



UNIVERSITAT DE
BARCELONA

Trabajo fin de grado

GRADO DE INGENIERÍA
INFORMÁTICA

Facultad de Matemáticas e Informática
Universidad de Barcelona

ANÁLISIS DE DATOS PARA MEJORAR LOS HÁBITOS DEL SUEÑO

Autor: Ainhoa Iragorri Aizpuru

Directora: Carme Zambrana Seguí

Codirectora: Eloisa Vargiu

Realizado en: Departamento de
Matemàtiques i Informàtica (UB) y
eHealth (Eurecat)

Barcelona, 27 de junio de 2019

Resumen

Para vivir más, ser más saludables y más activos, es necesario vivir un estilo de vida saludable incluyendo: una dieta saludable, actividad física, sueño reparador, evitar tóxicos y tener una conciencia plena. El proyecto CarpeDiem tiene como objetivo ayudar a las personas que quieran llevar un estilo de vida saludable enfocándose en cuatro pilares: dieta, composición corporal, actividad física y sueño. Este trabajo está enmarcado dentro de la primera fase del proyecto donde se validan las variables relacionadas con cada pilar. En particular, se estudia el pilar del sueño. Las variables estudiadas son la estación del año y la actividad física. Como resultados de dicho análisis se han visto pequeñas diferencias entre estaciones para las variables del tiempo total de sueño y el tiempo despierto durante la noche. La variable de actividad física no ha mostrado ninguna correlación con las variables del sueño estudiadas.

Agradecimientos

Ante todo, me gustaría expresar mi sincera gratitud a mis supervisoras, Carme Zambrana Seguí y Eloisa Vargiu, por su inmenso conocimiento, orientación profesional y apoyo valioso. Estoy realmente agradecida con ella y con los investigadores de Eurecat por darme la oportunidad de analizar este conjunto de datos.

Además de mis supervisoras, me gustaría agradecer a todo el departamento de e-health de Eurecat por su apoyo y amabilidad que han hecho estos meses más llevaderos.

También a mi familia por todo el apoyo y a los amigos.

Índice

1. Introducción	1
1.1. Motivación	1
1.2. Proyecto CarpeDiem	2
1.3. Objetivo del TFG	4
2. Análisis Exploratorio	6
2.1. Motivación	6
2.2. Dataset	6
2.3. Resultados	8
2.4. Conclusiones	13
3. Materiales y Métodos	19
3.1. Materiales	19
3.2. Métodos	23
4. Resultados	25
4.1. Estación	25
4.2. Actividad Física (Pasos)	28
5. Conclusiones y Trabajo Futuro	34

1. Introducción

Este apartado introduce la motivación por la cuál surge este trabajo, el proyecto en el cual se ha desarrollado, así como, el objetivo concreto del mismo.

1.1. Motivación

Las Enfermedades No Transmisibles (ENT), como las enfermedades cardiovasculares, la diabetes, el cáncer y las enfermedades respiratorias crónicas, son la principal causa de muerte en el mundo entero, causando el 70 % de las muertes. Más del 40 % de ellas son muertes prematuras ocurridas antes de los 70 años de edad [1]. En España, se calcula que las ENT son la causa del 91 % de todas las muertes. En el 2016, el riesgo por mortalidad prematura en España era del 15 % en los hombre y del 5 % en las mujeres [2].

La preocupación por este tipo de enfermedades es tan alta que la Organización Mundial de la Salud (OMS) puso en marcha un *Plan de Acción 2008-2013 para la Estrategia mundial para la prevención y el control de las enfermedades no transmisibles* [3] el cual fue seguido por el *Plan de Acción Global para la prevención y el control de las enfermedades no transmisibles 2013-2020* [4]. En estos dos documentos se recogen 9 metas globales y una hoja de ruta de como conseguirlas.

Varias de estas metas están enfocadas a reducir los factores de riesgo relacionados con los comportamientos modificables como la inactividad física, el uso nocivo del alcohol, el consumo de tabaco y las dietas malsanas (con un alto consumo de comidas preprocesadas y excesivo consumo de sal). Estos factores conducen al sobrepeso y la obesidad y aumentan la presión arterial y el colesterol lo que a su vez aumentan el riesgo de las ENT [1].

En España:

- Se estima que un 26,8 % de la población tiene una actividad física insuficiente [5].
- Durante el 2016 el consumo medio de alcohol per cápita fue de 8,58l, de los cuales la mitad son cervezas y hasta un cuarto son destilados [6].
- La prevalencia del consumo de tabaco ha bajado durante los últimos años de 45,2 a 31,3 en los hombre y 32,3 a 27,1 en las mujeres, entre los años 2000 y 2015 [7].
- Se estima que el consumo de sal es de 9,7g por día. Entre el 70 – 75 % de la sal consumida proviene de comidas procesadas [8].

Además de los factores expuestos por la OMS, la Sociedad Española de Sueño¹ destaca el sueño no saludable como otro factor de riesgo de las ENT. Solo un tercio de la población española duerme las horas recomendadas, un 32 % se despierta con

¹<http://ses.org.es/>

la sensación de no haber descansado y un 35 % se encuentra cansado durante el día [9].

Por consiguiente, para vivir más, ser más saludables y más activos, las personas de cualquier edad deben seguir sugerencias simples y claras que promuevan un estilo de vida saludable incluyendo: una dieta saludable, actividad física, sueño reparador, evitar tóxicos y tener una conciencia plena.

1.2. Proyecto CarpeDiem

El proyecto CarpeDiem tiene como principal objetivo crear un sistema de soporte inteligente y automático para personas que quieran llevar un estilo de vida saludable, permitiendo al usuario gestionar diferentes aspectos denominados pilares: dieta, composición corporal, actividad física y sueño.

El sistema CarpeDiem tendrá una perspectiva “holística” teniendo en cuenta los cuatro pilares en su conjunto. Al hacerlo, se definirá un enfoque novedoso con el objetivo final de implementar un sistema personalizado para cada tipo de usuario, considerando los hábitos de trabajo y de la vida diaria, la edad, el género, el estado de salud y sus preferencias. También se tendrán en cuenta factores ambientales como la ciudad donde habita, la zona y las condiciones climáticas.

Una aplicación móvil será el puente entre el sistema CarpeDiem y el usuario. Sus funciones principales serán: recoger todos los datos, ya sea directamente a través de la app o bien desde otros dispositivos, y enviarlos al sistema a través de internet; y mostrar al usuario las recomendaciones seleccionadas por el sistema.

El sistema de recomendación integral será una confederación de sistemas de recomendación donde cada uno estará especializado en un solo pilar. Este enfoque de *divide y vencerás* se seguirá para permitir la escalabilidad y el mantenimiento del sistema. Cada sistema incorporará un modelo de aprendizaje automático capacitado para abordar el tipo específico de cambio de comportamiento requerido.

En el proyecto CarpeDiem participan diferentes instituciones especializadas en los pilares estudiados. Uno de sus principales cometidos es especificar las variables a estudiar en sus correspondientes pilares. El *Institut de Recerca Biomèdica de Lleida* (IRBLL) se encarga de dar soporte en el pilar del sueño. Los expertos del departamento de *Nutrició i Salut* de Eurecat ayudan en los pilares de dieta y composición corporal.

Dieta

Para poder determinar si una dieta es saludable o no, hace falta mirarla desde un punto de vista nutricional. No obstante, las variables que permiten identificar una dieta saludable no solo están relacionadas con los alimentos consumidos y sus nutrientes, sino que también tiene en cuenta variables antropométricas como el gasto energético total, la tasa metabólica basal, el gasto energético relacionado con la actividad física y el índice de masa corporal.

La monitorización del primer grupo de variables, alimentos ingeridos y sus nutrientes, se hará a través de diferentes sistemas: un sistema inteligente capaz de detectar los ingredientes del plato a partir de una foto, un lector de código de barras asociado a los datos de productos, una espátula inteligente que detecta la cantidad de sal (Sodio y Potasio) y un vaso de agua inteligente que mide la cantidad de líquidos consumidos.

Las variables antropométricas se adquirirán a través de dispositivos como una báscula inteligente (p.e., Aria de Fitbit) y una pulsera (p.e., Charge 2 de Fitbit) que permite calcular el gasto energético relacionado con la actividad física.

Una vez recogidas todas las variables el sistema estará en disposición de recomendar nutrientes y/o grupos de alimentos que ayudarán al usuario a mejorar su dieta.

Composición corporal

Más allá de un peso y un índice de masa corporal adecuados, hay otras variables sobre la composición corporal que también se tiene que tener en cuenta a la hora de tener una vida saludable son: masa muscular, masa osea, cantidad de grasa y cantidad de agua.

Todas estas variables se medirán a través de una báscula inteligente (p.e., Aria de Fitbit). Estas básculas se basan en la impedancia bioeléctrica que mide la resistencia que un cuerpo opone al paso de una corriente, de baja frecuencia (50 kHz). En función de esta resistencia se puede determinar la cantidad de grasa que puede haber, ya que ese valor relaciona de manera aproximada la proporción de grasa y músculo que está presente.

Con estos datos el sistema orientará al usuario hacia unos valores saludables y le dará recomendaciones de como conseguirlos.

Actividad física

Se podría decir que el pilar de actividad física es el más estudiado. La OMS publicó en 2010 unas recomendaciones generales de actividad física para los diferentes rangos de edad. Para una vida sana recomienda al menos 60 minutos de actividad moderada y/o vigorosa al día para niños y adolescentes menores de 17 años y al menos 150 minutos a la semana de actividad moderada para adultos (mayores de 17 años) [10].

La actividad física se monitorizará a través de una pulsera (p.e., Charge 2 de Fitbit) que proporciona información sobre pasos, tipo y nivel de actividad física (sedentaria, ligera, moderada y vigorosa), y pulso.

Con estas variables el sistema no solo alentará a los usuarios para que consigan las recomendaciones generales de actividad física sino que las personalizará para cada uno de ellos. Además, motivará y aconsejará actividades a los usuarios para conseguir dichas metas.

Actividad del sueño

Un sueño saludable no solo consiste en dormir las horas recomendadas, que varían dependiendo de la edad, sino que también incluye otros aspectos como: la satisfacción con el sueño; estar alerta durante todo el día; que el sueño sea eficiente, es decir, que nuestro sueño no se vea interrumpido; y que el periodo de sueño sea por la noche.

Hay varios factores que pueden influir en nuestro sueño. Estos factores se pueden dividir en tres grupos de variables: extrínsecas, aquellas relacionadas con factores externos como la calidad del aire, la contaminación acústica, la temperatura y luminosidad, la estación del año, las horas de sol diarias, la humedad y la comodidad de nuestra cama; intrínsecas, las propias de la persona: edad, género y grado de obesidad; y las relacionadas con los hábitos de la vida como el uso de dispositivos móviles antes de acostarse, el tipo de dieta y el tipo de alimentos tomados antes de ir a dormir, fumar, la actividad física y el consumo de cafeína.

Según los clínicos el sueño de la población general se puede medir a través de las siguientes variables: hora de acostarse, hora de levantarse, tiempo dormido, tiempo despierto durante el período de sueño y tiempo dormido durante el día.

Las variables propias del sueño se monitorizarán a través de una pulsera (p.e., Charge 2 de Fitbit) que mide la cantidad y eficiencia del sueño. Las variables extrínsecas se obtendrán a través de fuentes externas (p.e. API), las variables intrínsecas se obtendrán directamente del usuario y las de los hábitos serán reportadas por los usuarios a través de cuestionarios.

El sistema se basará en todas estas variables para dar recomendaciones al usuario para tener un sueño saludable. Estas recomendaciones estarán basadas en cambios de estilo de vida.

Fases del proyecto

El proyecto se divide en tres fases. En la primera se validaran todas las variables relacionadas con los diferentes pilares a través de datos históricos. Así, se verificarán cuales van a ser utilizadas en los modelos finales. En la segunda fases los modelos correspondientes a cada pilar serán entrenados con datos de voluntarios. Finalmente, el sistema de recomendación final “holístico” se probará en un estudio de caso específico que se definirá durante la vida del estudio.

CarpeDiem se inició en Enero del 2019. Durante este medio año se ha desarrollado la primera fase, donde se ubica este TFG, y se ha empezado a recolectar los datos de voluntarios para la segunda fase.

1.3. Objetivo del TFG

Como se menciona en el apartado anterior, el presente trabajo se ha desarrollado durante la primera fase del proyecto CarpeDiem. El objetivo de esta fase es

corroborar a través del análisis de datos, si las variables seleccionadas influyen en los pilares estudiados. Más concretamente, este trabajo está centrado en el pilar del sueño.

El objetivo principal de este trabajo es validar si las variables seleccionadas por los clínicos están relacionadas con el sueño. En particular, se analizarán la estación del año, como variable extrínseca y la actividad física, contada como número de pasos, como variable relacionada con el estilo de vida. Para poder conseguir dicho objetivo se han establecido los siguientes hitos a cumplir:

- Analizar los datos que se obtienen mediante la pulsera Fitbit para obtener las variables especificadas por los médicos: hora de acostarse, hora de levantarse, tiempo dormido, tiempo despierto durante el período de sueño y tiempo dormido durante el día.
- Explorar todas estas variables y ver su relación con la estación del año y los pasos.
- Analizar si dicha relación se ve afectada por variables intrínsecas como el género y la edad.

2. Análisis Exploratorio

En este apartado se explica el proceso de análisis exploratorio llevado a cabo. Empezando por su motivación, con siguiendo los datos utilizados, los resultados principales y finalizando con las conclusiones obtenidas.

2.1. Motivación

En el proyecto CarpeDiem los datos relacionados con la actividad del sueño están recogidos a través de pulseras comerciales de la marca Fitbit. Este tipo de pulseras son denominadas *fitness trackers*, pues su cometido es monitorizar las variables relacionadas con el fitness como son la actividad física incluyendo pasos, niveles de actividad (sedentaria, ligera, moderada y vigorosa), el pulso y, también, la actividad del sueño. Al ser pulseras comerciales los datos que recogen están orientados a su propio propósito que es monitorizar e informar al usuario de la manera más simple posible su estado de *fitness*. Por este motivo, no siempre coinciden las variables que la aplicación comercial muestra al usuario final con las que se pueden obtener directamente para el desarrollo de otros proyectos basados en ellas.

El objetivo de este análisis exploratorio es indagar el tipo de datos relacionados con la actividad del sueño que se obtienen a través de este tipo de pulsera comercial, pues son las utilizadas en el proyecto CarpeDiem. Este primer análisis permitirá sacar conclusiones que posteriormente serán aplicadas a los datos utilizados para validar las variables relacionadas con el pilar del sueño.

2.2. Dataset

Como se menciona en el apartado 1.2, los datos relacionados con el sueño se obtienen mediante una pulsera de monitorización de la marca Fitbit². En este apartado se detallan el tipo de datos que proporciona dicha pulsera y como se obtienen. Además, se detalla el dataset utilizado para el análisis exploratorio de este tipo de datos.

Fitbit tiene dos modos de calcular la actividad del sueño dependiendo de si se utiliza el ritmo cardíaco o no. Cuando el dispositivo dispone de ritmo cardíaco este es combinado con los patrones de movimiento para un mejor seguimiento de la actividad del sueño. Este modo utiliza la variabilidad de la frecuencia cardíaca para estimar las fases del sueño: profundo, ligero, rem y despierto (en inglés: *deep*, *light*, *rem* y *wake*). Cuando no tiene la información del ritmo cardíaco, ya sea porque el dispositivo no tiene la opción o porque no se ha podido medir bien, la actividad del sueño se basa en el movimiento capturado por el acelerómetro. El algoritmo asume que el usuario está dormido si no se ha movido durante una hora. De hecho, debido a esta suposición surgen algunos registros de sueño falso si el usuario no se movió durante una hora, por ejemplo, si estuvo viendo la televisión.

²<https://www.fitbit.com/home>

Este modo, denominado clásico, clasifica la actividad del sueño en tres niveles: durmiendo, inquieto y despierto (en inglés: *asleep*, *restless* y *awake*).

Para cada período de actividad del sueño, Fitbit reporta los minutos y el conteo de veces que el usuario ha permanecido en cada una de las etapas o niveles, indicando con la variable *infoCode* en qué modo ha funcionado. Además, reporta la hora de inicio y fin del período (*startTime* y *endTime*); el tiempo total (*timeInBed*); la eficiencia (*efficiency*), es decir, el tiempo que ha dormido realmente el usuario; los minutos que han transcurrido después de despertarse y antes de quedarse dormido (*minutesAfterWakeUp* y *minutesToFallAsleep*); y finalmente, el tiempo total en minutos que ha pasado durmiendo y despierto (*minutesAsleep* y *minutesAwake*). La Tabla 1 resume todas las variables del sueño reportadas por Fitbit, divididas entre las variables comunes, medidos en ambos modos, y las específicas a cada modo.

Variables comunes	Variables modo fases	Variables modo clásico
<i>dateOfSleep</i>	<i>deepCount</i>	<i>asleepCount</i>
<i>startTime</i>	<i>deepMinutes</i>	<i>asleepMinutes</i>
<i>endTime</i>	<i>lightCount</i>	<i>awakeCount</i>
<i>efficiency</i>	<i>lightMinutes</i>	<i>awakeMinutes</i>
<i>infoCode</i>	<i>remCount</i>	<i>restlessCount</i>
<i>timeInBed</i>	<i>remMinutes</i>	<i>restlessMinutes</i>
<i>minutesAfterWakeUp</i>	<i>wakeCount</i>	
<i>minutesAsleep</i>	<i>wakeMinutes</i>	
<i>minutesAwake</i>		
<i>minutesToFallAsleep</i>		

Tabla 1: Resumen de las variables sin procesar que reporta Fitbit.

La recopilación de los datos se ha llevado a cabo mediante distintos interfaces de programación de aplicaciones, conocidos por las siglas API (del inglés *Application Programming Interface*). Gracias a estos API se obtienen las variables que se necesitan de manera que se evita iniciar la tarea de programación desde cero. Más concretamente, las variables del sueño se han obtenido mediante la API de Fitbit³.

Para explorar los datos del sueño que proporciona Fitbit se han utilizado los datos históricos de dos voluntarios, Voluntario 1 (V1) y Voluntario 2 (V2). Estos voluntarios son usuarios habituales de Fitbit y llevan su pulsera en su día a día, hecho que nos ha permitido disponer de más de un año de información con datos continuos. Se dispone de datos desde el 1 de Enero de 2018 hasta el 31 de Marzo de 2019 de ambos voluntarios. Ambos voluntarios son mujeres de entre 35 y 45 años y tienen un Fitbit Charge 2. Esta pulsera dispone de frecuencia cardíaca, como se ha comentado anteriormente en estos casos la actividad del sueño puede ser reportada en el modo de fases del sueño o en el modo clásico.

³<https://dev.fitbit.com/build/reference/web-api/sleep/>

2.3. Resultados

En este apartado se presentan los resultados del análisis exploratorio. Las variables se han explorado dependiendo de su naturaleza y su posible relación con las otras variables. Las siguientes variables comunes *efficiency*, *infoCode*, *timeInBed*, *minutesAfterWakeUp*, *minutesAsleep*, *minutesAwake* y *minutesToFallAsleep*, se han visualizado mediante una time serie para poder ver posibles diferencias entre los usuarios y posibles alteraciones a lo largo del tiempo. Las variables de ambos modos se han visualizado con gráficas de dispersión relacionando los minutos y el conteo de cada tipo de variable: *deep*, *rem*, *light*, *wake*, *asleep*, *awake* y *restless*. Finalmente se han analizado las variables *starTime* y *endTime* conjuntamente.

Como primer paso se han explorado las variables comunes a nivel visual. Se ha utilizado una visualización en forma de time series para poder ver el rango de valores de cada variable sin perder la temporalidad. Así, en las Figuras 1 y 2 se muestran las variables comunes de los dos voluntarios en forma de time series, respectivamente. La variable *efficiency* está entre los rangos esperados (0 a 100) ya que se trata de un porcentaje. Vemos que la variable *infoCode* varía notablemente entre los dos voluntarios. Aunque ambos voluntarios tiene la misma pulsera, la cual dispone de medidor de ritmo cardíaco, el 88,19% de los períodos del V1 son en modo fases de sueño (*infoCode* = 0) mientras que solo un 64,5% de los períodos del V2 son en este modo. Las variables *timeInBed*, *minutesAfterWakeUp*, *minutesAsleep*, *minutesAwake* están dentro de los rangos esperados, aunque en algunos días la variable *timeInBed* sobrepasa las 10h (600min). Entre estas 4 variables la que más diferencias tiene entre usuarios es la *minutesAfterWakeUp*, el V1 tiene una tendencia de no quedarse en la cama después de despertarse mientras que el V2 casi siempre se queda algunos minutos en la cama. Finalmente, se puede apreciar que la variable *minutesToFallAsleep* es siempre cero en ambos casos. Esto es debido a que esta variable solo es utilizada por Fitbit cuando el usuario selecciona manualmente el modo dormir, utilizado sobretodo en modelos más antiguos.

Para analizar las variables correspondientes a cada tipo de modo se han utilizado gráficas de dispersión, pues para cada variable disponemos de los minutos totales (*minutes*) y cuantas veces se ha estado en esa fase/nivel (*counts*). De este modo se analiza si hay una clara correlación entre los minutos y el conteo para cada variable y si esta correlación varía entre los usuarios. Las Figuras 3 y 4 muestran las gráficas de dispersión para las variables del modo de fases y clásico, respectivamente. Ambas Figuras contienen las gráficas del V1 (izquierda) y V2 (derecha).

En la Figura 3 se puede ver que para cada variable de fases del sueño sus rangos de valores son similares para los dos voluntarios, por ejemplo, en la fase de *rem* tanto el V1 (Figura 3a) como el V2 (Figura 3b) tienen el rango de *remCount* entre 0 y 20; y el *remMinutes* entre 0 y 175 minutos.

En cambio si analizamos los rangos de las variables vemos que *rem* y *deep* tienen valores similares tanto para el *count* como para *minutes* y que estas variables tienen valores menores que las otras dos. Entre ellas, *light* y *wake*, tienen valores similares en el *count*, sin embargo el rango de la variable *lightMinutes* es mayor que el de *wakeMinutes*. Estos rangos son coherentes con lo esperado pues la fase *light* (que

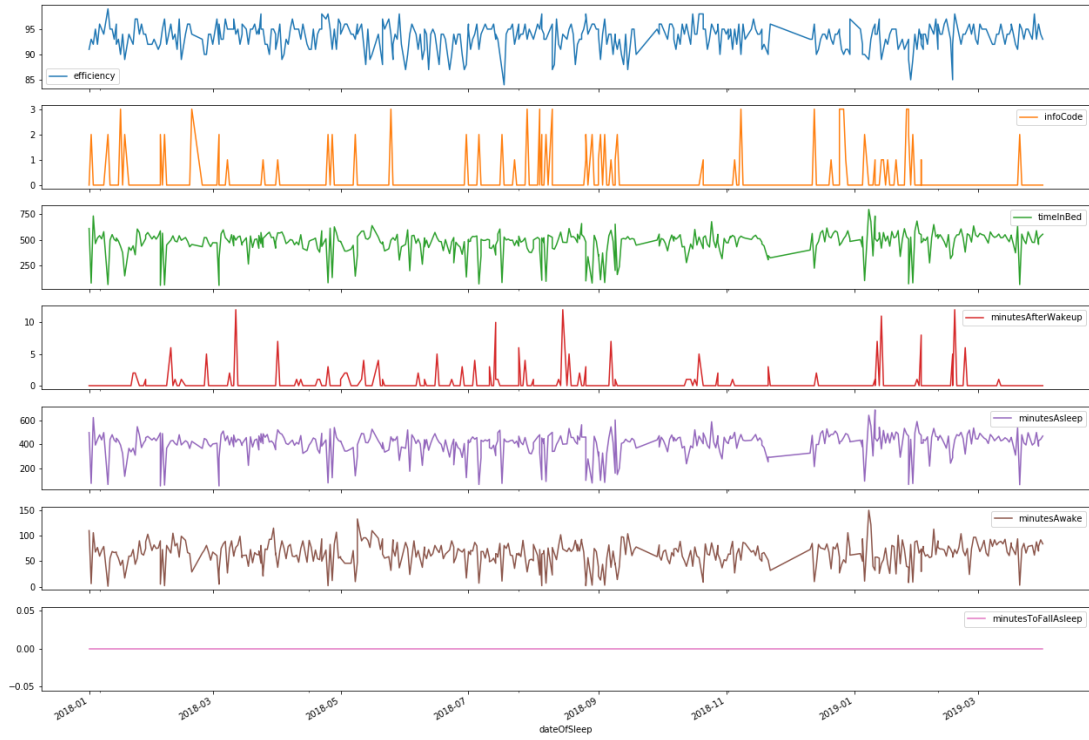


Figura 1: Time series mostrando las variables comunes del V1.

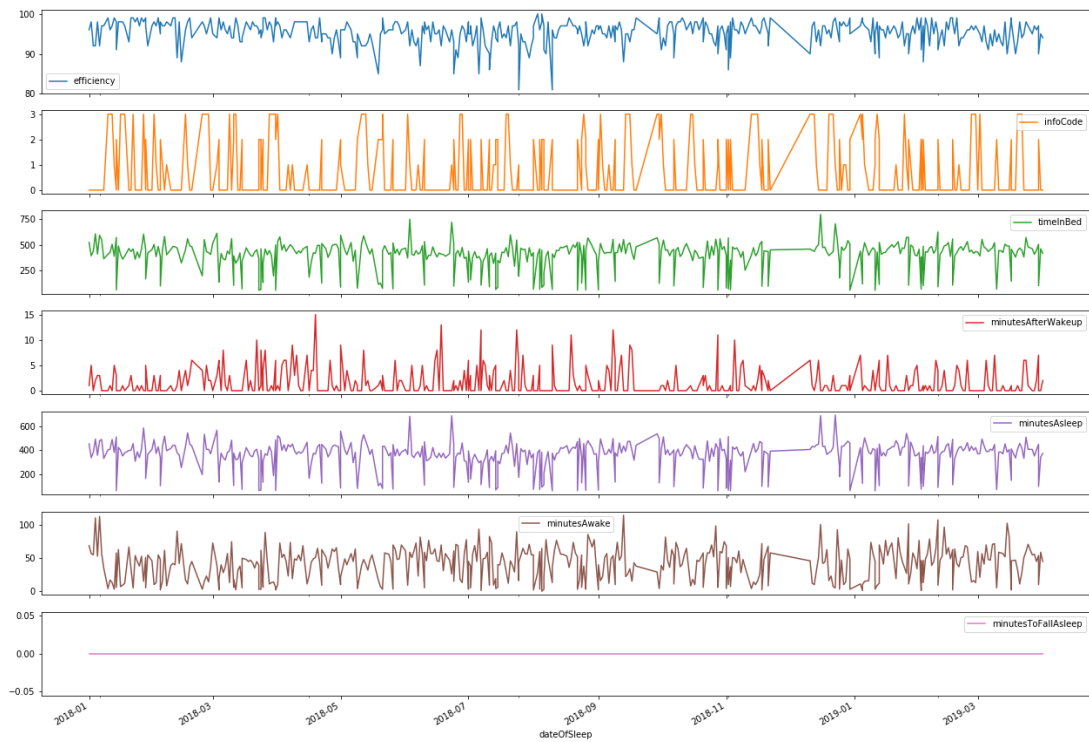


Figura 2: Time series mostrando las variables comunes del V2.

incluye las fases 1 y 2) es la más larga⁴.

Si nos fijamos en la correlación entre las variables *count* y *minutes*, vemos que todas las variables tienen una correlación lineal positiva. En cuanto a la dispersión, se ve que hay más en las variables *wake* y *light*. La variable con más dispersión es *wake* del V2.

Finalmente, en las Figuras 3a y 3c vemos que hay un punto en el (0,0) en ambas gráficas. Después de explorar estos datos no se ha podido sacar ninguna conclusión pues pertenecen a días diferentes y las duraciones del período al que pertenecen son 3,35h y 4,43h, para el período con *remCount* = 0 y *deepCount* = 0, respectivamente.

En la Figura 4 se puede ver que la variable *asleepCount* es siempre cero en ambos usuarios, no obstante *asleepMinutos* sí que varía. En la página de información de la API de Fitbit⁵ avisa que este campo, *asleepCount*, no debería usarse. Si analizamos las otras dos variables vemos que los rangos de valores son similares entre los dos voluntarios, por ejemplo, en el nivel de *restless* tanto el V1 (Figura 4c) como el V2 (Figura 4d) tienen el rango de *restlessCount* entre 0 y alrededor de 4-6; y el *restlessMinutes* entre 0 y alrededor de 15-20 minutos.

En cambio si analizamos los rangos entre las variables vemos que *restless* tienen valores menores que *awake* tanto para la variable *count* como para *minutes*.

Si nos fijamos en la correlación entre las variables *count* y *minutes*, vemos que las dos variables, *restless* y *awake* tienen una correlación lineal positiva. Ninguna de las dos variables muestra mucha dispersión.

Para analizar las variables *starTime* y *endTime* se ha utilizado el método Kmeans. Éste es un algoritmo de clasificación no supervisada (clusterización) que agrupa los objetos en k grupos basándose en sus características. Para ello se minimiza la suma de distancias entre cada objeto y el valor medio del cluster (centroide). Primero se escoge el número de clusters, k, y se establecen k centroides en el espacio de los datos. La selección de número de clusters se ha realizado mediante la técnica del *Elbow*.

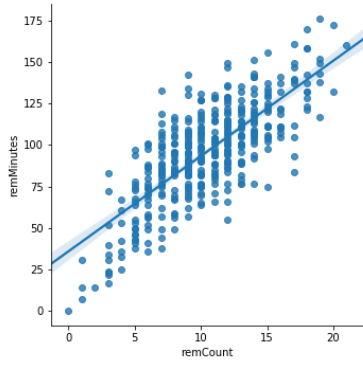
El método del *Elbow* consiste en representar el porcentaje de varianza respecto al número de clusters. El número adecuado de clusters se determina cuando la varianza no cambia mucho al aumentar el número de clusters. Esto se observa con un cambio brusco (en forma de codo) en la gráfica, el cual da nombre al método (ejemplos en las figuras 5, 6, 7 y 8 (a)).

Una vez seleccionado el número de clusters, cada objeto de los datos se asigna a su centroide más cercano. Seguidamente se actualiza la posición del centroide de cada grupo tomando como nuevo centroide la posición del promedio de los objetos pertenecientes a dicho grupo. Estos dos últimos pasos se repiten hasta que la diferencia entre los nuevos y anteriores centroides varíe poco. Los resultados del Kmeans se han visualizado utilizando gráficos de dispersión.

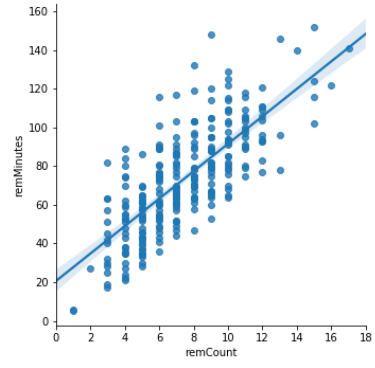
La Figura 5 muestra los resultados del análisis de las variables *starTime* y *endTime* del V1. La Figura 5a muestra los resultados obtenidos aplicando el Kmeans

⁴<https://blog.fitbit.com/sleep-stages-explained/>

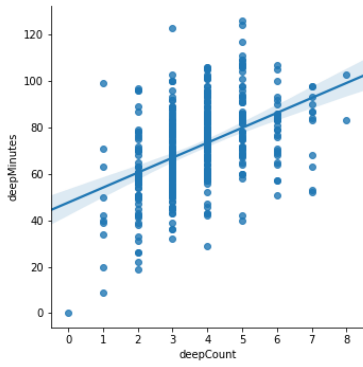
⁵<https://dev.fitbit.com/build/reference/web-api/sleep/>



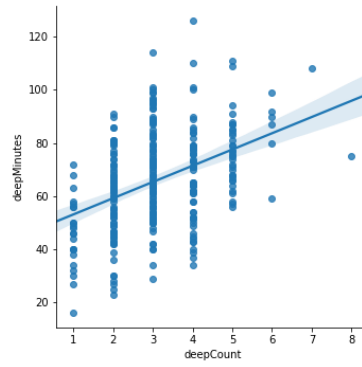
(a) Rem V1



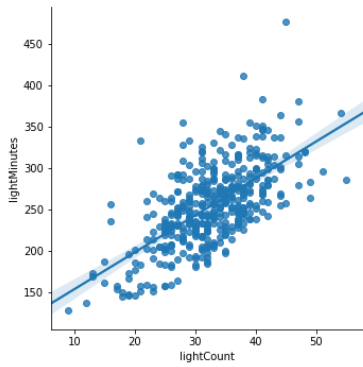
(b) Rem V2



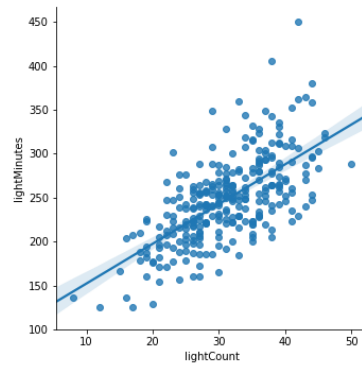
(c) Deep V1



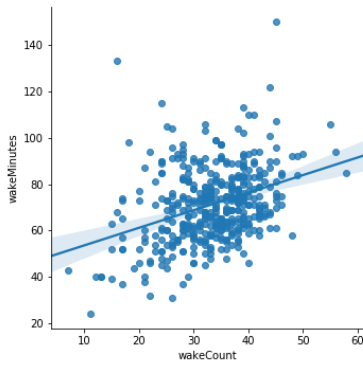
(d) Deep V2



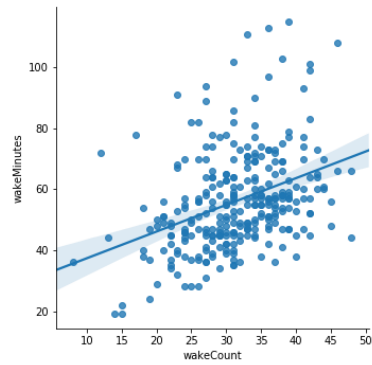
(e) Light V1



(f) Light V2

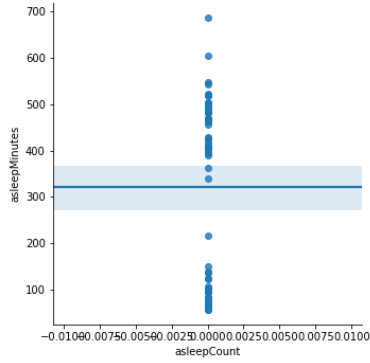


(g) Wake V1

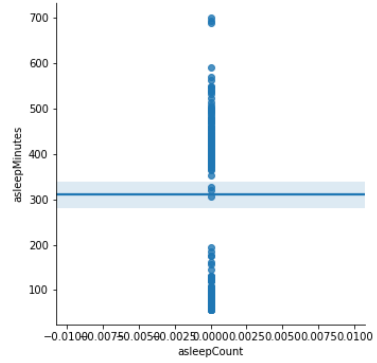


(h) Wake V2

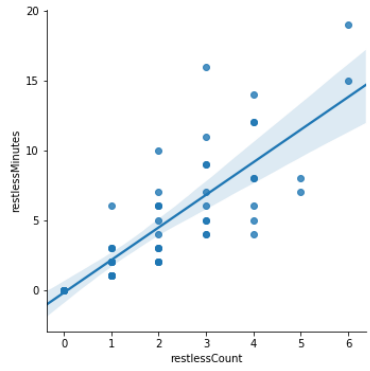
Figura 3: Gráficos de dispersión de la fases del sueño de V1 y V2, visualizando los minutos y el conteo de cada una.



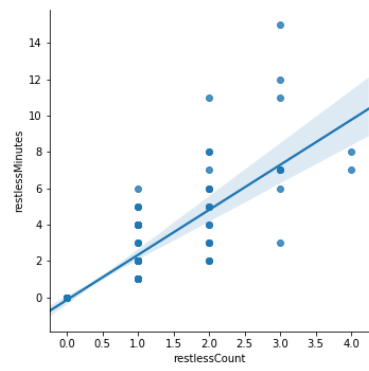
(a) Asleep V1



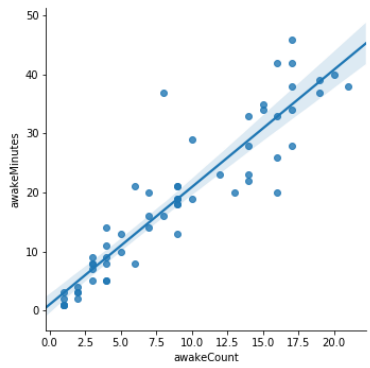
(b) Asleep V2



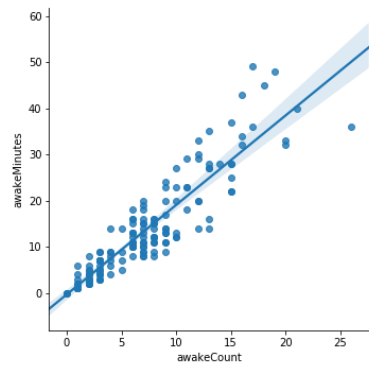
(c) Restless V1



(d) Restless V2



(e) Awake V1



(f) Awake V2

Figura 4: Gráficos de dispersión de los niveles del modo clásico del sueño de V1 y V2, visualizando los minutos y el conteo de cada una.

con diferente número de clusters. Utilizando la técnica del *Elbow*, vemos que el número óptimo de clusters a utilizar en estos datos es 2. Así, la Figura 5b muestra el resultado del Kmeans con 2 clusters sobre un gráfico de dispersión de las variables *startTime* (eje x) y *endTime* (eje y) para el V1. Cada color representa un cluster creado por el Kmean. Se puede observar que los dos grupos se encuentran en distintos horarios: se acuesta antes o después de medianoche, cluster rojo y azul, respectivamente. Ahora bien, se observa que el Kmeans se ve influenciado por la decisión arbitraria de empezar y acabar los días a las 12 de la noche.

La Figura 6 muestra los resultados del análisis de las variables *starTime* y *endTime* del V2. Estos resultados son muy parecidos a los obtenidos con V1. La Figura 6a se observa que utilizando la técnica del *Elbow* el número óptimo de clusters es 2. La Figura 6b muestra dos grupos con los mismos horarios que el V1: se acuesta antes o después de medianoche, cluster azul y rojo, respectivamente. En este caso el Kmeans también se ve influenciado por la decisión arbitraria de empezar y acabar los días a las 12 de la noche.

En las Figura 5b y Figura 6b, se observa claramente que parte de los puntos del grupo de la derecha, especialmente los que su *startTime* es después de las 20h, están apartados. Cuando se mira la actividad del sueño la hora más intuitiva para dividir los días dependerá de cada país e incluso de cada usuario. Basándose en estos datos es lógico determinar que el día acabe más bien sobre las 20h que no a medianoche. Por este motivo, se han vuelto a entrenar los modelos de Kmeans, definiendo que el día se acaba a las 20h.

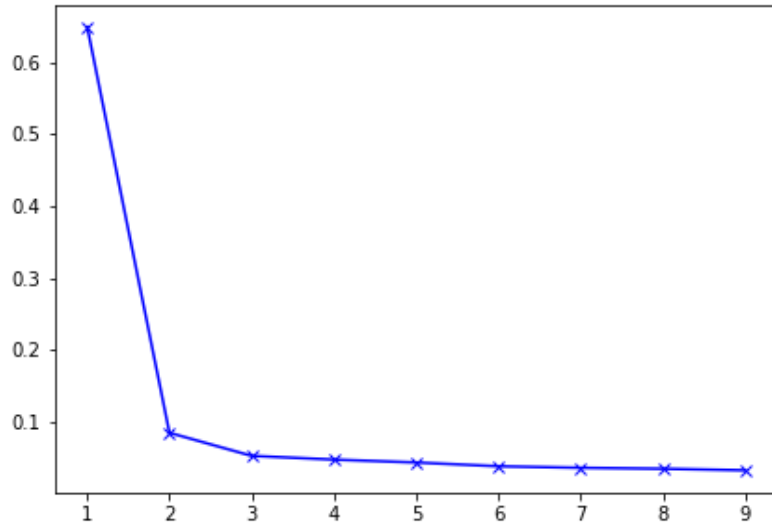
Las Figuras 7b y 8b muestran los resultados del reentrenamiento del Kmeans con los datos del V1 y V2, respectivamente. Como se muestra en las Figuras 7a y 8a, en ambos casos el número óptimo de clusters es dos. En las Figuras 7b y 8b se puede apreciar que los clusters encontrados son períodos de sueño por la noche (azul) y períodos de sueño durante el día (rojo).

La exploración de las variables *startTime* y *endTime* ha mostrado que Fitbit no reconoce solo los períodos de tiempo durante la noche como actividad del sueño. Con estos datos se comprueba que Fitbit es capaz de reportar más de una actividad del sueño por día. De hecho, según la información de la API de Fitbit, cada período de sueño debería tener una variable llamada *isMainSleep* indicando si ese período es el período de actividad de sueño principal de aquel día o no. No obstante, en los datos de los V1 y V2 no aparece dicha variable.

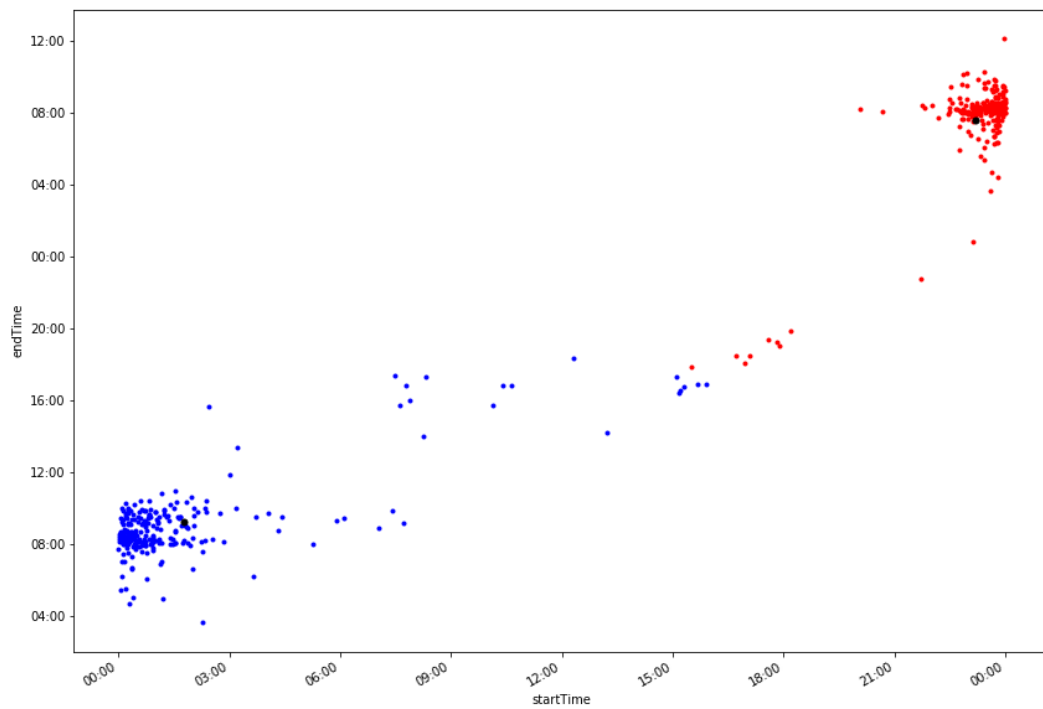
2.4. Conclusiones

El análisis exploratorio hecho con los datos históricos de los V1 y V2 ha permitido entender mejor los datos de actividad del sueño que proporciona Fitbit. A continuación se detallan las conclusiones extraídas en este análisis que serán utilizadas para la validación de las variables relacionadas con el sueño.

Cuando el dispositivo detecta automáticamente la actividad del sueño, las variables *minutesFallToSleep* y *asleepCount* son siempre nulas, por tanto, no aportan información. Los dispositivos que disponen de medidor de frecuencia cardíaca fun-

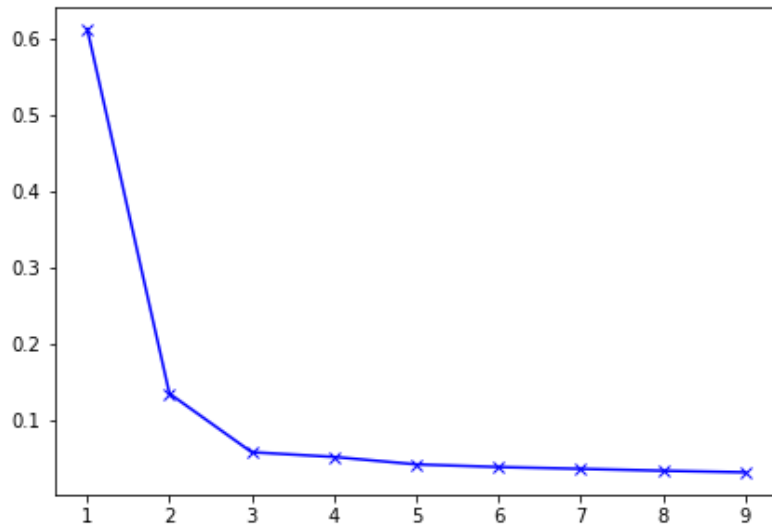


(a) Elbow plot

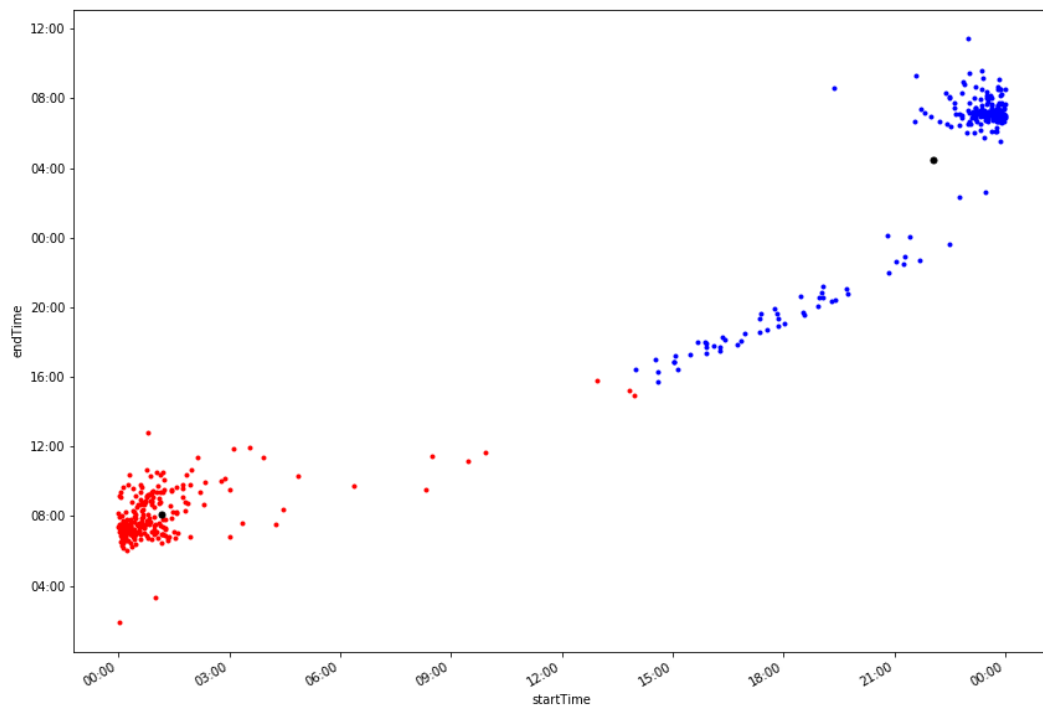


(b) Kmeans con 2 clusters

Figura 5: Resultado del Kmeans aplicado a las variables *startTime* y *endTime* del V1.

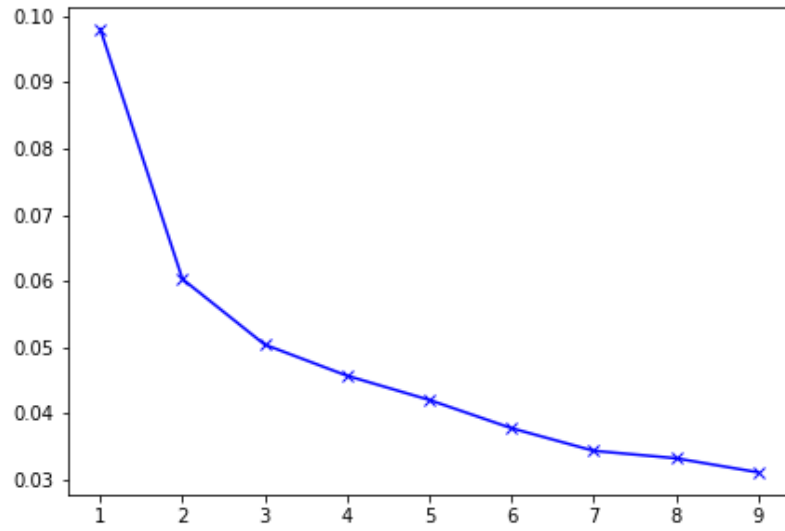


(a) Elbow plot

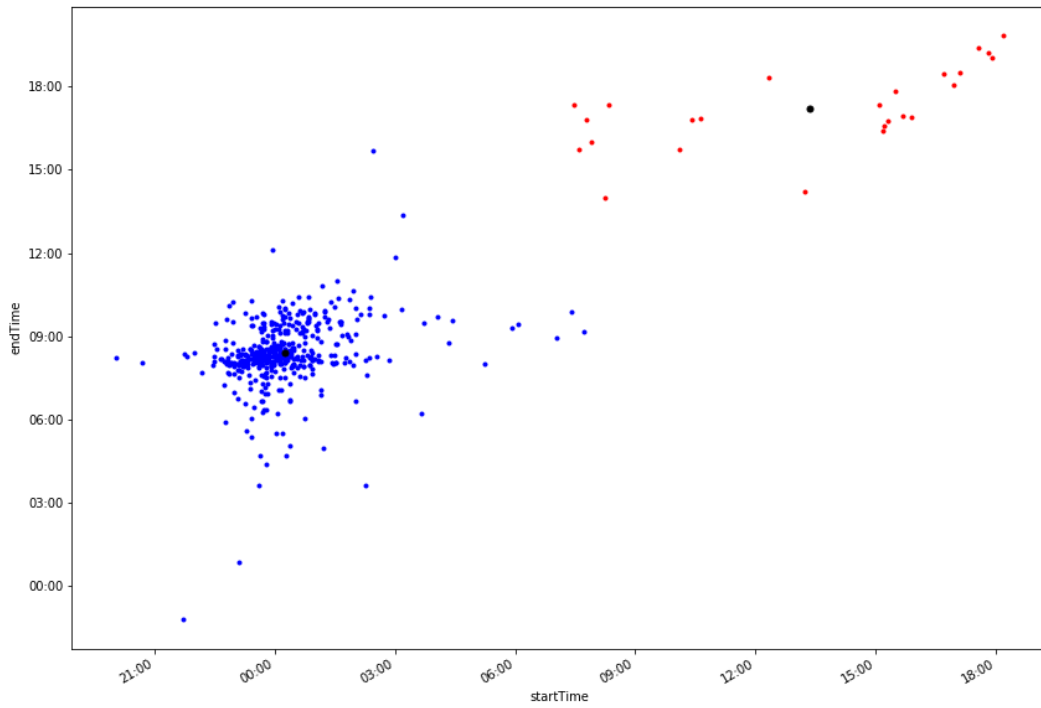


(b) Kmeans con 2 clusters

Figura 6: Resultado del Kmeans aplicado a las variables *startTime* y *endTime* del V2.

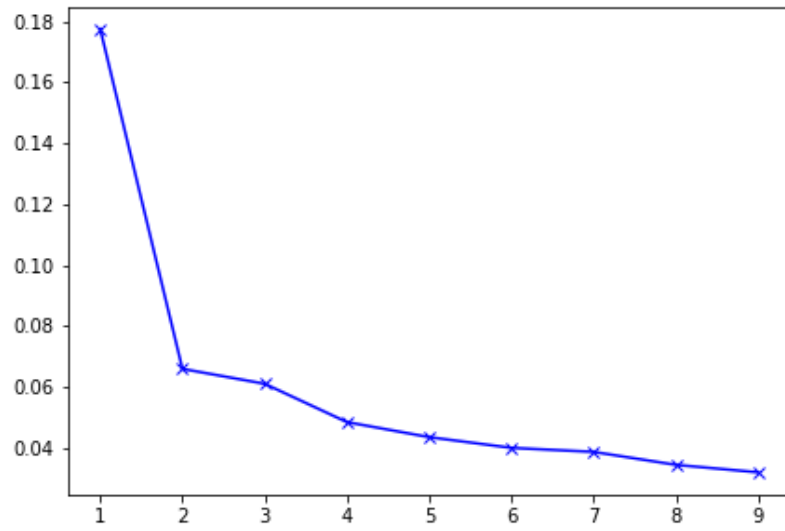


(a) Elbow plot

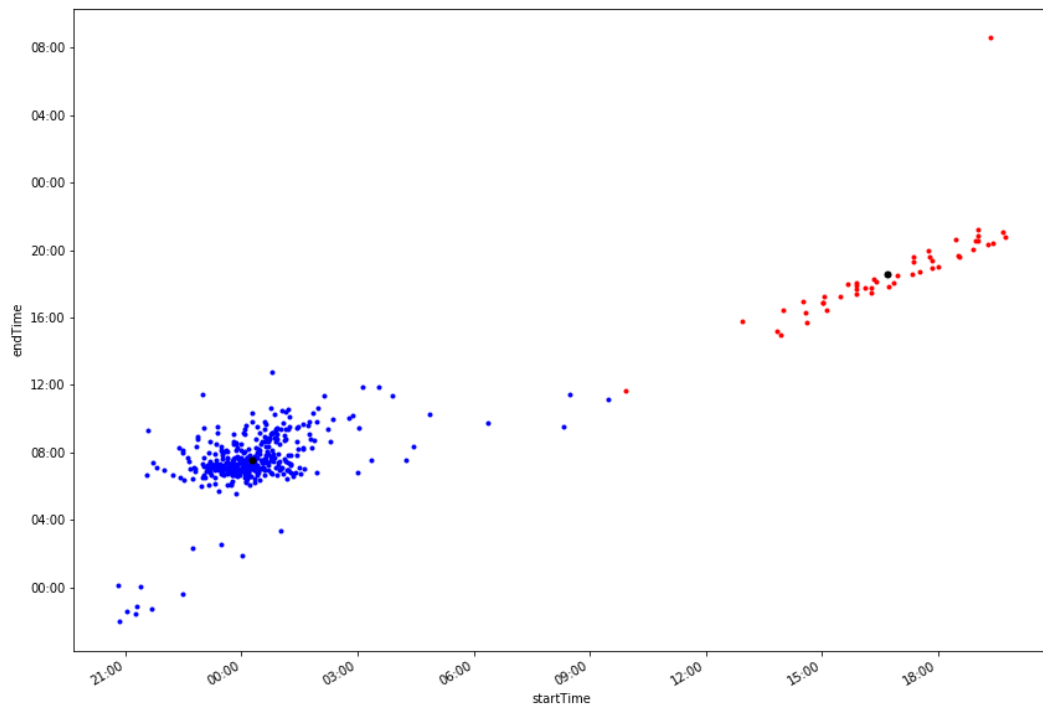


(b) Kmeans con 2 clusters

Figura 7: Resultado del Kmeans aplicado a las variables *startTime* y *endTime* definiendo el final del día a las 20h del V1.



(a) Elbow plot



(b) Kmeans con 2 clusters

Figura 8: Resultado del Kmeans aplicado a las variables *startTime* y *endTime* definiendo el final del día a las 20h del V2.

cionan en dos modos: modo fases del sueño y modo clásico. La variable *infoCode* indica para cada período de sueño qué tipo de modo se ha utilizado: 0 para el modo fases del sueño y otro número (1, 2, 3) para modo clásico. Si las variables correspondientes al modo se quieren usar deben ser agregadas para que todos los períodos de sueño tengan las mismas variables.

Como ocurre con los datos analizados del V1 y V2, es posible que para un mismo día haya más de un período de sueño. En el caso de querer analizar los datos de actividad del sueño se deberían agregar todos los datos de actividad del sueño por usuario y por día. Para tal propósito se debería escoger cuál es el período de sueño principal de cada día.

3. Materiales y Métodos

En este apartado se describen los datos y métodos utilizados para estudiar las variables relacionadas con el pilar del sueño. Para hacer este análisis se han utilizado datos históricos de usuarios reales. Estos datos fueron recogidos en el proyecto Europeo CONNECARE⁶.

3.1. Materiales

El objetivo del proyecto CONNECARE es crear un sistema integrado que proporcione soporte en la gestión de casos médicos mediante herramientas que permitan monitorizar las actividades y el estado de los pacientes. El sistema creado en CONNECARE ha sido validado en un piloto que incluye pacientes crónicos complejos mayores, de más de 65 años, de cuatro *sites* clínicos: el IRBLI y el *Institut d'Investigacions biomèdiques August Pi i Sunyer* (IDIBAPS), en España, *Assuta*, en Israel y *University Medical Center Groningen* (UMCG), en Holanda. Además, el sistema fue validado a través de dos casos de estudio: pacientes con hospitalización ordinaria (Case 1) y paciente que requieren una operación (Case 2). En tres de los cuatro *sites*, IRBLI, Assuta y UMCG, se han utilizado pulseras de la marca Fitbit para monitorizar las actividades de los pacientes. Por consiguiente, en este trabajo se han utilizado los datos de estos tres *sites*.

En este trabajo se analiza si el sueño se ve afectado por: la estación del año como variable extrínseca y la actividad física, medida por los pasos, como variable relacionada con el estilo de vida. Las variables intrínsecas como la edad y el género son analizadas conjuntamente con las variables anteriores. El sueño es medido a través de las variables indicadas por los clínicos: hora de acostarse, hora de levantarse, tiempo dormido, tiempo despierto durante el período de sueño y tiempo dormido durante el día.

Para calcular la actividad física, Fitbit, utiliza los datos de ritmo cardíaco, las calorías quemadas y el tiempo transcurrido. De ese modo diferencia la actividad física en cuatro niveles distintos: sedentaria, ligera, moderada y vigorosa (en inglés: *sedentary*, *light*, *moderate* y *vigorous*). Durante el día se registran los minutos que realiza el usuario en cada nivel, junto con los pasos que ha andado (*steps*).

Como se ha comentado en el apartado 2.4 para poder utilizar los datos de actividad del sueño de Fitbit hace falta hacer un preprocesamiento que agregue los períodos de sueño que corresponden al mismo día. En este dataset, como pasaba con los datos del V1 y V2 la variable *isMainSleep* no está presente en los datos. En este trabajo, se ha definido como período de sueño principal el período de sueño más largo. Así, las variables definidas por los clínicos se han obtenido de la siguiente manera: las cuatro primeras (hora de acostarse, hora de levantarse, tiempo dormido, tiempo despierto durante el período de sueño) corresponden a las variables *startTime*, *endTime*, *minutesAsleep*, *minutesAwake* del sueño principal; y la variable que recoge el tiempo dormido durante el día, denominada *minutesOtherSleep*, se ha

⁶<https://www.connecare.eu/>

calculado agregando la variable *minutesAsleep* de los períodos no principales.

Contamos con datos de 132 usuarios, la Tabla 2 muestra su distribución por género, rango de edad, *site* y caso de estudio. En esta tabla se puede ver que los usuarios están balanceados tanto por género, *site* y caso de estudio, en cambio el grupo de edad más numeroso es el de 65-75 que cuenta con casi la mitad de los usuarios, seguido de los rangos 55-65 y 75-85 contando con un cuarto de los usuarios cada uno, finalmente el rango con menos usuarios es el de 85-95.

	quantity (%)
Género	
FEMALE	60 (45.45)
MALE	71 (53.79)
UNKNOWN	1 (0.76)
Grupo edad	
55-65	27 (20.45)
65-75	59 (44.6)
75-85	34 (25.76)
85-95	12 (9.09)
<i>site</i>	
assuta	43 (32.57)
irbll	44 (33.33)
umcg	45 (34.09)
Caso de estudio	
1	65 (49.24)
2	67 (50.75)

Tabla 2: Descripción de los datos por género, grupo de edad, *site* y caso de estudio.

Debido a que la inclusión de los usuarios fue escalonada y discontinua no todos los usuarios tienen el mismo período de datos. Para analizar como afecta la estación del año al sueño necesitamos usuarios con un mínimo de días, definido como 30 días, en más de una estación. Después de aplicar este filtro quedan 15 usuarios. La Tabla 3 muestra su distribución por género, rango de edad, *site* y caso de estudio de estos 15 usuarios. En esta tabla se puede ver que en este dataset los datos no están equilibrados: dos tercios de los usuarios son hombres, la mitad de los usuarios están en el rango de edad 65-75 y no hay usuarios del rango 85-95, casi dos tercios de los usuarios son de Assuta y el 60 % de los usuarios pertenecen al caso de estudio 2. El heatmap de la Figura 9 muestra los datos disponibles para cada usuario. Las fechas son representadas en el eje x y cada usuario es una línea del eje y. Podemos ver que el periodo de monitorización varía según los usuarios, desde 2 a 5 meses. También podemos ver que los datos no son continuos, pero en general hay suficiente datos para todos los 15 usuarios.

Género	quantity (%)
FEMALE	5 (33.33)
MALE	10 (66.66)
Grupo de edad	
55-65	5 (33.33)
65-75	8 (53.33)
75-85	2 (13.33)
85-95	0 (0.00)
<i>site</i>	
assuta	9 (60.00)
irbll	3 (20.00)
umcg	3 (20.00)
Caso de estudio	
1	6 (40.00)
2	9 (60.00)

Tabla 3: Descripción de los datos utilizados para analizar la variable estación por género, grupo de edad, *site* y caso de estudio.

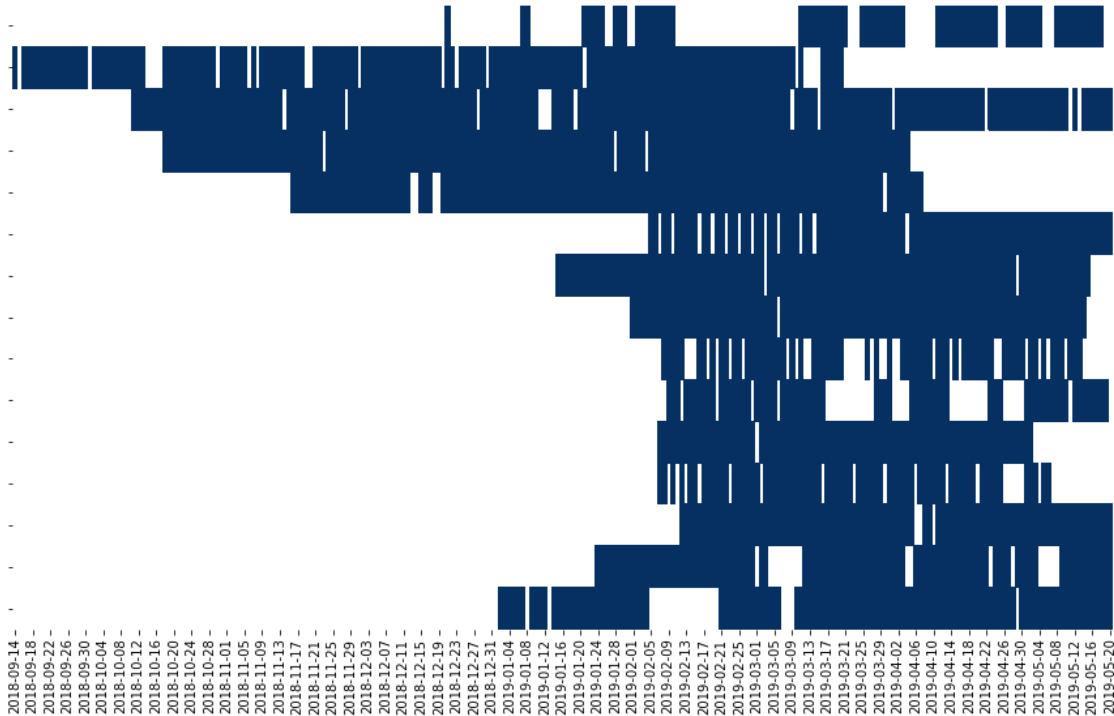


Figura 9: Heatmap visualizado los datos utilizados para analizar la variable estación.

Debido a que los datos pertenecen a usuarios reales los datos provenientes de

las pulseras son discontinuos y para algunos días no tenemos datos de actividad física o de actividad del sueño. Esto se puede deber a que el usuario se olvida de ponerse la pulsera o bien porque le molesta la pulsera al dormir y se la quita. Por este motivo, se han seleccionado solo aquellos usuarios de los cuales se dispone de un número mínimo de días, definido como 30 días, en los cuales se tiene información de pasos y de actividad del sueño. Tras esta selección solo quedan 81 usuarios. La Tabla 4 muestra su distribución por género, rango de edad, *site* y caso de estudio de estos 81 usuarios. Esta tabla muestra que la distribución por grupo de edad, por *site* y por caso de estudio sigue es la misma que presentaba el total de usuarios, en cambio en este caso un 60 % de los usuarios son hombre. El heatmap de la Figura 10 muestra los datos disponibles para cada usuario. Podemos ver que el período de monitorización varía según los usuarios, desde 1 mes hasta unos 5 meses. Como con los datos para analizar la variable de la estación, estos datos no son continuos.

	quantity (%)
Género	
FEMALE	32 (39.51)
MALE	49 (60.49)
Grupo de edad	
55-65	19 (23.46)
65-75	37 (45.68)
75-85	19 (23.46)
85-95	6 (7.410)
<i>site</i>	
assuta	29 (35.80)
irbll	26 (32.10)
umcg	26 (32.10)
Caso de estudio	
1	33 (40.74)
2	48 (59.25)

Tabla 4: Descripción de los datos utilizados para analizar la variable de pasos por género, grupo de edad, *site* y caso de estudio.

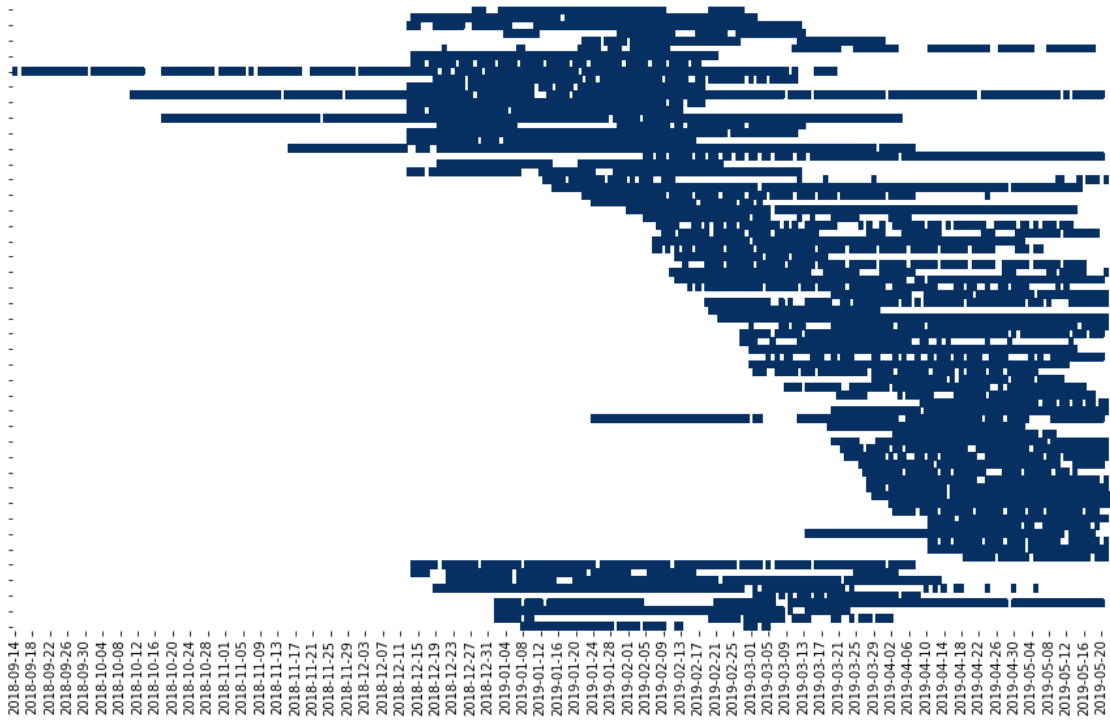


Figura 10: Heatmap visualizado los datos de los usuarios utilizados para analizar la variable de pasos.

3.2. Métodos

Para analizar los datos expuestos en el apartado anterior 3.1, se han aplicado los métodos de diagrama de violín y la correlación. Estos métodos son descritos a continuación.

Diagrama de Violín

Los diagramas de violín son una forma de representar datos numéricos. Permiten comparar la clasificación de varios grupos y su distribución, además de la densidad de probabilidad de los datos.

La distribución de los datos se representa a través de sus cuantiles, parecido a los diagramas de caja. Donde el punto blanco es la mediana (*Median*), la barra negra en el centro representa el intervalo intercuartil (*Interquartile Range (IQR)*) y la barra negra fina que se extiende desde ella representa el 95 % de los intervalos de confianza (*95 % Confident Interval*). La forma (*Density Plot*) representa la densidad de probabilidad de la variable: cuantos más datos haya en un rango específico, más grande es el violín para ese rango (*Higher Probability*) y, recíprocamente, cuantos menos datos haya en un rango específico, más pequeño es el violín para ese rango (*Lower Probability*).

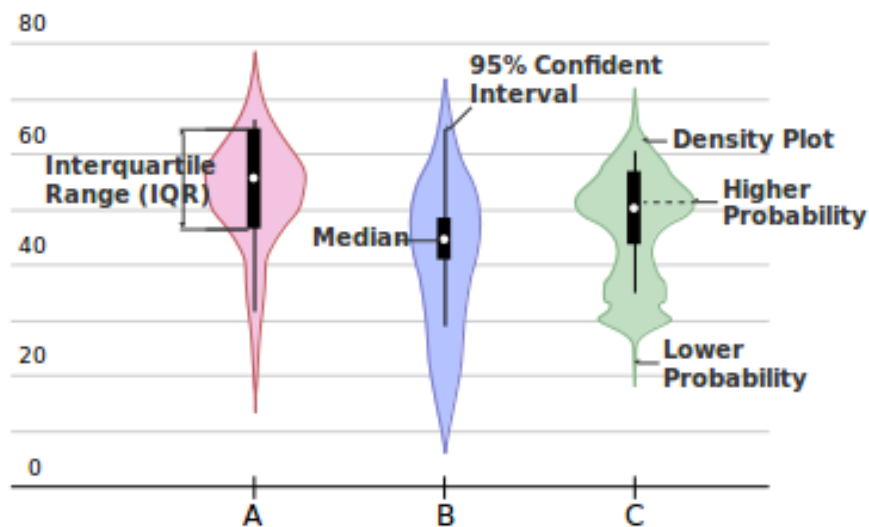


Figura 11: Términos del diagrama de Violín

Este método ha sido utilizado para visualizar las distribuciones de las variables estudiadas y así poder fácilmente comparar entre sí.

La Correlación Estadística

La correlación se utiliza para medir la fuerza y la dirección de la asociación entre dos variables. Cuando el cambio en una variable está acompañado de un cambio en la otra, entonces se dice que las variables están correlacionadas. Se mide mediante el coeficiente de correlación, que varía de -1.0 a 1.0 indicando la fuerza de relación entre las dos variables. En general, el coeficiente mayor que 0 indica una relación positiva y menor que 0 indica una relación negativa, mientras que el coeficiente igual a 0 indica que no hay relación. Cuanto más cerca estén los coeficientes de -1.0 y 1.0, mayor es la fuerza de la relación entre las variables. Además cuando el coeficiente es positivo quiere decir que si una de las variables se aumenta, la otra también se aumentará y viceversa. En cambio si el coeficiente es negativo, si una de las variables se aumenta la otra tenderá a disminuir y viceversa.

Este método ha sido utilizado para ver si hay correlaciones entre las variables de pasos y las diferentes variables de la actividad del sueño, pues tenemos por cada día un valor de ambas variables.

4. Resultados

En este apartado se presentan los resultados obtenidos al analizar si el sueño se ve afectado por la estación y por la actividad física, representada con número de pasos.

4.1. Estación

Como se ha descrito en el apartado 3.1 solo disponemos de 15 usuarios con datos en más de una estación. Este hecho hace que podamos analizar los datos en más profundidad y visualizándolos por usuario. Así, se ha utilizado el diagrama violín para visualizar la distribución de cada variable del sueño por usuario y por estación, pudiendo así comparar si hay diferencias entre las estaciones para cada usuario.

Las Figuras 12, 13, 14, 15 y 16 muestran los gráficos violín de los usuarios que tienen datos en solo dos estaciones. Vemos que para estos usuarios las estaciones son invierno y primavera para todos los usuarios. En la Figura 12 vemos que en general cada usuario tiene su hora de acostarse y que esta no varía dependiendo de la estación. En cambio, en la Figura 13 vemos que hay una ligera diferencia entre estaciones en la hora de levantarse. En esta Figura se pueden ver tres grupos: 4 usuarios se levantan una hora más tarde en primavera que en invierno, 5 usuarios se levantan una hora más temprano y dos se levantan más o menos a la misma hora. Respecto a las horas dormidas, en la Figura 14 se aprecia que la mayoría de usuarios tienen una pequeña diferencia entre estaciones: 3 duermen alrededor de una hora más en primavera que en invierno y 5 una hora menos. En este caso vemos que hay un usuario que pasa de dormir alrededor de unas 6 horas en invierno a dormir solo 3 horas en primavera, este sería un caso aislado. En el caso de los minutos despiertos durante la noche, la Figura 15 muestra que la mayoría de los usuarios no varían, mientras que 5 se despiertan más en invierno que en primavera y solo dos se despiertan menos. Finalmente, en la Figura 16 vemos que en general los usuarios no suelen dormir durante el día en ninguna de las dos estaciones. Solo tres de ellos tienen su media más alta que 0, en invierno. Se deben remarcar que todas las diferencias encontradas son pequeñas y que no se ha encontrado ningún patrón que las asocie todas a usuarios de un mismo grupo.

Las Figuras 17, 18, 19, 20 y 21 muestran los gráficos violín de los usuarios que tienen datos en solo tres estaciones. Vemos solo hay 4 usuarios y tres de ellos tienen las mismas estaciones, otoño, invierno y primavera, el otro usuario tiene verano, otoño y invierno. En la Figura 17 vemos que como en el caso de solo dos estaciones en general cada usuario tiene su hora de acostarse y que esta no varía dependiendo de la estación. En cuanto a hora de levantarse, como se puede ver en la Figura 18 también es constante para cada usuario, solo se visualiza una ligera diferencia en otoño para tres de los usuarios, en este caso se levantan un poco más tarde. Respecto a las horas dormidas, en la Figura 19 se aprecia que es constante para cada usuario sin diferencias significativas entre estaciones. Lo mismo ocurre en el caso de los minutos despiertos durante la noche, Figura 20. Finalmente, en la Figura 21 vemos que en general los usuarios no suelen dormir durante el día en ninguna de

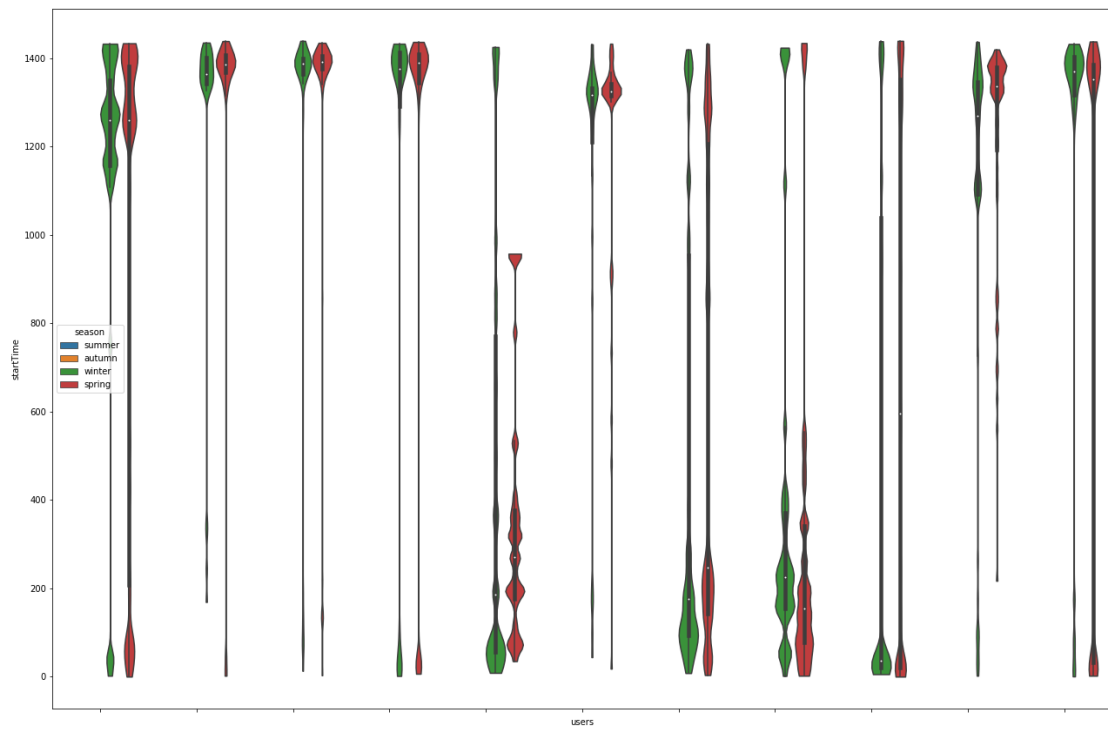


Figura 12: Gráfico violín comparando la variable de la hora de acostarse para los usuarios con datos en dos estaciones diferentes.

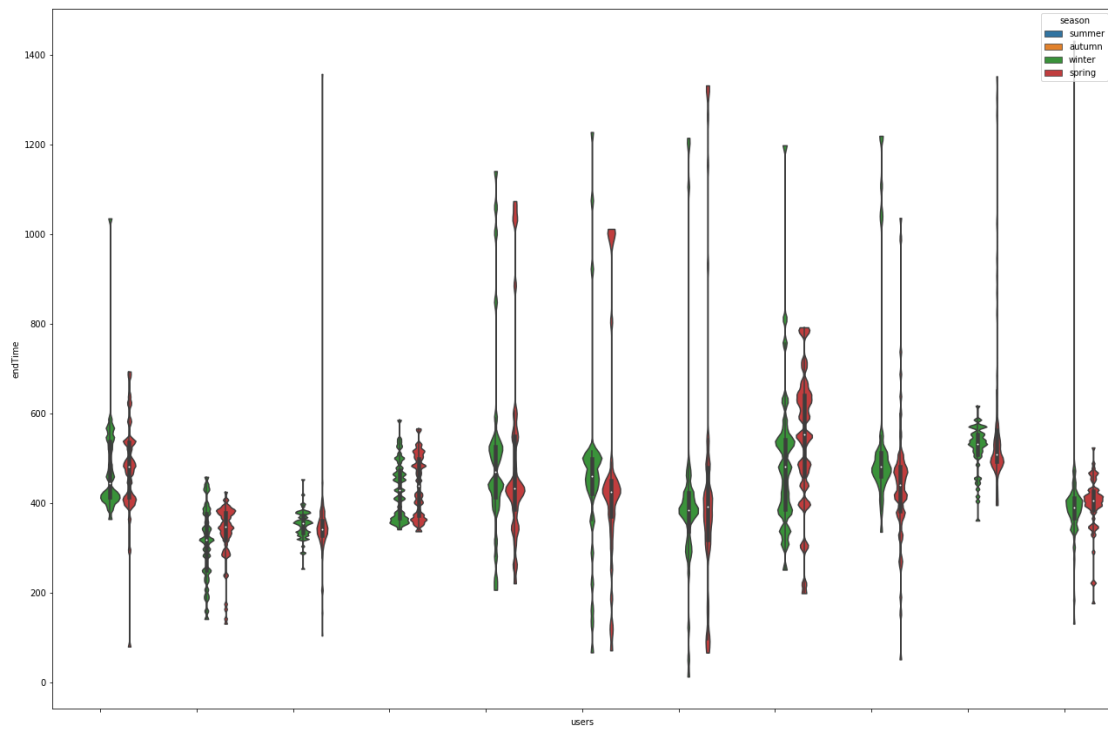


Figura 13: Gráfico violín comparando la variable de la hora de levantarse para los usuarios con datos en dos estaciones diferentes.

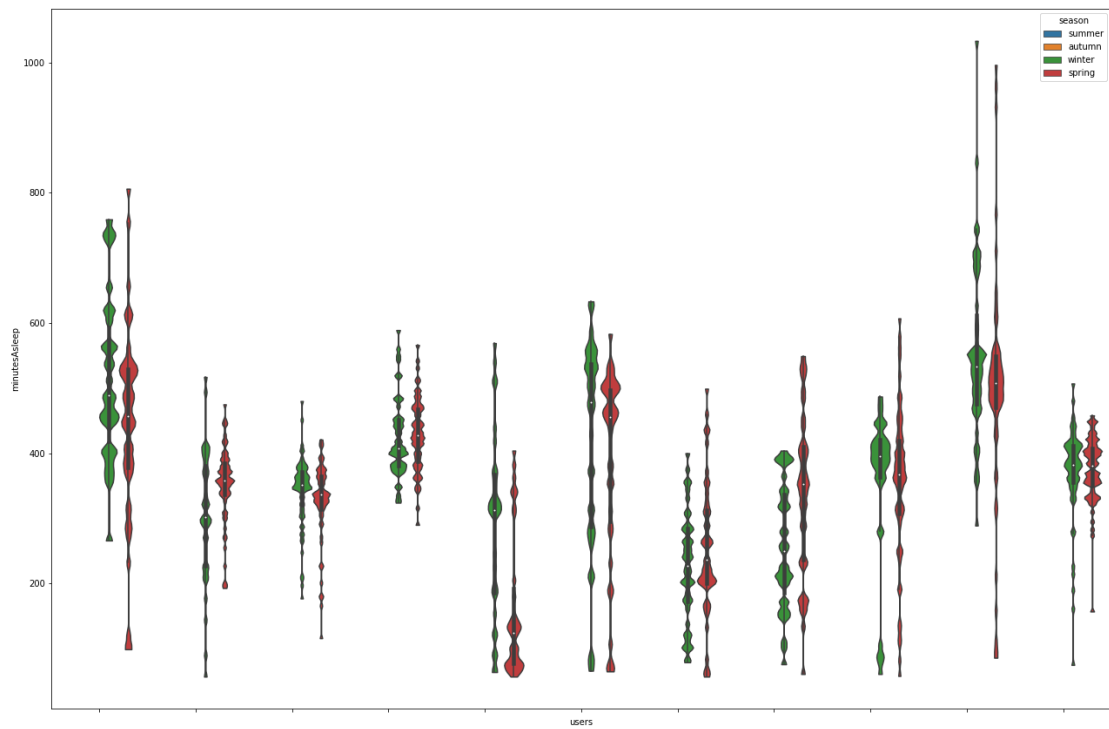


Figura 14: Gráfico violín comparando la variable de la tiempo durmiendo para los usuarios con datos en dos estaciones diferentes.

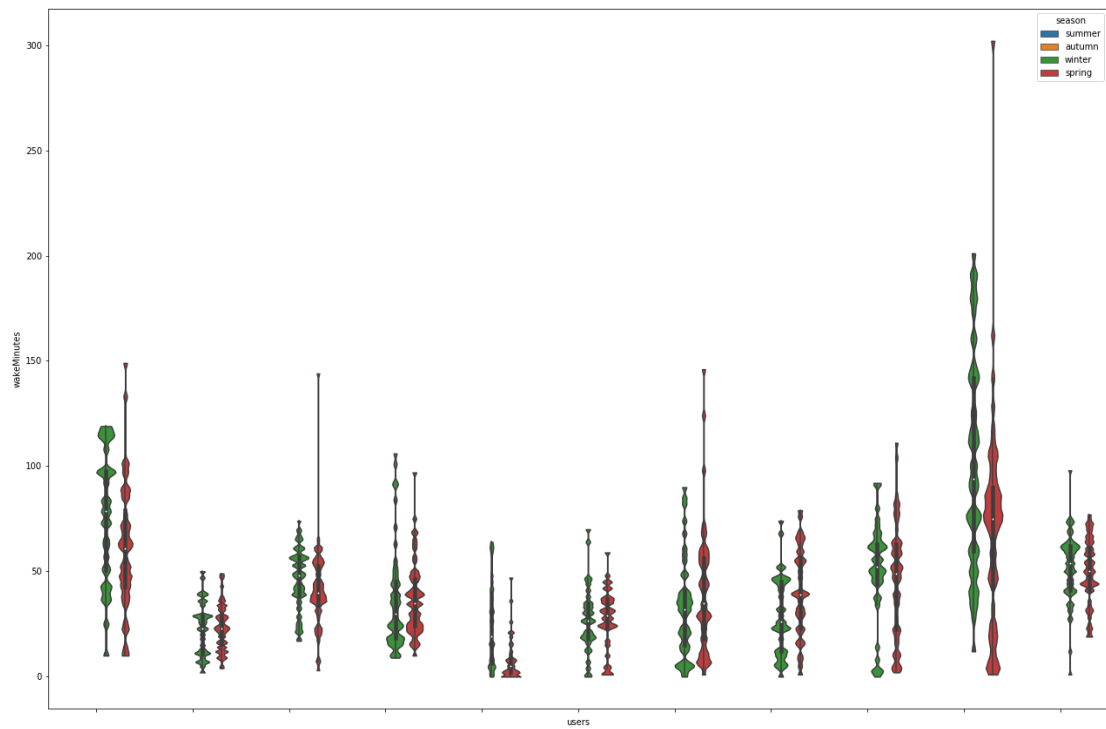


Figura 15: Gráfico violín comparando la variable de la tiempo despierto para los usuarios con datos en dos estaciones diferentes.

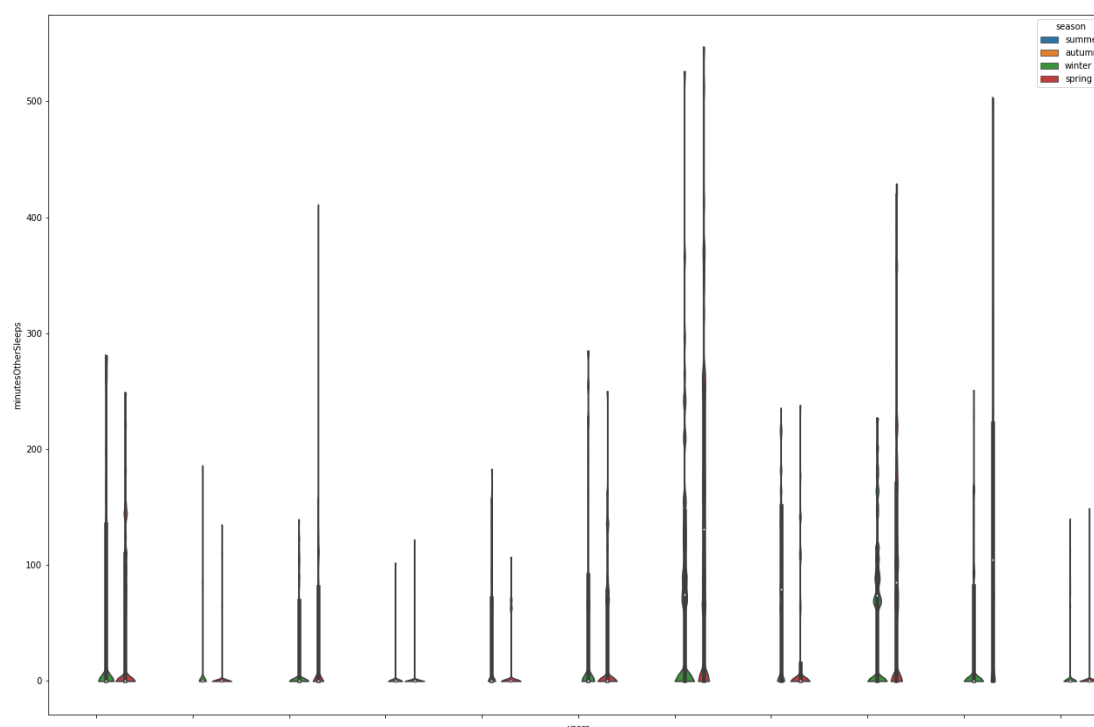


Figura 16: Gráfico violín comparando la variable de la tiempo durmiendo durante el día para los usuarios con datos en dos estaciones diferentes.

las dos estaciones. Solo uno de ellos tienen su media más alta que 0, en otoño y en invierno. Una vez más se deben remarcar que todas las diferencias encontradas son pequeñas y que no se ha encontrado ningún patrón que las asocie todas a usuarios de un mismo grupo.

4.2. Actividad Física (Pasos)

En el caso de los datos de pasos tenemos un valor para cada usuario y día, esto nos permite calcular la correlación que hay entre las variables que definen el sueño y el número de pasos. De tal modo que para cada usuario y cada variable del sueño tenemos un valor de correlación. Posteriormente, se han visualizado estas correlaciones mediante gráficos de violín, para todos los usuarios y agrupándolos para poder ver si alguna de las variable tiene alguna correlación significativa.

La Figura 22 muestra el gráfico de violín de las correlaciones para las variables de la actividad del sueño. En ella se puede ver que ninguna de las variables está correlacionada con el sueño pues todas las medias están alrededor de cero y el rango está entre 0,5 y $-0,5$. Las Figuras 23, 24, 25 y 26 muestran los gráficos de violín de las correlaciones para las variables de la actividad del sueño separadas por género, rango de edad, *site* y caso de estudio, respectivamente. En ninguna de estas figuras se puede apreciar un cambio significativo entre grupos.

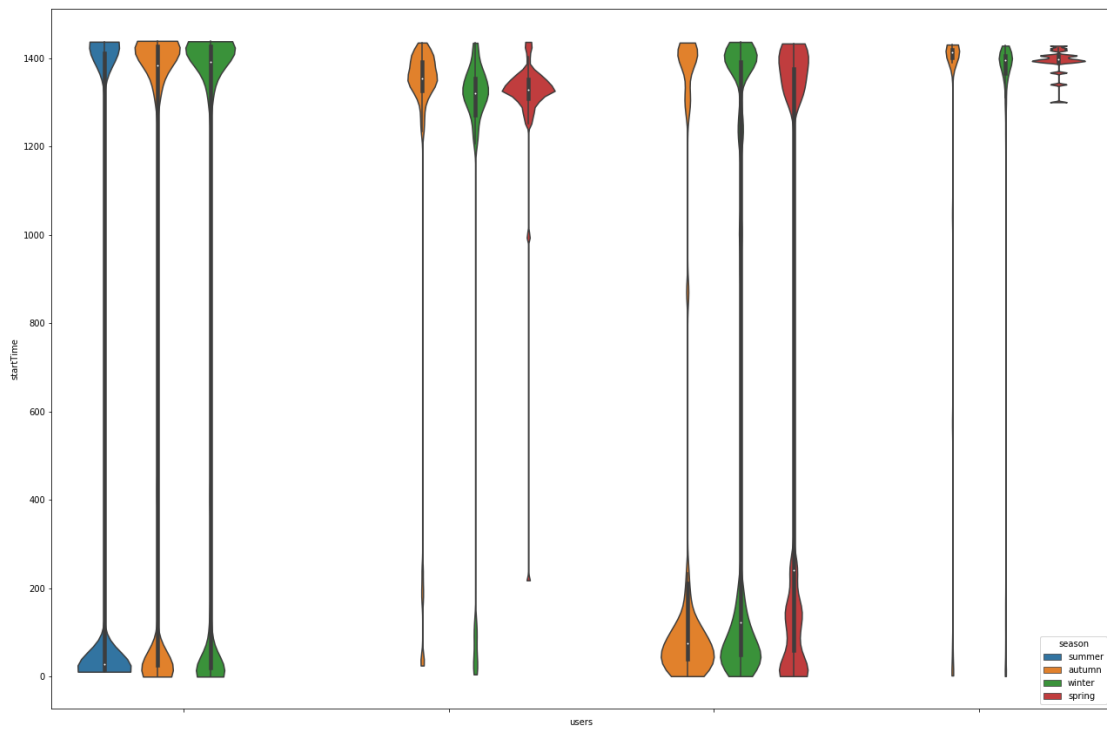


Figura 17: Gráfico violín comparando la variable de la hora de acostarse para los usuarios con datos en tres estaciones diferentes.

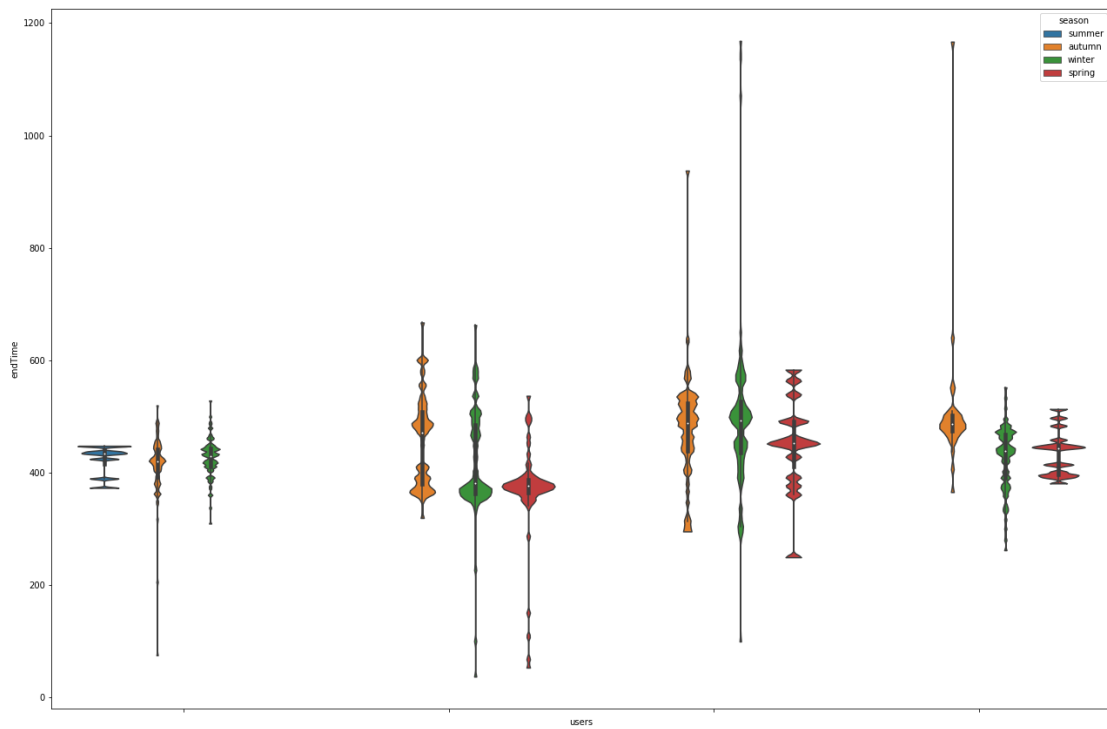


Figura 18: Gráfico violín comparando la variable de la hora de levantarse para los usuarios con datos en tres estaciones diferentes.

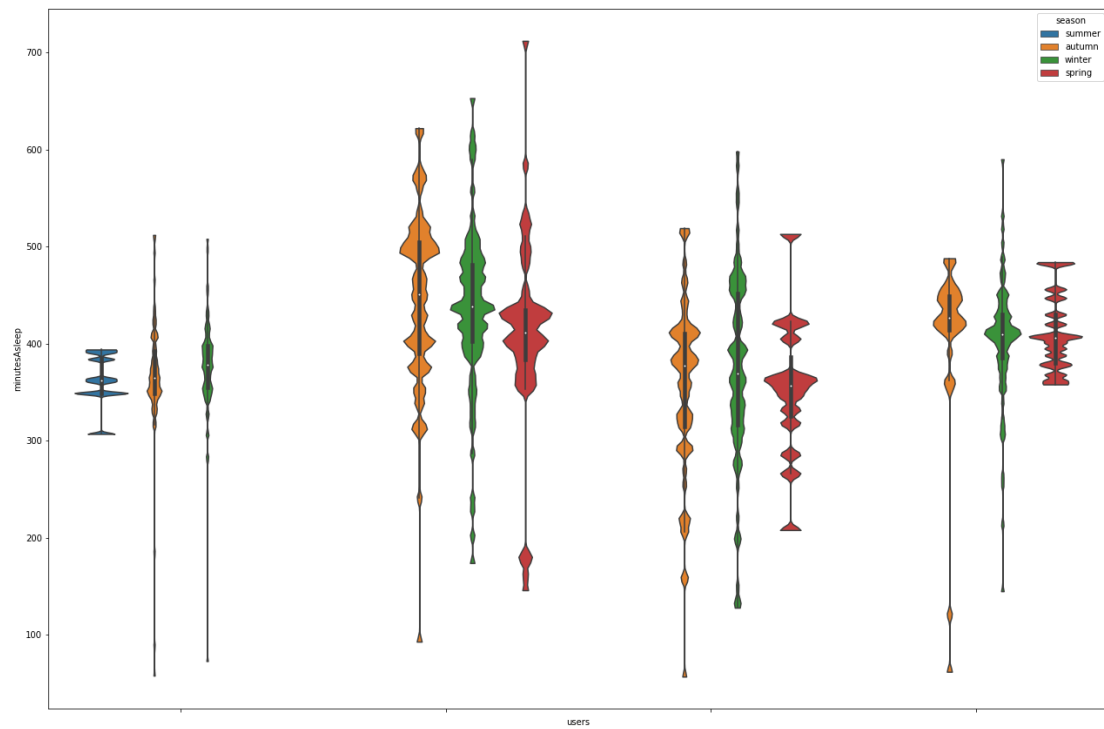


Figura 19: Gráfico violín comparando la variable de la tiempo durmiendo para los usuarios con datos en tres estaciones diferentes.

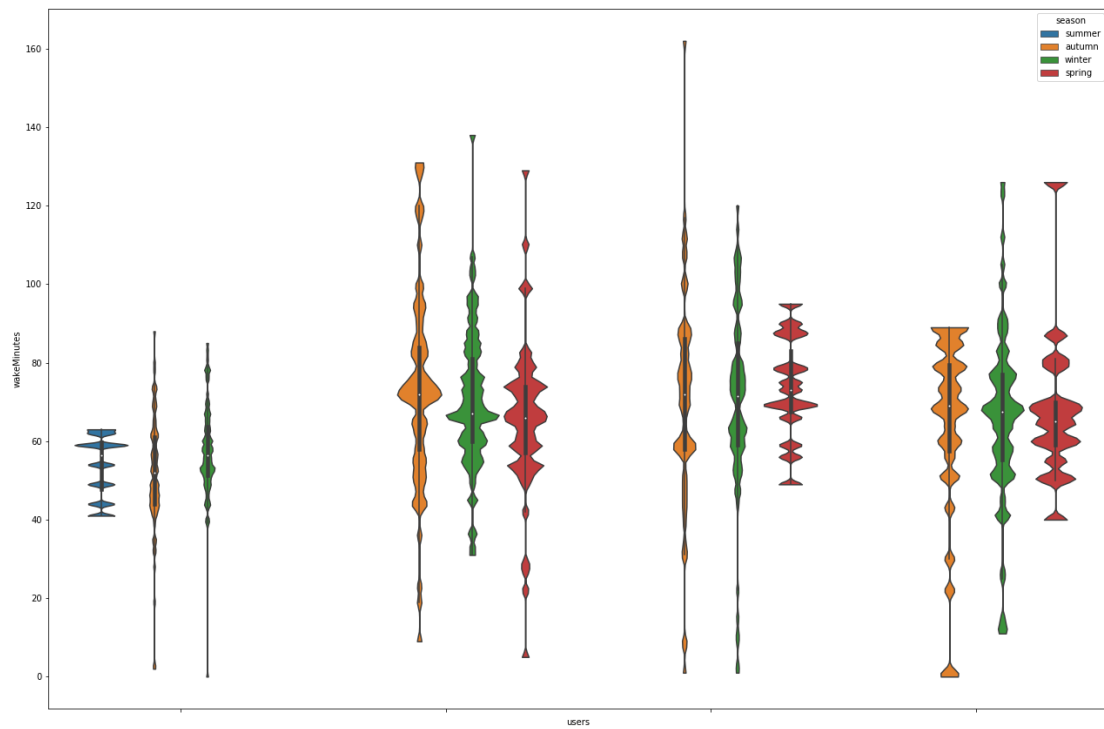


Figura 20: Gráfico violín comparando la variable de la tiempo despierto para los usuarios con datos en tres estaciones diferentes.

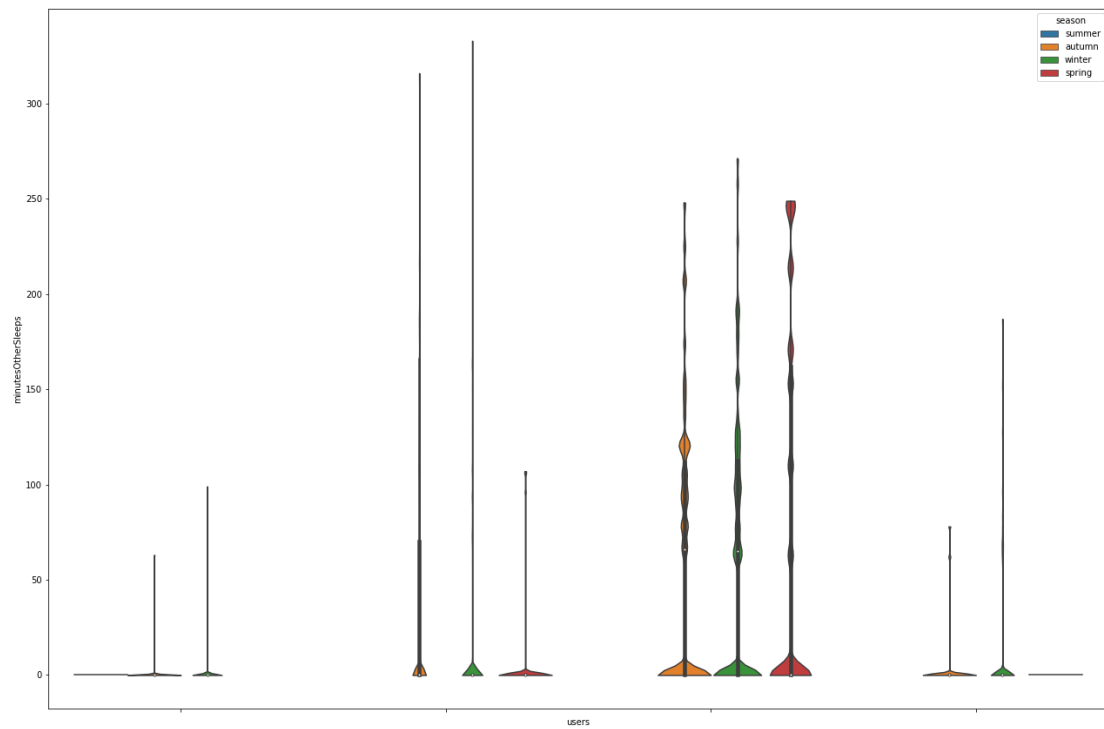


Figura 21: Gráfico violín comparando la variable de la tiempo durmiendo durante el día para los usuarios con datos en tres estaciones diferentes.

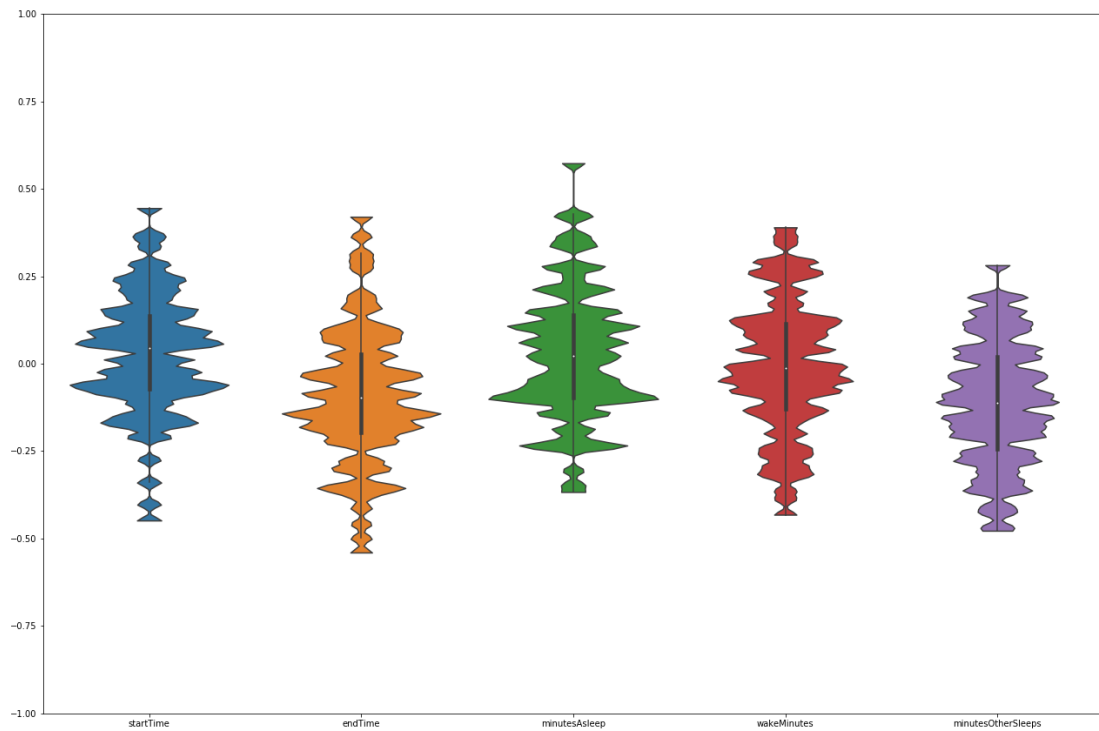


Figura 22: Gráfico violín comparando la correlación de las variables del sueño con el número de pasos.

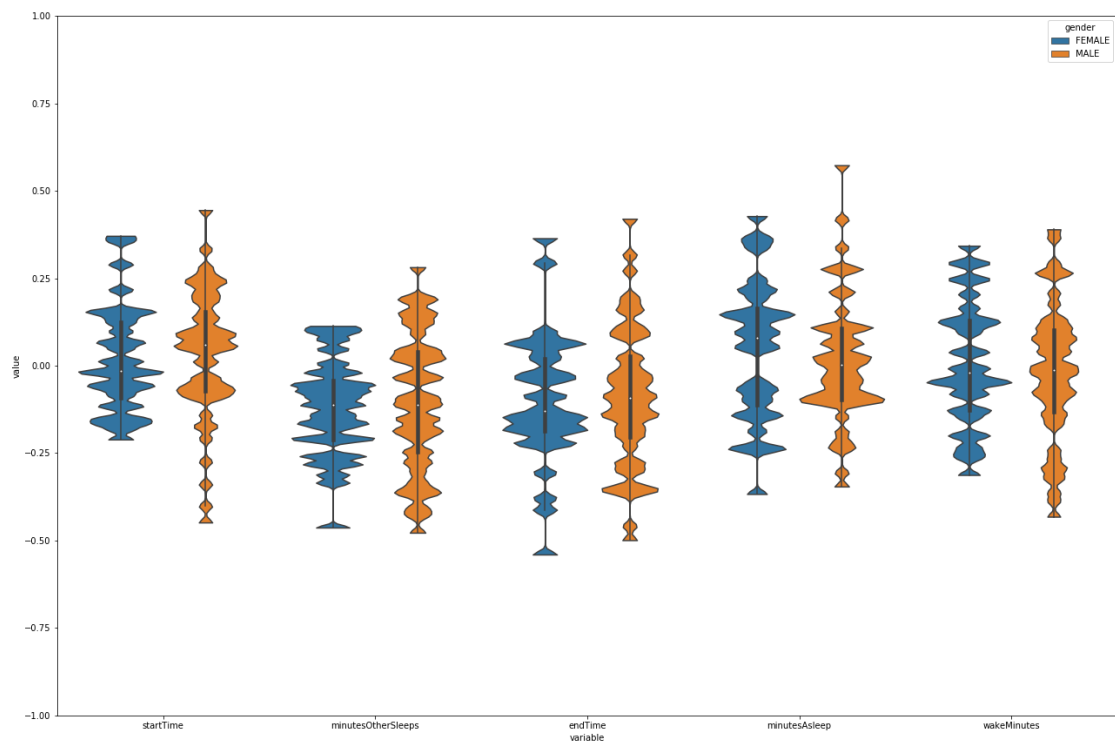


Figura 23: Gráfico violín comparando la correlación de las variables del sueño con el número de pasos, separando los datos por género.

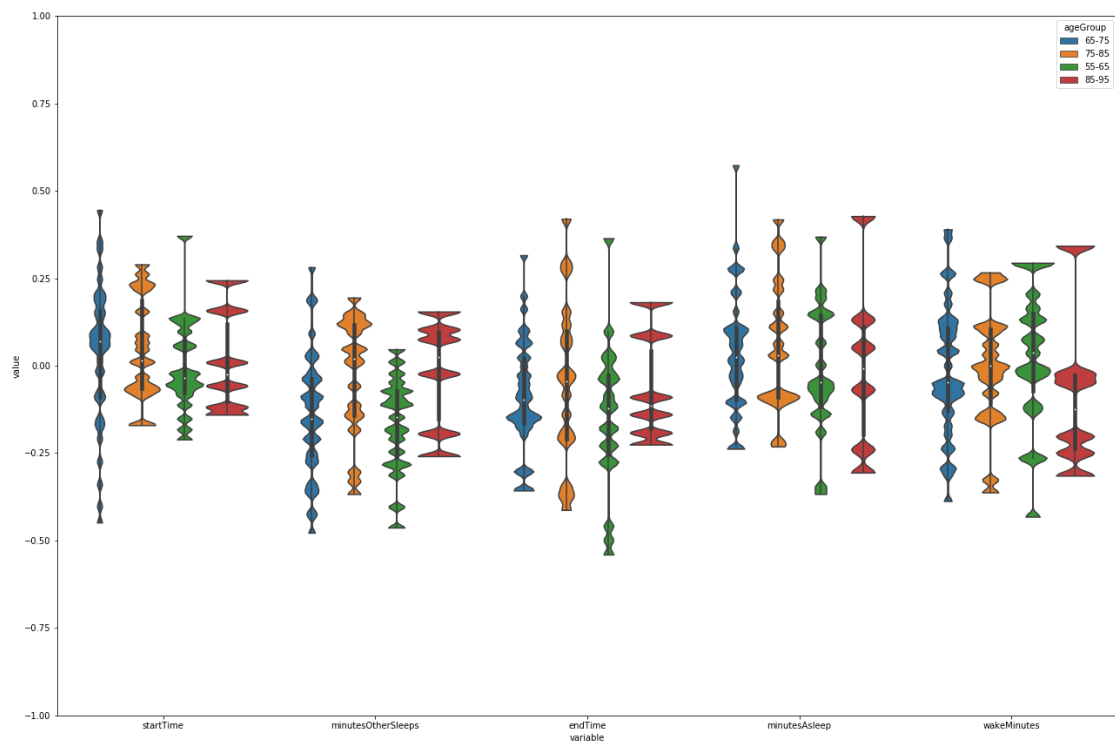


Figura 24: Gráfico violín comparando la correlación de las variables del sueño con el número de pasos, separando los datos por rango de edad

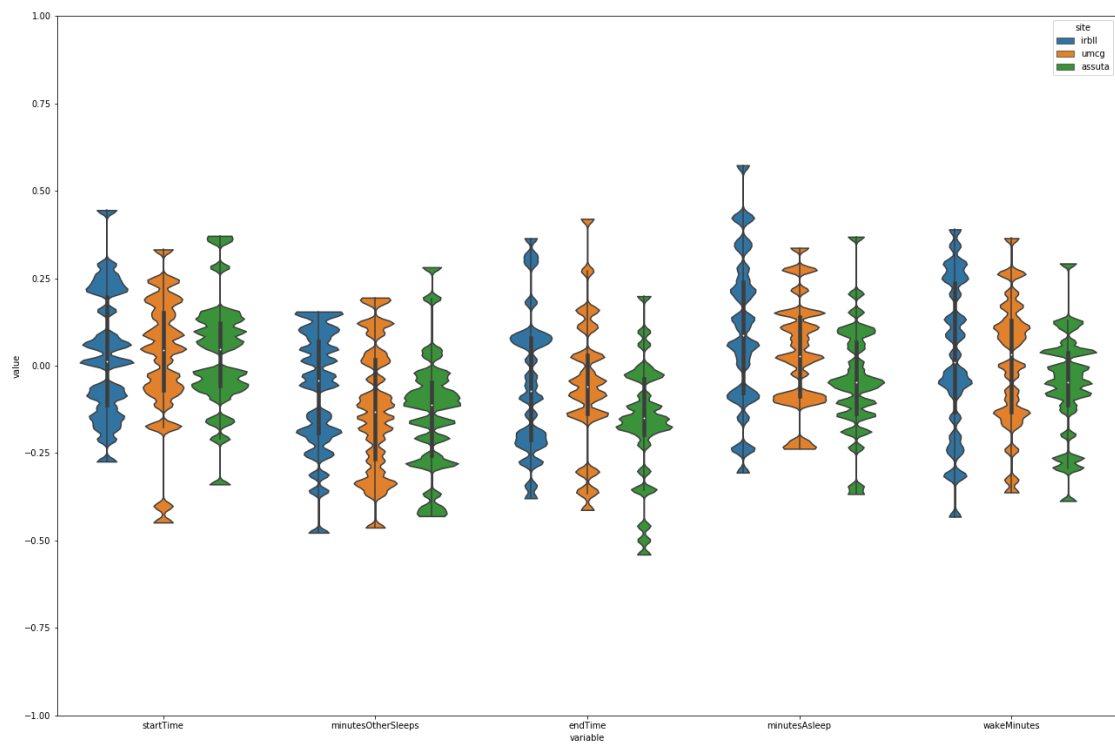


Figura 25: Gráfico violín comparando la correlación de las variables del sueño con el número de pasos, separando los datos por *site*.

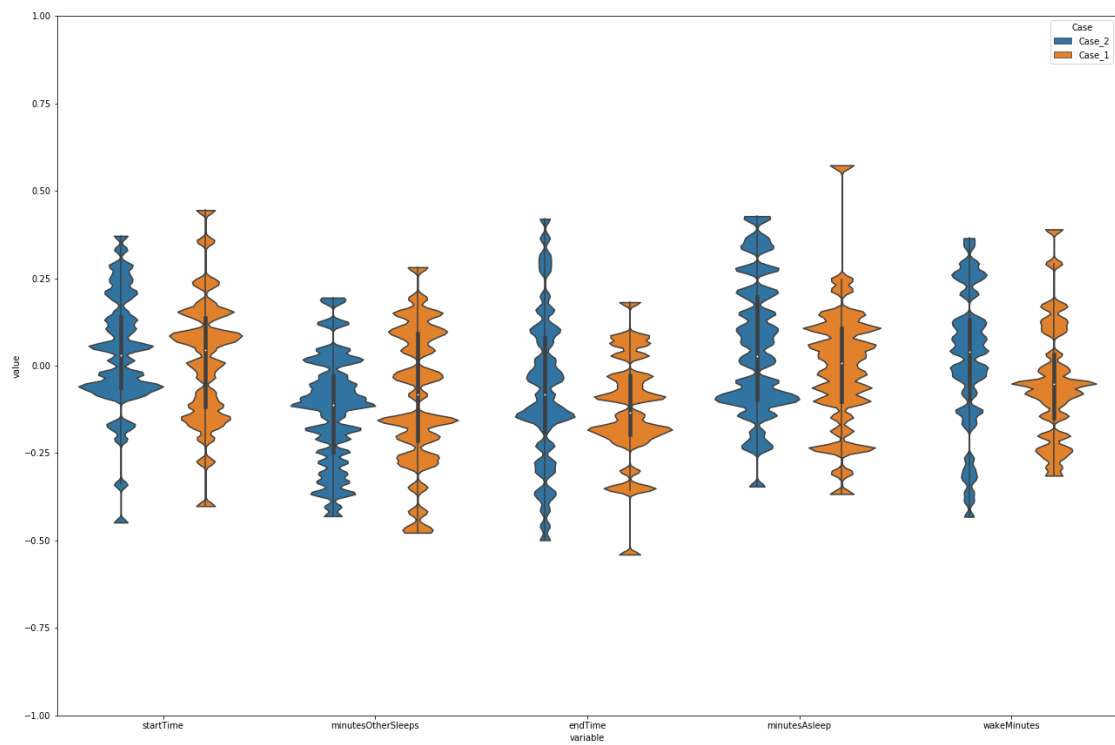


Figura 26: Gráfico violín comparando la correlación de las variables del sueño con el número de pasos, separando los datos por caso de estudio.

5. Conclusiones y Trabajo Futuro

Este apartado concluye el trabajo resumiéndolo y describiendo las conclusiones obtenidas. También incluye las limitaciones encontradas durante el análisis y describe posibles líneas de trabajo futuro.

Para vivir más, ser más saludables y más activos, es necesario vivir un estilo de vida saludable incluyendo: una dieta saludable, actividad física, sueño reparador, evitar tóxicos y tener una conciencia plena. Con las nuevas tecnologías, aplicaciones móviles y dispositivos, hay los recursos necesarios para ayudar a las personas que quieran llevar un estilo de vida saludable. Para que esta ayuda sea realmente efectiva debería ser personalizada y basada en el cambio de hábitos. Otra parte fundamental es tener una visión amplia y que englobe todas las partes que incluye una vida saludable. El análisis realizado en este trabajo va en esa dirección, enfocándose en sueño. Pues la investigación de qué variables afectan al sueño es fundamental para poder ayudar a que las personas tengan un sueño saludable.

El primer paso de este trabajo ha sido realizar un análisis exploratorio de los datos históricos de la actividad del sueño de dos voluntarios, recopilados a través de pulseras de la marca Fitbit. Los datos de actividad del sueño que proporciona Fitbit se han explorado mediante la visualización de gráficos en forma de series de tiempo y gráficas de dispersión. En particular, las variables *startTime* y *endTime* han sido exploradas con el algoritmo Kmeans. Gracias a este análisis exploratorio se han podido sacar conclusiones primordiales para el preproceso de los datos de Fitbit para el posterior análisis. Seguidamente se enumeran las conclusiones obtenidas:

- las variables *minutesFallToSleep* y *asleepCount* no aportan información;
- los dispositivos que disponen de frecuencia cardíaca pueden funcionar en dos modos: modo fases del sueño y modo clásico. Si las variables correspondientes al modo se quieren usar deben ser agregadas para que todos los períodos de sueño tengan las mismas variables;
- en el caso de querer analizar los datos de actividad del sueño se deberían agregar todos los datos de actividad del sueño por usuario y por día. Para tal propósito se debería escoger cual es el período de sueño principal de cada día.

A continuación, se han estudiado dos variables relacionadas con el pilar del sueño. Para dicho estudio se han utilizados datos históricos del proyecto CONNECARE. Las variables estudiadas son la estación del año y la actividad física. Los datos de la estación del año han sido analizados más profundamente, pues solo había 15 usuarios de los cuales se disponían datos suficientes de más de una estación. Así se han utilizado el diagrama violín para visualizar la distribución de cada variable del sueño por usuario y por estación. Los resultados han mostrado que solo en el tiempo total de sueño y el tiempo despierto durante la noche hay pequeñas variaciones entre las estaciones. No obstante, dichas variaciones son pequeñas y no son homogéneas para los usuarios de los mismos grupos, pues no podemos concluir que la estación afecte al sueño.

Por otro lado se ha analizado si la actividad física afecta al sueño, comparando el número de pasos con las variables del sueño. En este caso se ha calculado para cada usuario la correlación entre las variables a estudiar y después se han visualizado los resultados mediante gráficas de violín. Los resultados de este análisis muestran que para todos los usuarios la correlación entre todas las variables del sueño y los pasos está entre 0,5 y $-0,5$, con la media en el cero. La conclusión extraída con estos resultados es que no hay una asociación entre la actividad física y la actividad del sueño.

Las conclusiones obtenidas en este trabajo sobre la relación entre la estación del año y la actividad física con el sueño pueden ser debidas a las limitaciones que presenta el análisis hecho. Al tratar con datos reales, ha habido limitaciones en el número de datos que se tenía a disposición. Aún de tener más de 100 usuarios, algunos no aportaban suficientes datos o eran discontinuos para poder hacer un análisis adecuado. Al final se reducía considerablemente la cantidad de la muestra ha utilizar.

Por otra parte, en este trabajo solo se han podido analizar dos de las variables que según la explicación de los clínicos, son factores que pueden influir en el sueño. A continuación se detallan los factores no analizados: factores externos como las horas de sol diarias, la calidad del aire, la contaminación acústica, la temperatura y luminosidad, la humedad y la comodidad de la cama; y factores relacionadas con hábitos de la vida como el tipo de dieta, el uso de dispositivos móviles antes de acostarse, el consumo de cafeína y fumar. Al utilizar datos históricos de otro proyecto estas variables no fueron recogidas y por este motivo no se han podido analizar.

Por el motivo expuesto, como trabajo futuro queda analizar todas las variables definidas por los clínicos que no se han podido analizar en este trabajo. Estas variables se han empezado a recolectar en la fase dos de este proyecto y se podrían analizar con las mismas metodologías utilizadas en este trabajo.

Finalmente, también se podría seguir investigando la relación de la actividad física con las variables del sueño des de dos puntos de vista. Por un lado, se podría avanzar el análisis utilizando las otras variables de la actividad física (sedentaria, ligera, moderada y vigorosa) para compararlas con las variables de la actividad de sueño, de la misma manera que se ha explorado la variable de pasos con las del sueño. Por otro lado, se podrían aplicar otros métodos para calcular la correlación entre dos series temporales como por ejemplo el algoritmo *Dynamic Time Warping* (DTW).

Referencias

- [1] W. H. Organization *et al.*, “Global status report on noncommunicable diseases 2017,” tech. rep., World Health Organization, 2017.
- [2] “Noncommunicable diseases country profiles 2018 - spain.” https://www.who.int/nmh/countries/esp_es.pdf. *Accedido* : 2019 - 06 - 25.
- [3] W. H. Organization *et al.*, “2008-2013 action plan for the global strategy for the prevention and control of noncommunicable diseases: prevent and control cardiovascular diseases, cancers, chronic respiratory diseases and diabetes,” 2009.
- [4] W. H. Organization *et al.*, “Global action plan for the prevention and control of ncds 2013-2020,” 2014.
- [5] “Prevalence of insufficient physical activity among adults data by country.” <http://apps.who.int/gho/data/view.main.2463>. *Accedido*: 2019-06-25.
- [6] “Recorded alcohol per capita consumption, from 2010 updated may 2018.” <http://apps.who.int/gho/data/node.main.A1039>. *Accedido*: 2019-06-25.
- [7] “Tobacco use data by country.” <http://apps.who.int/gho/data/view.main.1805>. *Accedido*: 2019-06-25.
- [8] “World action on salt health - spain.” <http://www.worldactiononsalt.com/worldaction/europe/spain/>. *Accedido*: 2019-06-25.
- [9] M. M. Andréu, A. Á. R. de Larrinaga, J. A. M. Pérez, M. Á. M. Martínez, F. J. P. Cuesta, A. J. A. Guerra, O. R. Santo-Tomás, M. J. J. Luque, F. J. S. Isern, T. C. Sanz, *et al.*, “Sueño saludable: evidencias y guías de actuación. documento oficial de la sociedad española de sueño,” *Rev Neurol*, vol. 63, no. Suplemento 2, pp. S1–S27, 2016.
- [10] “Global recommendations on physical activity for health.” https://apps.who.int/iris/bitstream/handle/10665/44399/9789241599979_eng.pdf. *Accedido*: 2019-06-25.