Do the firm characteristics, financial performance and macroeconomic variables can predict future returns of stock exchange ?

Summary

Abstract

In this study, the firm characteristics, financial performance and macroeconomic variables were investigated to predict the future returns of several stock exchanges between the years 2017 to 2021.
Results indicated a significant relationship among the daily volume, the MACD and the stock prices.

Introduction

The purpose of this project was to analyse the relationship between firm characteristics, financial performance and macroeconomic variables (1-2-3) and future returns on several stock exchanges between the years 2017 to 2021.
Many researchers have analyzed the impact of firm characteristics (1), financial performance (2) and macroeconomic variables (3) on stock returns and found that many variables can predict stock returns like ‘Firm size’ or ‘Inflation’.

I tried to verify these affirmations on several stock exchanges with recent data.
My research question is «Do the firm characteristics, financial performance and macroeconomic variables can predict future returns of stock exchange ?»
I tried to answer the research question by analysing 23 stock exchanges, 6 variables in the period 2017-2021.

I analysed 1 **firm characteristic** :

- **Total assets** of the last 5 years, which estimates the **firm size**

I analysed 2 financial performance variables :

- **Cash flow** of the last 5 years

- **Daily volume** exchanged for the stock : which estimates the illiquidity

I also analyzed 3 **macroeconomic variables**:
- **Inflation**
- **Volatility** : represented by the VIX index
- **EPU** : Economic policy uncertainty

I also added 1 technical data **Moving Average Convergence Divergence (MACD) and EMA.**
MACD is a trend and momentum indicator that calculates the change and the speed of the evolution of the asset price. It was developed by Gerald Appeal in the late 1970s.
The **EMA** is the moving average that gives more weight to the most recent price points. This allows this type of moving average to react more strongly to recent price changes.

Based on prior empirical studies an economic theories, my hypothesis were :
H1 : Common stock returns is inversely related to firms size
H2 : Book to market ratio has a positive impact on stock returns
H3 : Price earnings ratio has a negative impact on stock returns
H4 : Stocks with low multiples P/B, P/C, P/D, P/E (value stocks) have higher returns than stocks with high multiples P/B, P/C, P/D, P/E (growth stocks).
H5 : Expected returns were positively correlated with illiquidity
H6 : Cash flow has an impact on stock returns.
H7 : Volatility indicators : VIX and EPU have an impact on stock returns

I tried to answer my research question in the following 5 chapters:
Chapter 1 'Literature review' : presents the historical research on this subject based on academic resources.
Chapter 2 'Methodology' : presents the hypothesis, the data, the variables and the methods.
Chapter 3 'Results' : describes my main findings.
Chapter 4 'Analysis / Discussion / interpretation' : explains the results and the limitations of the solution.
Chapter 5 'Conclusion' : summarizes the project and concludes.

(1) **Firm characteristics** = **firm size**, leverage, sales growth, asset growth and turnover, age of the firm, dividend pay-out, profitability, access to capital markets and growth opportunities.
(2) **Financial performance** = financial performance is usually measured using financial ratios, the categories of ratios include: **liquidity**, activity, profitability, debt or solvency, **cash flow**
(3) **Macroeconomic variables** = related to a country : **inflation**, gross domestic product, growth rate, **volatility**, ...

Literature Review

Many researchers have analyzed the impact of firm characteristics (1), financial performance (2) and macroeconomic variables (3) on stock returns.

Regarding the firm characteristics, the size effect has been well analyzed.
Gordon (2010) found that firms size has a negative impact on common stock returns.
Rutledge et al (2008) studied this effect on Chinese stock market and demonstrated that small firms have higher returns.
Drew et al (2003) indicated that small firms with high growth have higher returns than bigger companies.

Other studies demonstrated the same result : Davis & Desai (1998); Rouwenhorst (1998); Fama & French (1992); Banz (1981) and Ringanom (1981); ...

In the field of financial performance, the ratios have been well analyzed.
The book to market ratio and earnings to price ratio are well known to be able to predict stock returns.
Many studies indicated that high book to market ratio results to high stock returns : Hoang et al (2015); Drew and Veeraraghavan (2003); Lam (2002); Ashiq & Hwang (2002); Rouwenhorst (1998); Fama & French (1996); Maroney (1995); Rosenberg et al (1985).

Other ratios (multiple) have also a significant impact on stock returns : P/B (4), P/C (5), P/D (6), P/E (7).
In several studies the conclusion was that stocks with low multiples (value stocks) had higher returns than stocks with high multiples (growth stocks) :
Fama & French (1992a, 1993, 1996, 1998); Chan, Hamaoa & Lakonishok (1991); Reid & Lanstein (1985); Stattman (1980) and Rosenberg .

In the same field of financial performance, the illiquidity effect has also been demonstrated.
For many researchers the conclusion was that expected returns were positively correlated with illiquidity.
Amihud (2002); Alaraini and Stephens (1999); Brennan and Subrahmanyam (1996); Amihud and Mendelson (1986)

Fiinally the macroeconomic variables like inflation and volatility have also been highligted as good predictors for stock returns.
A famous volatility indicator VIX which is sometimes called the 'Fear factor' has been introduced by Brenner et al (1989).
Another index related to volatility is EPU (economic policy uncertainty) created by Baker et al (2012).
Several studies were conducted to analyze the relationship between VIX or EPU and stock returns.
On VIX : Giot (2002); Brenner et al (1989); Chen et al (1986);
On EPU : Gao et al. 2019; Brogaard and Detzel (2015); Antonakakis et al. (2013)

(1) **Firm characteristics** = **firm size**, leverage, sales growth, asset growth and turnover, age of the firm, dividend pay-out, profitability, access to capital markets and growth opportunities.
(2) **Financial performance** = financial performance is usually measured using financial ratios, the categories of ratios include: **liquidity**, activity, profitability, debt or solvency, **cash flow**
(3) **Macroeconomic variables** = related to a country : **inflation**, gross domestic product, growth rate, **volatility**, ...
(4) **P/B** = Price to book value per share
(5) **P/C** = Price-to-cash flow
(6) **P/D** = Price-to-dividend
(7) **P/E** = Price to earnings per share.

Methodology

What is the question/goal ?

Define a model which can predict a local stock market behavior and then define an efficient trading strategie.
It’s a supervised learning problem.
The performance should be higher than random prediction : more than 50%

«Do the firm characteristics, financial performance and macroeconomic variables can predict future returns of stock exchange ?»

Hypotheses of this research

Based on the concepts and findings of previous studies, the hypotheses of this research are as follows:
H1 : Common stock returns is inversely related to firms size
H2 : Book to market ratio has a positive impact on stock returns
H3 : Price earnings ratio has a negative impact on stock returns
H4 : Stocks with low multiples P/B, P/C, P/D, P/E (value stocks) have higher returns than stocks with high multiples P/B, P/C, P/D, P/E (growth stocks).
H5 : Expected returns were positively correlated with illiquidity
H6 : Cash flow has an impact on stock returns.
H7 : Volatility indicators : VIX and EPU have an impact on stock returns

Existing solution

I noted them for information, the aim is for me to train my new skills on ML not to do a benchmark !
Some products (**SentimentTrader, SentiTrade**) offer their services for a time-limited subscription fee, other websites (**MarketWatch, DataMinr**) allow free registration but their approach still remains a secret.
A little more transparent solution is the online sentiment analysis tool **Sentdex.com** maintained by Harrison Kinsley who also runs a YouTube channel providing tutorials on data analysis.
Although the Sentdex product has not been described by any academic paper, the tutorial videos and the product homepage provide basic information about the used data and processes.
The algorithm running behind is written using Python's Natural Language Toolkit and crawls input data from over 20 of the most famous American journals (Reuters, Bloomberg, WSJ, etc.) [Kin15].
[Kin15] Harrison Kinsley. Sentiment Analysis. 2015.

Methods

Overview

I'll go through the KDD process and follow the necessary steps until I find a useful and understandable pattern which makes sense to answer to the question.

Developing an understanding of the application domain

Done previously by reading and analysing documentation (about 80 hours). See references.

Data

Dataset

Our dataset consists of all available stocks listed on below Stock Exchanges in the period 2017-2021.

Market	MIC	Market place
Australian Securities Exchange	XASX	Sydney
Istanbul Stock Exchange	BIST	Istanbul
Brazil Stock Market	BS	São Paulo
Chile Stock Market	BVS	Santiago
European Stock Exchange	XAMS	Amsterdam
European Stock Exchange	XBRU	Brussels
European Stock Exchange	XMSM	Dublin
European Stock Exchange	XLIS	Lisbon

European Stock Exchange	XOSL	Oslo
European Stock Exchange	XPAR	Paris
Frankfurt Stock Exchange	FWB	Frankfurt
Warsaw Stock Exchange	WSE	Warsaw
Indonesia Stock Exchange	IDX	Jakarta
Johannesburg Stock Exchange	XJSE	Johannesburg
London Stock Exchange	XLON	London
Nasdaq Stock Market	XNAS	New York City
New-York Stock Exchange	XNYS	New York City
NYSE American	AMEX	New York City
Stockholm Stock Exchange	XSTO	Stockholm
Helsinki Stock Exchange	XHEL	Helsinki
Russian Trading System	RTS	Moscow
SIX Swiss Exchange	XSWX	Zurich
Shanghai Stock Exchange	XSHG	Shanghai
Shenzhen Stock Exchange	XSHE	Shenzhen
Tel-Aviv Stock Exchange	TASE	Tel-Aviv
Saudi Stock Exchange	XSAU	Riyadh
Toronto Stock Exchange	XTSE	Toronto

Data were obtained from Yahoo Finance for historical of price and firm characteristics, financial performance data.
The macroeconomic data 'Inflation' and 'Volatility' were obtained from the below Web site :
https://fred.stlouisfed.org/

Variables

What I want to determine at the end it's to know if I have to buy, sell or keep a stock.
So my dependent variable needs to be categorical.

The **dependent variable** is based on the close price. It's a **trend**. Therefore, I'll use classification methods.

I calculate the percentage of increase or decrease of the close price of the day versus the close price of the day – 1.
If this percentage is greater or equal than 0 then the **trend** is equal to 1.
Otherwise it's equal to -1.

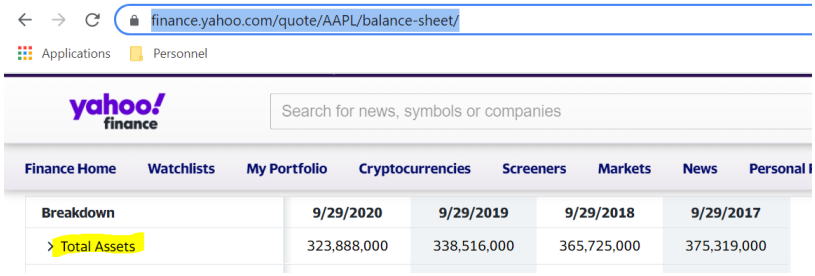
```
if ((Close price of the day - close price of the day-1)/close price of the day-1) >= 0
        trend = 1
elif
        trend = -1
```

Below are the **independent variables**:

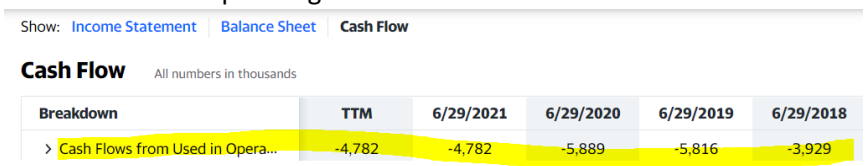
Firm characteristics data

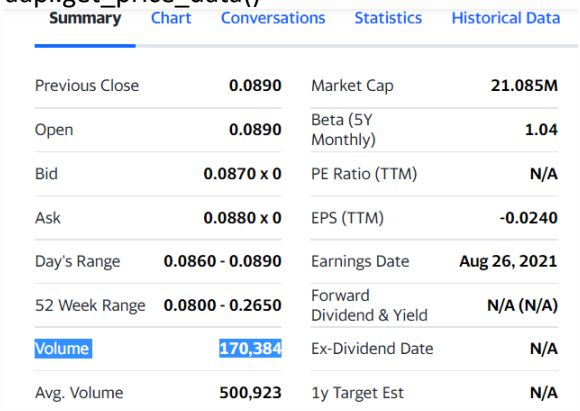
The data has been extracted from Yahoo Finance. Only date available on this website has been taken into account.

- **Firm size :** it can be estimated from the **total assets** of the firm.

- Total assets is available in Yahoo here :
- https://finance.yahoo.com/quote/AAPL/balance-sheet/
- 

Financial performance data

- **Cash flow** : a simple definition of a cash flow statement is how money, that is cash and cash equivalents, enters and exits a company
  - See Financials – Cash flow / In Yahoo Financials get field « totalCashFromOperatingActivities »
  - 

- **Illiquidity :** it can be estimated with the daily **volume** of the stock.
  - aapl.get_price_data()
  - 

- **Close price :** Can be check on Yahoo finance on 'Historical data'.

Macroeconomic variables data

- **Volatility :** how fast prices change, is often seen as a way to gauge market sentiment, and in particular the degree of fear among market participants.
  - VIX is the ticker symbol and the popular name for the Chicago Board Options Exchange's CBOE Volatility Index, it's the most famous volatility index
  - VIX, is a real-time market index representing the **market's expectations** for volatility **over the coming 30 days**.
  - There is a VIX for specific stocks market, for emerging markets, VIX based on gold, silver, ... : VIX CAC40, …
  - It's a **daily** data
  - The data were obtained from : https://fred.stlouisfed.org/

Date	VIX
02/01/2017	#N/A
03/01/2017	12.85
04/01/2017	11.85
05/01/2017	11.67
06/01/2017	11.32
09/01/2017	11.56

- **Inflation :** it's the loss of the purchasing power of the currency which results in a general and lasting increase in price
  - It's calculated **monthly** by **country**
  - The data were obtained from : https://fred.stlouisfed.org/

Date	Inflation
01/01/2017	-0.23845008
01/02/2017	0.11951001
01/03/2017	0.63662588
01/04/2017	0.08895918
01/05/2017	0.04937784

- **EPU** : is the Economic policy uncertainty
  - The data were obtained from : http://www.policyuncertainty.com/
  - It's calculated **monthly** by **country**

Date	EPU
01/01/2017	462.965607
01/02/2017	379.425964
01/03/2017	521.564148
01/04/2017	574.633179
01/05/2017	288.260437
01/06/2017	308.565308

Technical variable

Moving Average Convergence Divergence (MACD) and EMA.

MACD is a trend and momentum indicator that calculates the change and the speed of the evolution of the asset price. It was developed by Gerald Appeal in the late 1970s.

I calculated it from the price historic.

The **EMA** is the moving average that gives more weight to the most recent price points. This allows this type of moving average to react more strongly to recent price changes.

Results

These are the complete results :

Accuracy&profit_by_mkt.xlsx

This is an extract :

Market	NS='X'	Skewed	Skewness	Best Classifier&params	Accuracy best classifier	Important features (eli5)	Important features (feature_importances)	Second best Classifier&params	Accuracy second classifier	Important features (eli5)	Important features (feature_importances)
							0 **Volume** 0.535737				0 **Volume** 0.2
							5 **Calc_ema** 0.326845				7 **EPU_rate** 0.1
							19 month 0.038087				19 **month** 0.1
							20 day 0.035642				11 **VIX3_rate** 0.1
							14 VIX6_rate 0.025771				1 **Asset** 0.1
							3 Calc_macd 0.012336				10 **VIX2_rate** 0.1
							17 VIX9_rate 0.009240				5 **Calc_ema** 0.1
				DecisionTreeClassifier(max_dep			1 Asset 0.007701			0.1000 ± 0.0000 **Calc_ema**	3 **Calc_macd** 0.1
GPW	No	NA	NA	th=5, min_samples_leaf=5)	**0.90%**	**None**	2 Cashflow 0.004258	AdaBoostClassifier(n_es		0.0600 ± 0.0980 **Asset**	20 **day** 0.1
							16 VIX8_rate 0.002898	timators=10)	**0.90%**	0.0400 ± 0.0980 **Calc_macd**	
							0 **Volume** 0.155951				
							5 **Calc_ema** 0.141293				
							3 **Calc_macd** 0.111598				
						0.1000 ± 0.1789 **VIX2_rate**	4 **Calc_signal** 0.108821				
						0.0800 ± 0.0800 **Asset**	14 VIX6_rate 0.036056				
						0.0600 ± 0.0980 **Calc_macd**	10 VIX2_rate 0.035603				
						0.0400 ± 0.0980 **Volume**	11 VIX3_rate 0.035287			0.1000 ± 0.1265 **VIX1_rate**	
						0.0200 ± 0.0800 VIX8_rate	16 VIX8_rate 0.034884			0.1000 ± 0.0000 **VIX8_rate**	
				RandomForestClassifier(min_sa		0.0200 ± 0.0800 VIX6_rate	8 VIX0_rate 0.034313	LogisticRegression(C=0.1		0.0600 ± 0.0980 **Calc_signal**	
TASE	No	NA	NA	mples_leaf=5, n_estimators=20)	**0.90%**	0.0200 ± 0.0800 Calc_signal	17 VIX9_rate 0.034211	, penalty='none')	**0.90%**	0.0200 ± 0.0800 VIX6_rate	NA
						0.2571 ± 0.2138 **VIX5_rate**					
						0.2286 ± 0.1400 **Calc_macd**					
						0.1714 ± 0.2138 **VIX9_rate**					
						0.1714 ± 0.1143 **VIX0_rate**				0.1714 ± 0.1143 **VIX0_rate**	
						0.1429 ± 0.0000 **VIX1_rate**	0 **Volume** 0.171946			0.1143 ± 0.1143 **VIX2_rate**	
						0.1429 ± 0.0000 **VIX3_rate**	5 **Calc_ema** 0.136010			0.1143 ± 0.1143 **VIX8_rate**	
						0.1143 ± 0.2138 **Volume**	4 **Calc_signal** 0.109578			0.1143 ± 0.1143 **VIX6_rate**	
						0.1143 ± 0.1143 **VIX7_rate**	3 **Calc_macd** 0.108251			0.1143 ± 0.1143 **VIX5_rate**	
						0.0857 ± 0.2286 **Calc_ema**	14 VIX6_rate 0.036696			0.1143 ± 0.1143 **VIX3_rate**	
						0.0857 ± 0.2286 Cashflow	8 VIX0_rate 0.035390			0.0857 ± 0.1400 day	
						0.0857 ± 0.1400 VIX2_rate	11 VIX3_rate 0.035194			0.0857 ± 0.1400 VIX9_rate	
						0.0857 ± 0.1400 VIX8_rate	17 VIX9_rate 0.035037			0.0571 ± 0.1400 VIX7_rate	
				RandomForestClassifier(min_sa		0.0857 ± 0.1400 Asset	15 VIX7_rate 0.034693			0.0286 ± 0.1143 Volume	
SIX	No	NA	NA	mples_leaf=5, n_estimators=20)	**0.86%**	0.0571 ± 0.1400 day	13 VIX5_rate 0.034147	LogisticRegression(C=1)	**0.86%**	0.0286 ± 0.1143 VIX4_rate	NA

We can see the accuracy of the 2 best classifiers by market, with details on which classifier and the important variables.

Analysis / Discussion / interpretation

Analysis

I tested 23 markets.
We can disctinct 3 groups :

- 9 markets over 23 had an accuracy superior to 0.78%
- 7 markets over 23 had an accuracy equal to 0.70%
- 7 markets over 23 had an accuracy between 0.50% and 0.60%

For the first group with the highest accuracy, classifiers appear as the best classifier on several markets:

- RF 4 times
- DT 3 times
- LR 2 times

For the first group with the highest accuracy, classifiers appear as the second best classifier on several markets:

- Adaboost 5 times
- LR 4 times
- DT 2 times
- Bernouilli 2 times

The important features for RF are in this order :

- Volume 4 times
- Calc_macd 3 times
- Calc_ema 3 times

The important features for DT are in this order :

- Calc_ema 4 times
- Volume 3 times

The important features for Adaboost are in this order :

- Calc_ema 3 times

For the second group with the accuracy equal to 0.70%, classifiers appear as the best classifier on several markets:

- LR 4 times

For the second group with the accuracy equal to 0.70%, classifiers appear as the second best classifier on several markets:

- Bernouilli 5 times
- Adaboost 2 times
- RF 2 times

The important features for RF are in this order :

- Calc_ema 1 time

The important features for DT are in this order :

- Calc_ema 2 times
- Volume 2 times

The important features for Adaboost are in this order :

- Calc_ema 1 time
- Volume 1 time

Based on these results, we can say that the best classifiers are LR, DT, Adaboost and RF.
The important features for all these classifiers are calc_ema, volume, calc_macd.

Thes results confirm the fith hypothesis «Expected returns were positively correlated with illiquidity».

Unfortunately I was not able to confirm the others hypotheses :
H1 : Common stock returns is inversely related to firms size
H2 : Book to market ratio has a positive impact on stock returns
H3 : Price earnings ratio has a negative impact on stock returns
H4 : Stocks with low multiples P/B, P/C, P/D, P/E (value stocks) have higher returns than stocks with high multiples P/B, P/C, P/D, P/E (growth stocks).
H6 : Cash flow has an impact on stock returns.
H7 : Volatility indicators : VIX and EPU have an impact on stock returns

I can partially reply to the research question :
«Do the firm characteristics, financial performance and macroeconomic variables can predict future returns of stock exchange ?»
Financial performance data (Volume) and technical indicators (calc_ema, calc_macd) based on close price, have a significant impact on the future prices prediction.

Discussion

To validate these results it could be valuable to analyse the same markets with **more recent data**.

Also to extend the predictions to others fields of research already highlighted by many academic resources it could be good to investigate the impact on the below subjects on the future stock price:

- **Calendar**
- **Twitter**
- **Internet search**
- **Google trend**

Acting on the discovered knowledge

If we want to use the discovered knowledge we have to adapt the program to extract day to day data and to focus only on variables which have an explanation power.

Conclusion

The main objective of this study was to define a model which can predict a local stock market behavior and then define an efficient trading strategie.
The performance should be higher than random prediction : more than 50%

The model trained has an accuracy superior to 50% on all of the 23 markets and 9 of them have an accuracy superior to 78%.
The goal is achieved.

References

Firm characteristics and financial performance data

- An investigation of pricing multiples' ability to predict abnormal returns on the Oslo Stock Exchange
- Defining and Designing a Model to Predict the Performance of Mutual Funds by Using Macroeconomic Variables in Tehran Stock
- Does the fear gauge predict downside risk more accurately than econometric models Evidence from the US stock market
- Stocks as Lotteries Can Extreme Positive Returns Predict Future Returns
    - firm size, cash flows, illiquidity
- The Ability Of Earnings, Cash Flow To Predict Future Earnings, Cash Flow And Stock Price

Macroeconomic variables data

- Can uncertainty predict stock market returns
    - volatility, inflation
- Do Based-Market Data Predict Stock Return Better Than Accounting Data The Case of Tehran

Appendix

Glossary

Return is the percentage change in the asset value. I used the ~~adjusted~~ close prices to calculate returns.
~~**Adjusted close price** is the official closing price adjusted for capital actions and dividends. Nevertheless returns provide useful information about the probability distribution of asset prices.~~

~~This is essential for investors and portfolio managers as they use this information to value assets and manage their risk exposure.~~
~~The one of models that is used to explain the stock returns is the Capital Asset Pricing Model (CAPM). Capital Asset Pricing Model definite the returns of stock as function of the systematic risk of a stock.~~

Cash flow : a simple definition of a cash flow statement is how money, that is cash and cash equivalents, enters and exits a company.

Volatility is a statistical measure of the dispersion of returns for a given security or market index.
In most cases, the higher the volatility, the riskier the security.
Volatility is often measured as either the standard deviation or variance between returns from that same security or market index.
In the securities markets, volatility is often associated with big swings in either direction.
For example, when the stock market rises and falls more than one percent over a sustained period of time, it is called a "volatile" market. An asset's volatility is a key factor when pricing options contracts.

Trading volume is the number of shares traded in each day during a trading session. Volume can be used to measure stock liquidity, which in turn has been shown to be useful in asset pricing as several theoretical and empirical studies have identified a liquidity premium.
Liquidity can help to explain the cross-section of expected returns.

Accounting data (Firm Size, Return on Equity, Return on Assets, profit margin ratio, Financial Leverage ratio).

Based-market data (Price to Earnings ratio, book to market ratio and Dividend yield).

Size : The natural log of total asset (actifs) at the end of the year

Book to Market value (BM) : Book value of stock over stock price at the end of the year.
Valeur comptable du stock par rapport au prix du stock à la fin de l'année

Book-to-market ratio : The book-to-market ratio compares a company's book value to its market value. The book value is the value of assets minus the value of the liabilities. The market value of a company is the market price of one of its shares multiplied by the number of shares outstanding.

Return on Asset (ROA) : Earning after tax over total asset at the end of the year

Return on Equity (ROE) : Earning after tax over equity at the end of the year.
The **coefficient of return on equity** reflects how many turns it takes to pay bills for the date of analysis.

Margin Profit (MP) : Earning after tax over Sales.

Price to earnings ratio (PE) : Stock price over earnings per share.

Financial Leverage (LEV) : Total debt over total asset.
The **financial leverage** is the amount of debt a firm uses to finance assets.

Dividend Yield (DY) : Dividend per share over stock price.

Abnormal Return: A term used to describe the returns generated by a given security or portfolio over a period of time that is different from the expected rate of return. The expected rate of return is the estimated return based on an asset pricing model(…).

Earnings for price ratio (EP) : A measure indicating the rate at which investors will capitalize a firm's expected earnings in the coming period. This ratio is calculated by dividing the projected earnings per share by the current market price of the stock. A relatively low E/P ratio anticipates higher-than-average growth in earnings. Earnings-price ratio is the inverse of the price-earnings ratio. Also called earnings capitalization rate, earnings yield.

The **idiosyncratic volatility** is the difference between total risk and the systematic risk of a stock, I define idiosyncratic volatility as the standard deviation of the regression residual of the Fama and French three-factor model.

The **Short-Term Reversal** Effect : The short-term reversal anomaly, the phenomenon that stocks with relatively low returns over the past month or week earn positive abnormal returns in the following month or week, and stocks with high returns earn negative abnormal returns, is well-researched.

Illiquidity : illiquidity refers to assets that cannot be easily exchanged for money.
This may be due to the fact that there are not enough investors willing to buy them.

MACD Moving Average Convergence Divergence : MACD is a trend and momentum indicator that calculates the change and the speed of the evolution of the asset price. It was developed by Gerald Appeal in the late 1970s.

Program documentation

Data collection

Data collection is the ultimate first step, it is not part of the KDD process itself.

First I get a list of stock markets :

LIST-OF-APPROVED
-REGULATED-STOCK

Here are the markets I retained :

Market	MIC	Market place
Australian Securities Exchange	XASX	Sydney
Istanbul Stock Exchange	BIST	Istanbul
Brazil Stock Market	BS	São Paulo
Chile Stock Market	BVS	Santiago
European Stock Exchange	XAMS	Amsterdam
European Stock Exchange	XBRU	Brussels
European Stock Exchange	XMSM	Dublin
European Stock Exchange	XLIS	Lisbon
European Stock Exchange	XOSL	Oslo
European Stock Exchange	XPAR	Paris
Frankfurt Stock Exchange	FWB	Frankfurt

Warsaw Stock Exchange	WSE	Warsaw
Indonesia Stock Exchange	IDX	Jakarta
Johannesburg Stock Exchange	XJSE	Johannesburg
London Stock Exchange	XLON	London
Nasdaq Stock Market	XNAS	New York City
New-York Stock Exchange	XNYS	New York City
NYSE American	AMEX	New York City
Stockholm Stock Exchange	XSTO	Stockholm
Helsinki Stock Exchange	XHEL	Helsinki
Russian Trading System	RTS	Moscow
SIX Swiss Exchange	XSWX	Zurich
Shanghai Stock Exchange	XSHG	Shanghai
Shenzhen Stock Exchange	XSHE	Shenzhen
Tel-Aviv Stock Exchange	TASE	Tel-Aviv
Saudi Stock Exchange	XSAU	Riyadh
Toronto Stock Exchange	XTSE	Toronto

Second I found a list of tickers for each market by using different website listed in the below file.

data_stock_exchanges.xlsx

Here is an example of the ticker file by market :

Market	Company	Ticker	Industry	Category1	Category2	Category3
CMF	EMPRESAS COPEC S.A.	COPEC.SN	Energy	Global Equity	Common stocks	Emerging markets
CMF	BANCO DE CHILE	CHILE.SN	Financials	Global Equity	Common stocks	Emerging markets
CMF	BANCO SANTANDER-CHILE	BSANTANDER.SN	Financials	Global Equity	Common stocks	Emerging markets
CMF	BANCO DE CREDITO E INVERSIONES	BCI.SN	Financials	Global Equity	Common stocks	Emerging markets
CMF	EMPRESAS CMPC S.A.	CMPC.SN	Materials	Global Equity	Common stocks	Emerging markets
CMF	CENCOSUD S.A.	CENCOSUD.SN	Consumer Staples	Global Equity	Common stocks	Emerging markets
CMF	COMPANIA SUD AMERICANA DE VAPORES	VAPORES.SN	Industrials	Global Equity	Common stocks	Emerging markets
CMF	Cia Cervecerias Unidas SA	CCU.SN	Consumer Staples	Global Equity	Common stocks	Emerging markets
CMF	COLBUN S.A.	COLBUN.SN	Utilities	Global Equity	Common stocks	Emerging markets
CMF	CAP S.A.	CAP.SN	Materials	Global Equity	Common stocks	Emerging markets

I get macroeconomic variables data from the below sources :

- **Volatility :** how fast prices change, is often seen as a way to gauge market sentiment, and in particular the degree of fear among market participants.
    - VIX is the ticker symbol and the popular name for the Chicago Board Options Exchange's CBOE Volatility Index, it's the most famous volatility index
    - VIX, is a real-time market index representing the **market's expectations** for volatility **over the coming 30 days**.
    - There is a VIX for specific stocks market, for emerging markets, VIX based on gold, silver, ... : VIX CAC40, …
    - It's a **daily** data
    - The data were obtained from : https://fred.stlouisfed.org/

Date	VIX
02/01/2017	#N/A
03/01/2017	12.85
04/01/2017	11.85
05/01/2017	11.67
06/01/2017	11.32
09/01/2017	11.56

- **Inflation :** it's the loss of the purchasing power of the currency which results in a general and lasting increase in price
  - It's calculated **monthly** by **country**
  - The data were obtained from : https://fred.stlouisfed.org/

Date	Inflation
01/01/2017	-0.23845008
01/02/2017	0.11951001
01/03/2017	0.63662588
01/04/2017	0.08895918
01/05/2017	0.04937784

- **EPU** : is the Economic policy uncertainty
  - The data were obtained from : http://www.policyuncertainty.com/
  - It's calculated **monthly** by **country**

Date	EPU
01/01/2017	462.965607
01/02/2017	379.425964
01/03/2017	521.564148
01/04/2017	574.633179
01/05/2017	288.260437
01/06/2017	308.565308

I specified for each market which inflation file and EPU file has to be used :

Market	Suffix	Inflation_file	File	EPU
ASX	No	Growth Rate Previous Period Quaterly Australia.csv	ASX.csv	EPU_index_Australia.csv
BIST	No	Growth Rate Previous Period Monthly Turkey.csv	BIST.csv	EPU_index_Global.csv
BS	No	Growth Rate Previous Period Monthly Brazil.csv	BS.csv	EPU_index_Brazil.csv
CMF	No	Growth Rate Previous Period Monthly Chile.csv	CMF.csv	EPU_index_Chile.csv
EURONEXT	OL	Growth Rate Previous Period Monthly Norway.csv	EURONEXT.csv	EPU_index_Global.csv
EURONEXT	PA	Growth Rate Previous Period Monthly France.csv	EURONEXT.csv	EPU_index_France.csv
EURONEXT	LS	Growth Rate Previous Period Monthly Portugal.csv	EURONEXT.csv	EPU_index_Global.csv
EURONEXT	BR	Growth Rate Previous Period Monthly Belgium.csv	EURONEXT.csv	EPU_index_Belgium.csv
EURONEXT	IR	Growth Rate Previous Period Monthly Ireland.csv	EURONEXT.csv	EPU_index_Ireland.csv
EURONEXT	AS	Growth Rate Previous Period Monthly Netherlands.csv	EURONEXT.csv	EPU_index_Netherlands.csv
EURONEXT	No	Growth Rate Previous Period Monthly Belgium.csv	EURONEXT.csv	EPU_index_Belgium.csv
FWB	No	Growth Rate Previous Period Monthly Germany.csv	FWB.csv	EPU_index_Germany.csv
GPW	No	Growth Rate Previous Period Monthly Poland.csv	GPW.csv	EPU_index_Global.csv
IDX	No	Growth Rate Previous Period Monthly Indonesia.csv	IDX.csv	EPU_index_Global.csv
JSE	No	Growth Rate Previous Period Monthly South_Africa.csv	JSE.csv	EPU_index_Global.csv
LSE	No	Growth Rate Previous Period Monthly UK.csv	LSE.csv	EPU_index_UK.csv
NASDAQ	No	Growth Rate Previous Period Monthly US.csv	NASDAQ.csv	EPU_index_US.csv
NYSE	No	Growth Rate Previous Period Monthly US.csv	NYSE.csv	EPU_index_US.csv
NYSE_MKT	No	Growth Rate Previous Period Monthly US.csv	NYSE_MKT.csv	EPU_index_US.csv
OMXH	No	Growth Rate Previous Period Monthly Finland.csv	OMXH.csv	EPU_index_Global.csv
OMXS	No	Growth Rate Previous Period Monthly Sweden.csv	OMXS.csv	EPU_index_Sweden.csv
RTS	No	Growth Rate Previous Period Monthly Russia.csv	RTS.csv	EPU_index_Russia.csv
SIX	No	Growth Rate Previous Period Monthly Switzerland.csv	SIX.csv	EPU_index_Global.csv
SSE	No	Growth Rate Previous Period Monthly China.csv	SSE.csv	EPU_index_China.csv
SZSE	No	Growth Rate Previous Period Monthly China.csv	SZSE.csv	EPU_index_China.csv
TASE	No	Growth Rate Previous Period Monthly Israel.csv	TASE.csv	EPU_index_Global.csv
TASI	No	Growth Rate Previous Period Monthly Saudi_Arabia.csv	TASI.csv	EPU_index_Global.csv
TSX	No	Growth Rate Previous Period Monthly Canada.csv	TSX.csv	EPU_index_Canada.csv

Same for VIX files, I specified several VIX file per market :

A	F	G	H	I	J	K
Market	**VIX0**	**VIX1**	**VIX2**	**VIX3**	**VIX4**	**VIX5**
ASX	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
BIST	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
BS	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
CMF	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
EURONEXT	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
EURONEXT	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
EURONEXT	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
EURONEXT	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
EURONEXT	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
EURONEXT	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
EURONEXT	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
FWB	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
GPW	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
IDX	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
JSE	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
LSE	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
NASDAQ	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
NYSE	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
NYSE_MKT	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
OMXH	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
OMXS	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
RTS	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
SIX	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
SSE	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
SZSE	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
TASE	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
TASI	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv
TSX	CBOE Volatility Index VIX.csv	CBOE Equity VIX on Amazon.csv	CBOE Equity VIX on Apple.csv	CBOE Equity VIX on IBM.csv	CBOE Equity VIX on Google.csv	CBOE Energy Sector ETF Volatility Index.csv

Market	VIX6	VIX7	VIX8	VIX9	VIX10
ASX	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	SP ASX 200 VIX Index.csv
BIST	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE Emerging Markets ETF Volatility Index.csv
BS	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE Emerging Markets ETF Volatility Index.csv
CMF	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE Emerging Markets ETF Volatility Index.csv
EURONEXT	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE EuroCurrency ETF Volatility Index.csv
EURONEXT	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE EuroCurrency ETF Volatility Index.csv
EURONEXT	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE EuroCurrency ETF Volatility Index.csv
EURONEXT	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE EuroCurrency ETF Volatility Index.csv
EURONEXT	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE EuroCurrency ETF Volatility Index.csv
EURONEXT	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE EuroCurrency ETF Volatility Index.csv
EURONEXT	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE EuroCurrency ETF Volatility Index.csv
FWB	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE EuroCurrency ETF Volatility Index.csv
GPW	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE Emerging Markets ETF Volatility Index.csv
IDX	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE Emerging Markets ETF Volatility Index.csv
JSE	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE Emerging Markets ETF Volatility Index.csv
LSE	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	
NASDAQ	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE NASDAQ 100 Volatility Index.csv
NYSE	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE NASDAQ 100 Volatility Index.csv
NYSE_MKT	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE NASDAQ 100 Volatility Index.csv
OMXH	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	
OMXS	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	
RTS	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE Emerging Markets ETF Volatility Index.csv
SIX	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	
SSE	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE Emerging Markets ETF Volatility Index.csv
SZSE	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE Emerging Markets ETF Volatility Index.csv
TASE	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	
TASI	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	CBOE Emerging Markets ETF Volatility Index.csv
TSX	CBOE Crude Oil ETF Volatility Index.csv	CBOE Gold ETF Volatility Index.csv	CBOE Gold Miners ETF Volatility Index.csv	CBOE Silver ETF Volatility Index.csv	

Market	VIX11	VIX12	VIX13
ASX			
BIST			
BS	CBOE Brazil ETF Volatility Index.csv		
CMF			
EURONEXT			
EURONEXT			
EURONEXT			
EURONEXT			
EURONEXT			
EURONEXT			
EURONEXT			
FWB			
GPW			
IDX			
JSE			
LSE			
NASDAQ	CBOE DJIA Volatility Index.csv	CBOE Russell 2000 Volatility Index.csv	CBOE S&P 500 3-Month Volatility Index.csv
NYSE	CBOE DJIA Volatility Index.csv	CBOE Russell 2000 Volatility Index.csv	CBOE S&P 500 3-Month Volatility Index.csv
NYSE_MKT	CBOE DJIA Volatility Index.csv	CBOE Russell 2000 Volatility Index.csv	CBOE S&P 500 3-Month Volatility Index.csv
OMXH			
OMXS			
RTS			
SIX			
SSE	CBOE China ETF Volatility Index.csv		
SZSE	CBOE China ETF Volatility Index.csv		
TASE			
TASI			
TSX			

I used python to develop extractors to get financial performance data related to each tickers (total asset, cash flows, historical of prices, …) from Yahoo Finance using YahooFinancials library.
I was not able to use below data for a lack of history in Yahoo Finance :

- Book-to-Market Ratio (priceToBook)
- Price to earnings ratio
- Book-to-Market Ratio
- Price to earnings ratio
- Earnings-to-price ratio (E/P ratio)

Below is the code used to extract data :
Get_stock_data.py extracts data related to the stock (assets and cash flows of the last 4 years).

get_stock_data.py.txt

Get_historical_stock_data.py extracts the historic of price and volume for each stock.

get_historical_stock_data.py.txt

Python programs were launched from **scripts** located on a Linux server and scheduled through **crontab**.

get_stock_data_CMF.sh.txt

get_historical_stock_data_CMF.sh.txt

crontab.txt

I duplicated scripts in order to parallelize the load; each script loaded a specific market.
But it was generating some error due to heavy workload on the server.
So it's better to handle the load sequentially; 1 script for all markets.
I had also some errors due to http request which is not responding. I deal with this error by grepping the error in the python program.

These are the data get from Yahoo Finance :

Stock data :

Market	Company	Ticker	Current_price	DateAssetN	AssetN	DateAssetN-1	AssetN-1	DateAssetN-2	AssetN-2	DateAssetN-3	AssetN-3	DateAssetN-4	AssetN-4
CMF	Aes Andes Sa	AESANDES.SN	82.77	31/12/2021	4476782000	31/12/2020	8120006000	31/12/2019	8442560000	31/12/2018	7869361000	31/12/2017	8159807000
CMF	AGUAS ANDINAS SA	AGUAS-A.SN	154	31/12/2021	2.22922E+12	31/12/2020	2.14444E+12	31/12/2019	2.00144E+12	31/12/2018	1.90605E+12	31/12/2017	1.79688E+12
CMF	Almendral Sa Common Stock Clp 0	ALMENDRAL.SN	25.8	31/12/2021	5.66832E+12	31/12/2020	5.15397E+12	31/12/2019	5.43118E+12	31/12/2018	4.22975E+12	31/12/2017	3.79698E+12
CMF	Embotelladora Andina S.A.	ANDINA-A.SN	1460	31/12/2021	2.94611E+12	31/12/2020	2.44806E+12	31/12/2019	2.39095E+12	31/12/2018	2.2145E+12	31/12/2017	2.11486E+12
CMF	EMBOTELLADORA ANDINA SERIES B PREF	ANDINA-B.SN	1640.7	31/12/2021	2.94611E+12	31/12/2020	2.44806E+12	31/12/2019	2.39095E+12	31/12/2018	2.2145E+12	31/12/2017	2.11486E+12
CMF	ANTARCHILE SA	ANTARCHILE.SN	6450	31/12/2021	25867334000	31/12/2020	25556906000	31/12/2019	25154173000	31/12/2018	24026380000	31/12/2017	22727714000
CMF	BANCO DE CREDITO E INVERSIONES	BCI.SN	25600	31/12/2021	6.91586E+13	31/12/2020	5.71563E+13	31/12/2019	5.03366E+13	31/12/2018	4.13497E+13	31/12/2017	3.38834E+13
CMF	BESALCO S.A.	BESALCO.SN	190.3	31/12/2021	9.25441E+11	31/12/2020	7.87861E+11	31/12/2019	7.7662E+11	31/12/2018	6.98389E+11	31/12/2017	6.20159E+11
CMF	Banco Santander Chile	BSANTANDER.SN	35.98	31/12/2021	6.3842E+13	31/12/2020	5.57761E+13	31/12/2019	5.05782E+13	31/12/2018	3.91974E+13	31/12/2017	3.58236E+13
CMF	BANCO SANTANDER-CHILE	BSANTANDER.SN	35.97	31/12/2021	6.3842E+13	31/12/2020	5.57761E+13	31/12/2019	5.05782E+13	31/12/2018	3.91974E+13	31/12/2017	3.58236E+13
CMF	CAP S.A.	CAP.SN	8245	31/12/2021	6612342000	31/12/2020	5866188000	31/12/2019	5478735000	31/12/2018	5341485000	31/12/2017	5550301000
CMF	Cia Cervecerias Unidas SA	CCU.SN	6718	31/12/2021	2.84675E+12	31/12/2020	2.52534E+12	31/12/2019	2.35369E+12	31/12/2018	2.40586E+12	31/12/2017	1.97623E+12

Market	DatecashflowN	CashflowN	DatecashflowN1	CashflowN1	DatecashflowN2	CashflowN2	DatecashflowN3	CashflowN3	DatecashflowN4	CashflowN4	Pricetobook	Pricetoearning	EarningtopriceVolume	Earningtoprice	Volume
CMF	31/12/2021	323281000	31/12/2020	1247697000	31/12/2019	739118000	31/12/2018	673284000	31/12/2017	340657000		No division by 0		No division by 0	2648837
CMF	31/12/2021	2.31199E+11	31/12/2020	1.85293E+11	31/12/2019	2.20759E+11	31/12/2018	2.45501E+11	31/12/2017	2.13469E+11	1.0708574	10.36757776		0.096454545	853798
CMF	31/12/2021	5.72145E+11	31/12/2020	5.60137E+11	31/12/2019	6.15513E+11	31/12/2018	4.72473E+11	31/12/2017	5.37058E+11	0.47619924	7.239057239		0.138139535	675337
CMF	31/12/2021	3.05055E+11	31/12/2020	2.78769E+11	31/12/2019	2.55148E+11	31/12/2018	2.35279E+11	31/12/2017	2.4796E+11	1.5916328	11.90058932		0.084029452	8015
CMF	31/12/2021	3.05055E+11	31/12/2020	2.78769E+11	31/12/2019	2.55148E+11	31/12/2018	2.35279E+11	31/12/2017	2.4796E+11	1.7885699	13.37349103		0.074774791	619186
CMF	31/12/2021	1853499000	31/12/2020	1902786000	31/12/2019	956504000	31/12/2018	1734337000	31/12/2017	1602319000	436.54825	10932.20339		9.15E-05	15602
CMF	31/12/2021	1.00359E+12	31/12/2020	4.23943E+12	31/12/2019	95774000000	31/12/2018	-1.48785E+12	31/12/2017	-1.25043E+12	0.95441574	10.27455126		0.097327852	135247
CMF	31/12/2021	57325371000	31/12/2020	37476901000	31/12/2019	16490614000	31/12/2018	-62068012000	31/12/2017	13697941000	0.55047727	6.489343564		0.154098791	34693
CMF	31/12/2021	-3.17943E+12	31/12/2020	-1.46672E+12	31/12/2019	4.39019E+11	31/12/2018	-6.52858E+11	31/12/2017	6.93439E+11	1.9021939	12.2131704		0.081878822	42343821
CMF	31/12/2021	-3.17943E+12	31/12/2020	-1.46672E+12	31/12/2019	4.39019E+11	31/12/2018	-6.52858E+11	31/12/2017	6.93439E+11	1.9016653	12.20977597		0.081901585	39093695
CMF	31/12/2021	1415970000	31/12/2020	878882000	31/12/2019	92011000	31/12/2018	334373000	31/12/2017	494057000	570.3514	2432.871053		0.000411037	240597

Historical prices :

Market	Company	Ticker	Current_price	Date	Formatted_date	Open	Close	Adjclose	low	high	Volume	Return_calculated
CMF	Aes Andes Sa	AESANDES.SN	129.51	946902600	03/01/2000	120.814621	119.323082	96.13249969	119.323082	120.814621	4764576	-1.491539001
CMF	Aes Andes Sa	AESANDES.SN	129.51	946989000	04/01/2000	119.323082	116.8371811	94.12974548	116.340004	119.323082	699043	-2.485900879
CMF	Aes Andes Sa	AESANDES.SN	129.51	947075400	05/01/2000	117.3343582	117.3343582	94.5302887	115.3456421	117.3343582	2365544	0
CMF	Aes Andes Sa	AESANDES.SN	129.51	947161800	06/01/2000	117.3343582	115.3456421	92.9280777	115.3456421	117.3343582	574138	-1.988716125
CMF	Aes Andes Sa	AESANDES.SN	129.51	947248200	07/01/2000	117.3343582	116.340004	93.72918701	115.3456421	117.3343582	3893764	-0.994354248
CMF	Aes Andes Sa	AESANDES.SN	129.51	947507400	10/01/2000	117.3343582	115.0970535	92.72780609	114.3612289	117.3343582	2443584	-2.237304688
CMF	Aes Andes Sa	AESANDES.SN	129.51	947593800	11/01/2000	117.3343582	113.8541031	91.72642517	113.8541031	117.3343582	1810110	-3.480255127
CMF	Aes Andes Sa	AESANDES.SN	129.51	947680200	12/01/2000	114.3512878	113.8541031	91.72642517	113.8541031	114.3512878	2971361	-0.497184753

Preparation and prediction program

I created a program to prepare the data and predict the dependent variable.
Python program : **Stock_market_predictions.ipynb**
Mettre en objet la dernière version du pgm

Main program

Overview

The main program executes the below macro steps which will be detailed in the next parts.

The macro steps belong to the KDD process.

1. Firstly, I'll create a target data set: selecting a data set, or focusing on a subset of variables or data samples, on which discovery is to be performed.
2. Secondly, I'll clean and preprocess the data.
3. Thirdly, I'll try to find useful features to represent the data depending on the goal of the task
    a. This is data reduction and projection.
4. Fourthly, I'll explore the data trying to find data patterns with different data-mining methods.

Data have been collected :

- Financial performance data : historical of price data for many tickers of different markets.
- Macroeconomic variables data : Volatility and Inflation data. For volatility I get EPU, (Volatility by country) and VIX indexes
- Technical data : I'll calculate MACD and EMA from historical of price data

Macro steps

I loop on markets.csv file which list all markets and files related to each market.

1. **Preparation** step is then run, see '**Data preparation'** for details.
2. **Clean and preprocess** step is run, see '**Data cleaning and preprocessing**' for details.
3. **Data reduction and projection** is then run, see '**Data reduction and projection'** for details.
4. **Data Mining,** I'll explore the data trying to find data patterns with different data-mining methods, see '**Data Mining'** for details.

* Variable V for Verbose will allow different level of details to be displayed. Level 1 is the more synthetic as level 3 is the more detailed.

Data preparation

Task	**Function(arguments)**
First part : "Creating a target data set"	
Drop unuseful columns	
Delete not useful columns : **Dataset_1** : Market, Company, Current_price, Pricetobook, Pricetoearning, Earningtoprice, Volume **Dataset_2** : Market, Company, Current_price, Formatted_date, Open, Adjclose, low, high, Return_calculated	drop_unusefull_columns(dataset_1, dataset_2, market) dataset_1 = stock data dataset_2 = historical prices/volumes
Merge datasets	
Merge stock data with historical prices/volumes based on ticker (stock)	merge_datasets_by_ticker(cleanuped_dataset1, cleanuped_dataset2, ticker)
Cleanup merged datasets	
1/ Drop rows when asset or cash flows were not found. **2/ Keep historical data related to asset and cash flows dates** (only the last 4 years were available for asset and cash	cleanup_merged_dataset(merged_set)

flows so I keep only historical prices of the last 4 years). **3/ Calculation of the return which will be used later to calculate the dependent variable (trend)** I calculate the **percentage** of increase or decrease of the close price of the day versus the close price of the day – 1. **'Calc_return'** : «close price of the day» - «close price of the day-1» / «close price of the day-1»	
MACD, signal and EMA calculation	
MACD = Price average at 12 days - Price average at 26 days Signal = MACD average at 9 days EMA = MACD - signal	MACD_calculated_dataset(merged_set, market)
Adding macroeconomic variables data (VIX, inflation, EPU)	
Adding macroeconomic variables data for side files specified in the markets.csv file.	Infos_added_set(ticker_with_MACD_set, market, suffix)

Data Cleaning and preprocessing

Task	**Function(arguments)**
Second part : "Cleaning the dataset"	
Split dataset into training and test	
Split dataset into chronological sets (training and test), a variable horizon will define the test size.	split_dataset(dataset)
Replace missing data	
I will use the interpolate method to fill the missing inflation rate, volume, VIX rate. I do this for **training and test set separately**.	repl_missing_data(dataset, settype)
Identify and remove noise (error and residuals = outliers)	
Convert alphanumeric into numeric variables	convert_to_numeric(dataset)
Identify outliers using LocalOutlierFactor. Add a column called 'outlier' to the dataset containing a -1/1 flag for outliers	identify_outliers(dataset, no_neigh, contam, typ)
Now remove residuals = outliers Apply separately on train and test dataset.	remove_outliers(dataset)
Check and address skewness	
Check skewness with kurtosis. If the value is not between -0.5 and 0.5 then I	Functions : check_and_address_skewness(dataset, typ)

address skewness with RandomOverSampler.	check_skewness(dataset, typ) address_skewness(dataset, typ)

Data reduction and projection

Task	**Function(arguments)**
Third part : "Data reduction and projection"	
Normalize or standardize variables	
Normalize dataset using MinMaxScaler. Standardize dataset using StandardScaler.	norm_std_variables(dataset)
Check correlation	
Check correlation between independent variables and dependent one ==> **line plot on all variables**	correlated_variables(dataset)
Check important variables according to Mutual information. Chi-squared,	return_variables_to_be_retained_or_removed(dataset)
Display a **heatmap** for all variables.	find_correlation(X)
Display scree-plots, loadings on principal components	find_pca_variables(X)

Data mining

I fit models on validation set with all the data related to a market. All stocks of this market will be used for fiitting.
Then I predict on the test set for each stock independently.

I keep building models until I find a suitable one that **works with the test set**.

Task	**Function(arguments)**
Fourth part : "Data mining"	
First I add a **trend** variable based on the return calculated. It will be our **dependent variable**. If the return calculated is greater or equal than 0 then the **trend** is equal to 1. Otherwise it's equal to -1. As the dependent variable is a trend which can take 2 values, I'll use **classification methods**.	
Fit and predict with **Simple Linear Regression and Multiple Linear Regression**	Functions : • evaluation_process1(dataset, variable) • evaluation_process2(dataset_train, dataset_test, variable) • calculate_vif(variables) • evaluation_process3(dataset_train) **Accuracy** evaluated with **R-squared**.
Fit and predict with **Clustering data**	
Fit and predict with **Decision trees**	**Accuracy** evaluated with **confusion matrix** : function calculate_cm(predicted, actual)
Fit and predict with **Random Forest and AdaBoost model**	**Accuracy** evaluated with **confusion** matrix : function calculate_cm(predicted, actual)

Fit and predict by **comparing different tree-based models**	**Accuracy** evaluated with **confusion** matrix : function calculate_cm(predicted, actual)
Fit and predict with **support vector machines**	**Accuracy** evaluated with confusion matrix : function calculate_cm(predicted, actual)
Fit and predict with **support vector regression model**	**Accuracy** evaluated with **confusion matrix** : function calculate_cm(predicted, actual)
Fit and predict with **MLP in Keras**	**Accuracy** evaluated with **R-squared**.
Fit and predict with **neural networks**	**Accuracy** evaluated with **confusion matrix** : function calculate_cm(predicted, actual)
Fit and predict with **ANN Hyperparameter exploration**	
Fit and predict with **Microsoft Light Gradient Boosting Machine model**	**Accuracy** evaluated with **confusion matrix** : function calculate_cm(predicted, actual)
Fit and predict with **Logistic Regression**	**Accuracy** evaluated with **R-squared**.
Fit and predict with **Naive Bayes**	**Accuracy** evaluated with **R-squared**.