(avg) for each month

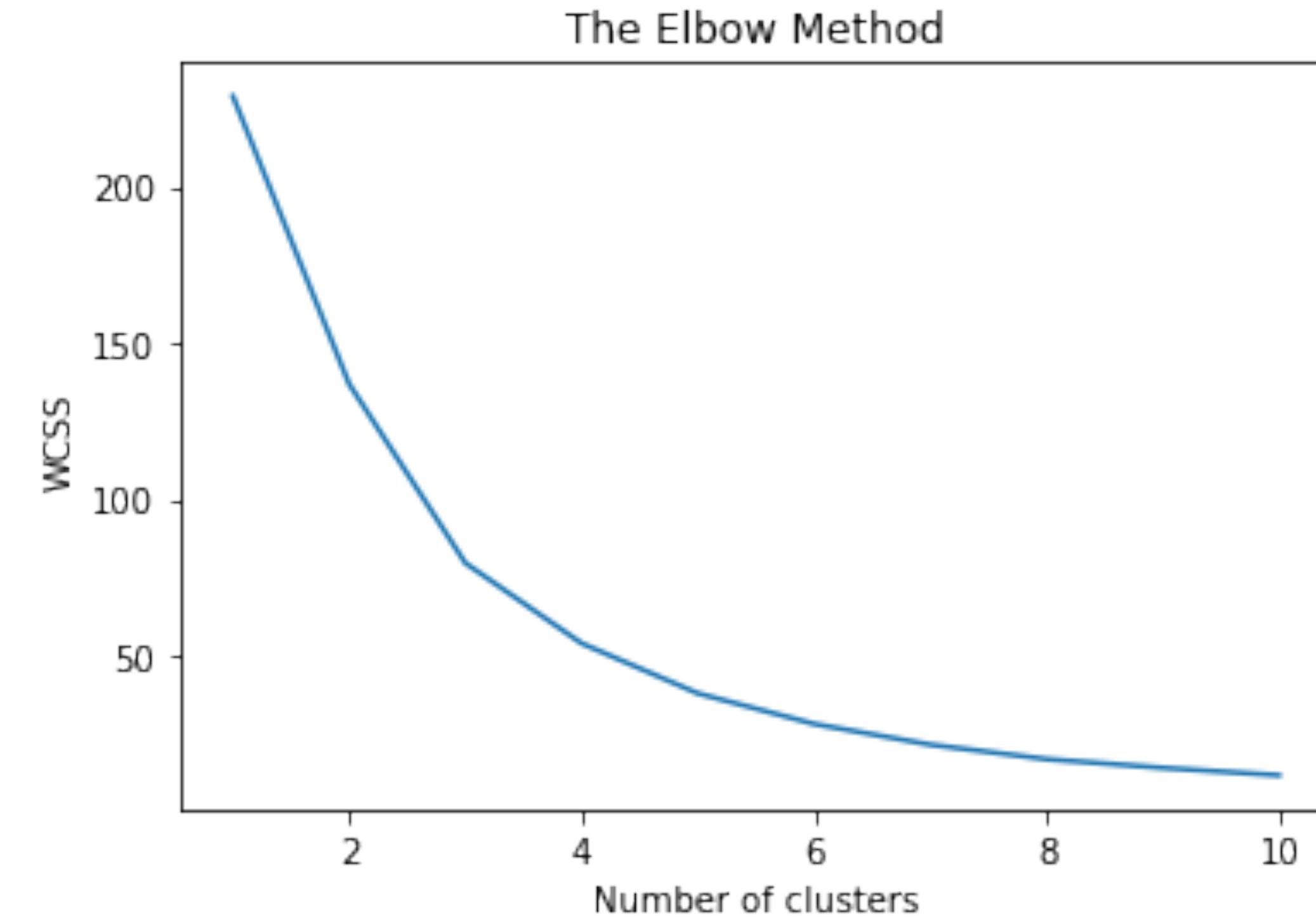


Returns and Class



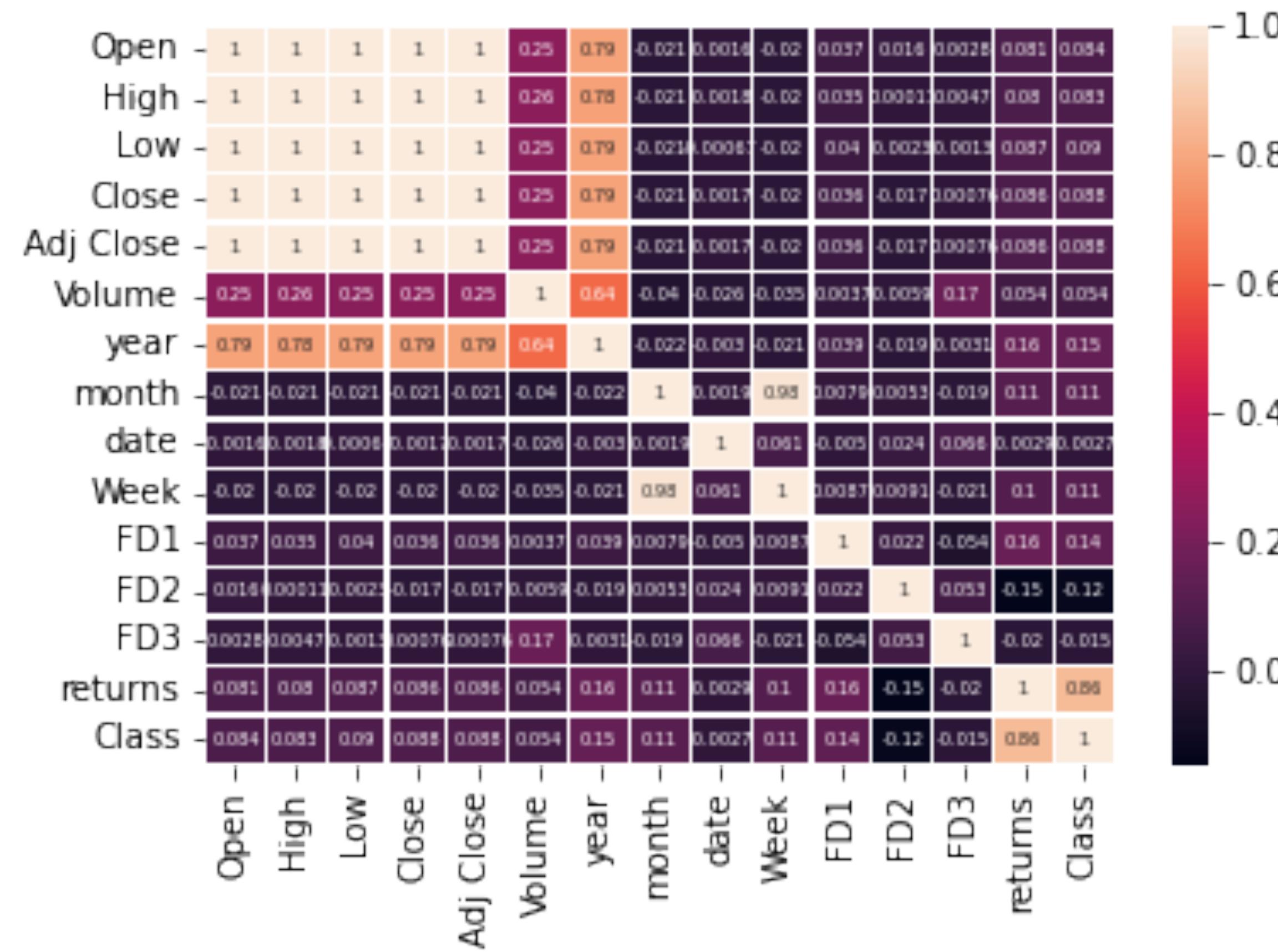
Optimum number of cluster for FD1 FD2 FD3

- Fractional difference data could be clustered
- FD1:fractional difference between current day open and previous day
- FD2:fractional difference between current day open and current day close
- FD3:fractional difference between current day volume and previous day



Correlation heat map

- We drop highly correlated data to build a model



DTClassifier error metrics

	precision	recall	f1-score	support
0	0.91	0.93	0.92	659
1	0.95	0.93	0.94	911
accuracy			0.93	1570
macro avg	0.93	0.93	0.93	1570
weighted avg	0.93	0.93	0.93	1570
[[611 48] [62 849]]				

Final DTclassifier tree

