Summary

The goal in this workshop is to implement a trading strategy with google trends information, the strategy is based from the paper Quantifying Trading Behavior in Financial Markets Using Google Trends.

Our data corresponds to the Google trends information from 2005 to 2011 of the word "debt" and the Down jones index from yahoo finance.

The following libraries were used in the code.

```python
import pandas as pd
import numpy as np
from datetime import datetime, timedelta
import matplotlib.pyplot as plt
import yfinance as yf
```

One of the drawbacks of the initial data is about the frequency of Google trends information, unfortunately the trend report only contains monthly information and the Down Jones information have a daily frequency.

For this exercise it was necessary to merge those databases in a weekly frequency table to calculate the signals of short or long investment.

Google trends information is always monthly and the performance of the strategies will be affected by the lack of details in trends in terms of weeks.

1. **Download Dow jones data from 2005 to 2011**

The data from yahoo finance starts in 2005-01-01 and finishes in 2011-12-31, the start time is 1 week before 2005-01-01 allowing a gat for handling the weekly data frequency.

```python
Stardate=datetime(2005, 1, 1)
Enddate=datetime(2011, 12, 31)
tck_DJ_data = yf.download('^DJI', Stardate-
timedelta(weeks=1),Enddate)
Close_DJ_data=pd.DataFrame(tck_DJ_data['Close'])
```

The daily data from Yahoo finance have some missing days because they are holidays, so it necessary to add those records and perform a forward fill.

```python
Close_DJ_data = (Close_DJ_data.set_index('Date')
    .reindex(pd.date_range(Stardate-
timedelta(weeks=1), Enddate, freq='D'))
    .rename_axis(['Date'])
    .fillna(method='ffill')
    .dropna()
    .reset_index())
```

2. **Load the monthly google trends data for term "debt"**

```python
Google_trends=pd.read_csv('debt_google_trend.csv')
Google_trends['Mes']=Google_trends['Mes'].astype('da
tetime64[ns]')
```

For Google trends monthly data base was transform to have all the days between 2005 and 2011 and filled with forward fill.

```python
Google_trends = (Google_trends.set_index('Mes')

    .reindex(pd.date_range(Stardate-
timedelta(weeks=1), Enddate, freq='D'))
    .rename_axis(['Mes'])
    .fillna(method='ffill')
    .dropna()
    .reset_index())
```

3. **Combine the Down Jones Close price data base with the google trends database and set date frequency as weekly starting every Monday.**

The function used was merge from pandas with an inner join.

```python
Cnsl_db_Gtr=pd.merge(Close_DJ_data,Google_trends,how
='inner',left_on='Date',right_on='Mes')
Cnsl_db_Gtr=Cnsl_db_Gtr.loc[Cnsl_db_Gtr['Date']>=Sta
rdate+timedelta(days=2)]
```

Once both databases with a daily frequency in dates were merged the frequency was updated to weekly.

```python
Cnsl_db_Gtr2 = (Cnsl_db_Gtr.set_index('Date')
    .reindex(pd.date_range(Stardate+timedelta(days
=2), Enddate, freq='W-MON'))
    .rename_axis(['Date'])
    .dropna()
    .reset_index())
```

4. Google trends plot
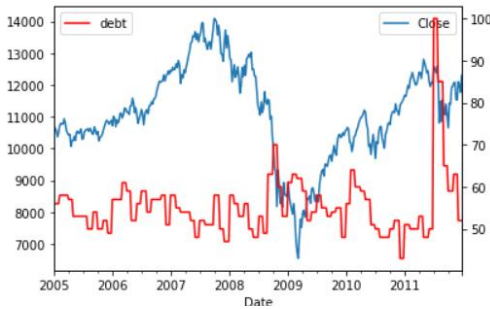
5. Strategy rules

If the current week Google trend search value is higher than moving average for previous 3 weeks select short on the other hand if the current week Google trend search value is lesser than the moving average for previous 3 weeks select long.

```
try:
    del StrategyDataframe
    StrategyDataframe=Cnsl_db_Gtr2.copy()
except:
    StrategyDataframe=Cnsl_db_Gtr2.copy()
StrategyDataframe['MovAvg']=StrategyDataframe['debt'
].rolling(3).mean()
StrategyDataframe['MovAvg']=StrategyDataframe['MovAv
g'].shift(1)
```

6. Create signal column in dataframe for trading strategy.

```
StrategyDataframe['Signal']=np.sign(StrategyDatafram
e['MovAvg']-StrategyDataframe['debt'])
StrategyDataframe['Signal']=StrategyDataframe['Signa
l'].shift(1)
StrategyDataframe['Signal']=StrategyDataframe['Signa
l'].replace(to_replace=0, method='ffill')
```

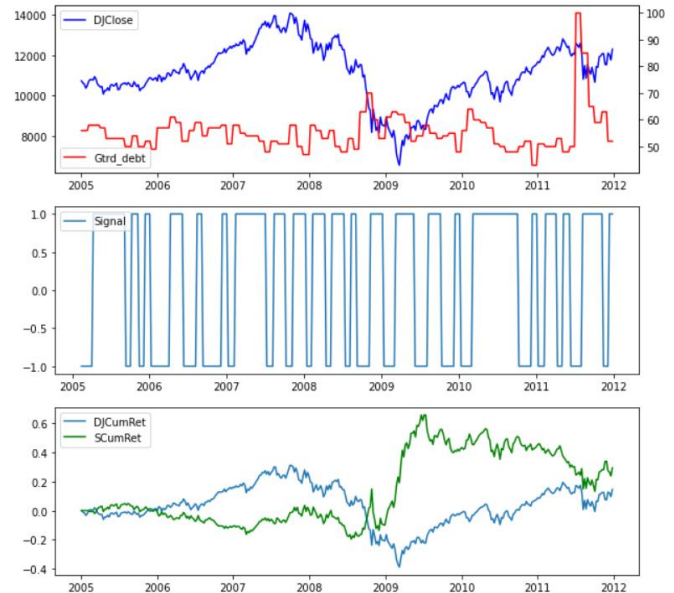### 7. Calculate weekly returns for buy and hold and the strategy.

```
StrategyDataframe['DJRet']=StrategyDataframe['Close'
]/StrategyDataframe['Close'].shift(1)-1
StrategyDataframe['SRet']=StrategyDataframe['Signal'
]*StrategyDataframe['DJRet']
```

8. Calculate cumulative returns

```
StrategyDataframe['DJCumRet']=(1+StrategyDataframe['
DJRet'] ).cumprod()
StrategyDataframe['SCumRet']=(1+StrategyDataframe['S
Ret'] ).cumprod()
StrategyDataframe['DJCumRet']=StrategyDataframe['DJC
umRet'].fillna(1)
```

```
StrategyDataframe['SCumRet']=StrategyDataframe['SCum
Ret'].fillna(1)
StrategyDataframe['DJCumRet']=StrategyDataframe['DJC
umRet']-1
StrategyDataframe['SCumRet']=StrategyDataframe['SCum
Ret']-1
```


By executing the same trading algorithm with the paper data, the cumulative return obtained in 2011-02-01 is 3, this figure is similar to the exercise performed in the paper by by Tobias Preis, Helen Susannah Moat and H. Eugene Stanley.

The chart below represents the execution in the paper
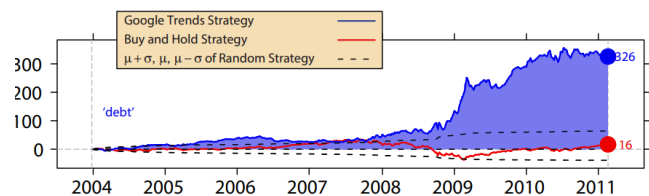
The execution of the paper strategy used the same code for the previous exercise but using the paper data provided. The result performed much better than our strategy from Google trends and Yahoo finance information, the difference may be caused due the additional transformation in Google trends figures by the paper authors.

Google trends strategy - Down Jones vs "debt" Paper data
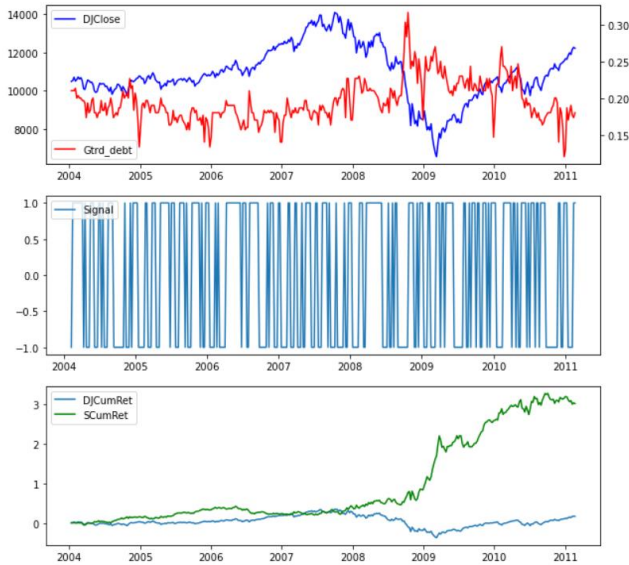
9. Repeating the strategy for S&P500 using the term "increase rates"

The period of time selected stars in 2020-01-01 and ends in 2022-10-01, the data is evaluated in a weekly strategy with same rules for signals, when there is an increase in search versus the average of 3 previous weeks go short if there is a decrease versus the average of 3 previous weeks go long.

Google trends strategy - S&P vs "increase rates"
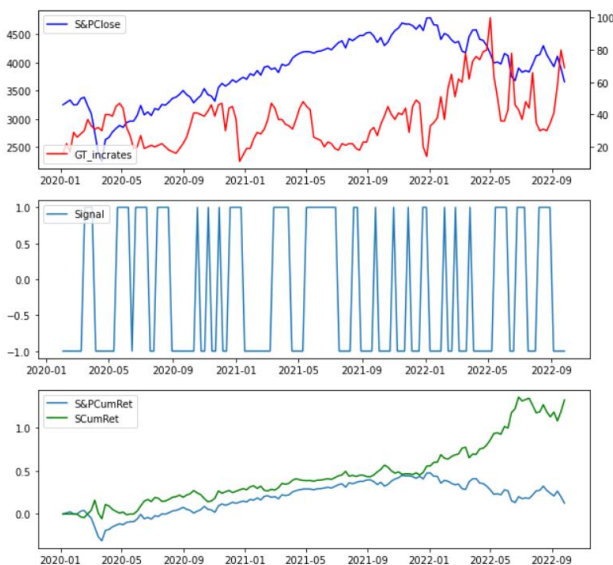
10. Conclusions

Using the Google trends data provided the trading strategy performed better than buy and hold right after the 2008 crisis.

The success for the paper strategy comes from the input data, they used the Google trends information but with some transformation that allow the model get a better perception of the market which translates in a higher return.

The trading strategy was used for S&P500 following the Google trend term "increase rates" did a great performance even though the period of time selected starts before the COVID pandemic. It very interesting to find that the strategy was always getting a hinger return.

This exercise is very interesting for understanding how the people trends are so related with the financial markets.

To implement this model in a real-life investment strategy is necessary to normalize the historical data with new entries. The Google trends figures are in a range of 0 to 100, that scale vary depending on the period of time consulted. Also it's very important to consider the impact of the transactions costs in the strategy performance.

References

- 2013 -Quantifying Trading Behavior in Financial Markets Using Google Trends by Tobias Preis, Helen Susannah Moat and H. Eugene Stanley.