APARTADO 4

# ¿Por qué Python?

La razón es la sencillez y capacidad de este lenguaje para el análisis y tratamiento de los datos gracias a librerías como Pandas o Numpy. Base gran parte de la metodología de análisis usada en este proyecto en el libro: LIBRO.

Las distintas posibilidades que se barajaron antes de empezar fueron Java, Android, R y Python. Al final me decanté por Python por diferentes motivos, el primero en ser descartado fue Java, tras un estudio inicial sobre lo que quería hacer y el como hacerlo, me di cuenta que necesitaba un lenguaje potente que me dejase tratar los datos con claridad para su análisis, y después de leer varios artículos como: REF\_ARTICULOS, me percaté de que Java no era la mejor opción, estos artículos siempre te orientaban hacia Python y R. Entonces, ¿Por qué Android?, por la sencilla razón de que este trabajo está desarrollado para unir el mundo de la dietoterapia, y las ciencias de la Salud con la informática para hacer que el usuario tenga un fácil aprendizaje de dicha metodología. A mi parecer la forma más clara y rápida de llevar al usuario dicha tecnología es a través de su SmartPhone, pero se acabó descartando debido a la falta de conocimientos sobre sistemas Android. A estas alturas ya solo me quedaba elegir entre Python y R, tras indagar superficialmente sobre ambos lenguajes para el análisis y el tratamiento de datos, se llego a la conclusión que ambos lenguajes tienen una forma de trabajar muy similares, y entonces la decisión fue clara, debido a que he trabajado en numerosas ocasiones con Python y la sintaxis de R para mi era desconocida, al final, me decante por Python.

# Metodología

## Introducción

En este subapartado explicaremos el cómo y porque se ha tratado los datos en este TFG además de los diferentes cálculos internos que se realizan para el sistema de recomendaciones, cálculos, etc.

## Excel y Pandas

Se usará la herramienta de Microsoft, Excel, para trabajar como una base de datos, y poder tratarlos en forma de DataFrame, esta particularidad, nos la da la librería de Pandas. Hay un Excel para la base de datos genérica, y otro para la base de históricos. Se leen los datos automáticamente en cuanto el usuario entra en la aplicación, pero solo se guarda si el usuario así lo desea.

## DataFrame

Se usan los DataFrame, para llevar un registro de todos los datos que el programa necesita, se tiene en cuenta tanto las bases de datos de los alimentos, usuarios y comidas como lo que el usuario lleva en el día.

Los alimentos se tratan en el momento en el que se registra el usuario, de manera que, se separan las comidas de la lista principal, creando 5 listas (Desayuno, Merienda, Comida, Almuerzo y Cena), se tratan por separado y se filtran por el umbral, el umbral es un medidor de calidad de la comida, para qué solo te muestre lo más sano en un principio, y se van actualizando sensibles a los cambios de la aplicación. Si se refresca la página aumenta el umbral para que sea mas amplia las posibilidades de comida, aunque esto signifique una peor alimentación. Con cada elección se actualizan el resto de las comidas, para que se ajuste al máximo la dieta a sus necesidades, se ordena por la formula que se hablará mas adelante, se crea la variable “dif” que sale de esta fórmula, y este es el eje principal de las recomendaciones, cada vez que se elige una comida, la comida se almacena para poder tratarla mas tarde, en cuanto se selecciona los botones se desactivan, y se cambia el botón seleccionar a editar, para que se pueda editar en cualquier momento. Cada vez que se elige una comida se reparte las calorías sobrantes/que faltan, y se recalculan las posibilidades.

# Librerías

## Pandas

La librería desarrollada por Wes McKinney, Pandas, es una librería usada para el tratamiento de los datos como estructuras. No fue la primera opción para este proyecto, debido a que este proyecto se enfoca en el análisis de datos, y no en el tratamiento de bases de datos, se decidió crea una base de datos local en un archivo Excel, la primera opción fue la librería “openpyxl”, la cual, es una librería de código abierto, que permite la carga y manejo de datos XLS, el problema vino que te devolvía objetos desestructurados difíciles de tratar, y complicaba él objetivo principal del proyecto, acto seguido basándome en él libro: LIBRO, decidí probar con Pandas, esta librería me permitía leer los datos de los archivos .XLS, en forma de dataframes. Esto simplificaba el análisis de los datos, permitiendo tratarles tanto como dataFrame como en forma de numpy.array, dicha librería será analizada a continuación.

## Numpy

Como ya he nombrado en varias ocasiones en este memoria, el libro: LIBROO, se basa principalmente en esta librería, como añadido, es la librería sobre la que he trabajado a lo largo de la carrera en cuanto a análisis de datos en Python, y la mas recomendada por los usuarios en la web, sobre esta librería se sustenta principalmente este proyecto, siendo la encargada, del tratamiento, procesamiento y calculo, de todos los datos que internamente realiza el programa.

## Interfaz gráfica

La librería utilizada para realizar y diseñar la interfaz gráfica fue Tkinter, tras buscar en diferentes páginas que encontraremos en la bibliografía, me decante por Tkinter debido a la falta de conocimientos en interfaces gráficas con Python, me decante por la mas fácil de aprender. Entre las opciones que baraje se encontraban: Tkinter, WxPython, PyQt y PyGTK.

Tkinter traía una serie de ventajas, entre ellas que viene preinstalada con Python, era fácil de aprender, y hay una documentación amplia y extensa. Pero también tiene sus desventajas, incluye pocos elementos gráficos, tiene un limitado control y la navegabilidad mas sencilla se hace complicada, y sobre todo lo que mas noto es su lentitud, cuanto mas elementos añadas, mas lento va la interfaz, no tiene alguna especie de cache o memoria que guarde lo que has “dibujado”, sino que los dibuja cada vez que salen en pantalla, lo hace sobre cada botón etiqueta, etc.

La segunda opción que se me paso por la cabeza, fue WxPython, tenia grandes ventajas, como la rapidez, la flexibilidad que este ofrecía y sobre todo su mayor ventaja son todas las opciones que tiene para crear una interfaz gráfica compleja y “profesional”, tras barajar, estas opciones y compararlas con las de Tkinter, me di cuenta que todo lo que te ofrecía WxPython, era innecesario para la interfaz tan simplista que necesitaba para mi proyecto, y el aprendizaje era más complejo, además de ser mas complicado encontrar información o documentación sobre esta librería. El mayor problema de esta librería era que tiene una comunidad muy activa, la cual, esta constantemente insertando cambios, esto era un problema para proyectos largos, pues creaba problemas de compatibilidad, pero para este proyecto, no era algo que me resultará un inconveniente.

El resto de librerías fueron descartadas, al poco de buscar información sobre ellas debido a que daban las mismas ventajas o similares que la WxPython, pero tenían mas inconvenientes, al menos para el tema que aborda este proyecto.

# Módulos

## CalculosDieta

### Introducción

Es el modulo principal sobre el que se sustenta el programa, hace todos los cálculos de recomendación, kilocalorías, repartos, etc.

En el manual del programador se extenderá mas sobre cada función, a continuación, se hará un breve comentario sobre cada función en base a la metodología.

### CalculoTMB, rapartoDeKcal y distribuciónDeMacronutrientes

Estas tres funciones trabajan de manera paralela, el **calculoTMB**, lo que hace es calcular el gasto calórico basal, de aquí se saca el objetivo diario, una vez calculada la cantidad de Kilocalorías diarias, se reparten las kcal (**repartoDeKcal** )según sea desayuno, almuerzo, merienda o cena, se reparten en base a unos porcentajes calculados en base a distintas fuentes de información encontrados a lo largo del estudio sobre el que se sustenta este proyecto, no son porcentajes exactos, Se distribuyen de la siguiente manera:

* Desayuno: 24,75% -- Alta carga calórica, es una de las tres comidas principales y la carga en la cual el cuerpo absorbe mejor los nutrientes.
* Almuerzo: 13,5% --Baja carga calórica, pero mayor que la merienda debido a la disposición del metabolismo por la mañana.
* Comida: 30,5% --Mayor carga calórica, por ser el eje central del día, y pese a los mitos que corren sobre el desayuno, es posiblemente la comida más importante del día.
* Merienda: 11,5% --Baja carga calórica, por ser una comida de paso y prescindible.
* Cena: 19,75% --Alta carga calórica pero con moderación, es una de las tres comidas principales, pero ocurre cuando el metabolismo esta mas ralentizado, por ello es la mas baja de las tres principales

A la par se calcula el número de Kilocalorías que se tiene que tomar de hidratos, grasas y proteínas, en base al tipo de dieta recomendable para tu patología en la función **distribuciónDeMacronutrientes**, Esta función toma como parámetros las kilocalorías diarias calculadas previamente por la función calculoTMB, y el tipo de dieta del usuario en base a su patología, y te devuelve una lista de los macronutrientes diarios, repartidos en gramos y las kilocalorías totales en el siguiente orden: ListMacDiarios [Kilocalorías, Hidratos, Proteínas, Grasas ], es importante saber que se dividen hidratos, proteínas y grasas, por 4,4 y 8 (respectivamente), para calcular los gramos, y luego de nuevo entre otro número que es la proporción que van a tener (una proporción normal, de una dieta normal sería: Hidrados:50%,proteínas:25%, grasas:25%.

## OrdMinimaDiferencia, formulDif, repartoKcalSobrantes

Son tres funciones que coexisten de alguna manera turnándose las interacciones con el programa. En verdad OrdMinimaDiferencia y formulDif, se usan juntas y se separaron para mayor claridad en caso de problemas, o posibles mejoras. Y repartoKcalSobrantes funciona independientemente, pero siempre de manera que cuando se llama a OrdMinimaDiferencia se llama a repartoKcalSobrantes.

**rapartoKcalSobrantes,** Recoge como parámetros la lista de kilocalorías por comida, la comida en la que nos encontramos y las kilocalorías totales. Se calcula de tal manera que según la comida en la que te encuentres, compara las kilocalorías que has comido con las que deberías haber comida y distribuye lo que sobre/falte entre el resto de las comidas.

Mientras que **OrdMinimaDiferencia** y **formulDif**, son el pilar base de esta aplicación, en cuanto al calculo de comidas se refiere, OrdMinimaDiferencia, coge como parámetros la lista de las comidas, el objetivo, el tipo de comida, lo que ya llevo comido y las kilocalorías diarias. En base a esto hace un calculo de lo que debería llevar de cada macronutriente concreto en cada comida concreta en el total del día(Siendo esto el sumatorio de las comidas anteriores), lo que debería llevar de cada macronutriente específico en esa comida específica, luego se recorre todas las comidas de las bases de datos calculando cual es la que mas se ajusta a lo que necesitamos, una vez cogemos la comida le pasamos todo lo calculado anteriormente, y se lo pasamos a **formulDif**, para llevar a cabo la fórmula, que consiste en:

INSERTAR FORMULA

Si los macronutrientes que debería llevar son mas de lo que en realidad llevo, sacamos la diferencia en base al macronutriente concreto, resto lo que debería llevar a lo que llevo, y dividiéndolo, entre los carbohidratos de la comida en la que nos encontramos, menos los datos del alimento que analizamos y más la resta de las kilocalorías que llevo menos las kilocalorías del alimento que se analiza.

De esta manera el resultado será mayor cuando la ausencia de ese macronutriente sea mayor, y por el lado del denominador, si las kilocalorías y macronutrientes del menú se ajustan a las necesidades del usuario, menos será este y mayor el resultado final. De esta manera se crea el mayor ajuste posible para el usuario.

Se suma el resultado de los 3 macronutrientes y se divide entre tres, creando así un valor por el que posteriormente se ordenará la lista provocando que las primeras recomendaciones sean las más adecuadas.

Cada vez que se elige una comida se hace el reparto de kilocalorías sobrantes y se recalcula el nuevo valor “diferencia”, en base a lo que el usuario a elegido.

## AdminBase y estructura de la base de datos

### Introducción

En este apartado hablaremos de como esta estructurada la base de datos, y de como se cargan, se guardan y se trabajan con los datos desde el módulo AdminBase, el cual, es un modulo bastante simple, necesario para el trato de datos.

### CargaBaseDeDatos y guardaDatos

Estas dos funciones cargan y guardan los datos de/en la base de datos respectivamente, como se decidió simular una base de datos con un Excel, se usa la librería Pandas, mencionada antes en este apartado para cargar las bases de datos en los respectivos arrays, nada mas arrancar el programa, te guarda las tres bases de datos principales, Alimentos, usuarios y patrologías en tres vectores independientes, que se trabajaran con cada uno de ellos sin tener en cuenta al resto y se guarda el resultado de los cambios producidos durante la ejecución del programa.

La base de datos se estructura en dos archivos Excel independientes, donde cada hoja seria lo equivalente a una colección de una base de datos, y cada fila seria un objeto dentro de dicha colección. A distinguir las siguientes hojas:

#### Alimentos

La hoja de alimentos, en verdad, sería una hoja de menús ya construidos los cuales tienen 10 campos, sobre los que actualmente se trabajan 8,pero se mantiene su estructura para posibles ampliaciones.

La clave primaria de esta colección sería el nombre, en un futuro esta pensado añadir un id, o depende la metodología de desarrollo de la base de datos, ese id se generaría de manera automática teniendo en cuenta que no sigue ningún tipo de normal en relación con el alimento.

Los campos son:



Nombre (PK), Calorias, Grasa, Saturadas (Sin uso), Hidratos, Azucares ( Sin uso), Proteína, Tipo, LRE, Calidad

De los primeros siete campos no hay mucho mas que añadir, que lo que el propio nombre indica, no obstante, se hará una pequeña aclaración de los últimos tres.

El **tipo**, marca el tipo de comida que es, se pensó en que metodología usar, teniendo en cuanta que un mismo menú puede ser a la vez, desayuno, almuerzo y merienda, cena y comida, y un largo etc. Por ello se pensaron dos posibilidades, la primera poner la inicial, letra o nombre de la comida para la que valga ese menú, tratarlo como un array y descomponerlo en su debido momento, pero en el futuro caso de que se diera la posibilidad al usuario de editar un alimento de manera que no le apareciese entre una comida específica se complicaba algo más, y la segunda opción era tratarlo como una cadena de bits: desayuno-almuerzo-comida-merienda-cena, donde el valor 1 sería si es valido para esa comida y 0 si no lo es. Por ejemplo:

1-1-0-1-0

Significa que es desayuno, almuerzo y merienda, pero no comida o cena, luego esto se traduce en el número decimal, que en este caso sería 26.

El **LRE**, recibe ese nombre por un juego de palabras con la gestión de procesos de un sistema operativo LRU (last reciently used), pues lo que hace es llevar una métrica de la frecuencia con la que el usuario toma esa opción, literalmente significa “last reciently eat”.

La **Calidad**, como el propio nombre indica, es la calidad del alimento, es la variable de indicar la diferencia entre 100 kilocalorías de ensalada y 100 kilocalorías de azúcar, tiene un rango de valores entre el uno y el cuatro, siendo cuatro la peor calidad de todos, y uno por el contrario la mejor, la calidad va estrechamente relacionada con el umbral del momento y la comida dentro del programa, esta variable, empieza con el valor uno (siendo este el mas bajo posible) y según vas refrescando va aumentando para que las posibilidades de menú aumenten con el número de veces que refresques la comida. El umbral sirve para cribar del DataFrame de la comida que estemos trabajando, todos los alimentos cuya calidad sea superior al valor del umbral en el momento de selección, de esta manera se le complicará al usuario la posibilidad de hacer una mala elección a la hora de organizar su menú, además que se verá reflejado en el momento en la barra de progresión.

#### Usuarios

La hoja de usuarios, alberga los datos de todos los usuarios que usan la aplicación la estructura de esta colección en si es muy básica y contiene toda la información necesaria para el calculo de la dieta, del usuario.

La clave primaria seria el DNI del usuario, el cual, se encuentra dentro del campo ID. La hoja se estructura de la siguiente manera:

Id (PK)-nombre-apellido-password-sexo-edad-altura-peso-actividad-patología (FK de la tabla Patologías) y tipo.

Tanto **id**, como **nombre**, **apellido** y **password**, son los datos privados del usuario, los cuales, son simplemente informativos y no tienen ningún valor adicional en el calculo de la dieta y los resultados. En cambio, el **sexo**, la **edad**, **altura**, **peso**, **actividad**, **patología** y **tipo**, influyen de manera directa con el calculo de la dieta, como pasaba en la hoja de alimentos en los primeros campos su nombre, explica su significado. A tener en cuenta, que el valor del campo **patología** es numérico para mayor privacidad y conexión con la base de datos de Patologías, pues es el id de cada patología y en base a su id se busca la información sobre dicha fila. En el caso de valer -1, significa que el usuario no tiene ningún tipo de patología y su uso es exclusivamente para el aprendizaje y seguimiento de la dietoterapia adecuada. La **actividad** tiene valores entre 0 y 4, donde cero significa el máximo nivel de sedentarismo y 4 el máximo nivel de actividad. Por ultimo el **tipo**, hace referencia, a los objetivos que el usuario tiene respecto a si mismo, si quiere mantenerse, subir de peso, o bajar.