

Instrucciones

- Tendrás que hacer un commit cuando te lo indique el profesor con el mensaje **checkpoint** y un número.
- Al final del examen harás un push en tu repositorio remoto.
- En la entrega del examen, tendrás que copiar el link de tu repositorio.
- Ante dos soluciones que aparenten ser copia, se calificarán a los implicados con un 0. En determinados casos, se podrá realizar una entrevista personal, pudiendo preguntar problemas diferentes.
- Se puede consultar código descargado así como los colabs, pero no se puede acceder a páginas donde se pueda compartir código, como Github, Google Chat, Whatsapp, así como otras páginas que puedan dar soluciones, como Chatgpt.

ANTES QUE NADA: Descarga el fichero `winequality.csv` y cópialo en el directorio donde tienes el repositorio.

1. (1.5 puntos). Crea un fichero `functions.py` donde implementarás y probarás las siguientes funciones y otro fichero `main.py` que tendrá un código principal donde irá llamando por orden a todas estas funciones. En caso de que alguna emita una excepción, ésta se capturará en este código principal y mostrará el mensaje por pantalla «Ha ocurrido la excepción XXX», donde XXX es el tipo de excepción devuelto. Con que haya una llamada correcta por cada función junto con el tratamiento de excepciones, será suficiente.
2. (2 puntos). Crea una función `read_data` que recibe el nombre de un fichero csv con muestras de vino (pruébalo con el fichero `winequality.csv` o con un csv similar) y devuelve un diccionario con el formato que aparece a continuación. A modo de ejemplo, se muestra cómo deberá devolver la información de la primera muestra del fichero (primera fila), siendo `dato1` una clave que se irá incrementando (es decir, `dato1`, `dato2`, `dato3`, etc.). Si hay alguna muestra que tiene algún atributo vacío, la muestra no se insertará en el diccionario. Por ejemplo, la segunda muestra del fichero no tiene valor en el atributo (`volatile acidity`) y por tanto, esta muestra (esta fila entera) no debería aparecer en el fichero devuelto. Si el fichero tiene menos de 10 líneas con valor en todos los atributos, se emitirá un error de tipo `ValueError`.

```
{
    'dato1': {'type': 'white',
              'fixed acidity': '7',
              'volatile acidity': '0.27',
              'citric acid': '0.36',
              'residual sugar': '20.7',
              'chlorides': '0.045',
              'free sulfur dioxide': '45',
              'total sulfur dioxide': '170',
              'density': '1.001',
              'PH': '3',
              'sulphates': '0.45',
              'alcohol': '8.8',
```

```

        'quality': '6'
    },
    ...
}
```

3. (2 puntos). Crea una función `split` que recibe un diccionario como el que devuelve el ejercicio anterior y devuelve dos diccionarios. El primero es un diccionario con las muestras que tengan el valor `white` en el atributo `type` y el segundo es un diccionario con los datos que tengan el valor `red` en este atributo. El atributo `type` se eliminará de cada dato en estos diccionarios devueltos.
4. (2 puntos). Crea una función `reduce` que recibe un diccionario con el formato de los que devuelve el ejercicio anterior y un string que corresponde al nombre de un atributo. Esta función devuelve una lista con los valores de ese atributo. Si el atributo que se le pasa no existe en el diccionario, se emitirá un error de tipo `ValueError`. A modo de ejemplo, si le pasamos un diccionario con los datos que son de tipo `white` y con el atributo `alcohol`, la lista devuelta será:

8.8	10.1	9.9	etc.
-----	------	-----	------

5. (2.5 puntos). Crea una función `silhouette` que recibe dos listas como la que devuelve el ejercicio anterior y devuelve el coeficiente de Silhouette de la primera de las listas. Este coeficiente se calcula siguiendo la siguiente fórmula:

$$Silhouette(lista) = media(S(i))$$

, donde $S(i)$ es un coeficiente para cada uno de los datos i de esta lista y que se calcula con la siguiente fórmula:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

, donde:

- $a(i)$ es la distancia media entre i y el resto de valores de su lista, donde la distancia entre i y un $j \neq i$ se calcularía como $\sqrt{|i - j|^2}$.
- $b(i)$ es la distancia media entre i y todos los valores de la otra lista, calculada de la misma manera.

Tiempo estimado: 120 minutos

ENTREGA: Los ficheros y la url de tu repositorio.