# 5 puntos: Dataset del Titanic (dificultad baja)

|  |  |
| --- | --- |
|  | El dataset train.csv contiene datos del accidente que sufrió el Titanic en la noche del 14 de abril de 1912.  Las variables que contiene son las siguientes:   * Passengerid: Nº identificador de cada pasajero. * Survived: 1 o 0 para indicar si sobrevivió o no. * Pclass: A qué categoría pertenecía el pasajero 1ª a 3ª * Name * Sex * Age * Fare: Precio del billete en dólares. * Cabin: Cabina en la que se alojaba el pasajero. * Embarked: En qué puerto embarcó.   Utiliza estos datos para contestar a las siguientes preguntas: |

Ejercicio 1

* Carga el set de datos en una variable llamada "titanic"
* Obtén el número de filas y columnas incluidas en el set de datos.
* Obtén las primeras 10 observaciones para ver qué pinta tienen.
* Obtén el listado de variables incluidas en el set de datos y sus tipos.
* Obtén la distribución de las variables categóricas (sex, embarked, survived, pclass) del set de datos. Es decir, valores únicos / valores totales.
* Obtén los estadísticos básicos de las variables del set de datos.

Ejercicio 2

* Indica en qué variables hay valores con NA, NULL o vacíos "" y cuál es su proporción dentro de la variable.
* Imputa la media del valor de la variable en el caso de las numéricas y "No disponible" en el caso de string.
* Vuelve a comprobar la existencia de NAs y NULL para verificar que los has eliminado.

Ejercicio 3

* ¿Fallecieron más mujeres u hombres? (Cuánto en porcentaje sobre el total de su género).
* ¿Qué clase de pasajeros (primera, segunda o tercera) sobrevivió más? Quiero que el resultado de, únicamente, el porcentaje y la clase que más sobrevivió (no el porcentaje de las 3 clases).
* ¿Cuál fue la edad media y máxima de los supervivientes en cada una de las clases?

Ejercicio 4

* ¿Cuál de los puertos de embarque es el que tiene la media de precio de billete más barata? Indica solo ese puerto, no una lista con todos.
* ¿Qué correlación hay entre la longitud del nombre de un pasajero y el importe de su billete? No modifiques el nombre del pasajero para hacer el cálculo (úsalo tal y como venga).

Ejercicio 5

* Obtén los nombres de los pasajeros que no sobrevivieron y el precio de su billete está en el decil superior.
* ¿En qué cabina deberías alojarte para tener una mayor probabilidad de sobrevivir siendo hombre de entre 30 y 40 años? Indica una única cabina.

# 3 puntos: Mi primer algoritmo de inversión (dificultad moderada)

|  |  |
| --- | --- |
|  | Ibex\_data.csv contiene las cotizaciones de las empresas del Ibex 35 desde el 2004, hasta el 2019, en un dataset limpio y libre de sesgos.  Ten presente que la composición del índice no siempre ha sido la misma. Este dataset refleja la composición y cotización del Ibex, a lo largo del tiempo. Si una empresa ha entrado, salido, y vuelto a entrar en el índice, se le añade un sufijo. Por ejemplo: NHH\_1 |
|  | price\_departure.csv contiene un indicador que utilizo en diversos algoritmos de inversión. Lo necesitarás, como referencia en los ejercicios 7 y 8. |

Ejercicio 6

* Desarrolla un algoritmo de mechas

Para cada activo, cada día, compra a precio de apertura y vende cuando ocurra el primero de los siguientes eventos:

El activo sube 3 céntimos (stop profit)

El activo cae 10 céntimos (stop loss)

Si no ocurre ninguno de los anteriores, vende a precio de cierre

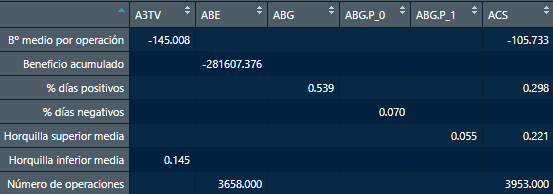
Ojo, habrá días positivos y negativos a la vez, en estos casos, supón que toca primero el stop loss

El capital que invertimos en cada activo, cada día, debe ser 30.000 €

La comisión de compra será de 0.0003 \* capital. Lo mismo para la venta.

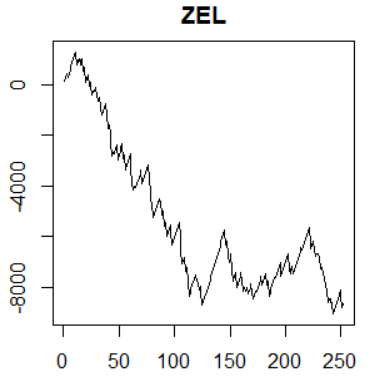
Comprueba que tienes al menos 30 datos para hacer los cálculos (si no es así descarta el activo)

Entregable: Código que genere un dataframe con la siguiente estructura (para todos los activos):



Horquilla superior media: (max – open) y horquilla inferior media (open – low)

Y grafique el beneficio acumulado por activo. Por ejemplo:



Ejercicio 7

* Partiendo del algoritmo de mechas anterior, añade el parámetro price\_departure.

Para cada activo, cada día, si el price\_departure es >= 0.75, compra a precio de apertura y vende cuando ocurra el primer de los siguientes eventos:

El activo sube 3 céntimos (stop profit)

El activo cae 10 céntimos (stop loss)

Si no ocurre ninguno de los anteriores, vende a precio de cierre

Ojo, habrá días positivos y negativos a la vez, en estos casos, supón que toca primero el stop loss

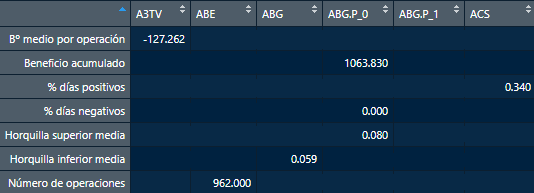
El capital que invertimos en cada activo, cada día, debe ser 30.000 €

La comisión de cada compra y venta será de 0.0003 \* capital

Homogeneiza los datos Ibex\_data y price\_departure (utiliza solo las fechas que existan en ambos DF)

Comprueba que tienes al menos 30 datos para hacer los cálculos, antes de aplicar el filtro del price\_departure (si no es así descarta el activo)

Entregable: Código que genere el mismo dataframe que en el ejercicio anterior y los mismos gráficos.



Ejercicio 8

* Optimización de la mecha y el capital por activo

Utilizar un stop profit, o un stop loss estático, no parece lo más eficiente.

Tampoco lo parece el utilizar el mismo capital para todos los activos.

Objetivo: Partiendo del algoritmo de mechas anterior

* + Modifica el capital asignado a cada activo: usa la media de datos de cierre y el 0,5% del volumen.
  + Modifica el stop profit de cada activo: utiliza el cuantil 30 de la mecha superior (max – open)
  + Modifica el stop loss de cada activo: utiliza el cuantil 80 de la mecha inferior (open – low)

Entregable: Código que genere un dataframe con la siguiente estructura (para todos los activos), y los mismos gráficos.



# 2 puntos: Más allá de lo visto en clase (dificultad elevada)

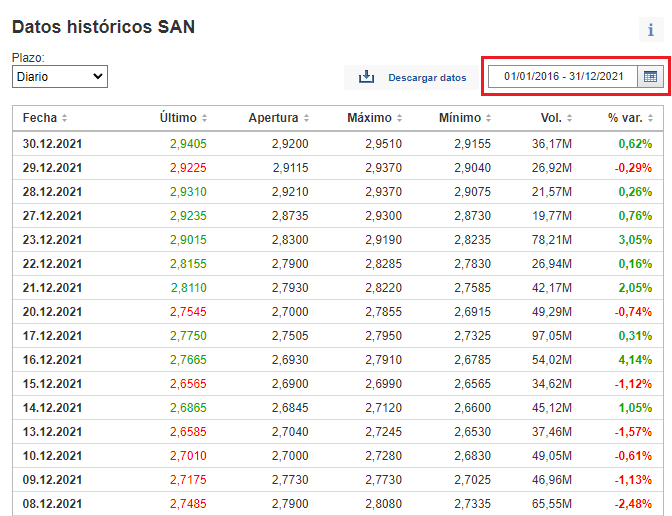
Ejercicio 9

El objetivo de este ejercicio es hacer web scraping de Investing.com

Concretamente, queremos obtener los datos de cotización diarios del Banco Santander, desde el 1 de enero de 2016, hasta el 31 de diciembre de 2021. La url de donde obtendremos la información será: <https://es.investing.com/equities/banco-santander-historical-data>

La dificultad de este ejercicio radica en el que la web no presenta directamente esta información. Únicamente podemos ver los 30 últimos días de cotización.

Para poder hacer el web scraping, deberemos modificar la fecha de consulta (recuadro en rojo), para que nos aparezca la información que queremos obtener.



Objetivo del ejercicio: programa una función que obtenga, en un dataframe, la información del Banco Santander de Investing.com. Todo el proceso debe estar programado, no puede haber ninguna interacción manual.

La función deberá recibir como input dos parámetros: fecha de inicio, fecha de fin.

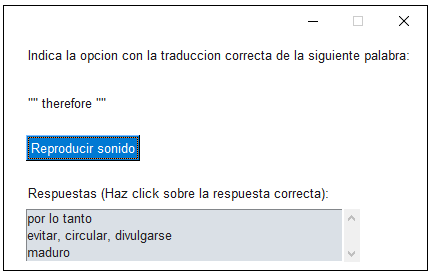
No se podrá hacer uso del API de Investing, ni de ninguna otra librería que haga el trabajo por vosotros. El proceso de web scraping debe estar programado por vuestra parte.

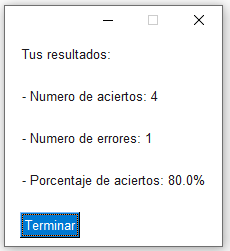
Ejercicio 10

El Excel Palabras.xlsx contiene palabras en inglés y su traducción

EL objetivo del ejercicio es hacer un “juego” en donde se elija al azar una palabra en inglés y se dé al usuario 3 posibles traducciones en español (y el juego contrario).

Una posible visualización del juego, sería la siguiente:





Contenidos obligatorios de la aplicación

* Número de rondas del juego: por defecto 10 (pero el usuario podrá elegirlo al inicio).
* El usuario podrá elegir si quiere que le aparezcan palabras en inglés y opciones en español (o al contrario).
* La palabra que aparezca deberá escucharse por voz al aparecer en pantalla (las opciones no).
* Se le mostrarán al usuario 3 opciones de traducción, debiendo elegir una de ellas antes de pasar a la siguiente ronda.
* Al finalizar el juego se mostrará al usuario su porcentaje de aciertos en el juego.

# Entregables y normas aplicables

* La práctica se realizará de manera individual.
* Se deberá entregar el script de R. Código suelto o en cualquier formato que no sea un script de R no será evaluado.
* La entrega se realizará en la plataforma de Instituto BME antes de que acabe el plazo establecido. Cualquier entrega fuera de la plataforma no será evaluada. Recomendación: no esperes a la última hora para entregar la práctica (si internet falla en el último momento echarás a perder 3 semanas de trabajo).
* El nombre del fichero deberá ser nombre\_1erapellido.zip (no hacerlo restará un punto en la evaluación).
* Los ejercicios deberán estar claramente diferenciados por su numeración y se incluirá el enunciado en la solución entregada. No hacerlo restará 1 punto en la evaluación.

Recomendación: Aunque disponéis de 3 semanas de plazo, no os pongáis a trabajar dentro de 2 semanas porque no os dará tiempo.

Advertencia: Si observo dos códigos muy similares, ambos alumnos defenderán su trabajo en un examen oral, para determinar la autoría de los mismos.