

```
# Veera Venkata Lakshmi Srikanth Marni
# IT-586 Correlation Assignment
```

1) Before analyzing the file, what do you think is the largest indicator to predict 'Tomorrow's Percent Change'? Purely your opinion, try to form some intuition on features.

I think the volume of trades is the biggest indicator of how a stock is going to perform in the future. When a stock volume increases/decreases suddenly, it means some kind of interest increase/decrease is happening which impacts the future price of that stock.

```
# Global modules
import numpy as np
import pandas as pd
import os
```

```
# Data source
stocks = 'Data_files'+os.sep+'stocks.csv'
```

```
def correlation(data_source):
    data = pd.read_csv(data_source)
    print(data)
    del data['Open']
    return data.corr()
```

```
correlation(stocks)
```

2) Were there any strong indicators (positive or negative)?

"Positive: Low and High, Close and High, Low and Close, Volume and Yesterday Volume"

3) What in the data is missing that might be a stronger indicator from your own knowledge of stocks?

"Moving Average Convergence/Divergence (MACD) might be a strong indicator to predict the future price of a stock."

4) Come up with one new feature based off the current data that gives a stronger (positive or negative) correlation than the existing values (stronger just means slightly higher positive or slightly lower negative correlation)

"To improve the correlation, I added a Low_plus_one feature by adding one to the Low feature."

```
def improve_correlation(data_source):
    data = pd.read_csv(data_source)
    data['Low_plus_one'] = data['Low'] + 1
    print('\n***** After adding Low_plus_one feature.\n')
    return data.corr()
```

```
improve_correlation(stock)
```