Data-driven portfolio management

Back-testing and analysis of investing strategies

Programming for Data Processing Final Project

Author: Guillermo Vigueras & Raúl Gutiérrez

Institute: Computer Science Faculty - UPM

Date: February, 2021

Contents

Disclaimer

It should be noted that concepts, investment methods and assets data described in this document are provided with education purposes only and are not intended to provide specific advice or recommendations for any individual or on any specific security or investment product. Thus, it is only intended to provide an academic example of data harvesting, manipulation and analysis.

Chapter 1 Data-driven portfolio management

This chapter introduces and states the context and main concepts that will be used for completing the tasks of the project.

1.1 Introduction

Smallville Asset Management (or SAM in abbreviated form) is an investment advice firm, that wants to improve their advice services provided to their customers, mainly individual investors. SAM managers tasks mainly consist of defining the type of financial assets a customer should invest in depending on the personal profile of the customer. Thus, after some interviews with the customer, the manager obtains a profile with aspects like return or profit expected, risk tolerance, ... Based on this profile, the manager proposes to the customer the type of assets to invest in, with an explanation according to customer expectations and profile.

So far, Smallville Asset Management relays heavely on the intuition and long experience of their portfolio managers to control and define the investment strategies used in the firm. However, with recent rise of Big Data and Data Analytics, they are thinking to complement the human expertise and perform data-driven portfolio management.

In order to leverage the power of Data Analytics, at SAM are planning to develop a *Back testing* tool. This tool basically allows to evaluate how an investment strategy has performed in the past, taking into account the previous behaviour (e.g. price evolution) of the financial assets included in the investment strategy. For that reason, managers at SAM have selected the different assets they daily work with. Main aspects of these assets are described in following section.

1.2 Asset description

At Smallville Asset Management are considering the following family of assets to carry out the study of data-driven portfolio construction:

Family of assets

- ❏ *Stocks*
- ❏ *Corporate bonds*
- ❏ *Public bonds*
- ❏ *Gold*
- ❏ *Cash*

Following sections describe main characteristics of each type of asset and the investment vehicle chosen by SAM for each asset family.

1.2.1 Global stocks

One of the assets chosen by Smallville Asset Management is a set of shares of companies publicly traded worldwide. The investment vehicle for the set of shares is a fund that automatically replicates the MSCI world index. For that reason, this way of investing is called passive since the fund simply replicates what the index does (buys/sells) as opposed to an active fund, where the investing strategy is delegated to the fund manager. For that reason, costs for passive funds are lower (i.e. « 1.0% of invested capital) with respect to active funds (typically 2.0% of invested capital) .

Definition 1.1. Fund

Funds pool money from interested investors and use that money to buy other securities, usually public company stocks and/or bonds. A fund issues a number a shares, that can be bought by people or institutions interested in that particular fund.

The value, over time, of the fund shares depends on the performance of the stocks the fund decides to buy. So, when investors buy a unit or share of a fund, they are buying the performance of its portfolio or, more precisely, a part of the portfolio's value. For that reason, the value of fund shares varies over time (up/down), according to the evolution of the stocks/bonds the fund is invested in. On the other hand, stocks/bonds share price varies (up/down) depending on how (well/bad) the business is performing (earnings, growth of sales, ...), economic expectations, market sentiment, ...

When trading (sell/buy) fund shares, its value is computed once a day, when markets close. Such value is the price at which fund shares are bought or sold, according to investors commands.

♣

Information of the stock fund chosen by Smallville Asset Management can be found in Table 1.1. For more information, check the URL in the table:

Table 1.1: Asset description of the stock fund

Feature	Value
ISIN	LU0996182563
Name	Amundi Index Msci World Ae-c
Inv. Vehicle	Fund
AssetType	Stocks
Cost	0.4 % of invested capital
URL	https://www.investing.com/funds/amundi-msci-wrld-ae-c

1.2.2 Corporate bonds

Another asset chosen by Smallville Asset Management is the debt issued by publicly traded companies worldwide, in the form of bonds. SAM has added this asset to the list of products as investors highly demand Corporate bonds since the perception of less risk with respect to company stocks. In reality, this is not always the case as explained following.

Definition 1.2. Corporate bonds

Corporate bonds are made up of the debt issued by companies to bondholders in order to raise capital. In this way, bondholders receive a fixed interest rate for the money lent during a period of time. Once the time expires, the bondholder receives the money lent to the company plut the interest agreed. However, the bondholder can potentially lose the investment in case of bankruptcy of the borrower company.

Due to the fixed interest paid by bonds, they are considered to be less risky than company stocks, since there is no a-priori knowledge of stock rentability as these are influenced by many factors (business performance, economic forecast, market sentiment, ...). However, bonds are also publicly traded, and thus the bond price can rise or fall due to different reasons: business performance, changes in interest rates, ... Potentially, a bondholder can lose the whole investment in case of bankruptcy, as mentioned before.

♣

Since Smallville Asset Management wants to diversify geographically the Corporate bonds offered to its customers, the investment vehicle chosen for Corporate bonds is an ETF managed by Blackrock investment corporation.

Definition 1.3. ETFs

Exchange-traded funds (ETFs) are similar to funds but traded on the public market like stocks. In the same way as funds, ETFs issue units or shares that are traded at a given price. Additionally, ETF shares can be assigned a price at a fraction of time (e.g. milliseconds), as opposed to funds shares that are assigned a price in a daily basis.

♣

Information about the ETF for Corporate bonds chosen by Smallville Asset Management can be found in Table 1.2. For more information, check the URL in the table:

Table 1.2: Asset description for Corporate bonds

Feature	Value
ISIN	IE00B7J7TB45
Name	iShares Global Corporate Bond UCITS (CRPS)
Inv. Vehicle	ETF
AssetType	Corporate Bonds
Cost	0.2 % of invested capital
URL	https://www.investing.com/etfs/ishares-global-corporate-bond-$

1.2.3 Public or government bonds

Smallville Asset Management has also selected as an asset the public debt issued by governments worldwide, in the form of bonds.

Definition 1.4. Public bonds

Government or public bonds are for investors looking to put their money away in low-risk investments through Treasury securities, such as Treasury bonds, or agency-issued debt. As for Corporate bonds, public bonds are considered as one of the least risky assets. The main risk associated to public bonds is the possibility of default (i.e. bankruptcy) of the government/agency issuing the bonds. Although, the probability of sovereign default is low for some countries, history has demonstrated that such probability can not be considered to be zero, forever.

♣

The investment vehicle chosen for Public bonds is an ETF managed by DWS Investments UK Limited. Information about the ETF for Public bonds can be found in Table 1.3. For more information, check the URL in the table:

Table 1.3: Asset description for Public bonds

Feature	Value
ISIN	LU0908508731
Name	Xtrackers II Global Government Bond UCITS ETF 5C (XG7S)
Inv. Vehicle	ETF
AssetType	Public Bonds
Cost	0.2 % of invested capital
URL	https://www.investing.com/etfs/db-x-trackers-ii-global-sovereign-5

1.2.4 Gold

Gold is an asset that has equally supporters and haters. The main arguments for haters is the fact that gold does not produce earnings and cannot pay dividends, some others argue that gold is a highly speculative asset. On the contrary, supporters consider that gold is a true store of value (as oppose to fiat money) and thus can provide profits to an investor in the long run.

The investment vehicle chosen for gold is an ETF managed by World Gold Council. Information about the ETF for gold can be found in Table 1.4. For more information, check the URL in the table:

Table 1.4: Features description for Gold

Feature	Value
ISIN	US78463V1070
Name	SPDR® Gold Shares (GLD)
Inv. Vehicle	ETF
AssetType	Gold
Cost	0.1 % of invested capital
URL	https://www.investing.com/etfs/spdr-gold-trust

1.2.5 Cash

Although cash is not a financial asset *per se*, SAM managers have also identified it as an asset since having some cash in an investment portfolio, allows to take advantage of buying opportunities that arise in markets.

Nevertheless, cash suffers from inflation (i.e. devaluation) and benefits from deflation (i.e. value appreciation). The fact that these economic phenomena exist should be taken into account when managing an investment portfolio.

The investment vehicle chosen for cash is an Index that represents the evolution of US dollar. At SAM has considered this currency since the rest of investment vehicles uses this currency. Information about the dollar index can be found in Table 1.5. For more information, check the URL in the table:

Table 1.5: Features description for Cash

Feature	Value
ISIN	US78463V1070
Name	US Dollar Index (DXY)
Inv. Vehicle	Index
AssetType	Cash
Cost	0.0 % of invested capital
URL	https://www.investing.com/indices/usdollar

Note *The US dollar index represents the evolution of the dollar value over time. Daily values registered by the US dollar index are expressed in base 100. This means, for example, if for a given day the index value is 96, it should be divided by 100 to compute the value of 1 US dollar for that day. Thus, for an index value of 96, the real value of* $1\$ = 1\$ * 0.96 = 0.96\$$

Chapter 2 Data harvesting

Smallville Asset Management is interested in generating different investment strategies for evaluating several aspects of portfolio performance. The main input data required by SAM is the price over time of each asset.

In order to reduce costs of accessing financial data, SAM has decided to freely obtain the data from investing.com. This website offers financial information about different types of assets and investment vehicles, publicly traded. Since `investing.com` has the same format for webpages for different assets, SAM has realized that data harvesting can be programmatically performed through web scrapping.

Following sections describe website locations in order to browse and get the required financial data. Locations are explained for the Corporate bonds ETF, but can be extrapolated for the rest of investment vehicles.

2.1 Asset main page

When accessing the main page for a given investment vehicle, a similar screen is shown as the one in Figure 2.1. Such screen contains a link to the historical data associated to the asset. Figure 2.1, magnifies the area of the webpage where the link to `historical data` is.

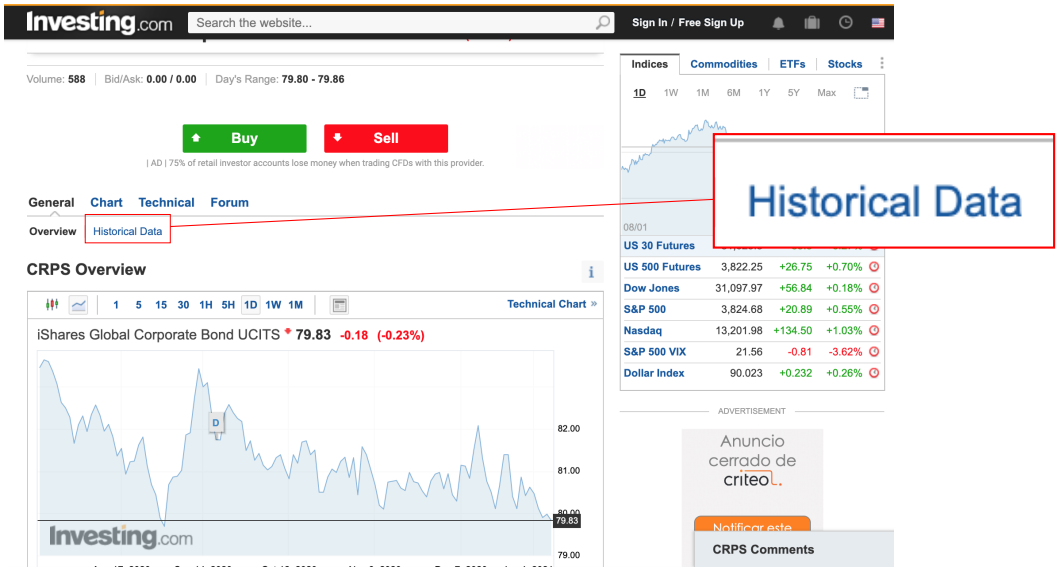

Figure 2.1: Main page of asset and access to historical data

2.2 Asset historical data

The page containing asset historical data is shown in Figure 2.2. When setting the `Time frame` to a daily basis (default set), the page shows historical data (prices, volume of trading and

% of change in price) for the last month. The range of dates can be changed by clicking on a button with a calendar icon. Such button is highlighted in Figure 2.2

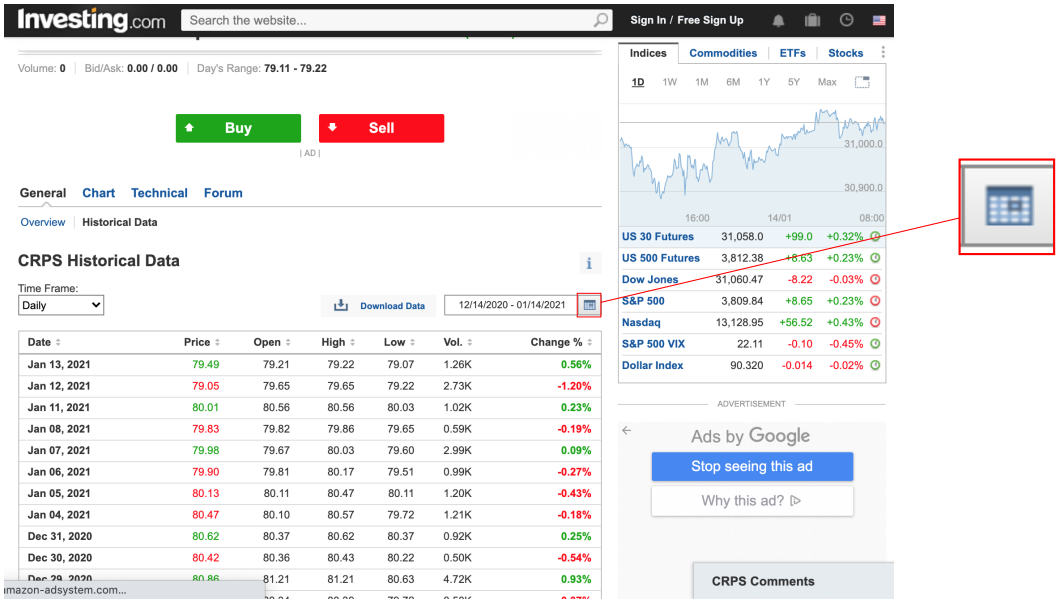

Figure 2.2: Webpage displaying asset historical data and detail of the button for date change

Once displayed the form for specifying the date range, the starting and ending dates can be changed using the input text fields `Start Date` and `End Date`. These fields are highlighted in Figure 2.3, where dates can be specified using the format `MM/DD/YYYY`.

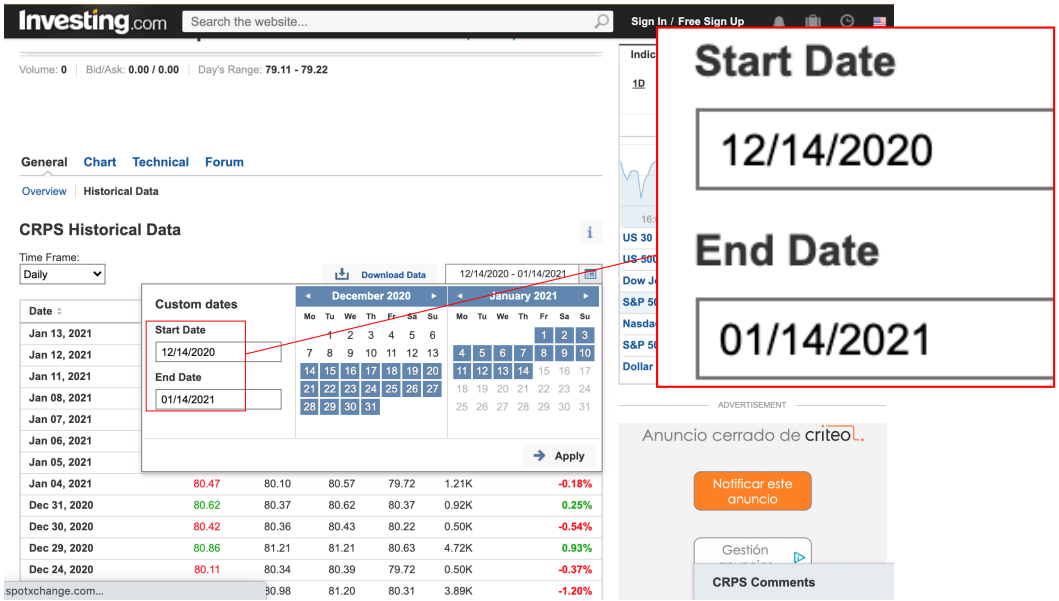

Figure 2.3: Webpage displaying the form for updating date ranges for asset prices

After input text fields for dates have been updated, these changes can be set by clicking on button `Apply`. This button is highlighted in Figure 2.4.

Once button `Apply` has been clicked, the table showing asset information over time is

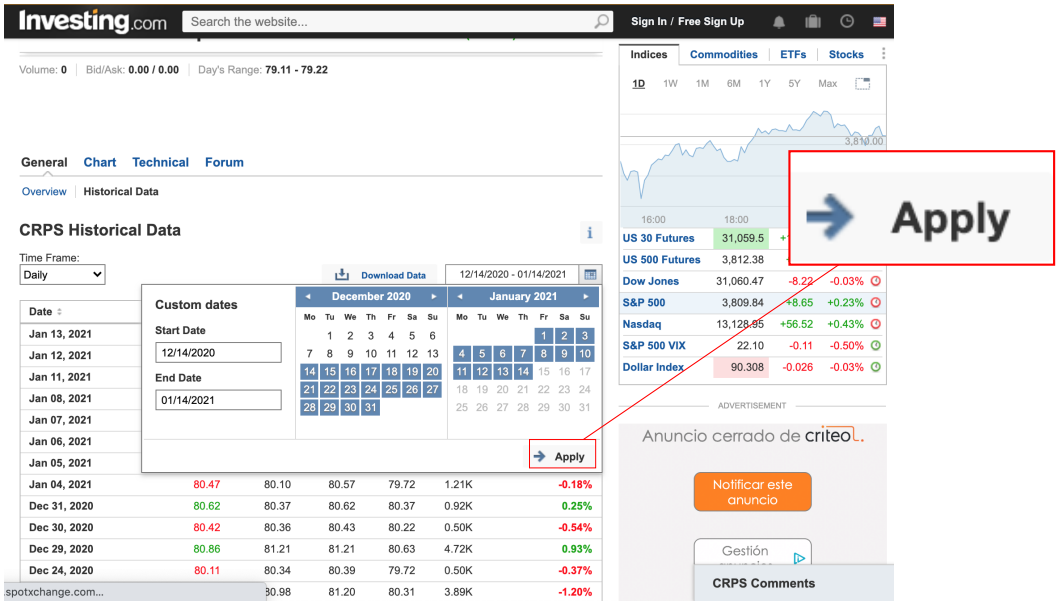

Figure 2.4: Detail of the `Apply` button for updating date ranges for asset data

updated to reflect the new date range set. This table is highlighted in Figure 2.5. Also, that table has several columns with asset price information. The column corresponding to the price for a given date is the one with column header `Price`. Columns with headers `Open`, `High` and `Low` can be discarded. The temporal evolution of asset prices can be obtained by accessing the DOM element corresponding to the HTML table mentioned before.

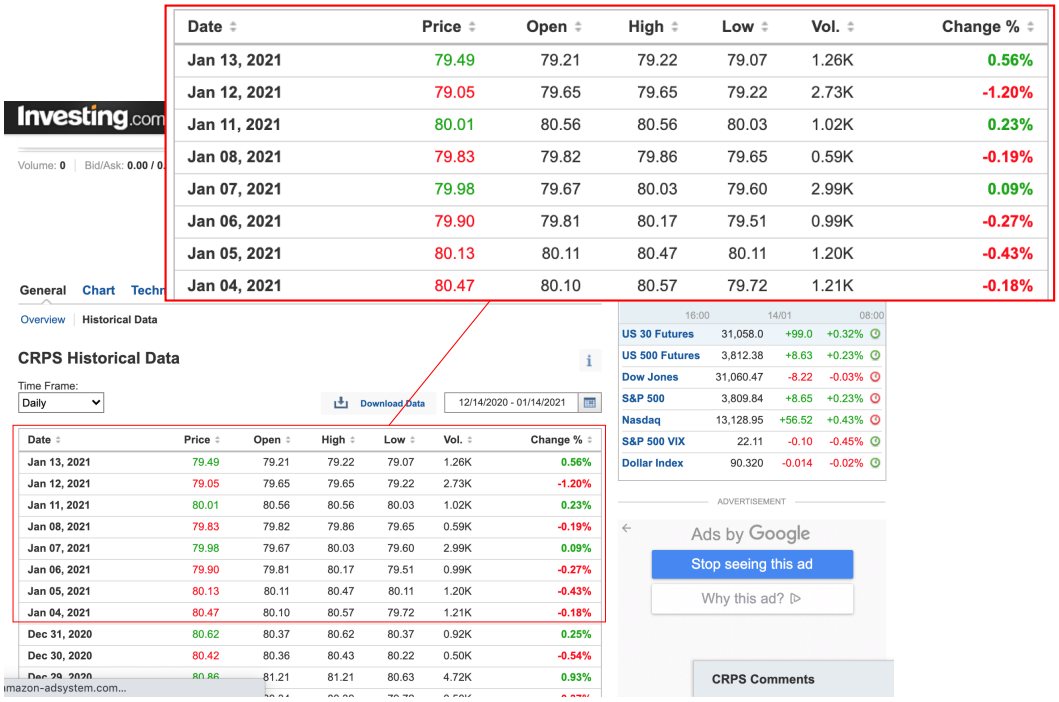

Figure 2.5: Detail of the table showing asset information over time

2.3 Assets prices

As a reference, Figure 2.6 shows price evolution over time of assets described in previous sections.

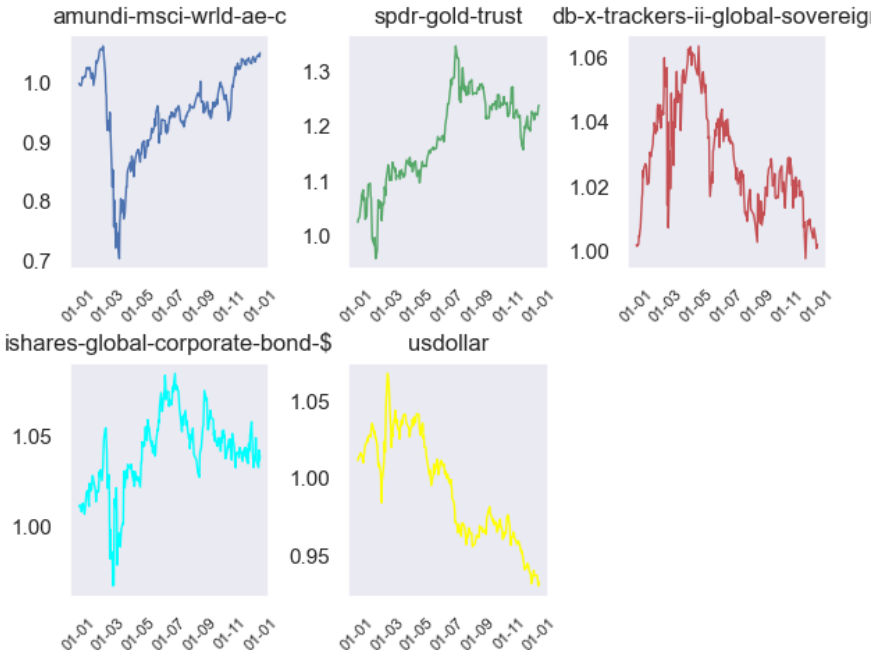

Figure 2.6: Price evolution of assets over time

2.4 Tasks related to data harvesting

Smallville Asset Management requires the harvesting of data from investing.com for obtaining the historical data of the assets described in Section 1.2. The dates SAM is interested in fall in the range: `01/01/2020` and `12/31/2020`, both inclusive.

For that purpose, historical data of assets should be harvested and stored in individual files (e.g. one file per asset) in order to be used in later tasks. This files should be stored in a Pandas friendly format (e.g. `.csv`) since tasks in following chapters are expected to be completed using Pandas.

Names of generated files should be:

- amundi-msci-wrld-ae-c.csv
- db-x-trackers-ii-global-sovereign-5.csv
- ishares-global-corporate-bond-$.csv
- spdr-gold-trust.csv
- usdollar.csv

Chapter 3 Investment strategies generation

Based on the assets data harvested, according to previous chapter, Smallville Asset Management wants to generate different investment strategies in order to later analyse and characterise each strategy.

An investment strategy is defined by two aspects that will impact the performance obtained by the strategy:

- **Investment portfolio**: consists of defining what type of assets and the weight assigned to each type in the portfolio.
- **Trading methodology**: consists of defining when to buy and sell the assets selected for the portfolio. Recall that the return of an investment depends on the price paid and the price obtained when selling.

3.1 Investment portfolio

An investment portfolio is defined by mainly two input arguments:

- **Type of assets**: assets selected to invest in, among all the assets considered by SAM (see Section 1.2).
- **Asset allocation**: the weight assigned to each asset selected for the portfolio. The weight is typically expressed as a percentage. Thus, the weights for all assets within a portfolio must sum 1.0 or 100%.

Following an example of portfolio definition is shown according to typical representation used in finance and also adopted in SAM .

Example 3.1 Portfolio definition

As mentioned before, Smallville Asset Management has decided to form portfolios, assigning weights to the following assets:

- Stocks (ST)
- Corporate bonds (CB)
- Public bonds (PB)
- Gold (GO)
- Cash (CA)

Thus, a portfolio is defined by a sequence of numbers separated by slash symbol (/). The order of each number in the sequence indicates the type of asset and each number in the sequence represents the weight for that type of asset in the portfolio. If an asset should not be included in the portfolio, the assigned weight is 0.0. Table 3.1 shows some examples of portfolio mix definitions:

Table 3.1: Examples of portfolio definition

Asset allocation					Portfolio mix
ST	CB	PB	GO	CA	ST / CB / PB / GO / CA
50%	20%	20%	0%	10%	50 / 20 / 20 / 0 / 10
10%	20%	20%	40%	10%	10 / 20 / 20 / 40 / 10
50%	40%	10%	0%	0%	50 / 40 / 10 / 0 / 0
0%	20%	20%	60%	0%	0 / 20 / 20 / 60 / 0

3.2 Trading methodology

The other dimension that defines an investment strategy is the trading methodology, that basically defines when to buy and sell shares of the given assets selected for a portfolio. The number of trading methodologies available in literature is massive, but SAM has chosen the following ones:

- **One off (1-OFF)**: All the funds are invested the first day. The money invested is split among the assets according to the portfolio allocation defined.

  Example 3.2 One off investment

  Let's suppose an investor has 100$ to invest using the following portfolio allocation: 50 / 20 / 0 / 20 / 10. For the one off investment, the 100$ are completely invested the first day according to weights. The following table shows how is computed the amount of shares bought, assuming an hypothetical price:

  Table 3.2: One off investment

Asset	Amount ($)	Asset price	# shares
ST	50	12.5	4
CB	20	4	5
PB	0	–	0
GO	20	10	2
CA	10	1	10

  It should be noted that the amount of shares the investor will buy of each asset will depend on the asset price, except for cash. Thus, for example, the investor would buy 4 shares of stocks since the price is 12.5$. However, for cash, the price is not obtained from the US Dollar index described before, but the buy price is considered to be 1$.

- **Dollar Cost Averaging (DCA)**: The funds are equally divided in as many amounts as months available in the investment period. Later, each monthly amount is invested the first day of each month, splitting such monthly amount among the assets according to the portfolio mix defined.

  Example 3.3 Dollar Cost Averaging investment

Let's suppose an investor has 100$ to invest along 5 months, using the following portfolio allocation: 50 / 20 / 0 / 20 / 10. For the DCA investment, the 100$ are equally divided in 5 periods to invest each of the 5 months. So each month, 20$ are invested according to the portfolio allocation. The following table shows how is invested each period amount (20$), based on the portfolio allocation:

Table 3.3: DCA investment

		Asset investment ($)				
Period	Period Amount ($)	ST	CB	PB	GO	CA
Month 1	20	10	4	0	4	2
Month 2	20	10	4	0	4	2
Month 3	20	10	4	0	4	2
Month 4	20	10	4	0	4	2
Month 5	20	10	4	0	4	2
5 Months	100	50	20	0	20	10

It should be noted that for each month a different asset price will be obtained, so the number of shares bought each month for each asset will vary. Only exception is cash, for which a price of 1$ is used each month.

Portfolio rebalance: Previous methodologies define when to buy assets. Regarding sales of assets a rebalance process is adopted. Rebalancing is the process by which an investor restores their portfolio to its target allocation. Rebalancing brings a portfolio back to the desired asset mix. This is done by divesting (i.e. selling) overperforming assets and investing (buying) underperforming ones.

Example 3.4 Portfolio rebalance

Let's suppose an investor has 100$ to invest using the following portfolio allocation: 50 / 20 / 0 / 30 / 0. Table 3.4 shows the value of the different assets at different instants (t_i) as well as after rebalancing (t_i^r)

Table 3.4: Portfolio rebalancing example

Asset	t_0		t_1		t_2^r	
	$	%	$	%	$	%
ST	50	50%	70	54%	65	50%
CB	20	20%	10	8%	26	20%
PB	0	0%	0	0%	0	0%
GO	30	30%	50	38%	39	30%
CA	0	0%	0	0%	0	0%
Total	100	100%	130	100%	130	100%

Table 3.4 shows how the portfolio changes in 3 different moments: initial (t_0), later (t_1) and

rebalance (t_2^r). From t_0 to t_1 Stocks increases by 20$, Corporate bonds decreases by 10$ and Gold increases by 20$. As a result the weights in portfolio at t_1 are: 54%, 8% and 38%. After rebalance, weights for Stocks, Corporate bonds and Gold return back to 50%, 20% and 30%, respectively. And the amounts for each asset are: 65$, 26$ and 39$, respectively.

It should be noted that portfolio rebalance is a result of selling overvalued assets for buying undervalued assets. Thus, in the previous example, the 16$ bought in t_2^r of CB shares, are obtained after selling 11$ of GO and 5$ of ST. Also note that the amount of shares of CB and GO that are required to be sold, will depend on the actual share price applied. In turn the share price will depend on the day the rebalance and the sell of shares is performed.

Rebalancing process can be combined with trading methodologies described above. Table 3.5 summarises the different trading methodologies obtained as a result of combining buying and rebalancing options.

Table 3.5: Examples of portfolio definition

		Trading operations dates	
Trading Meth.	Rebal.	Buys	Rebal.
1-OFF	NO	The first day of investment period	No
1-OFF	YES	The first day of investment period	Day 15th of each month
DCA	NO	The first day of each month	No
DCA	YES	The first day of each month	Day 15th of each month

3.3 Investment strategy performance

Smallville Asset Management has defined some metrics to analyse and classify a given portfolio that follows some trading methodology. These metrics are the following:

- **Cost**: This metric represents the cost of the portfolio, computed as the sum of each individual asset cost, multiplied by the weight of the asset in the portfolio. If the cost of ith asset is represented as c_i and the weight of the asset in the portfolio as w_i, the cost of the portfolio can be calculated as:

$$Cost\ of\ portfolio = \sum_{i\ \in\ Assets} c_i * w_i \quad where\ \ Assets = \{ST, CB, PB, GO, CA\}$$

- **Volatility**: This metric refers to the amount of uncertainty or risk related to the size of changes in an asset value. A higher volatility means that an asset value can potentially be spread out over a larger range of values. This means that the price of the asset can change dramatically over a short time period in either direction. A lower volatility means that an asset value does not fluctuate dramatically, and tends to be more steady.

Volatility can be measured with different metrics. For this matter is computed based on the standard deviation of the evolution of an asset price over time. So, suppose x_i is the different daily prices of an asset with $1 \leq i \leq N$ where N is the number of days considered in a sample. First the standard deviation (σ_X) of the sample (X) is computed, using the sample average (μ_X) as:

$$\sigma_X = \sqrt{\frac{1}{N}\sum_{i}^{N}(x_i - \mu_X)^2} \quad where \;\; X = \{x_0, x_1, ..., x_N\}$$

Then volatility of the asset is computed as a percentage with respect to the sample average (μ_X):

$$Volatility(X) = \frac{\sigma_X}{\mu_X} * 100$$

In order to characterise an investment strategy, Smallville Asset Management computes the yearly volatility of its associated portfolio. For that matter, a temporal sequence with the daily values of the portfolio ($Values = \{value_0, value_1, ..., value_N\}$) and the sequence values are computed as:

$$value_i = \sum_{j \in Assets} shares_i^j * price_i^j \quad where \;\; Assets = \{ST, CB, PB, GO, CA\}$$

Note *$shares_i^j$ represents the number of shares for the j-th asset of the portfolio on the i-th day. This value is obtained as a result of the trading methodology used in the portfolio (i.e. buys/sales). On the other hand, $price_i^j$ represents the share price for the j-th asset on the i-th day. This information is obtained from investing.com through web scraping (see Chapter 2).*

Based on the sequence of portfolio values defined above, the volatility of an investment strategy is computed using the portfolio value sequence defined above (i.e. $Values$):

$$Volatility\ of\ portfolio = Volatility(Values)$$

- **Return**: The last metric used by SAM is the return of profit obtained by a portfolio. Return is computed as a percentage taking into account the price paid for every *buy* transaction of shares and the current value of portfolio, calculated using share's price at the time return is computed. If the current portfolio value is higher than the amount paid for shares bought, then return would be positive, otherwise a negative return would be obtained. Additionally, portfolio return is computed for different time periods, measured in months: 1, 3, 6, 9 and 12 months. Nevertheless, return for all periods is computed in the same way, simply

changing the date at which current portfolio value is computed. Thus, assuming the set of buy operations performed within the portfolio is defined as:

$$B = \{b_i^j\} \quad with\, 1 \leq i \leq M,\ and \quad j \in Assets = \{ST, CB, PB, GO, CA\}$$

where each b_i^j is the number of shares of the j-th asset adquired in the i-th buy transaction. Thus, return can be computed as:

$$buy\ amount = \sum_{j \in Assets} \sum_{i}^{M} (b_i^j\ *\ p_i^j)$$

$$current\ value = \sum_{j \in Assets} (\sum_{i}^{M} b_i^j)\ *\ p_c^j$$

$$portfolio\ return = \frac{(current\ value - buy\ amount)}{buy\ amount} * 100$$

where p_i^j is the share price paid for the i-th buy transaction and p_c^j is the current share price, both for the j-th asset.

3.4 Tasks related to investment strategies generation

Smallville Asset Management requires three tasks to automatise the generation and evaluation of investment strategies:

1. **Portfolio allocation**: automatic generation of portfolio allocations must be generated, where the Δ or increment/decrement in each asset weight is 0.05 (5%). Note that the sum of asset weights for a portfolio mix must be always equal to 1.0 (100%). Table 3.6 shows the schema for portfolio generation.

Table 3.6: Portfolio rebalancing example

Asset Alloc.	ST	CB	PB	GO	CA
1	100%	0%	0%	0%	0%
2	95%	5%	0%	0%	0%
3	95%	0%	5%	0%	0%
4	95%	0%	0%	5%	0%
5	95%	0%	0%	0%	5%
6	90%	5%	5%	0%	0%
7	90%	5%	0%	5%	0%
		...			
i	90%	10%	0%	0%	0%
i+1	90%	0%	10%	0%	0%
i+2	90%	0%	0%	10%	0%
i+3	90%	0%	0%	0%	10%
		...			
i+n	0%	100%	0%	0%	0%
		...			

It should be noted that the number of different portfolio allocations can be computed a priori, since the 100% is split in 20 pieces, representing each piece a 5% of the portfolio. Table 3.7 tries to represent that:

Table 3.7: Description of portfolio generation with $\Delta = 5\%$

5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%	5%
ST	ST	ST	ST	ST	ST	ST	ST	ST	ST	ST	ST	ST	ST	ST	ST	ST	ST	ST	ST
CB	CB	CB	CB	CB	CB	CB	CB	CB	CB	CB	CB	CB	CB	CB	CB	CB	CB	CB	CB
PB	PB	PB	PB	PB	PB	PB	PB	PB	PB	PB	PB	PB	PB	PB	PB	PB	PB	PB	PB
GO	GO	GO	GO	GO	GO	GO	GO	GO	GO	GO	GO	GO	GO	GO	GO	GO	GO	GO	GO
CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA	CA
?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?

Thus according to Table 3.7, a portfolio with weights 50/10/0/20/20, would have the following distribution for the different asset types:

- ST: 10 'pieces' of 5% each, for a total weight of 50% in the portfolio.
- CB: 2 'pieces' of 5% each, for a total weight of 10% in the portfolio.
- PB: 0 'pieces' of 5% each, since the asset weight in the portfolio is 0%.
- GO: 4 'pieces' of 5% each, for a total weight of 20% in the portfolio.
- CA: 4 'pieces' of 5% each, for a total weight of 20% in the portfolio.

According to the previous formulation of the portfolio generation, the number of different portfolios can be computed as n-combinations with repetition of m elements (element = asset), where n=20 and m=5. Thus, the number of different allocations to be computed can be calculated as:

$$CR_m^n = \binom{m+n-1}{n} = \frac{(m+n-1)!}{n! * (m-1)!} = 10626\ portfolio\ allocations$$

Note that this number is just a reference for a simple check of the the python code for computing the portfolio allocations.

As part of this subtask, portfolio allocations generated should be stored in the file

`portfolio_allocations.csv` in order to be used in later tasks and also to allow individual validation of this subtask.

2. **Trading methodology**: for the different portfolio allocations generated previously, apply the trading methodologies, shown in Table 3.5, with and without rebalancing.

   As part of this subtask, trading methodologies generated should be stored in the file `trading_methodologies.csv` in order to be used in later tasks and also to allow individual validation of this subtask.

3. **Portforlio performance**: compute the performance metrics, described before, for the resulting investment portfolios obtained from the combination of asset allocations and trading methodologies, computed in previous tasks.

   As part of this subtask, portfolio performance metrics generated should be stored in the file `portfolio_metrics.csv` in order to be used in later tasks and also to allow individual validation of this subtask.

Note *It is important to note that, it might be case that price data obtained according to Chapter 2, might have missing values for the date ranges considered for this exercise. Missing values have to be properly accounted for in order to correctly compute portfolio metrics. Thus, how to deal with missing values and the best filling stratagy is a decision of your team.*

Chapter 4 Investment strategies analysis

With all the portfolio information generated in previous tasks, Smallville Asset Management wants to develop some model for automatic financial advisoring of its customers. But before starting with predictive models, SAM Data Science (DS) manager wants your team to perform some data analysis.

4.1 Tasks related to investment strategies analysis

Smallville Asset Management requires the following tasks to perform the analysis of investment strategies:

1. **Return**: SAM Data Science (DS) manager has asked you to analyse the distribution of returns obtained from the different portfolios generated. The analysis must be perform for the different time periods used previously: 1 month (1M), 3 months (3M), 6 months (6M), 9 month (9M), 12 months (12M). For such analysis, generate the plots and/or data your team considers necessary.

   According, to the analysis performed, does your team think it is more probable to obtain a positive or negative return? Also provide a reply to this question for the different time periods (1M, 3M, 6M, 9M and 12M). Use the previous generated plots and/or data to support your answer to the question.
2. **Return vs. risk**: Your team has been requested to analyse the relation among return and risk for the 12M period. Return vs. risk is typically analysed in finance displaying a 2D plot with a measure of risk in the 0X axis and return in the 0Y axis. In this case, risk will be quantified using volatility.

   According, to the analysis performed, does your team think it is **ALWAYS** true that the higher the risk, the higher the obtained return is? Use the previous generated plot and/or data to support your answer to the question.
3. **Financial advisoring**: The DS manager wants your team to provide evidence (i.e. plots/-data) to answer the following questions:

   A client is seeking to invest some spare money to obtain the highest possible return in the period from 01/01/2020 to 31/12/2020. However, the client only wants to invest in CB and PB, holding a 10% of cash in case of an event. So,

   - Do you think the customer requirements are feasible according to the investment strategies data your team has generated?
   - Either if your answer to the previous question is yes or no, provide a recommendation to the customer of an asset allocation along with the reasons for such recommendation.

   Provide plots/data supporting your answers to previous questions.

Chapter 5 Submission instructions

- One submission (and unique) by each group of students.
- Submission will be done by uploading (i.e. commit) your final version to a Git repository (more details to be provided later on)
- **Due date for submission is Friday 23/04/2021**.
- The following content should be included in your project's submission:
  - The Python code developed by your group, providing solution to the final project
  - A report regarding data analysis questions in Chapter 4. This report must include a brief answer to each question, supported by plots, tables or any other data representation you consider relevant. Please, be concise in your answer, taking into account that a longer answer is not always a good answer. The report must be a `.pdf` file, and the format can be freely selected by each group, either slides or document.
  - In addition, each group must upload to the project corresponding, a plain text file (e.g., README.txt) in which the following is briefly described:
    - Description of the content (files and dirs.) of the root directory of your practice, it is not necessary to describe the content of each subdirectory.
    - The final project statement is divided in 3 parts: web scraping, data generation and data analysis. Thus, a brief explanation of how to run your code for each part (either explaining the instructions for the steps to follow or a Makefile or equivalent file). Also indicate python packages and versions required by your code.
    - A list of the Python modules/files included in your project, briefly indicating a description of them. There is no need to describe each file on a function basis, simply a general description of the file. For example: file `File1.py` computes the different asset allocations for portfolios.
    - What functionality required have been addressed in the code, relating (if possible) project requirements with the code implemented. For example, indicate: functionality about web scraping has been addressed in files `File1.py`, `File2.py`, ...