



# Gestión de inversiones basada en los datos

**Análisis histórico de datos para gestión de carteras de inversión**

**Práctica de la asignatura Programación para Ciencia de Datos**

**Author:** Guillermo Viguera & Sergio Paraíso & Antonio Díaz

**Institute:** Computer Science Faculty - UPM

**Date:** September, 2021



# Contents

<b>1</b>	<b>Generación de estrategias de inversión</b>	<b>2</b>
1.1	Cartera de inversión . . . . .	2
1.2	Comportamiento de una estrategia de inversión . . . . .	3
<b>2</b>	<b>Tareas a realizar</b>	<b>6</b>
2.1	Tareas relacionadas de la funcionalidad obligatoria . . . . .	6
2.2	Tareas relacionadas de la funcionalidad extra u opcional . . . . .	8
<b>3</b>	<b>Evaluación y entrega de la práctica</b>	<b>10</b>
3.1	Evaluación de la práctica . . . . .	10
3.2	Entrega de la práctica . . . . .	10

## Disclaimer

Los conceptos, métodos de inversión y datos de activos descritos en este documento se proporcionan únicamente con fines educativos y no están destinados a proporcionar consejos o recomendaciones específicas para ninguna persona o sobre ningún producto de inversión o valor específico. Por lo tanto, solo tiene la intención de proporcionar un ejemplo académico de recolección, manipulación y análisis de datos.

## Chapter 1 Generación de estrategias de inversión

Basándose en los datos de activos recolectados, según lo realizado en la práctica 1, Smallville Asset Management quiere generar diferentes estrategias de inversión para luego analizar y caracterizar cada estrategia. La principal característica que marcará la rentabilidad proporcionada por una estrategia de inversión es:

- **Cartera de inversión:** Consiste en definir qué tipo de activos y el peso asignado a cada tipo en la cartera.

### 1.1 Cartera de inversión

Una cartera de inversiones se define principalmente por dos aspectos:

- **Tipo de activos:** activos seleccionados para invertir, de entre todos los activos considerados por SAM (ver enunciado de la práctica 1).
- **Asignación de activos:** el peso asignado a cada activo seleccionado para la cartera. El peso se expresa típicamente como porcentaje. Por tanto, las ponderaciones de todos los activos de una cartera deben sumar 100%.

A continuación, se muestra un ejemplo de definición de cartera de acuerdo con la representación típica utilizada en finanzas y también adoptada por SAM .

#### Example 1.1 Definición de una cartera de inversiones

Como se mencionó anteriormente, Smallville Asset Management define sus carteras de inversión asignando pesos a los siguientes activos:

- Acciones (Stocks o ST)
- Bonos Corporativos (Corporate Bonds o CB)
- Bonos Públicos (Public Bonds o PB)
- Oro (Gold o GO)
- Efectivo (Cash o CA)

**Table 1.1:** Ejemplos de definición de carteras

Asignación de activos					Cartera resultante
ST	CB	PB	GO	CA	ST / CB / PB / GO / CA
60%	20%	20%	0%	0%	60 / 20 / 20 / 0 / 0
20%	20%	20%	40%	0%	20 / 20 / 20 / 40 / 0
40%	40%	20%	0%	0%	40 / 40 / 20 / 0 / 0
0%	20%	20%	60%	0%	0 / 20 / 20 / 60 / 0

Por lo tanto, una cartera de inversión se define mediante una secuencia de números separados por un símbolo de barra (/). El orden de cada número en la secuencia indica el tipo de activo y cada número en la secuencia representa el peso de ese tipo de activo en la cartera. Si un activo

no debe incluirse en la cartera, el peso asignado es 0.0. La tabla 1.1 muestra algunos ejemplos de definiciones de combinación de carteras.

En base a la definición de carteras explicada, se asume que la inversión de los fondos disponibles en cada cartera se hace el primer día, es decir, el 01/01/2020. Así, el dinero invertido se divide entre los activos de acuerdo con la asignación de cartera definida. A continuación se muestra un ejemplo que explica la inversión para una cartera concreta:

### Example 1.2 Compra de activos en una cartera de inversión

Supongamos que un inversor tiene 100 \$ para invertir utilizando la siguiente asignación de cartera: 40 / 20 / 0 / 20 / 20. Esos 100\$ se invierten completamente el primer día de acuerdo con la asignación de activos en la cartera. La siguiente tabla muestra cómo se calcula la cantidad de participaciones de cada activo compradas, asumiendo un precio hipotético de los activos para el día 01/01/2020:

**Table 1.2:** Inversión en una cartera 50 / 20 / 0 / 20 / 10

Activo	Asignación (%)	Importe (\$)	Precio	# participaciones
ST	40	40	12.5	3.2
CB	20	20	4	5
PB	0	0	–	0
GO	20	20	10	2
CA	20	20	1	20

Destacar que la cantidad de participaciones que se comprarán de cada activo dependerá del precio del activo, excepto para el efectivo (CA). Así, por ejemplo, el inversor compraría 3,2 acciones (ST) ya que el precio es de 12,5\$. Por simplicidad, se asume que se puede adquirir un número decimal de participaciones. Respecto al uso de precios de participaciones, para el caso del efectivo (CA), el precio no se obtiene del US Dollar index (descrito en la práctica 1), sino que el precio (o coste) de compra se considera 1\$.

## 1.2 Comportamiento de una estrategia de inversión

Smallville Asset Management ha definido algunas métricas para analizar y clasificar una cartera determinada. Estas métricas son las siguientes:

- **Rentabilidad o Rendimiento:** Esta métrica utilizada por SAM es la rentabilidad obtenida por una cartera. Dicha rentabilidad se calcula como un porcentaje teniendo en cuenta el precio pagado por cada transacción de compra de acciones y el valor actual de la cartera, calculado utilizando el precio de la acción en el momento en que se calcula el rendimiento. Si el valor actual de la cartera es superior a la cantidad pagada por las acciones compradas, la rentabilidad sería positiva, de lo contrario se obtendría una rentabilidad negativa. Smallville Asset Management calcula la rentabilidad de una cartera

para un período de tiempo de 12 meses. Así, asumiendo que el conjunto de operaciones de compra realizadas dentro de la cartera se define como:

$$B = \{b^j\} \text{ con } j \in \text{Activos} = \{ST, CB, PB, GO, CA\}$$

donde cada  $b^j$  es el número de participaciones compradas del  $j$ -ésimo activo. Así, la rentabilidad se puede calcular como:

$$\begin{aligned} \text{importe\_de\_compra} &= \sum_{j \in \text{Activos}} (b^j * p_c^j) \\ \text{valor\_actual} &= \sum_{j \in \text{Activos}} (b^j * p_a^j) \\ \text{rentabilidad\_cartera} &= \frac{(\text{valor\_actual} - \text{importe\_de\_compra})}{\text{importe\_de\_compra}} * 100 \end{aligned}$$

donde  $p_c^j$  es el precio de compra y  $p_a^j$  es el precio actual de una acción, ambos para el  $j$ -ésimo activo.

- **Volatilidad:** Esta métrica se refiere a la cantidad de incertidumbre o riesgo relacionado con el tamaño de los cambios en el valor de un activo. Una mayor volatilidad significa que el valor de un activo se puede distribuir potencialmente en un rango más amplio de valores. Esto significa que el precio del activo puede cambiar drásticamente en un corto período de tiempo en cualquier dirección. Una volatilidad más baja significa que el valor de un activo no fluctúa drásticamente y tiende a ser más estable.

La volatilidad se puede medir con diferentes métricas. Para ello se calcula en base a la desviación estándar de la evolución del precio de un activo en el tiempo. Así, supongamos que  $x_i$  son los diferentes precios diarios de un activo con  $1 \leq i \leq N$  donde  $N$  es el número de días considerados en una muestra  $X = \{x_i\}$ . Primero se calcula la desviación estándar ( $\sigma_X$ ) de la muestra ( $X$ ), usando el promedio de la muestra ( $\mu_X$ ) como:

$$\sigma_X = \sqrt{\frac{1}{N} \sum_i^N (x_i - \mu_X)^2} \text{ donde } X = \{x_0, x_1, \dots, x_N\}$$

Entonces, la volatilidad del activo se calcula como un porcentaje con respecto al promedio de la muestra ( $\mu_X$ ):

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


Con el fin de caracterizar una estrategia de inversión, Smallville Asset Management calcula la volatilidad anual, es decir a 12 meses, de cada cartera. Por lo tanto la muestra  $X$

comprenderá los precios de un activo desde el 01/01/2020 hasta el 31/12/2020. Para ese rango de fechas se calcula una secuencia temporal (*Values*) con los valores diarios de la cartera:

$$Values = \{value_i\} = \{value_0, value_1, \dots, value_N\}$$


Donde cada valor diario  $value_i$  de la cartera de la secuencia temporal se calcula como:

$$value_i = \sum_{j \in \text{Activos}} particip_i^j * precio_i^j \quad \text{donde } \text{Activos} = \{ST, CB, PB, GO, CA\}$$

 **Note**  $particip_i^j$  representa el número de participaciones para el  $j$ -ésimo activo de la cartera en el  $i$ -ésimo día. Este valor se obtiene del número de participaciones que se compraron el primer día (01/01/2020), y se mantiene estable ya que no se compran ni se venden más participaciones. Por otro lado,  $precio_i^j$  representa el precio de la participación del  $j$ -ésimo activo en el  $i$ -ésimo día. Esta información de precios diarios de cada activo se ha de obtener de [investing.com](https://www.investing.com) tras el proceso de web scraping realizado en la práctica 1.

En base a la secuencia de valores de una cartera, definida anteriormente (*Valores*), la volatilidad de una estrategia de inversión se calcula utilizando dicha secuencia y la función *Volatility()* como:

$$Volatilidad \text{ de la cartera} = Volatility(Values)$$

 **Note** Es importante destacar que puesto que rentabilidad y volatilidad se calculan como porcentajes, no importa la cantidad inicial a invertir que se utilice (i.e. 1000\$, 10000\$, ...), puesto que los valores de rentabilidad y volatilidad deberían dar lo mismo.

## Chapter 2 Tareas a realizar

Con toda la información de carteras generada anteriormente, Smallville Asset Management quiere desarrollar algún modelo para la asesoría financiera automática de sus clientes. Pero antes de comenzar con modelos predictivos, el manager de Data Science (DS) de SAM quiere que vuestro grupo realice algunos análisis de datos. Este tarea solicitada por el manager de DS, define la funcionalidad requerida en esta práctica que se divide en 2 partes. Una primera parte incluye funcionalidad obligatoria y otra segunda parte que es funcionalidad extra u opcional, aunque cuenta para tener mejor nota.

### 2.1 Tareas relacionadas de la funcionalidad obligatoria

La funcionalidad incluye tareas de generación de carteras de inversión y análisis de la rentabilidad de dichas carteras generadas.

#### 2.1.1 Generación de carteras de inversión

Smallville Asset Management requiere las siguiente tarea para automatizar la generación y evaluación de estrategias de inversión:

1. **Asignación de carteras:** Se debe generar automáticamente carteras con distintas asignaciones de activos, donde el  $\Delta$  o incremento/decremento en cada peso de activo es 0.20 (es decir 20%). Tener en cuenta que la suma de las pesos de los activos para una combinación de cartera debe ser siempre igual a 1,0 (100%). La tabla 2.1 muestra el esquema para la generación de carteras.

**Table 2.1:** Generación de carteras con distintos pesos de activos

Asset Alloc.	ST	CB	PB	GO	CA
1	100%	0%	0%	0%	0%
2	80%	20%	0%	0%	0%
3	80%	0%	20%	0%	0%
4	80%	0%	0%	20%	0%
5	80%	0%	0%	0%	20%
6	60%	20%	20%	0%	0%
7	60%	20%	0%	20%	0%
...					
i	40%	60%	0%	0%	0%
i+1	40%	0%	60%	0%	0%
i+2	40%	0%	0%	60%	0%
i+3	40%	0%	0%	0%	60%
...					
i+n	0%	100%	0%	0%	0%
...					





Notar que el número de carteras diferentes obtenidas se puede calcular a priori, ya que la suma de pesos debe sumar el 100% que se divide en 5 segmentos, por lo que cada segmento es de un 20% de la cartera. La tabla 2.2 trata de representar lo descrito:

**Table 2.2:** Generación de carteras con  $\Delta = 5\%$

20%	20%	20%	20%	20%
ST	ST	ST	ST	ST
CB	CB	CB	CB	CB
PB	PB	PB	PB	PB
GO	GO	GO	GO	GO
CA	CA	CA	CA	CA
?	?	?	?	?

Por lo tanto, de acuerdo con la Tabla 2.2, una cartera con pesos 40/20/0/20/20, tendría la siguiente distribución para los diferentes tipos de activos:

- ST: 2 segmentos del 20% cada uno, para un peso total del 50% en la cartera.
- CB: 1 segmentos del 20% cada uno, para un peso total del 10% en la cartera.
- PB: 0 segmentos del 20% cada uno, ya que el peso del activo en la cartera es del 0%.
- GO: 1 segmentos del 20% cada uno, para un peso total del 20% en la cartera.
- CA: 1 segmentos del 20% cada uno, para un peso total del 20% en la cartera.

Así al representar esta cartera con pesos 40/20/0/20/20, conforme a la notación de la Tabla 2.2, obtendríamos una asignación como la que se muestra en la Tabla 2.3

**Table 2.3:** Generación de carteras con  $\Delta = 20\%$

20%	20%	20%	20%	20%
ST	ST	ST	ST	ST
CB	CB	CB	CB	CB
PB	PB	PB	PB	PB
GO	GO	GO	GO	GO
CA	CA	CA	CA	CA
ST	ST	CB	GO	CA

De acuerdo con la definición anterior de carteras basada en segmentos de 20% cada uno, el número de carteras diferentes se puede calcular como n-combinaciones con repetición de m elementos (elemento = activo), donde  $n = 5$  y  $m = 5$ . Por lo tanto, el número de asignaciones diferentes que se calcularán se puede calcular como:

$$CR_m^n = \binom{m+n-1}{n} = \frac{(m+n-1)!}{n! * (m-1)!} = 126 \text{ carteras distintas}$$

Este valor calculado sólo sirve para que comprobéis que vuestro código python de generación de carteras calcula 126 distintas carteras de un modo similar a como se muestra en la Tabla 2.1 y conforme a todo lo explicado antes.

## 2. Rentabilidad de carteras: Calcular la rentabilidad en 12 meses (12M) desde el 01/01/2020

hasta el 31/12/2020 de las 126 carteras generadas anteriormente. La rentabilidad de las carteras se debe calcular conforme a lo descrito en la Sección 1.2.

La información generada de las distintas carteras y la rentabilidad anual deben almacenarse en archivo con nombre `portfolio_returns.csv` para poder utilizarlo en el posterior análisis y también para permitir la validación individual de esta subtarea.

Dicho fichero `portfolio_returns.csv` debe tener 6 columnas y 126 filas, una fila por cada cartera generada. Las columnas deben estar en el orden:

ST,CB,PB,GO,CA,RETURN

Así, las 5 primeras columnas son para representar el peso de cada activo en cada cartera y la última columna para representar la rentabilidad anual de cada cartera.



**Note** Es importante señalar que, podría darse el caso que los datos de precios obtenidos por Web Scraping, puedan tener missing values para los rangos de fechas considerados para este ejercicio. Dichos missing values deben tratarse adecuadamente para calcular correctamente las métricas de la cartera (i.e. rentabilidad y volatilidad). Cómo tratar dichos missing values y completarlos es una decisión de cada grupo de prácticas.

### 2.1.2 Análisis de las carteras de inversión generadas

**Rentabilidad:** El manager de Data Science (DS) de SAM os ha pedido que analicéis la distribución de las rentabilidades obtenidas de las distintas carteras generadas. El análisis debe realizarse para el período de tiempo mencionado anteriormente, es decir 12 meses (12M) desde el 01/01/2020 hasta el 31/12/2020. Para tal análisis, debéis generar la(s) gráfica(s) y/o utilizar datos que consideréis necesarios.

Según la(s) gráfica(s) y datos obtenidos, tratar de dar respuesta a la siguiente pregunta:

Teniendo en cuenta la rentabilidad de **TODAS** las carteras generadas, **¿sería más probable obtener una rentabilidad positiva o negativa?**

Dar una respuesta justificada utilizando la(s) gráfica(s) y/o datos generados anteriormente para argumentar vuestra respuesta.

## 2.2 Tareas relacionadas de la funcionalidad extra u opcional

### 2.2.1 Métricas de carteras de inversión

**Volatilidad de carteras:** Calcular la volatilidad anual desde el 01/01/2020 hasta el 31/12/2020 de las 126 carteras generadas anteriormente. La volatilidad de las carteras se debe calcular conforme a lo descrito en la Sección 1.2.

La información generada de volatilidad se debe almacenar junto con las distintas carteras y la rentabilidad calculadas antes, y almacenarlo en un nuevo archivo con nombre `portfolio_volatility.csv` para poder utilizarlo en el posterior análisis y también para permitir la validación individual de esta subtarea.

Dicho fichero `portfolio_volatility.csv` debe tener 7 columnas y 126 filas, una fila por cada cartera generada. Las columnas deben estar en el orden:

ST,CB,PB,GO,CA,RETURN,VOLAT

Así, las 5 primeras columnas son para representar el peso de cada activo en cada cartera, y la sexta y séptima columna para representar la rentabilidad y volatilidad anual de cada cartera, respectivamente.

### 2.2.2 Análisis de las carteras de inversión generadas

**Return vs. risk:** El manager de DS también a pedido a vuestro grupo que analice la relación entre rentabilidad y riesgo para el período de 12M. El rendimiento frente al riesgo se analiza típicamente en finanzas mostrando un gráfico 2D con una medida de riesgo en el eje 0X y rendimiento (en %) en el eje 0Y. En este caso, el riesgo se cuantificará mediante la volatilidad (también en %).

Según el análisis realizado, tratar de dar respuesta a la siguiente pregunta:

¿Podríais afirmar que **SIEMPRE**, es decir para todas las carteras, **es cierto que cuanto mayor es el riesgo, mayor es la rentabilidad obtenida?**

Utilizad la(s) gráfica(s) y/o los datos generados anteriormente para argumentar vuestra respuesta a la pregunta.

## Chapter 3 Evaluación y entrega de la práctica

### 3.1 Evaluación de la práctica

Los pesos para la evaluación de la práctica son:

- **80%** - implementación y análisis de la parte obligatoria
- **20%** - implementación y análisis de la parte opcional

Respecto al código se evaluarán aspectos como el cumplimiento de la funcionalidad solicitada, así como aspectos del código como: legibilidad, buena estructuración, poner comentarios, ...

### 3.2 Entrega de la práctica

- Se hará una única entrega por grupo de prácticas.
- Las entregas se harán a través de una tarea específica para esta práctica en Moodle. Dicha tarea en Moodle se publicará más adelante.
- **La fecha límite de entrega será el domingo 19/12/2021 hasta las 23:59.**
- Cada proyecto se debe empaquetar en un fichero .zip con el siguiente contenido:
  - El código Python desarrollado por vuestro grupo de prácticas para implementar la funcionalidad requerida en la práctica. El código correspondiente a la funcionalidad obligatoria y la opcional, pueden estar en un mismo script o en scripts separados, pero se deben permitir ser ejecutadas de manera independiente.
  - Un informe dando respuesta a las preguntas de análisis de datos antes planteadas. Este informe debe incluir una respuesta breve a cada pregunta, respaldada por gráficos, tablas o cualquier otra representación de datos que consideréis relevante. Por favor, sed concisos en vuestra respuesta, teniendo en cuenta que una respuesta más larga no siempre es una mejor respuesta. El informe debe ser un archivo .pdf y cada grupo puede seleccionar libremente el formato, ya sea diapositivas o documento de texto.
  - Además, cada grupo debe proporcionar un archivo de texto plano (por ejemplo, README.txt) en el que se describa brevemente lo siguiente:
    - Descripción del contenido (archivos y carpetas) del directorio raíz de esta práctica, no es necesario describir el contenido que haya dentro de cada subdirectorio.
    - Una breve explicación de cómo ejecutar su código para cada parte (obligatoria y opcional), bien sea explicando las instrucciones para ejecutar directamente los scripts o bien a través de un fichero tipo Makefile o similar). También debéis indicar las dependencias de paquetes Python y las versiones requeridas

por vuestro código.

- Una lista de los módulos/archivos Python que habéis desarrollado, indicando brevemente una descripción de los mismos. No es necesario describir cada función implementada, simplemente una descripción general del archivo. Por ejemplo: *el fichero file1.py calcula las diferentes asignaciones de activos para las carteras.*
- Qué funcionalidad se ha implementado en el código de lo que se ha solicitado, relacionando (si es posible) los requisitos de la práctica con el código implementado. Por ejemplo, indicar: *la funcionalidad sobre el cálculo de asignación de pesos en las carteras se ha implementado en archivos file1.py, file2.py, ....*