Final Project

US Stock Trading Prediction Model

Majid Feiz

1-Pre Processing

In [2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [3]:
```
idx = pd.IndexSlice
```

In [4]:
```
#dir_root ='C:/Users/Owner/Documents/Majid/UConn/CSE5835_ML_for_Physical_Scienc_Systems/Final_Project/'
```

Load source data

In [5]:
```
df = pd.read_csv('data/us_stocks.csv',
                 parse_dates=['date'],
                 index_col=['ticker', 'date'])
```

In [5]:
```
df
```

Out[5]:

ticker	date	open	high	low	close	volume	ex-dividend	split_ratio	adj_open	adj_high	adj_low	adj_close	adj_volume
A	1999-11-18	45.50	50.00	40.0000	44.00	44739900.0	0.0	1.0	31.041951	34.112034	27.289627	30.018590	44739900.0
	1999-11-19	42.94	43.00	39.8100	40.38	10897100.0	0.0	1.0	29.295415	29.336350	27.160002	27.548879	10897100.0
	1999-11-22	41.31	44.00	40.0600	44.00	4705200.0	0.0	1.0	28.183363	30.018590	27.330562	30.018590	4705200.0
	1999-11-23	42.50	43.63	40.2500	40.25	4274400.0	0.0	1.0	28.995229	29.766161	27.460188	27.460188	4274400.0
	1999-11-24	40.13	41.94	40.0000	41.06	3464400.0	0.0	1.0	27.378319	28.613174	27.289627	28.012803	3464400.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
ZUMZ	2018-03-21	23.80	24.60	23.6058	23.95	354092.0	0.0	1.0	23.800000	24.600000	23.605800	23.950000	354092.0
	2018-03-22	23.90	24.35	23.3000	23.35	269607.0	0.0	1.0	23.900000	24.350000	23.300000	23.350000	269607.0
	2018-03-23	23.55	24.20	23.4500	23.55	301584.0	0.0	1.0	23.550000	24.200000	23.450000	23.550000	301584.0
	2018-03-26	23.75	24.80	23.7000	24.65	375320.0	0.0	1.0	23.750000	24.800000	23.700000	24.650000	375320.0
	2018-03-27	24.65	24.65	23.3500	23.60	403884.0	0.0	1.0	24.650000	24.650000	23.350000	23.600000	403884.0

15389314 rows × 12 columns

In [6]:
```
df.info(show_counts=True)
```

```
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 15389314 entries, ('A', Timestamp('1999-11-18 00:00:00')) to ('ZUMZ', Timestamp('2018-03-27 00:00:00'))
Data columns (total 12 columns):
 #   Column       Non-Null Count     Dtype  
---  ------       --------------     -----  
 0   open         15388776 non-null  float64
 1   high         15389259 non-null  float64
 2   low          15389259 non-null  float64
 3   close        15389313 non-null  float64
 4   volume       15389314 non-null  float64
 5   ex-dividend  15389314 non-null  float64
 6   split_ratio  15389313 non-null  float64
 7   adj_open     15388776 non-null  float64
 8   adj_high     15389259 non-null  float64
 9   adj_low      15389259 non-null  float64
 10  adj_close    15389313 non-null  float64
 11  adj_volume   15389314 non-null  float64
dtypes: float64(12)
memory usage: 1.4+ GB
```

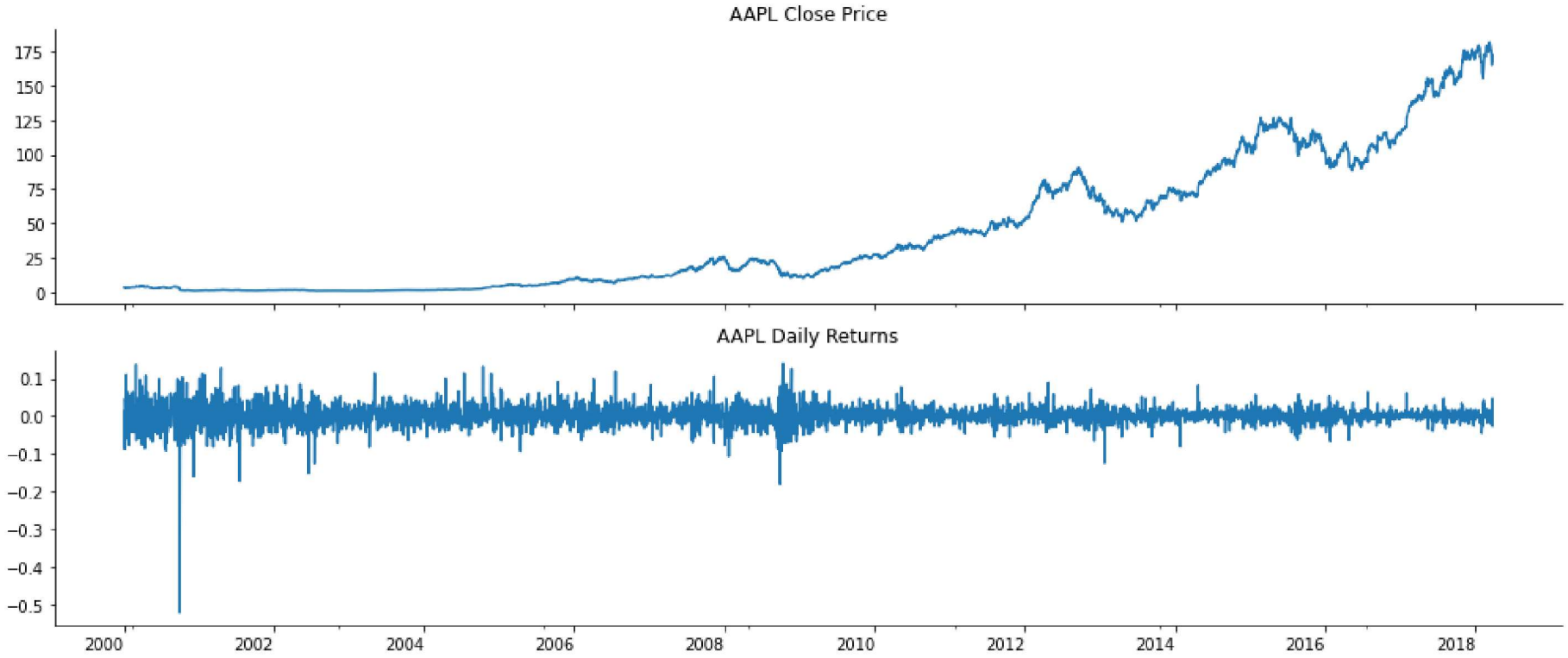
In [7]:
```
df = (df.loc[idx[:, '2000':], :]
      .filter(like='adj')
      .rename(columns=lambda x: x.replace('adj_', ''))
      .dropna())
```

In [8]:
```
df.info(show_counts=True)
```

```
<class 'pandas.core.frame.DataFrame'>
MultiIndex: 11343366 entries, ('A', Timestamp('2000-01-03 00:00:00')) to ('ZUMZ', Timestamp('2018-03-27 00:00:00'))
Data columns (total 5 columns):
 #   Column  Non-Null Count     Dtype  
---  ------  --------------     -----  
 0   open    11343366 non-null  float64
 1   high    11343366 non-null  float64
 2   low     11343366 non-null  float64
 3   close   11343366 non-null  float64
 4   volume  11343366 non-null  float64
dtypes: float64(5)
memory usage: 476.7+ MB
```

In [9]:
```
symbol = 'AAPL'
```

In [10]:
```
fig, axes = plt.subplots(nrows=2, figsize=(14, 6), sharex=True)
s = df.loc[symbol, 'close']
s.plot(rot=0, ax=axes[0], title=f'{symbol} Close Price')
s.pct_change().plot(rot=0, ax=axes[1], title=f'{symbol} Daily Returns')
axes[1].set_xlabel('')
sns.despine()
fig.tight_layout()
```



Store in HDF5 format

In [7]:
```
df.to_hdf('data/stock_prices.h5', 'us_stocks')
```

In [ ]:
```
```

In [ ]:

In [ ]: