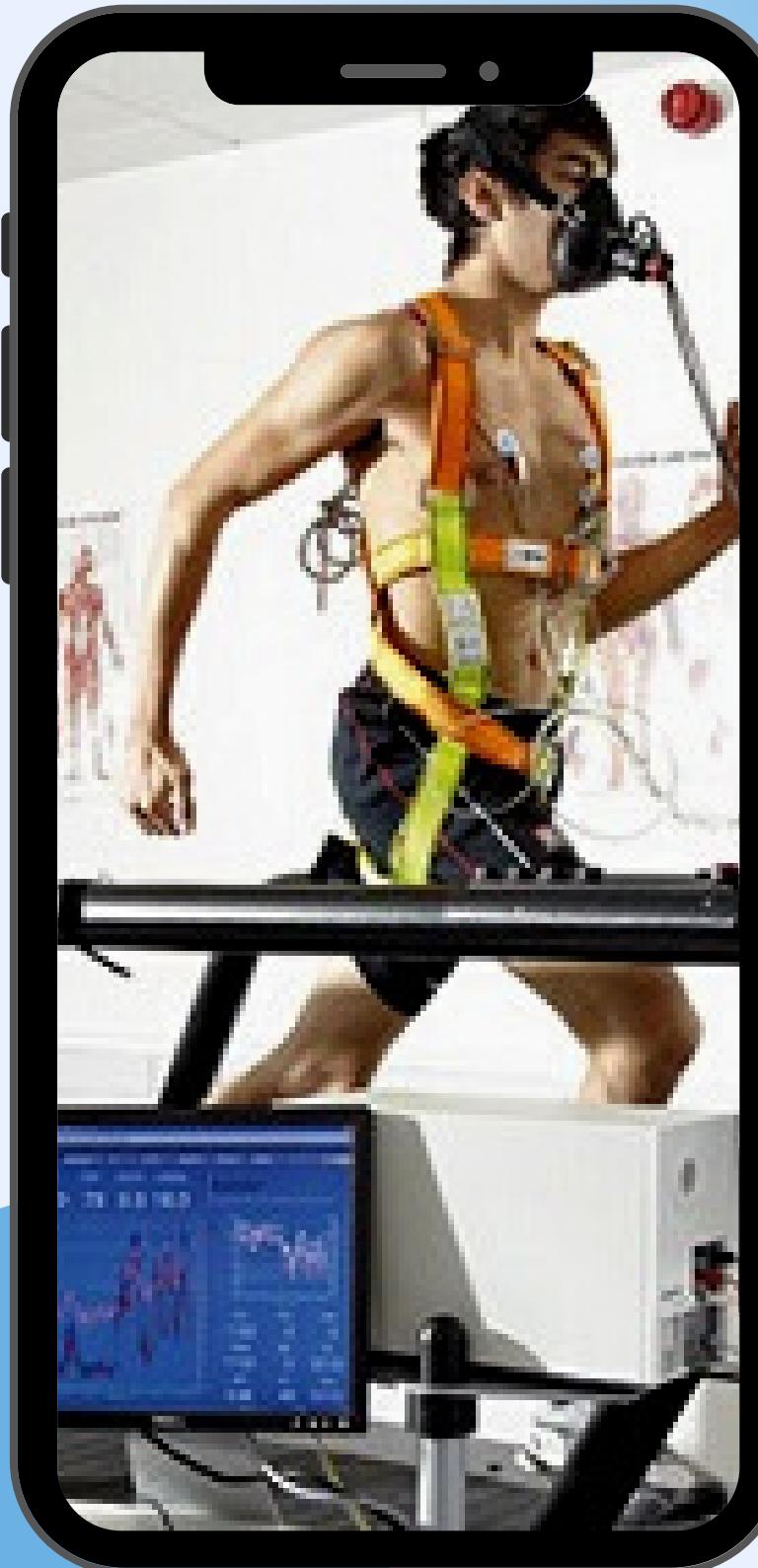


EDA & Statistics Project

Midiendo actividad física y estrés con monitores de salud

2 aproximaciones

Monica Duarte



julio| 2022

Planteamiento del proyecto

1

¿Cómo podemos medir el estrés a partir de los datos de smartwatch?

2

¿Se puede analizar variables de salud para identificar correlación entre ellas?

3

¿Cómo identificar patrones de actividad?



Metodología de trabajo

CRISP-DM



Carga de Datos -> XML

8 voluntarios exportaron sus datos de salud de iphone.



Limpieza de información

Datos duplicados, selección de variables, división de los datos.



Transformación de datos

Tipos de datos, fechas, nombre de columnas, valores numéricos.

Entendimiento de datos

Intervalos de tiempo recolectados, cantidad de datos por persona y por variable, data selection.



Análisis grupal

Detección de patrones, técnicas estadísticas, correlaciones.



Análisis individual

Detección de outliers, técnicas estadísticas, correlaciones.



VARIABLES MEDIDAS

para al menos 1 de las personas del estudio



VARIABLES MEDIDAS

para al menos 1 de las personas del estudio



Días medidos

Por persona por cada variable

Person	Count_days	first_day	Last_day	Heart Rate	Step Count	Distance Walking Running	Active Energy Burned	Body Mass	Distance Cycling	Resting Heart Rate	Heart Rate Variability	Respiratory Rate	Distance Swimming	Distance Downhill Snow Sports
1	171 days	2022-01-18	2022-07-08	21.0	171.0	150.0	171.0							
2	1607 days	2018-02-13	2022-07-09	6.0	1607.0	1607.0	6.0	1434.0	0.0	6.0	6.0			
3	1259 days	2019-01-29	2022-07-11	910.0	1259.0	1259.0	910.0	1195.0						
4	2094 days	2016-10-18	2022-07-13	365.0	2094.0	2094.0	840.0	2025.0						
5	1986 days	2017-01-30	2022-07-09	473.0	1555.0	1555.0	232.0	1538.0						
6	32 days	2022-06-10	2022-07-12	14.0	31.0	31.0	13.0	25.0						
7	1041 days	2019-09-04	2022-07-11		1041.0	1041.0								
8	2357 days	2016-01-28	2022-07-12	1608.0	2357.0	2357.0	2291.0	1984.0	1706.0	854.0	855.0	288.0	707.0	1.0

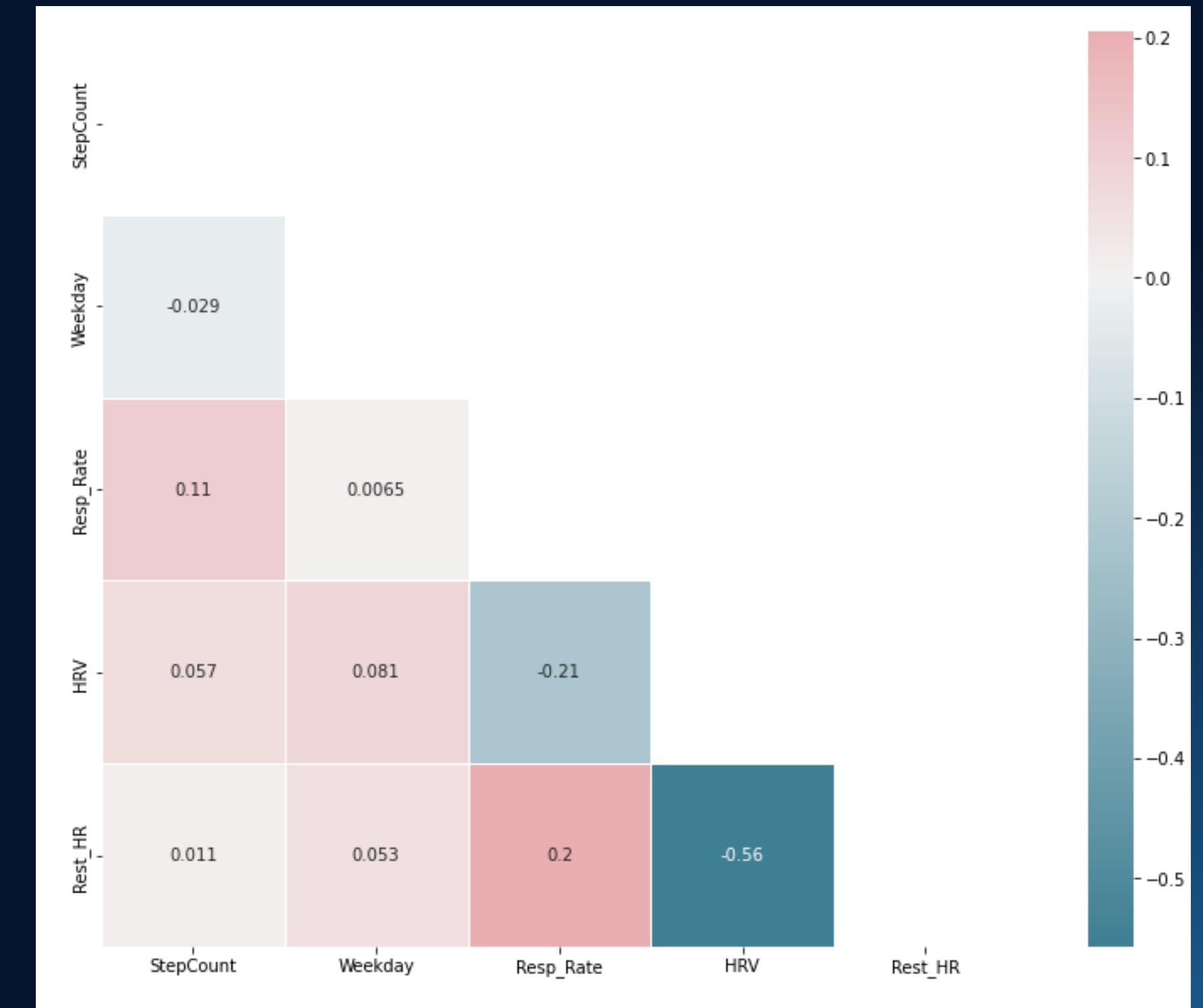
Observaciones

- Pocas variables con suficientes datos como para compararlos.
- Distintos periodos de tiempo para cada persona
- Un solo individuo tiene gran cantidad de datos significativos para análisis individual
- Las conclusiones difícilmente podrán generalizarse.

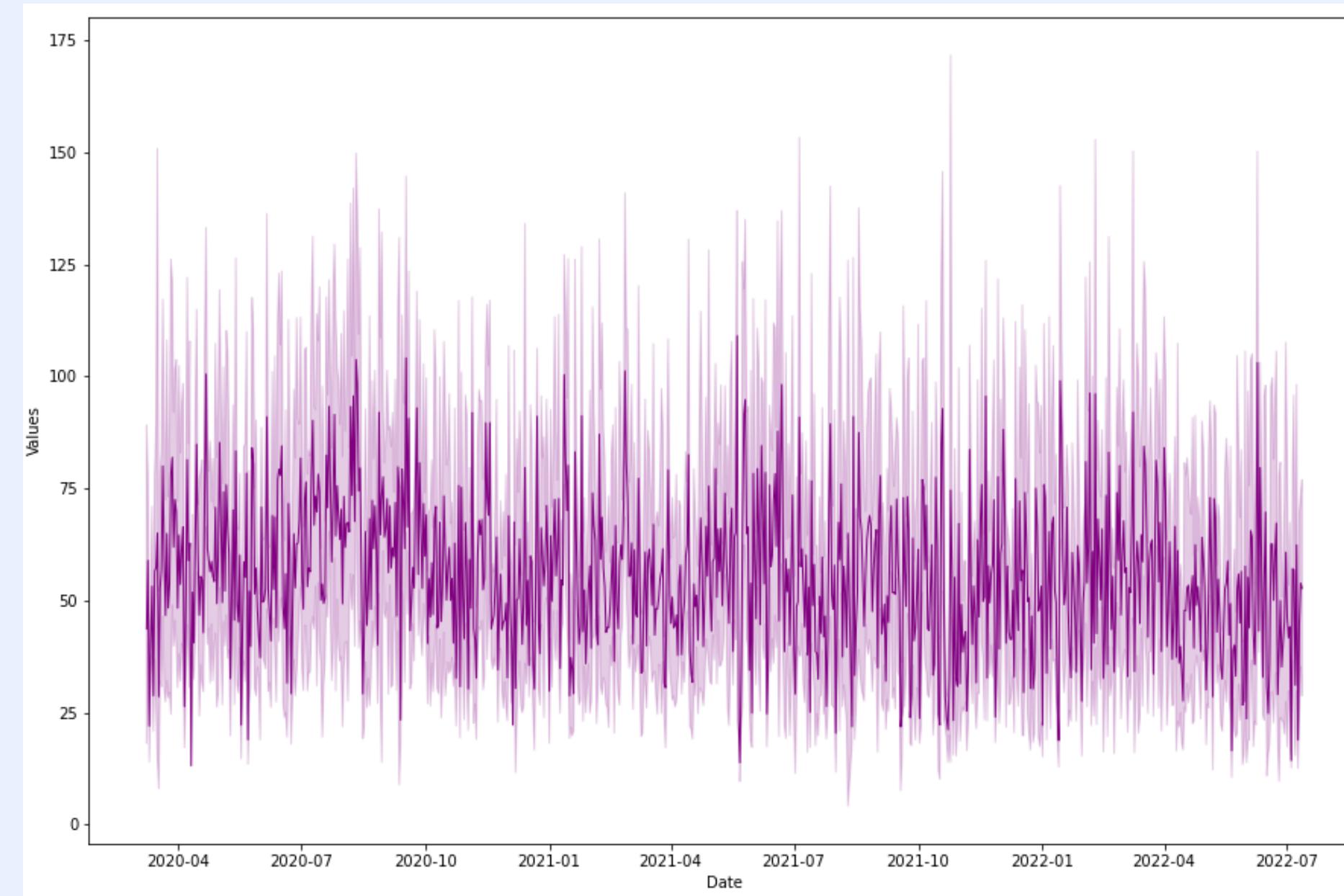
¿Cómo podemos medir el estrés a partir de los datos de smartwatch?

Variables que tomaremos en cuenta según la literatura médica:

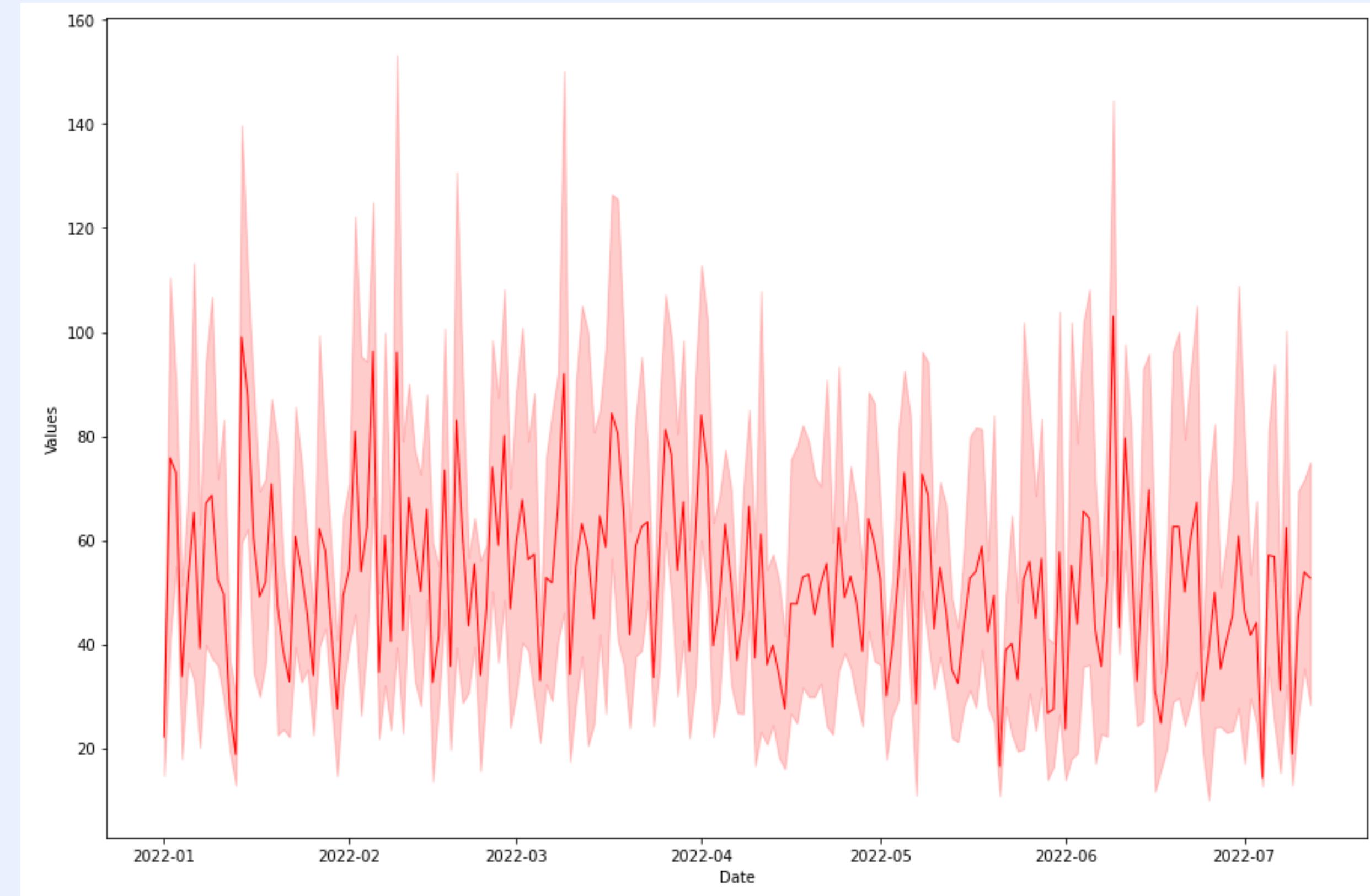
- Variabilidad del ritmo cardíaco
- Frecuencia cardíaca en reposo
- Ritmo respiratorio
- Pasos diarios
- Tiempo de sueño



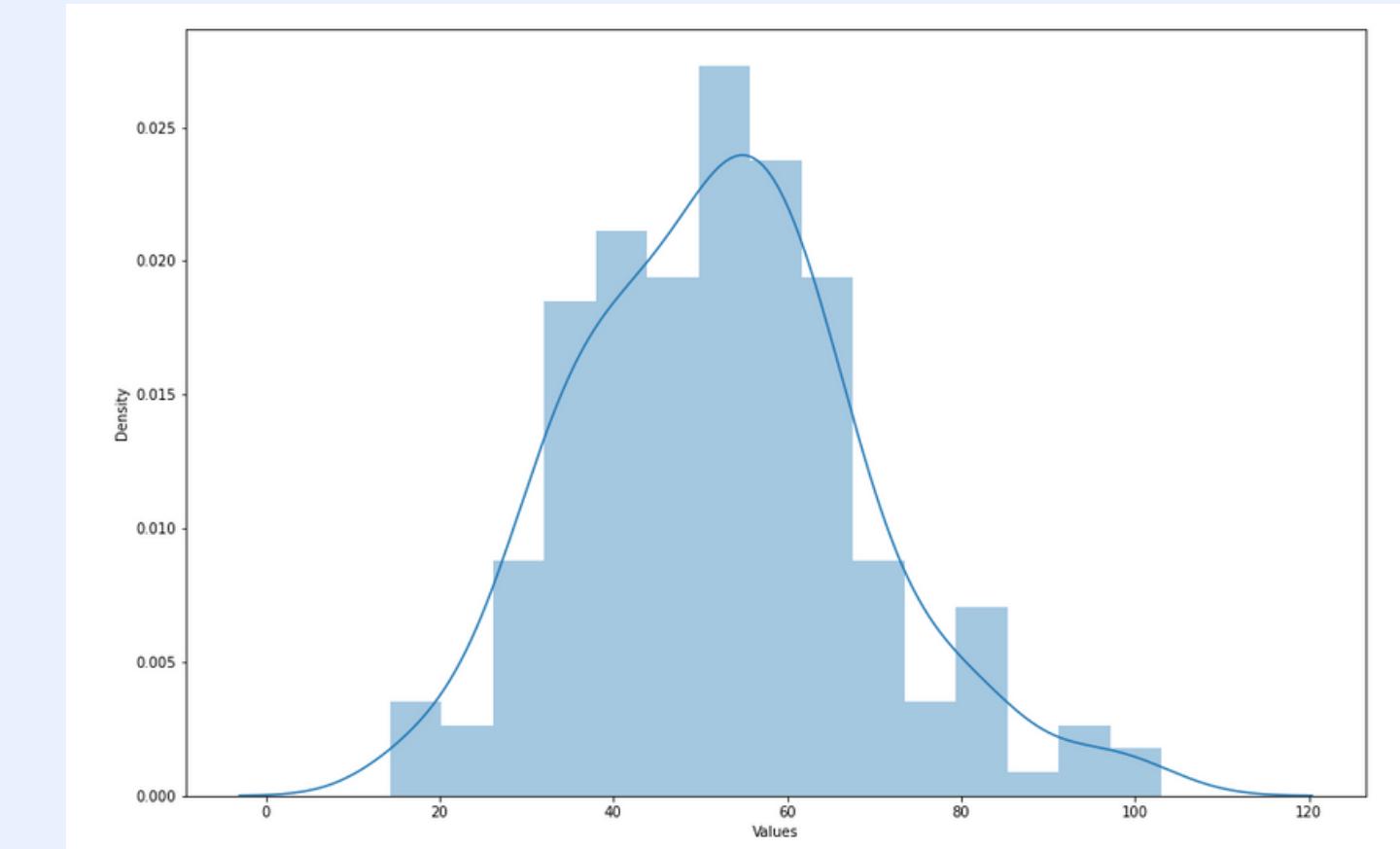
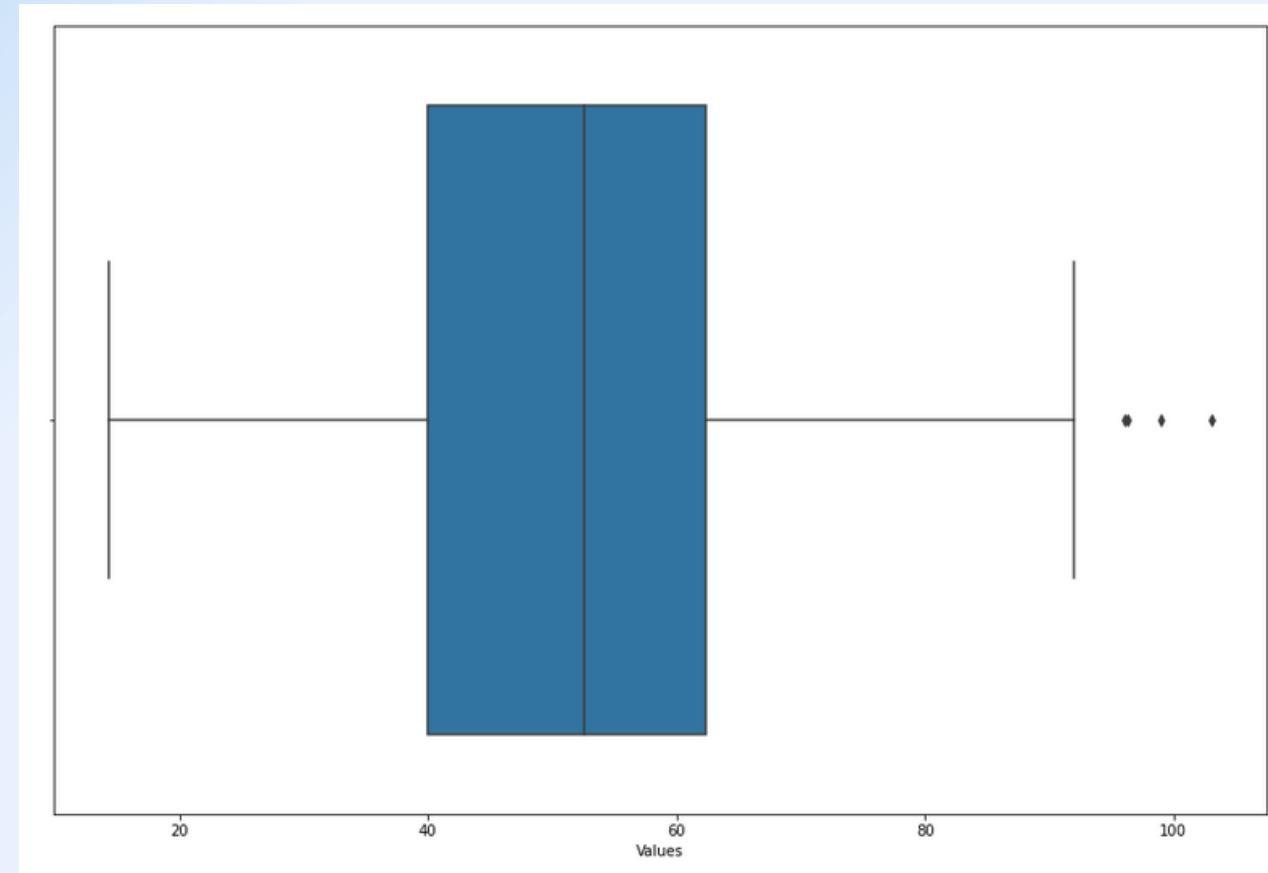
Variabilidad del ritmo cardiaco



Variabilidad del ritmo cardiaco



Variabilidad del ritmo cardiaco



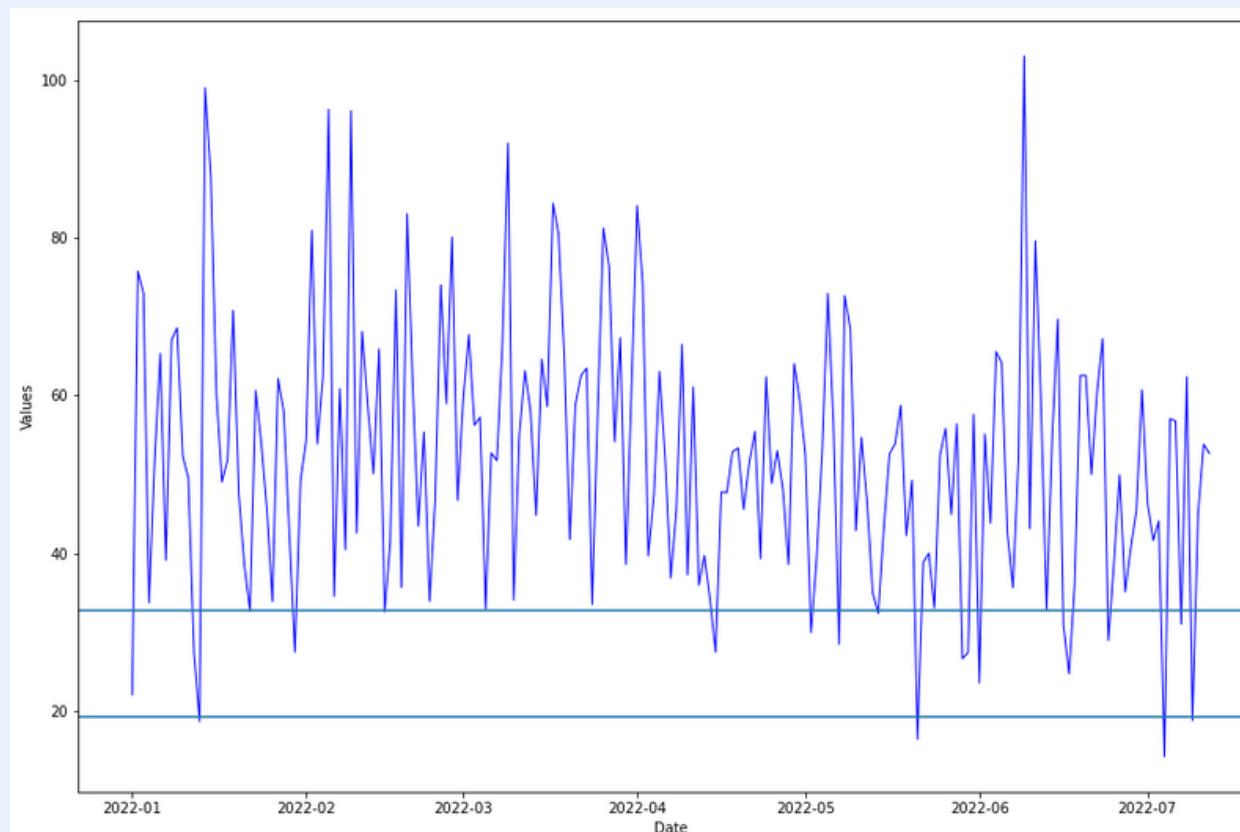
Media: 52.30

Desviación estándar: 16.52

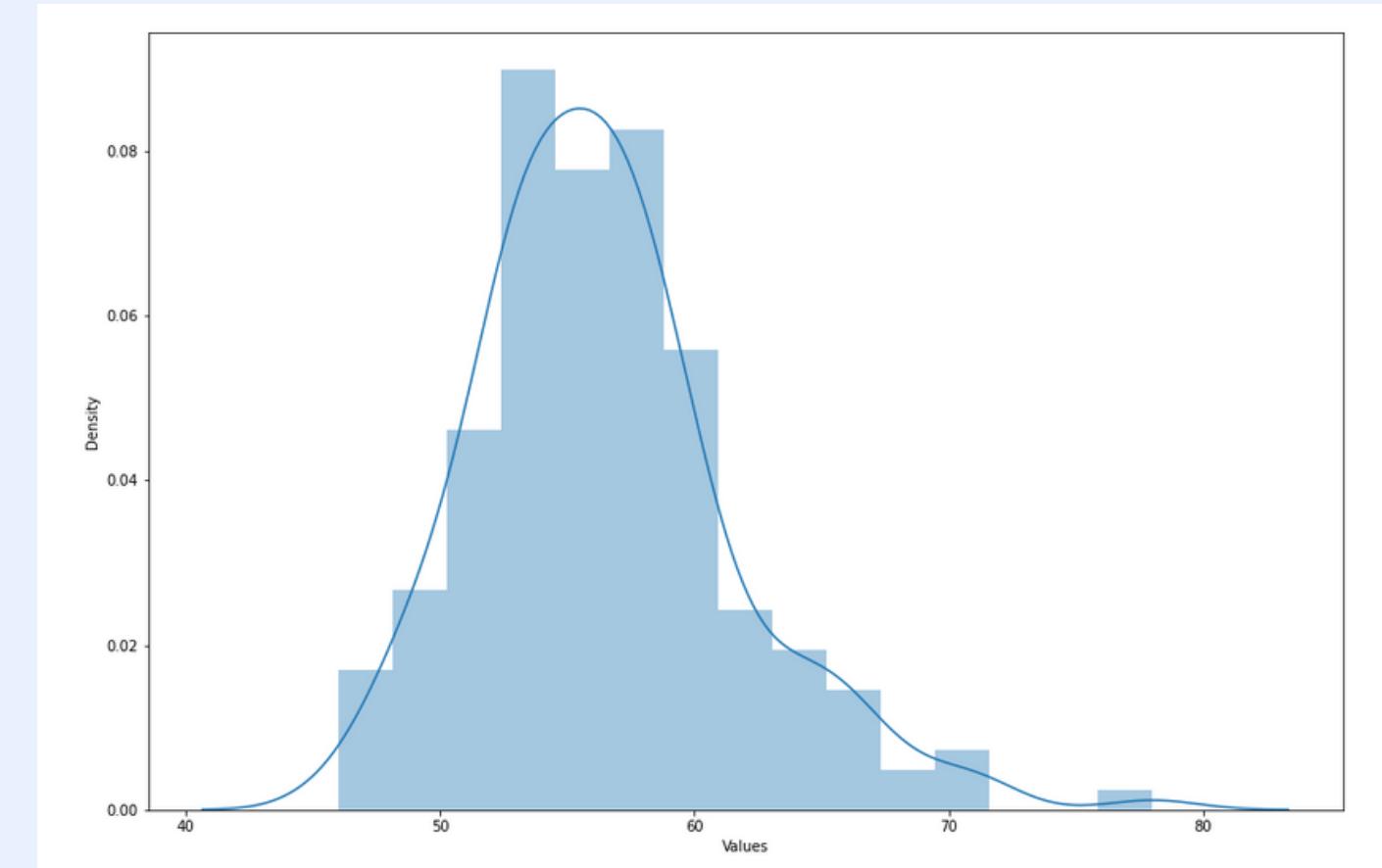
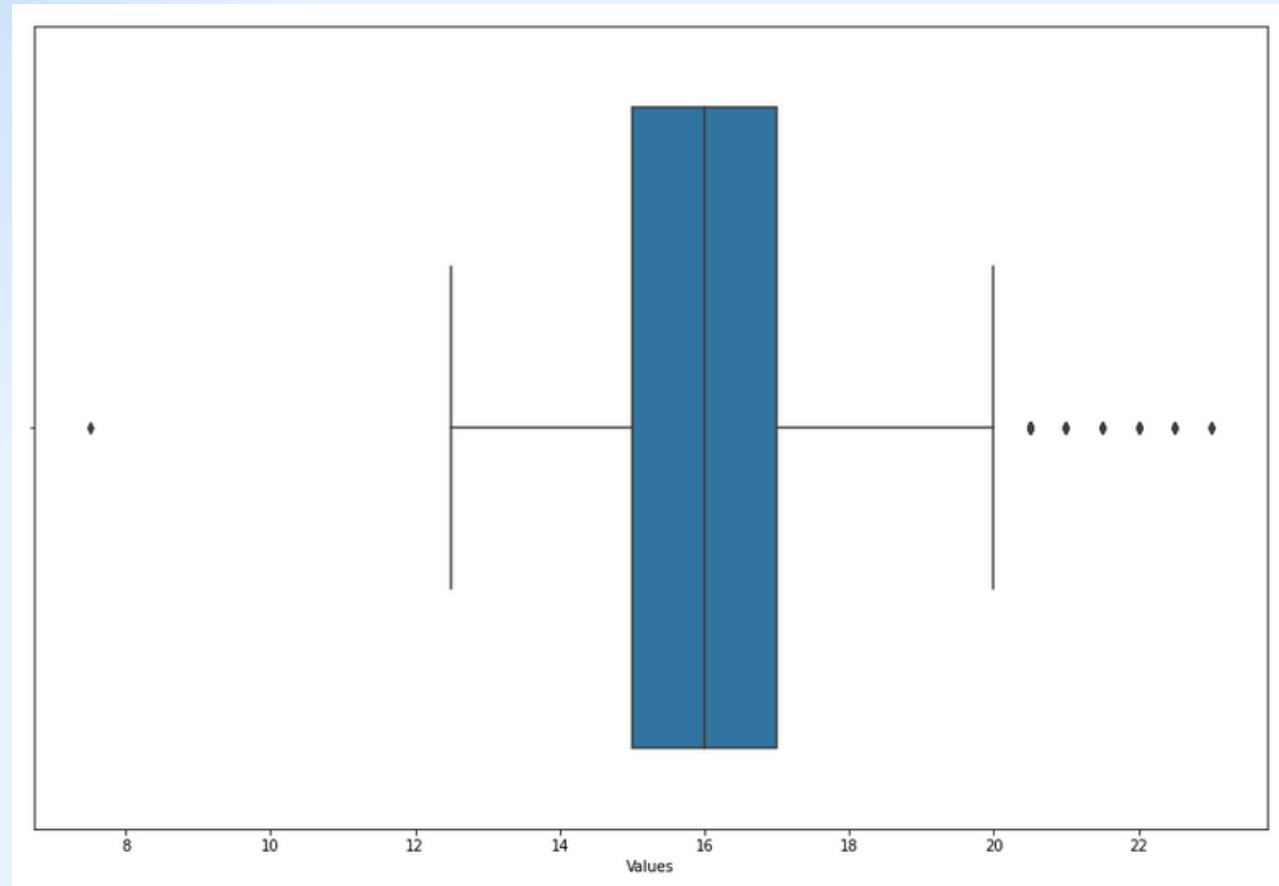
Líneas de outliers:

2 std = 19.25

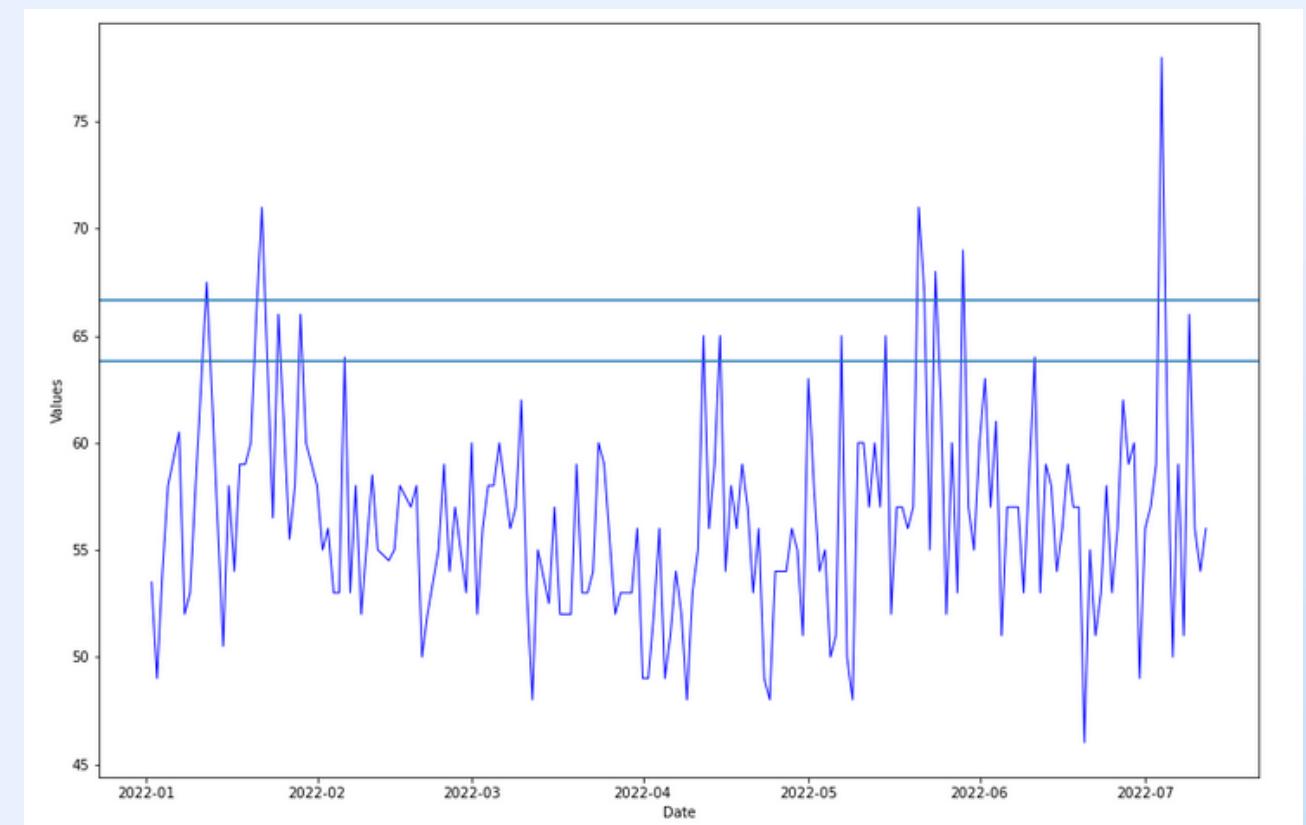
percentil 10 = 32.76



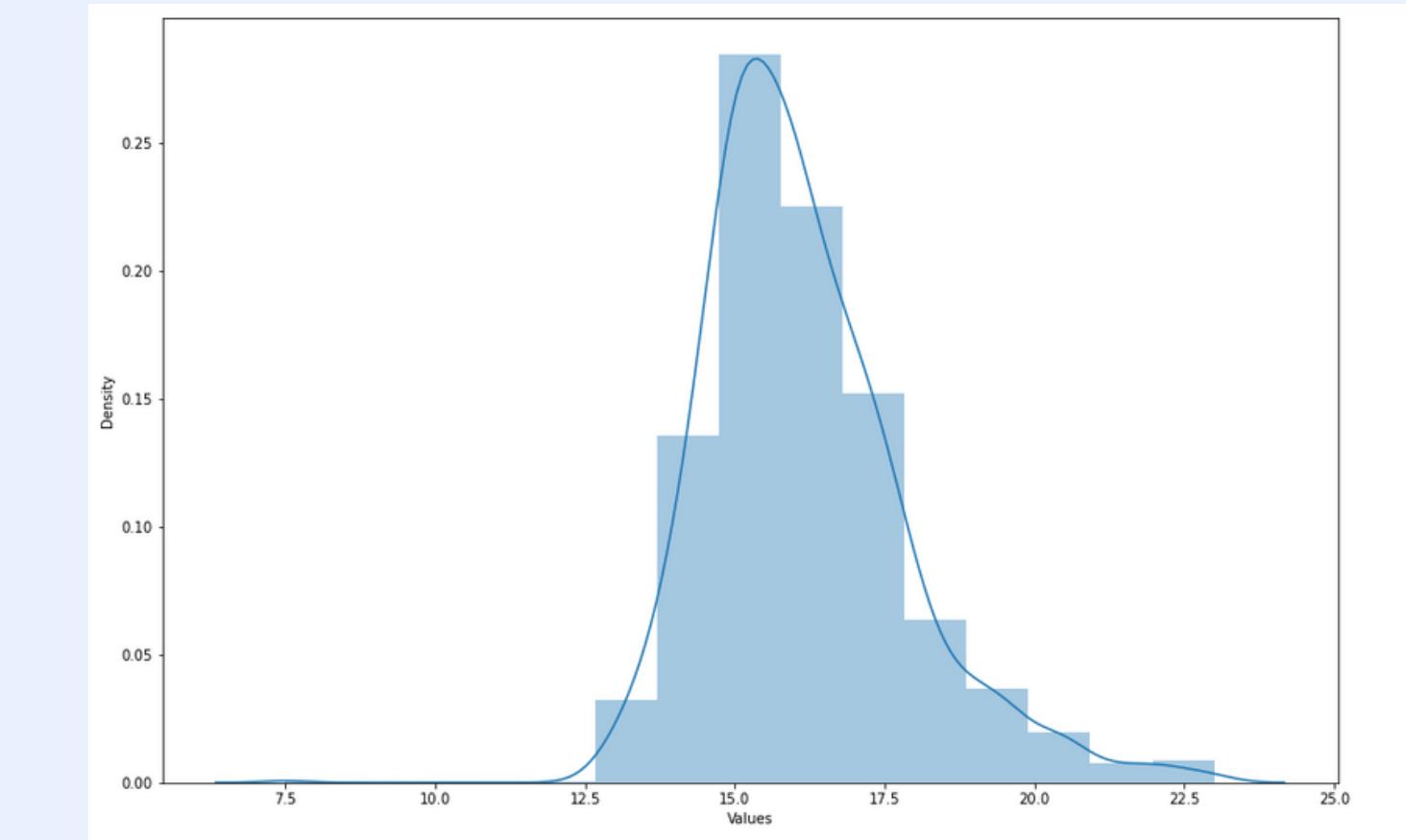
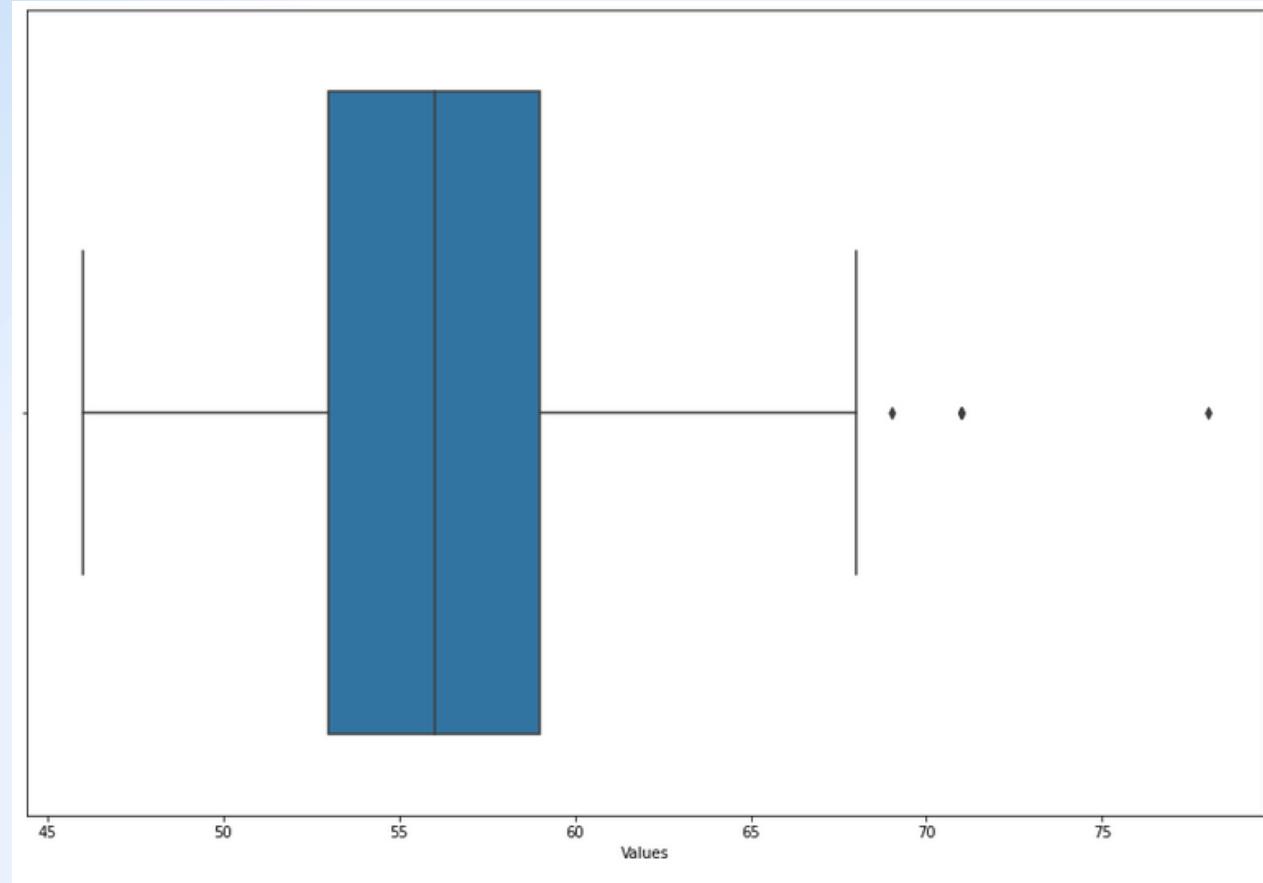
Frecuencia cardiaca en reposo



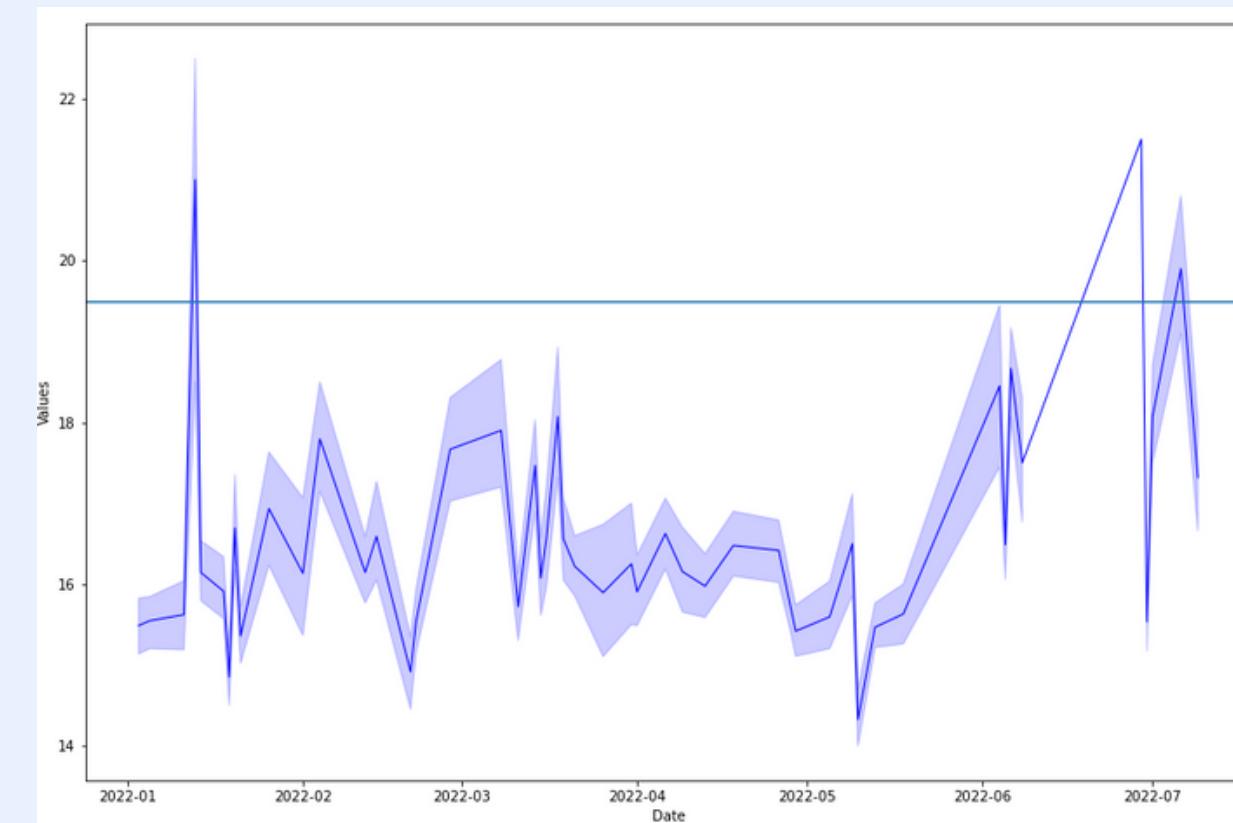
Líneas de outliers:
2 std = 66.65
percentil 90 = 63.80



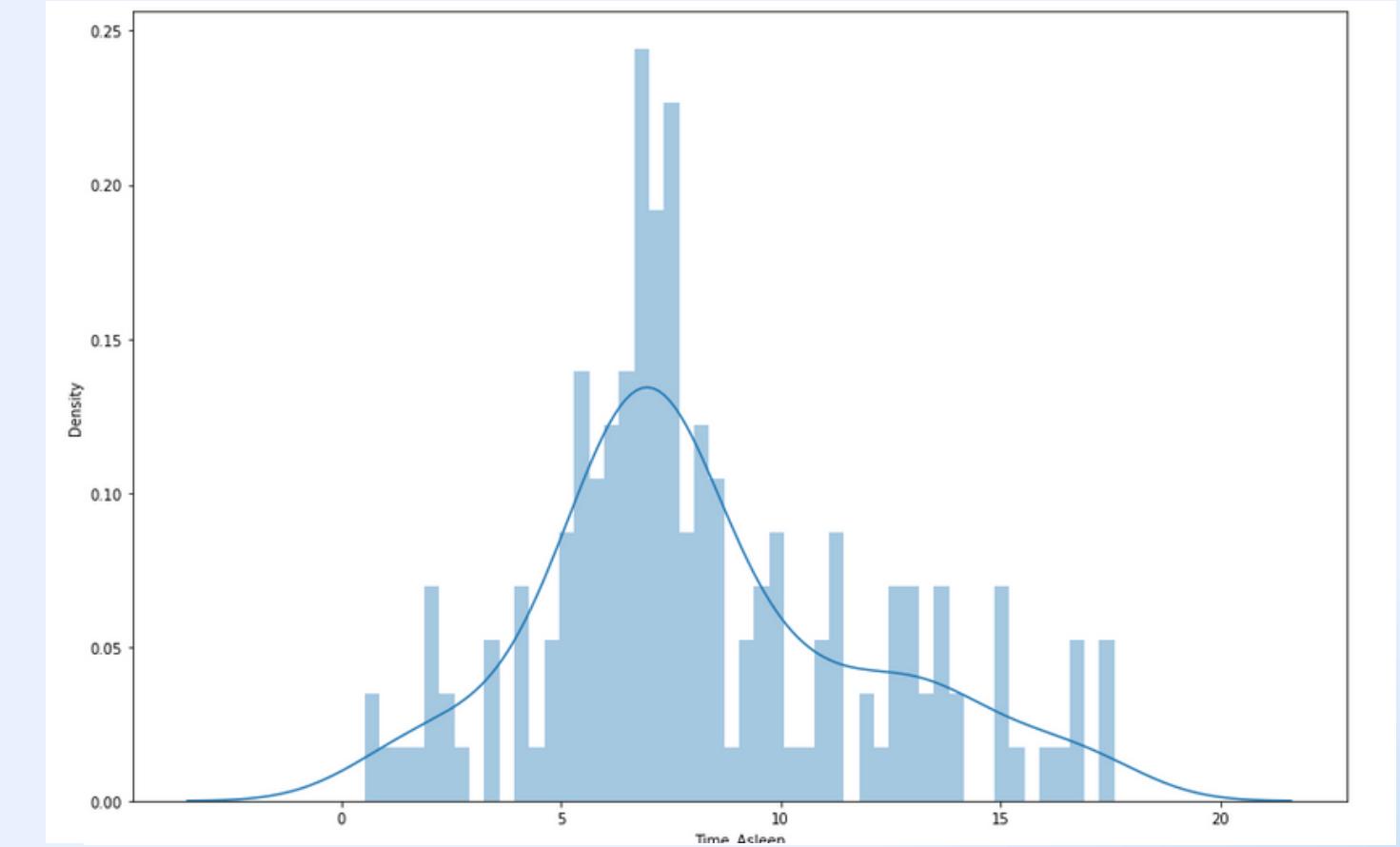
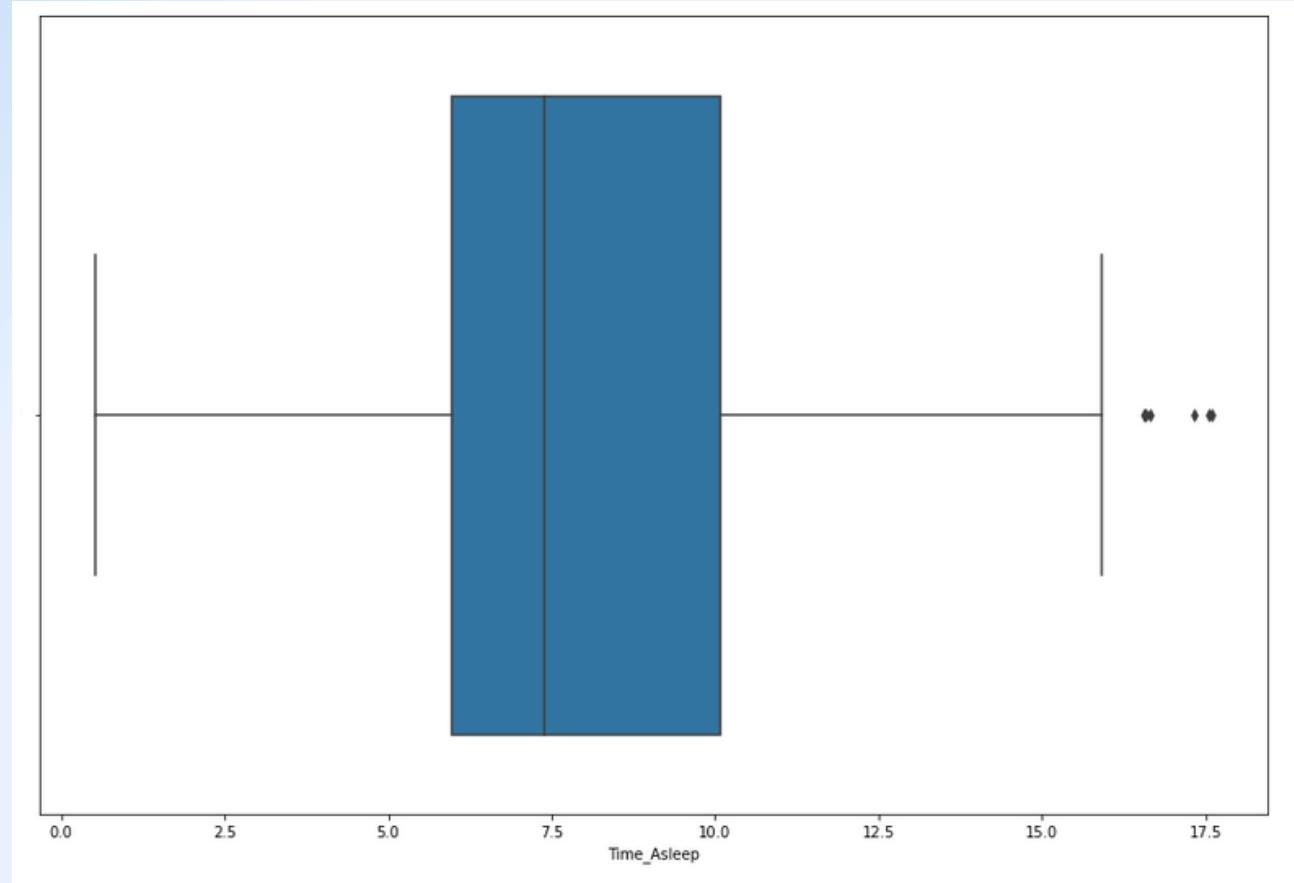
Ritmo respiratorio



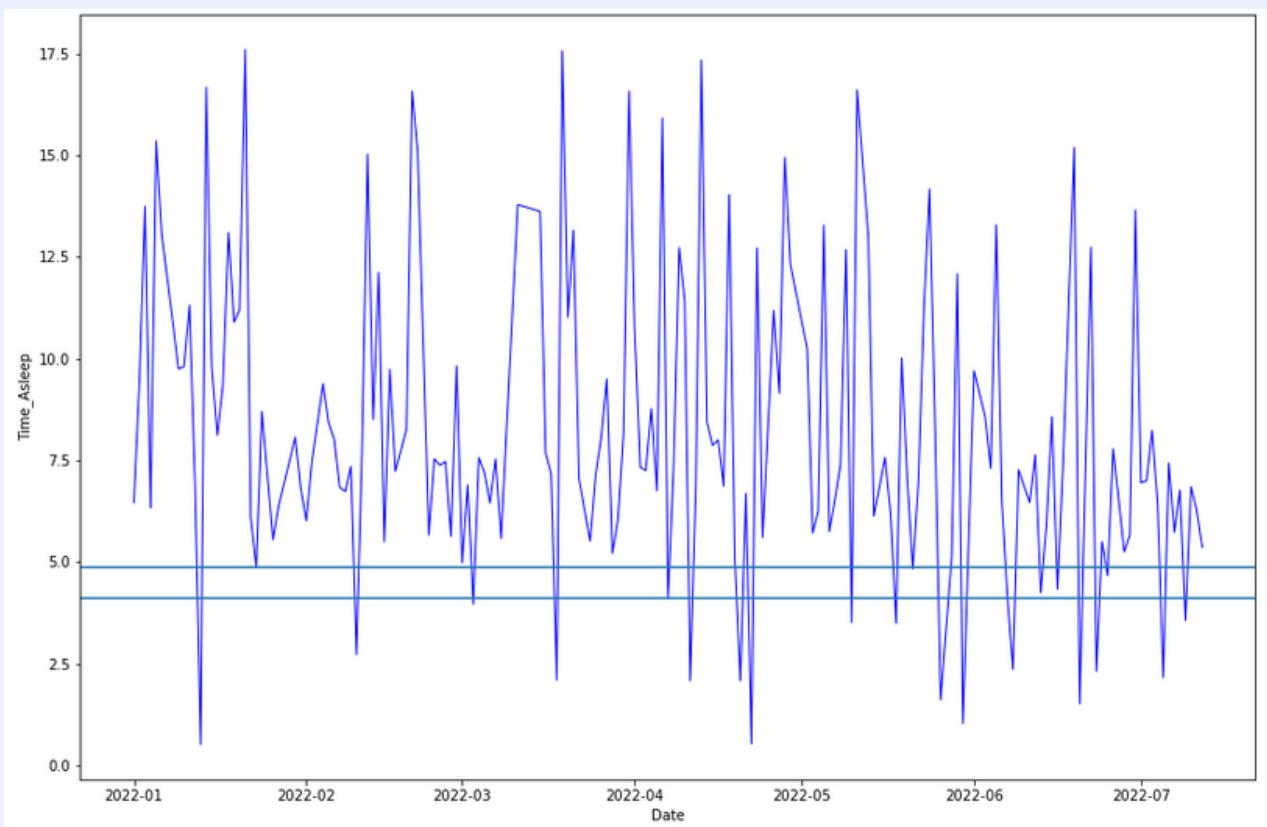
Líneas de outliers:
2 std = 19.484
percentil 90 = 18



Tiempo de sueño

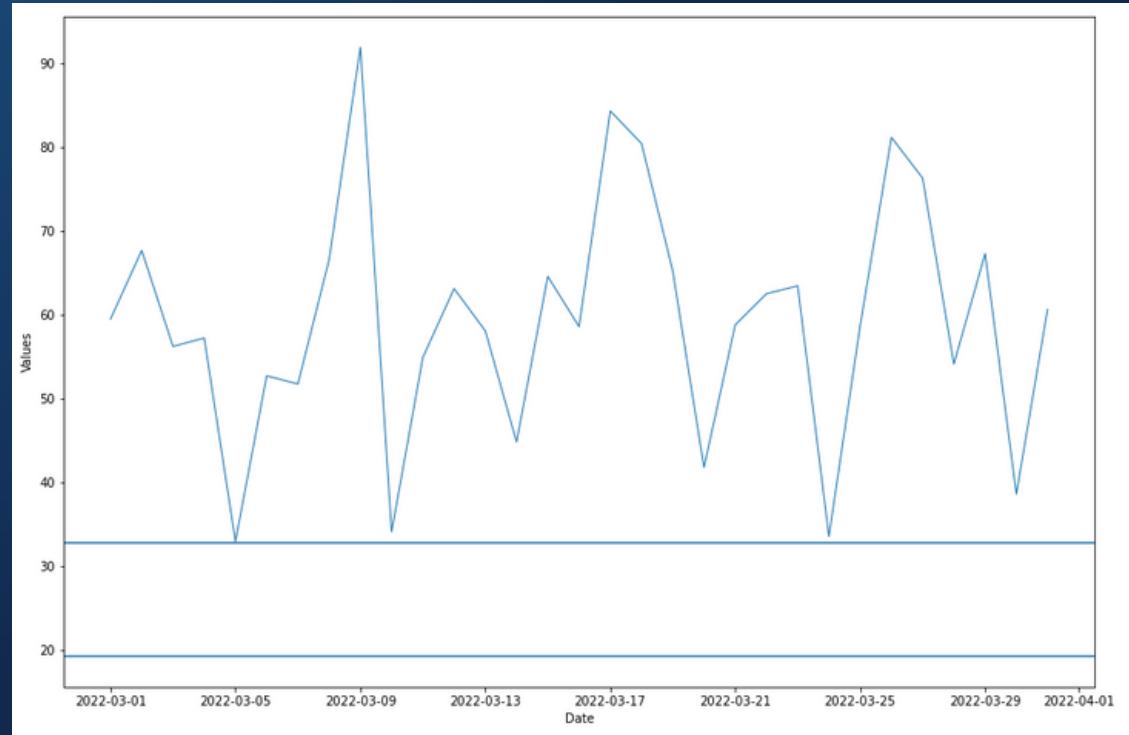


Líneas de outliers:
 $2 \text{ std} = 4.86$
percentil 10 = 4.12



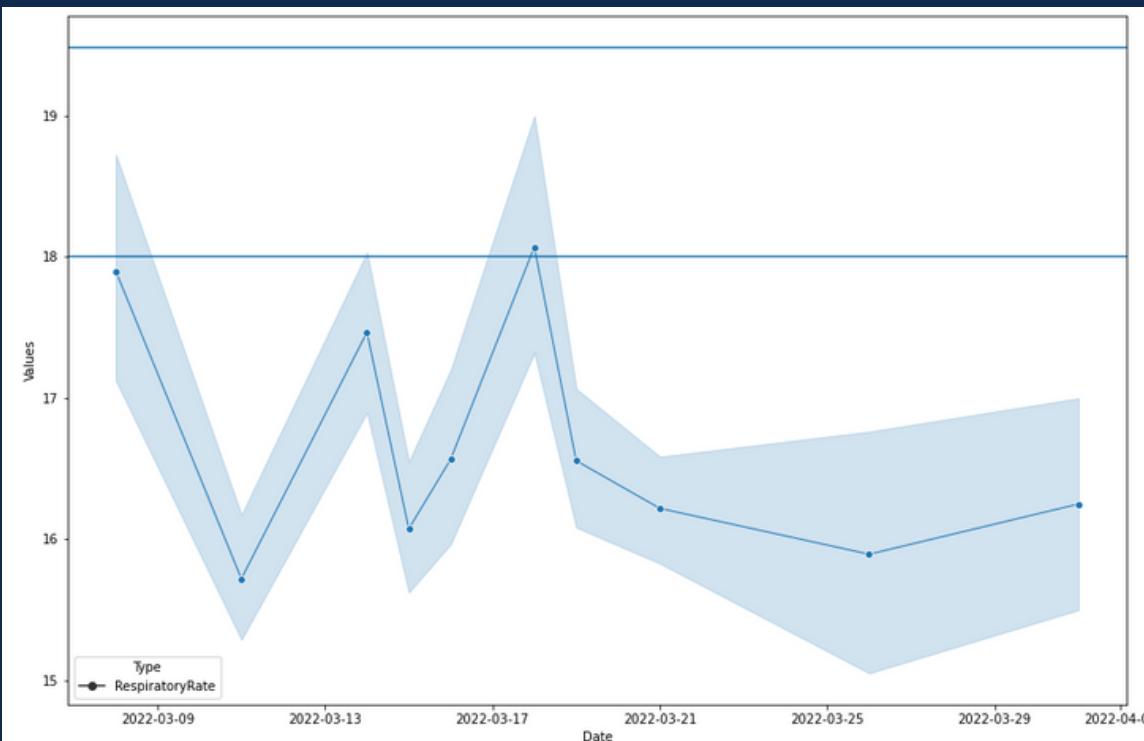
Marzo

VRC



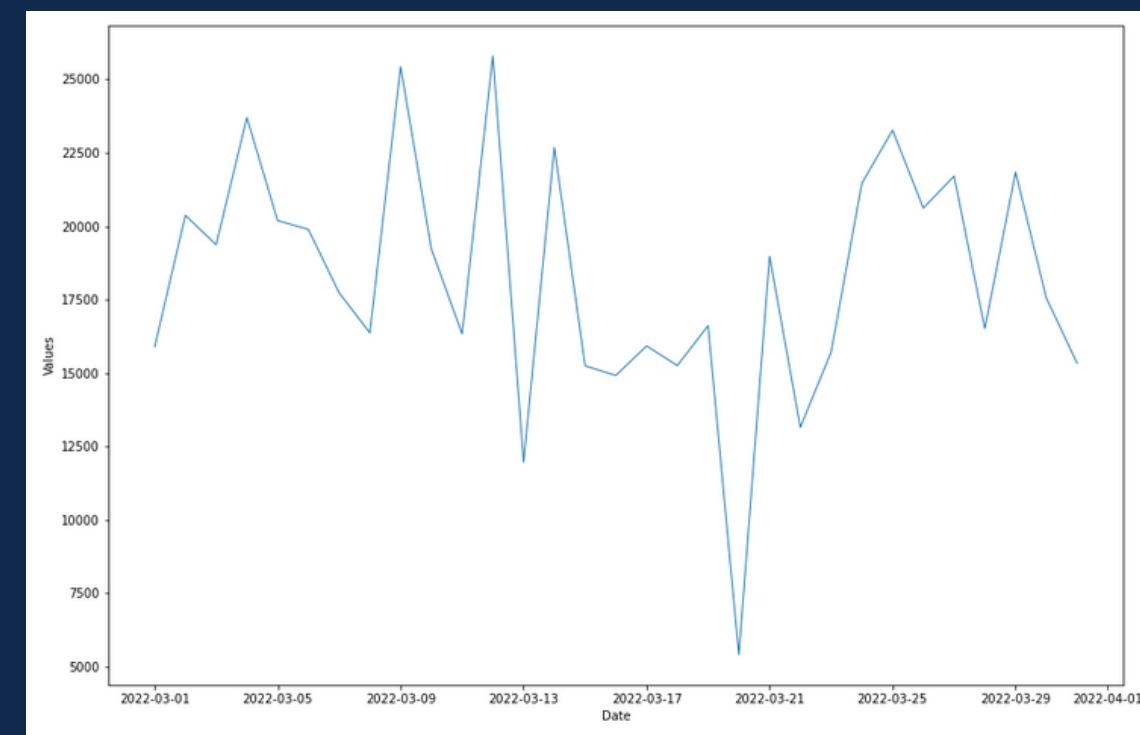
5/10/24

Ritmo respiratorio



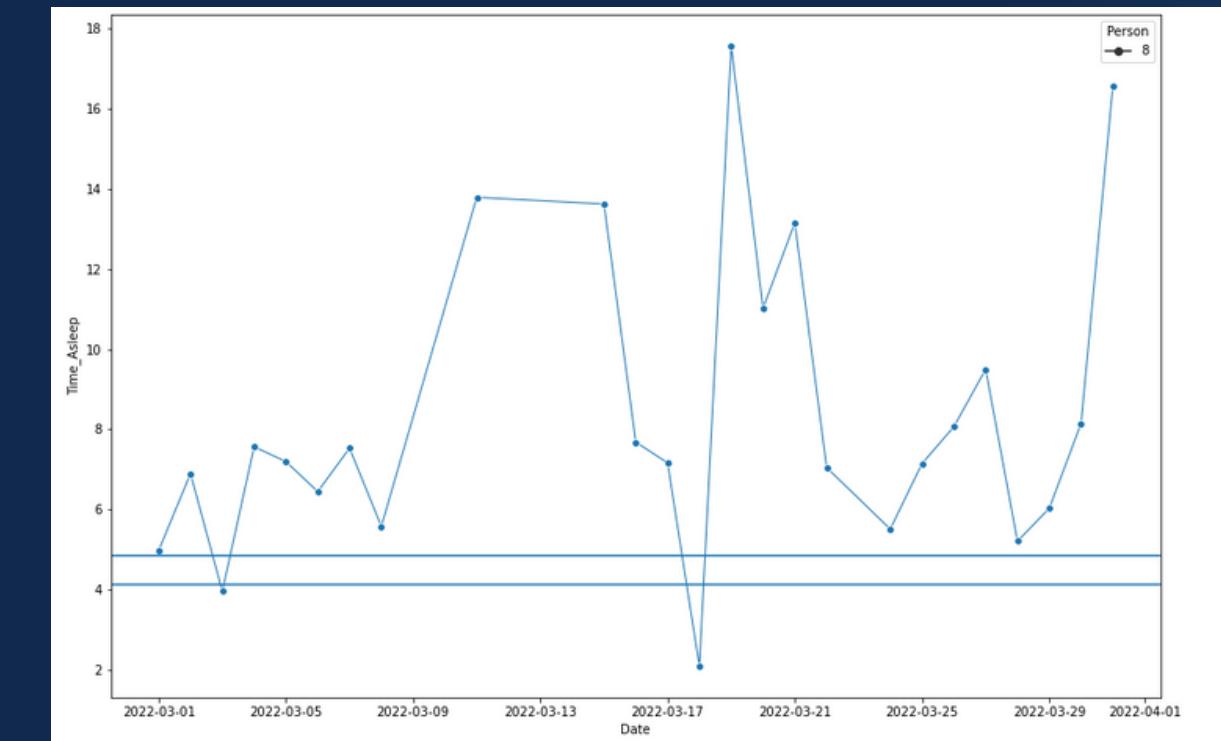
14 y 18

Pasos



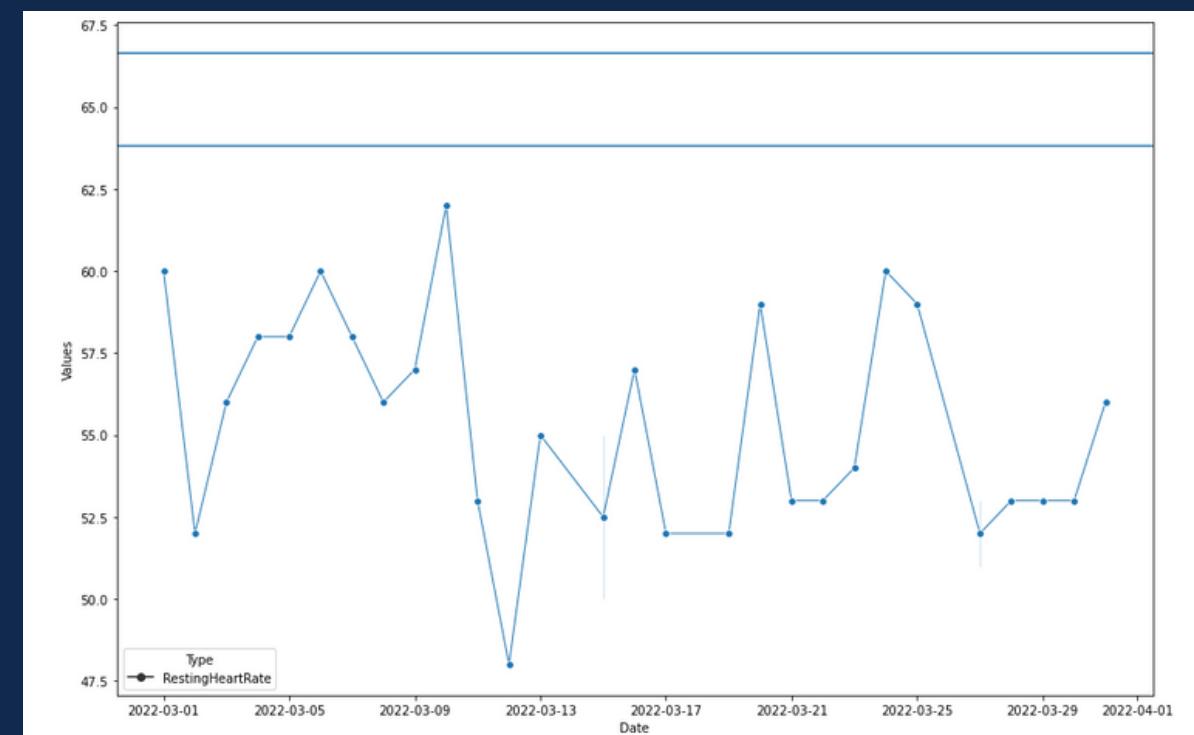
9/12/20

Tiempo de sueño



3 y 18

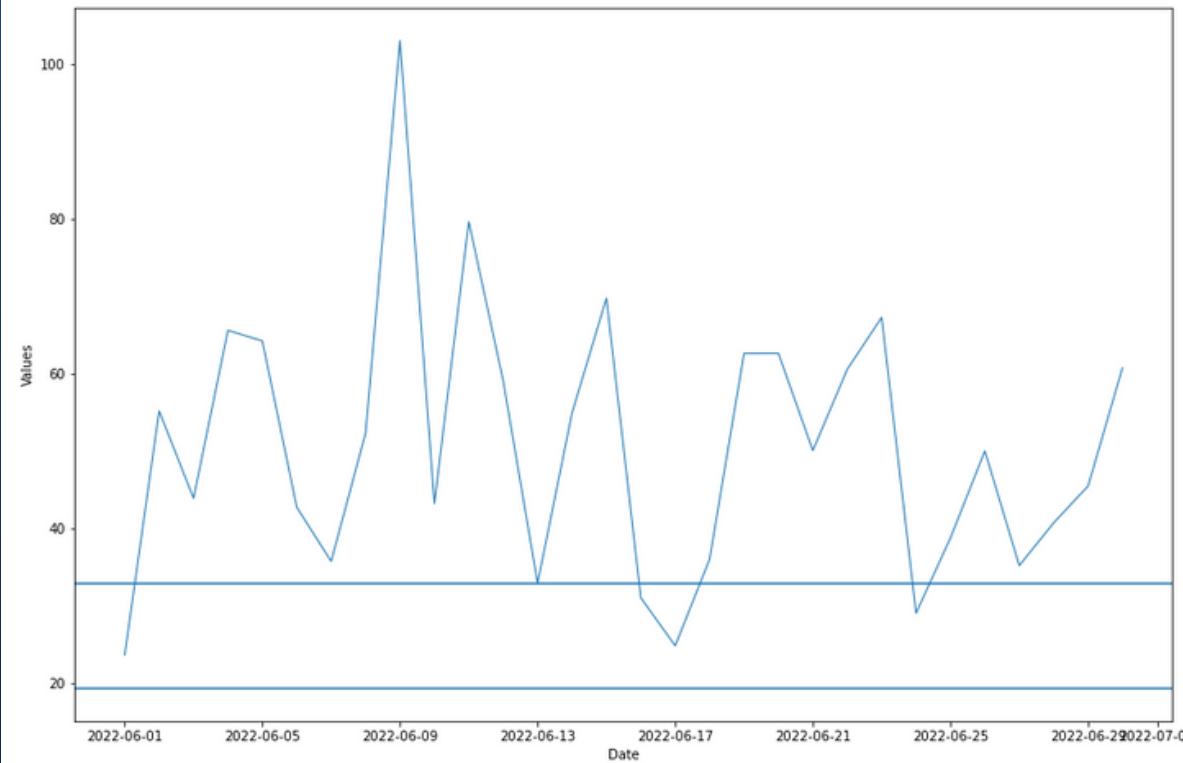
Frecuencia cardiaca en reposo



10/20 /24

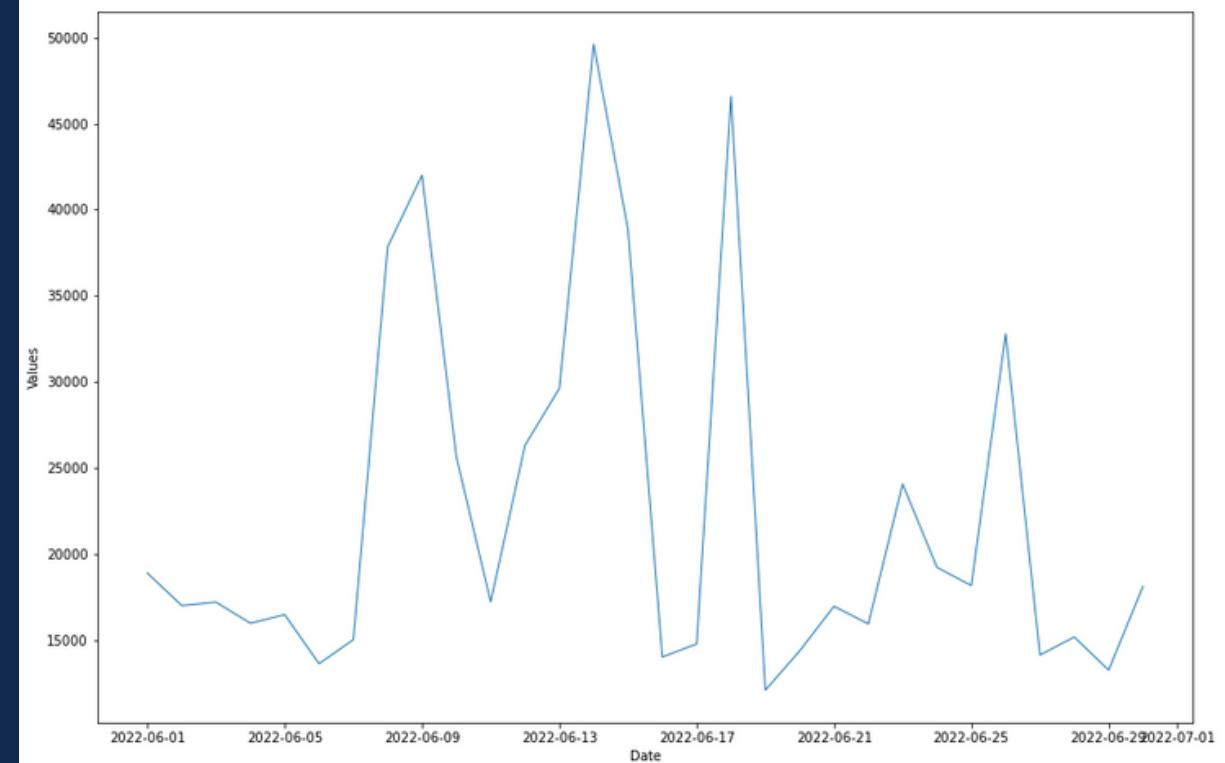
Junio

VRC



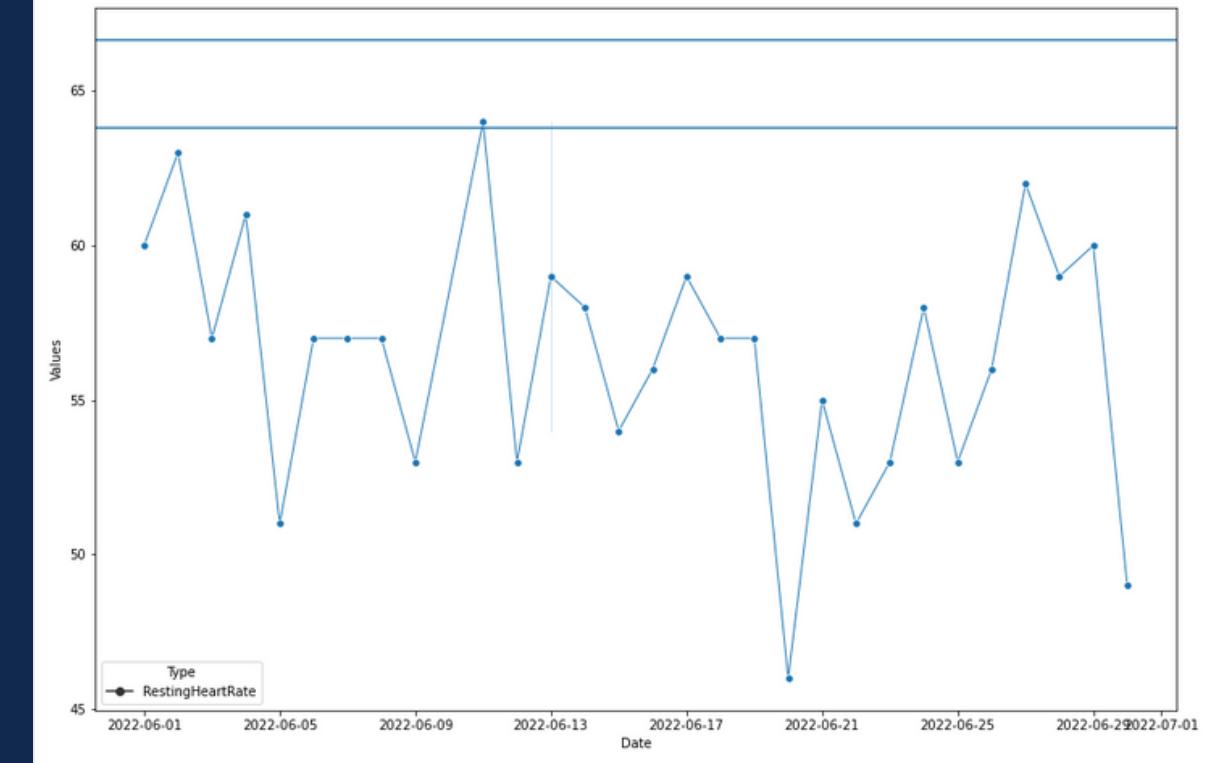
1/16/17/24

Pasos



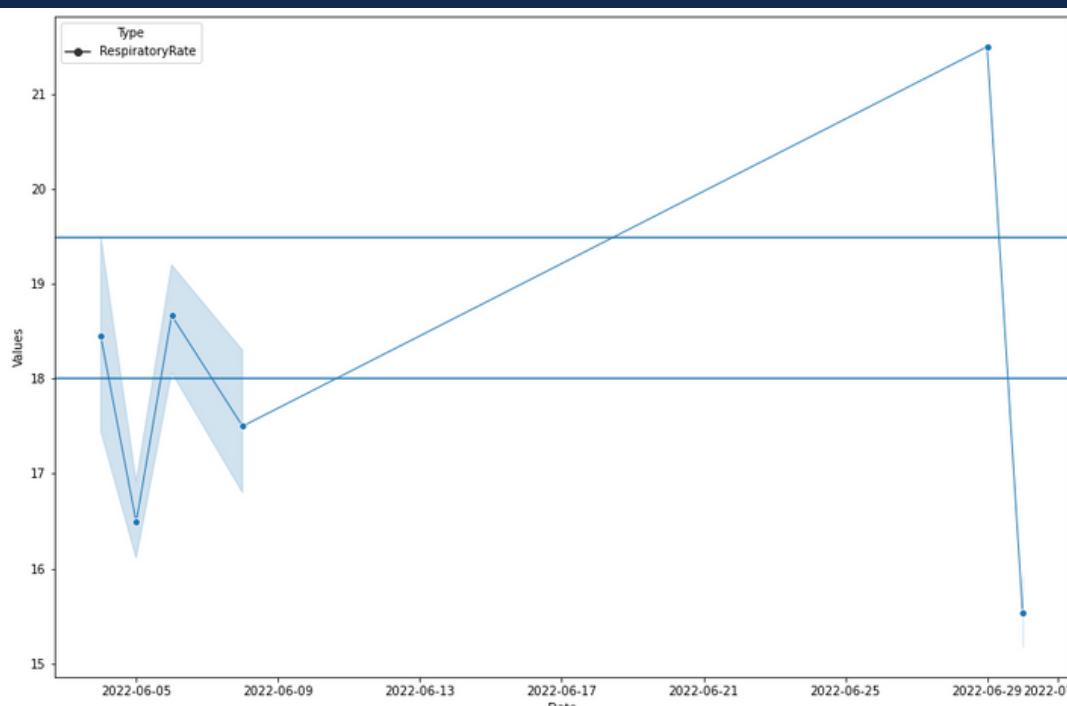
9/14/17/18

Frecuencia cardiaca en reposo



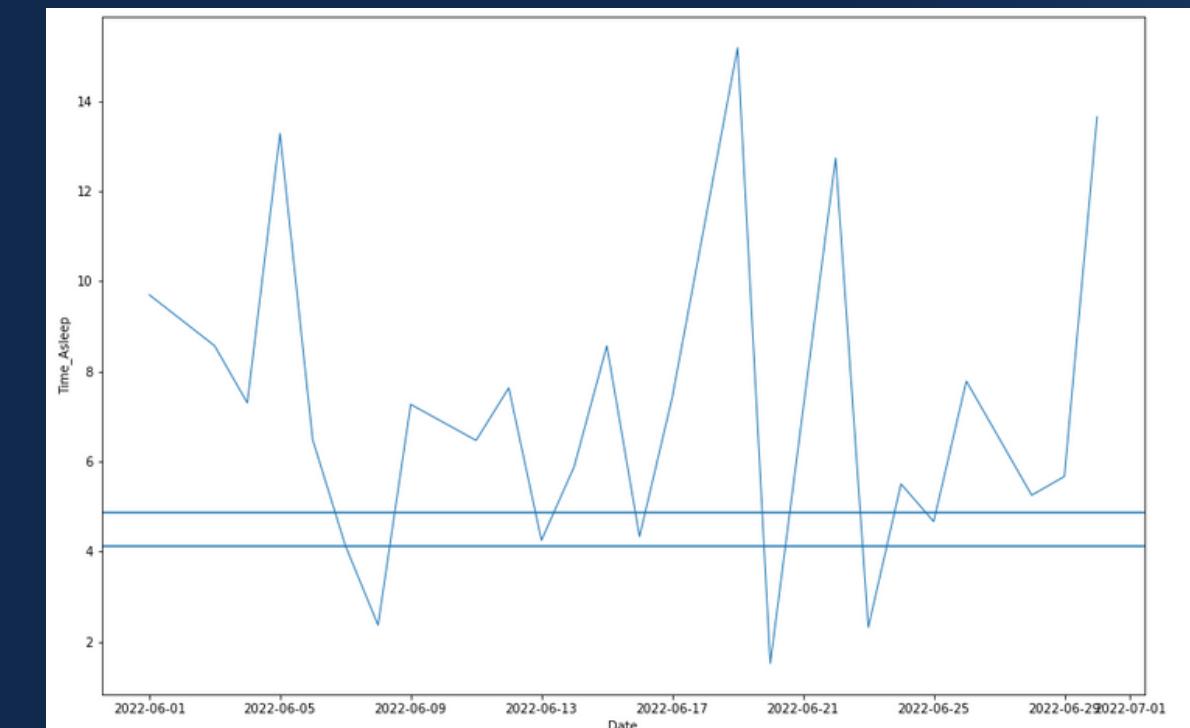
4/ 13/27

Ritmo respiratorio



4, 6 y 30

Tiempo de sueño

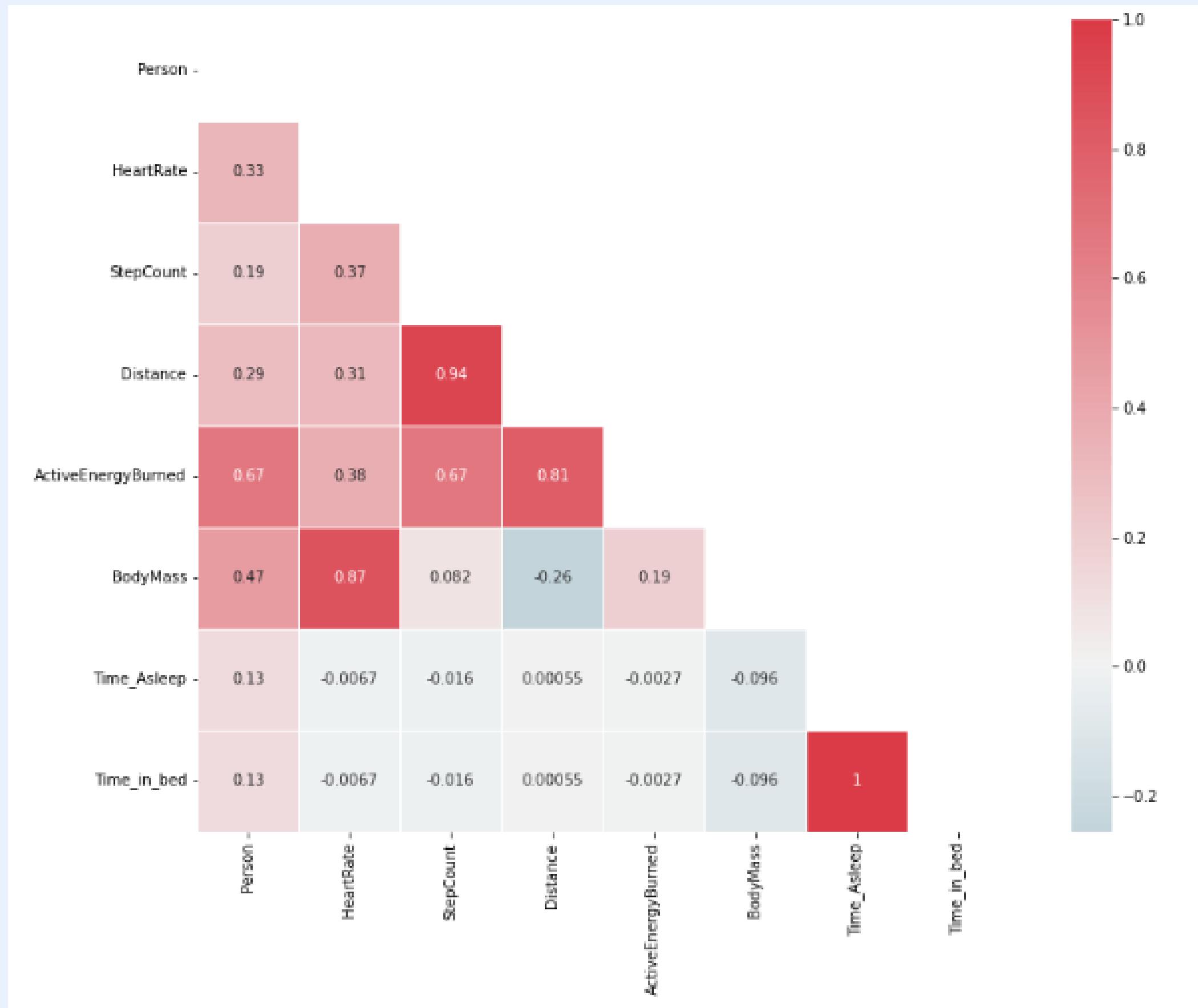


7, 8, 13, 16, 20, 23 y 25

Observaciones

- Al visualizar los datos de cada una de las variables en 2 meses diferentes vemos que los outliers de cada una de las mediciones coinciden en ciertos días, lo que puede indicar picos de estrés en jornadas específicas.
- En ocasiones un mismo pico de estrés puede encontrarse al observar mediciones de dos días consecutivos, por ejemplo, un día anterior con poco sueño lleva a un día con alto VRC.
- Para identificar mejor los patrones, es necesario cruzar los datos con variables dummies que puedan identificar días de mucho trabajo, vacaciones o jornadas especiales que puedan dar explicación a la variabilidad de los datos.

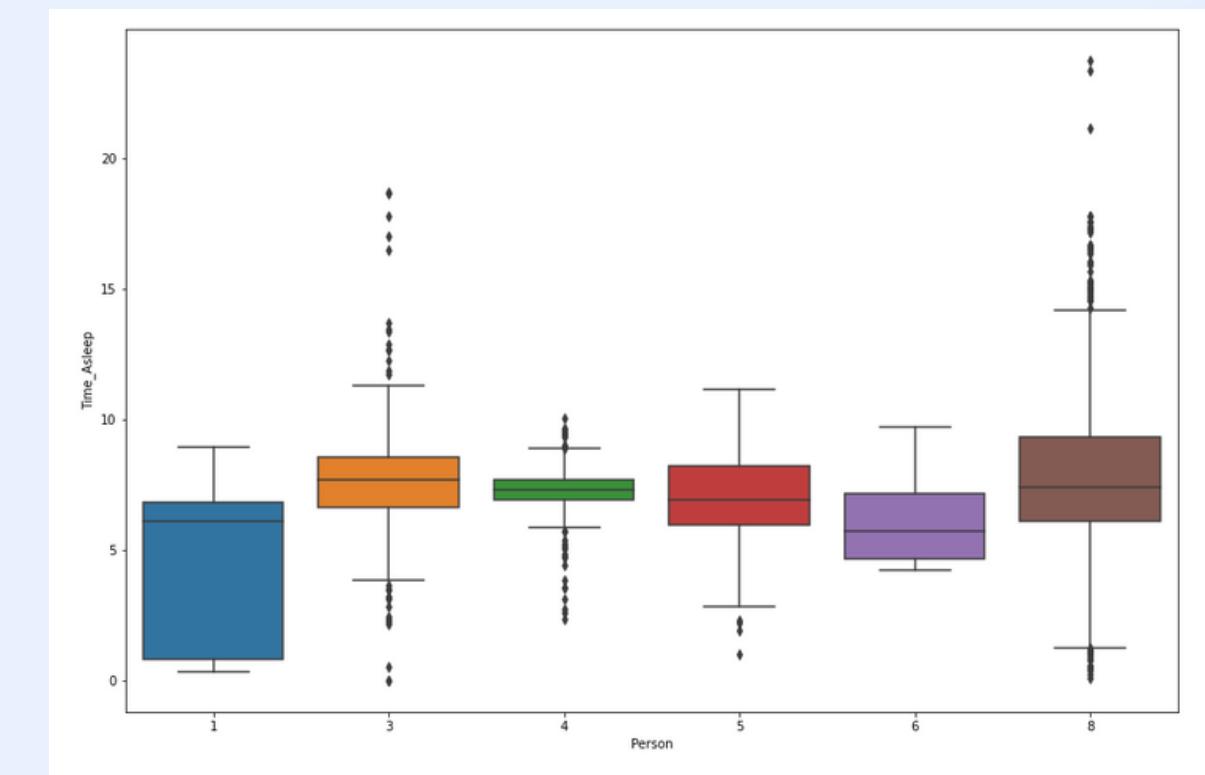
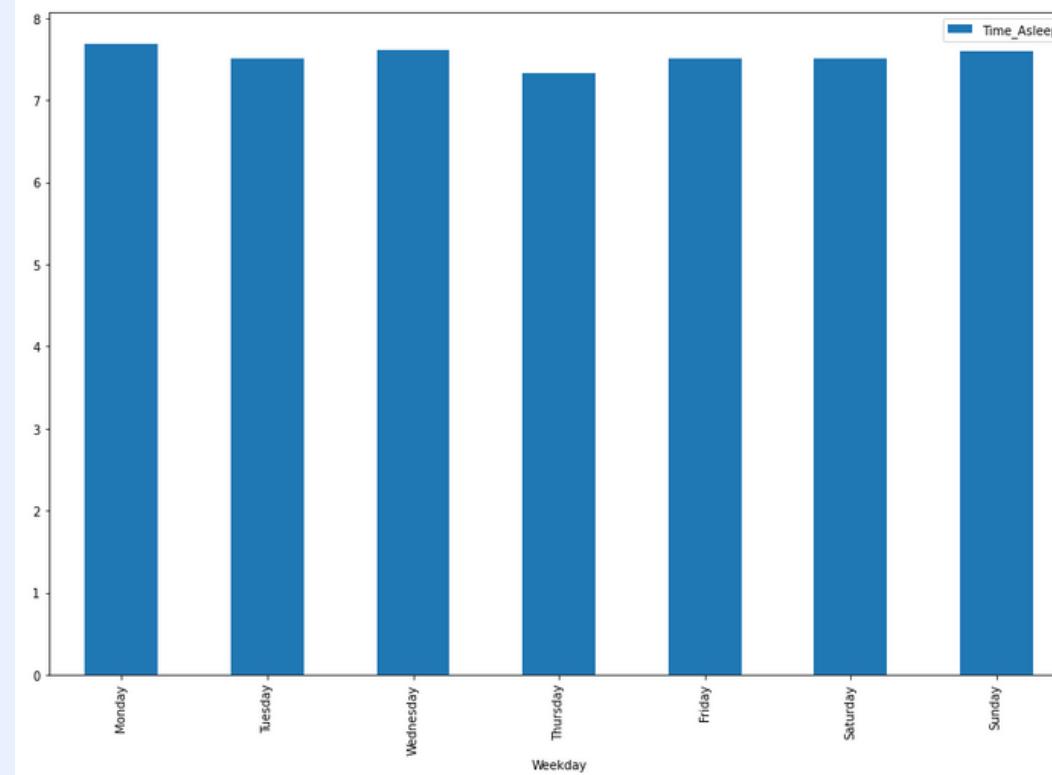
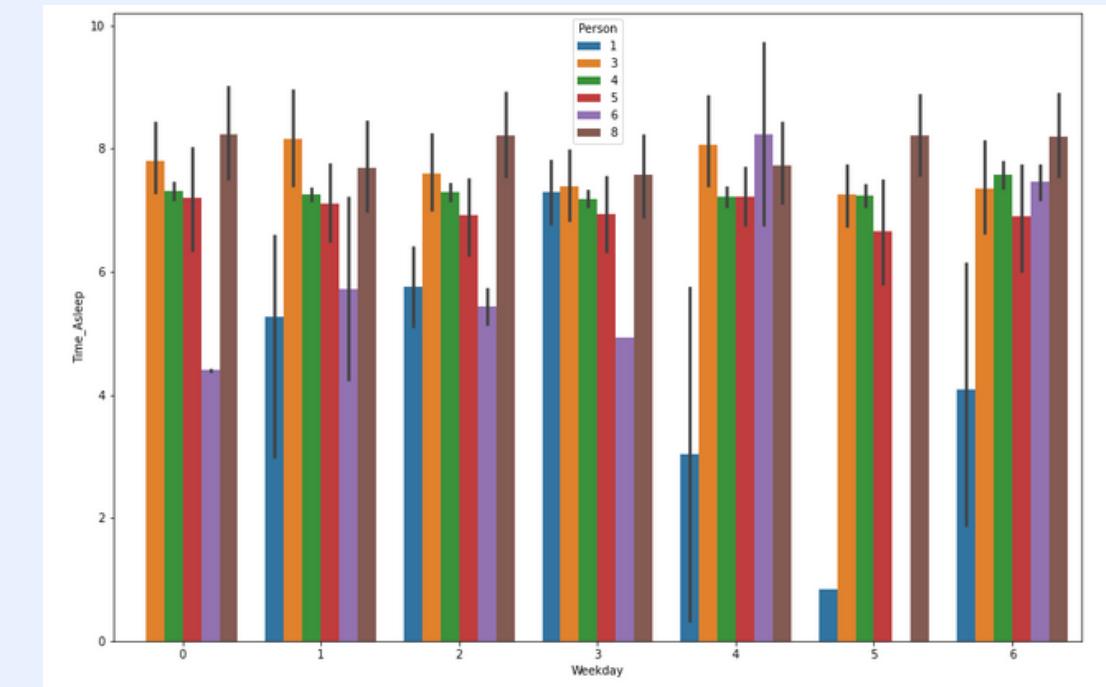
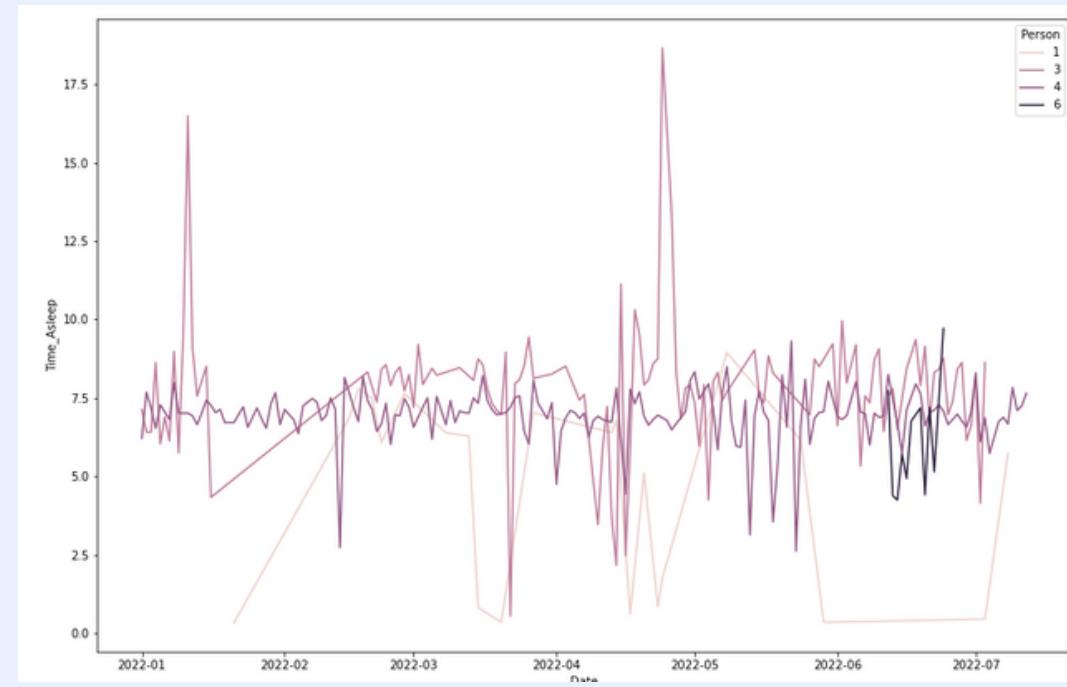
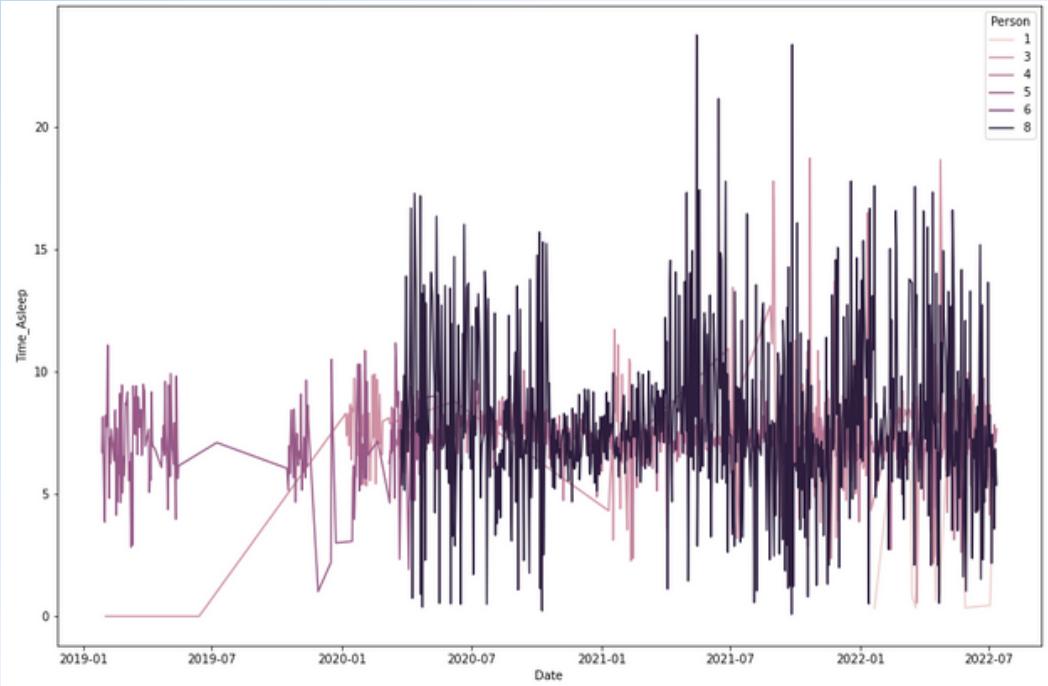
Análisis comparativo



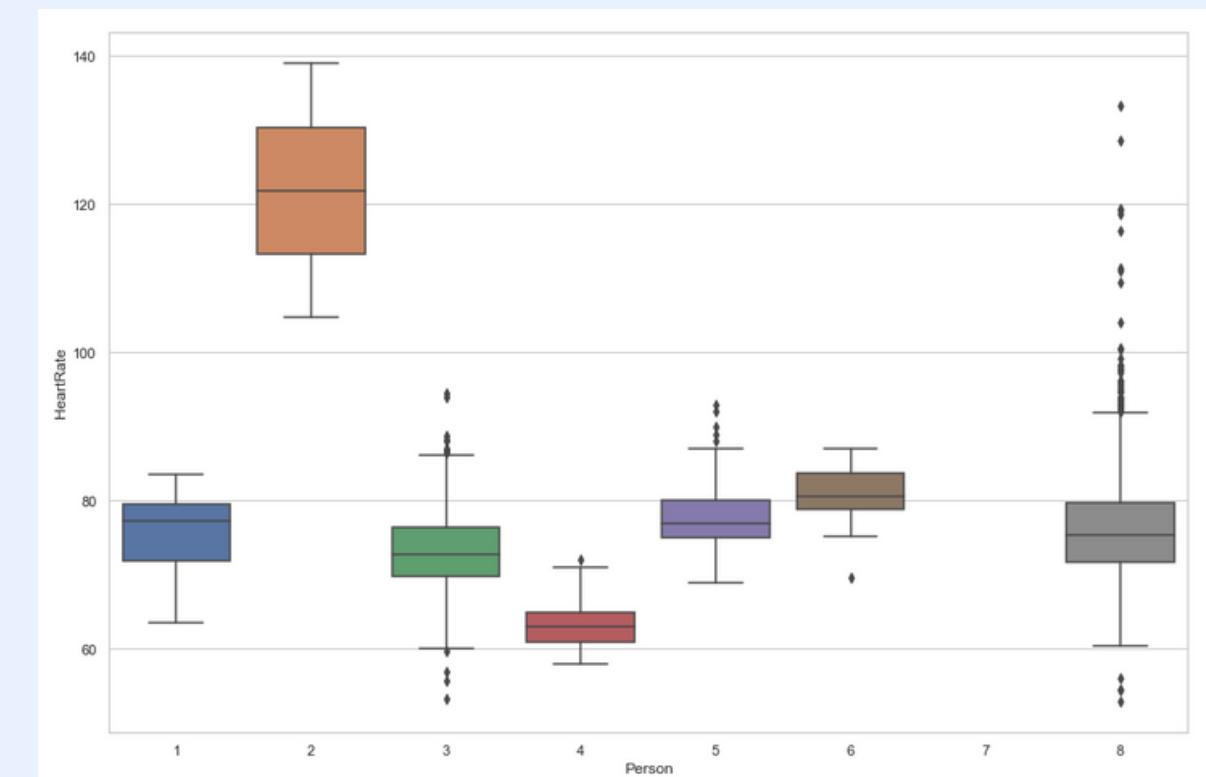
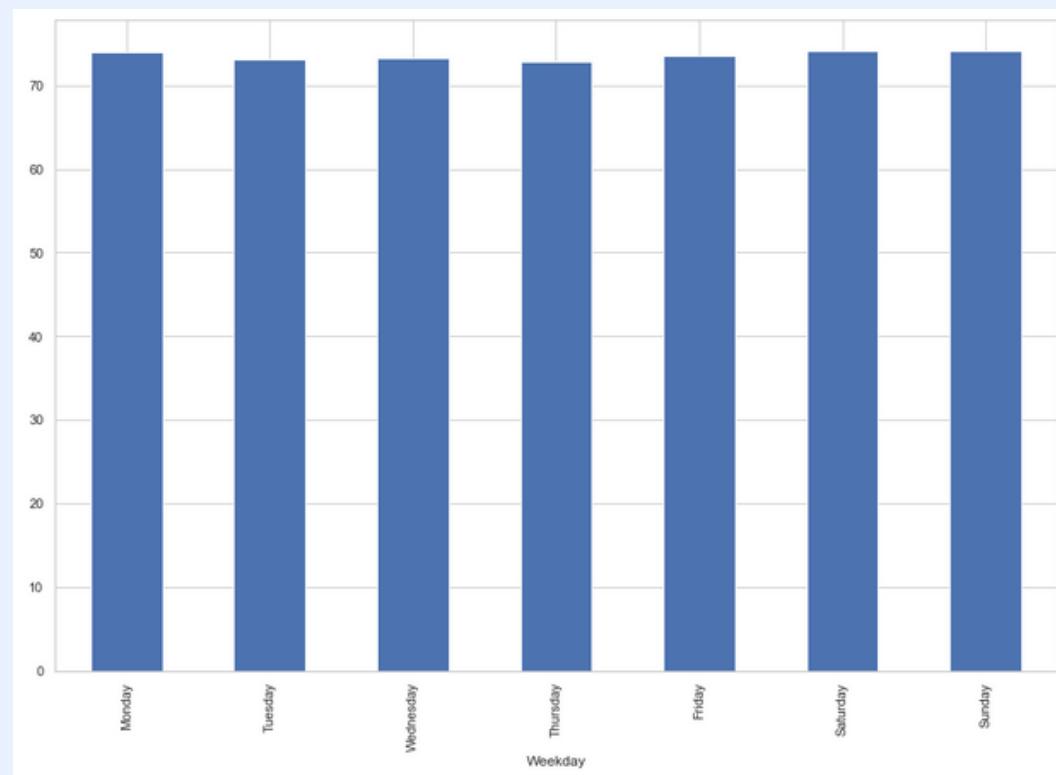
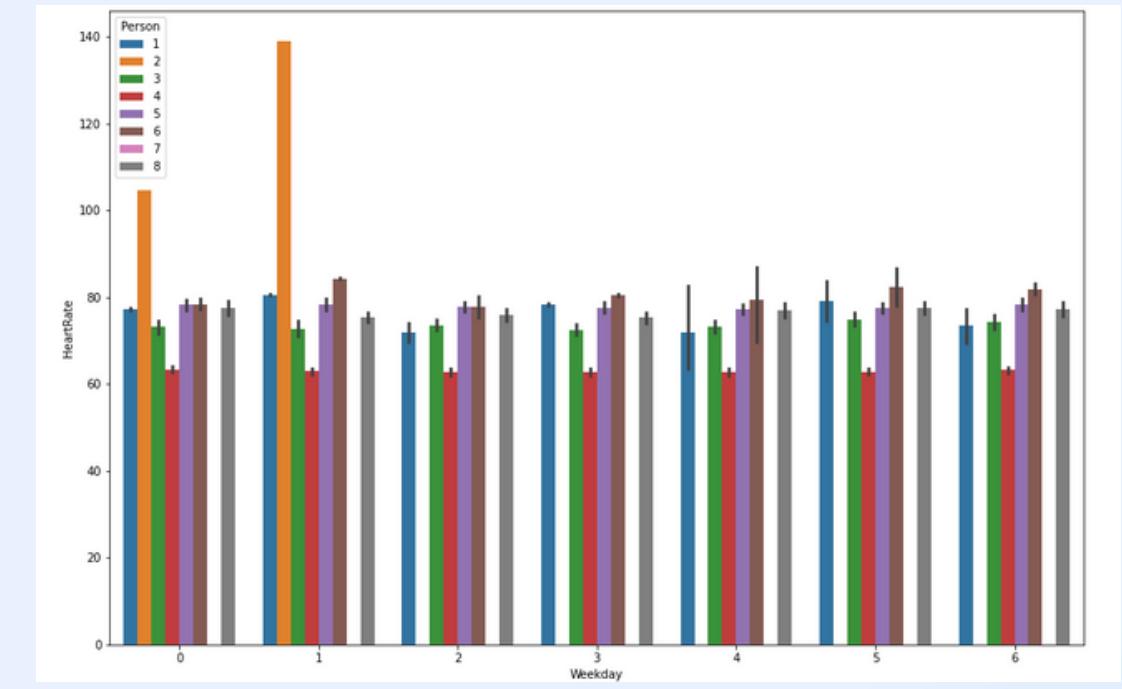
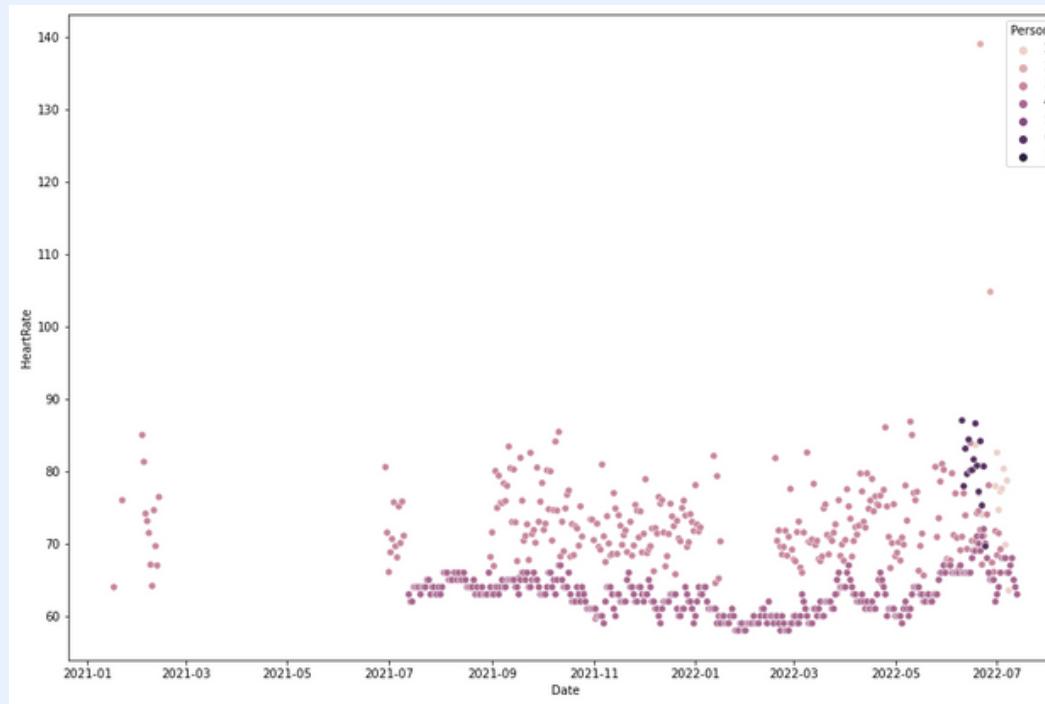
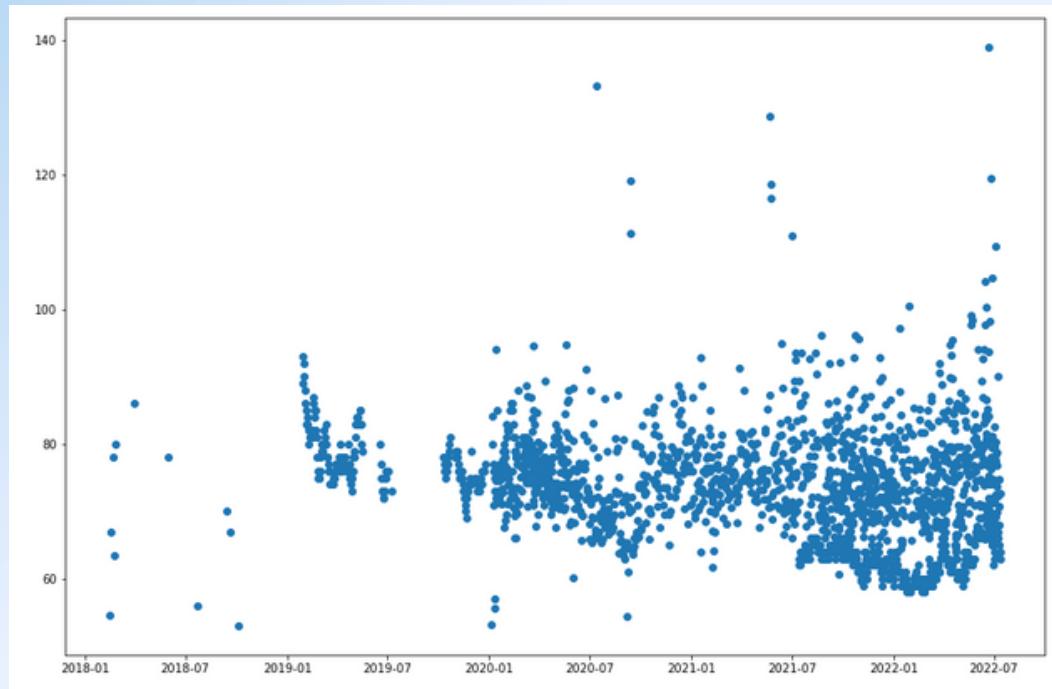
Pasos



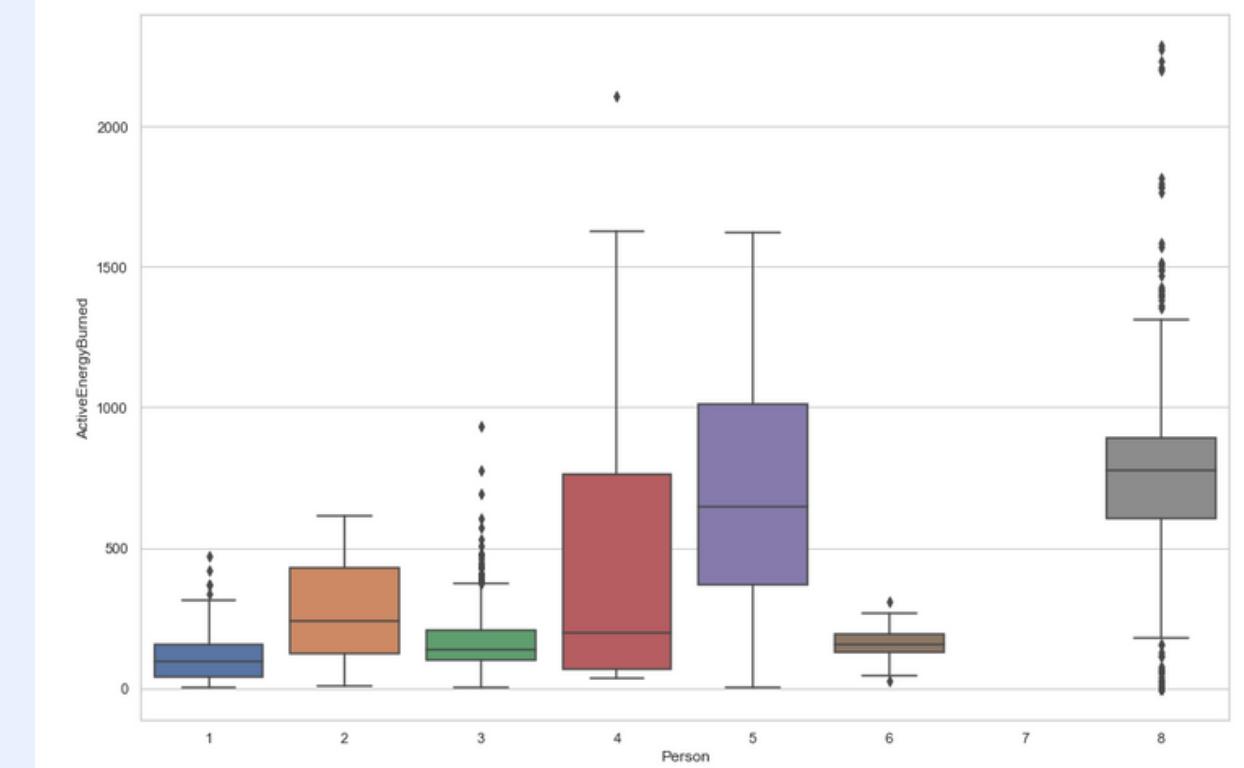
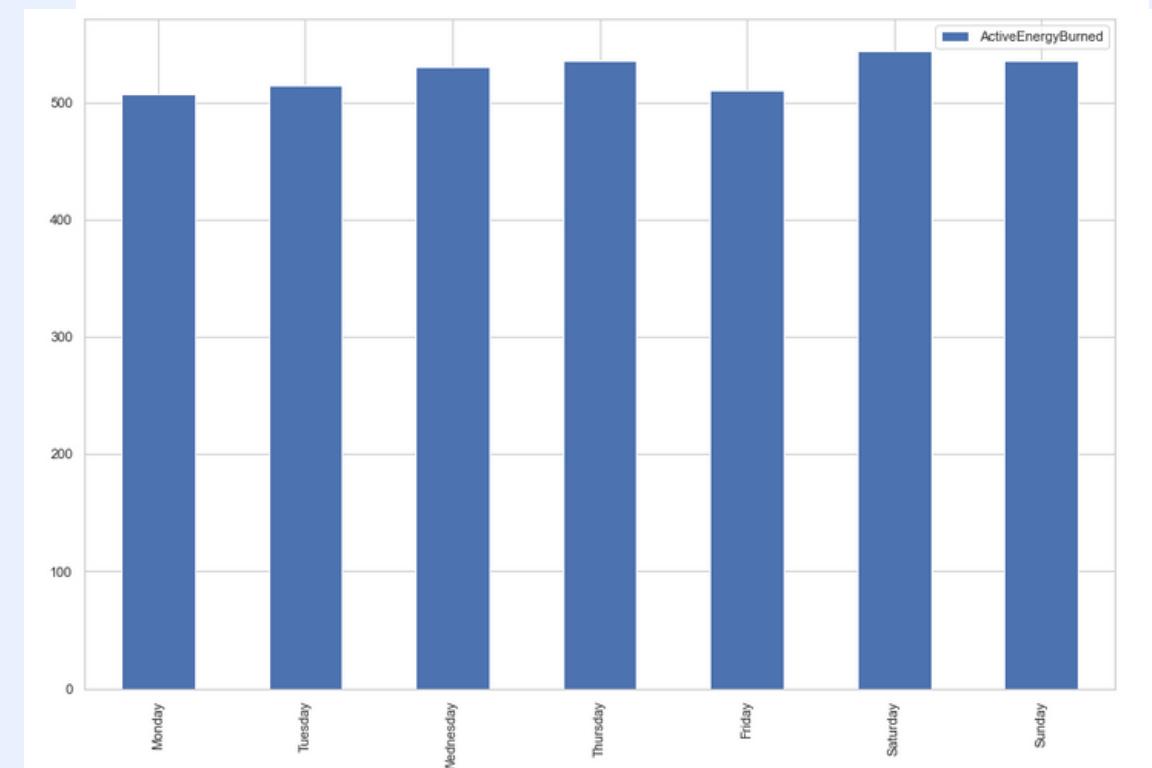
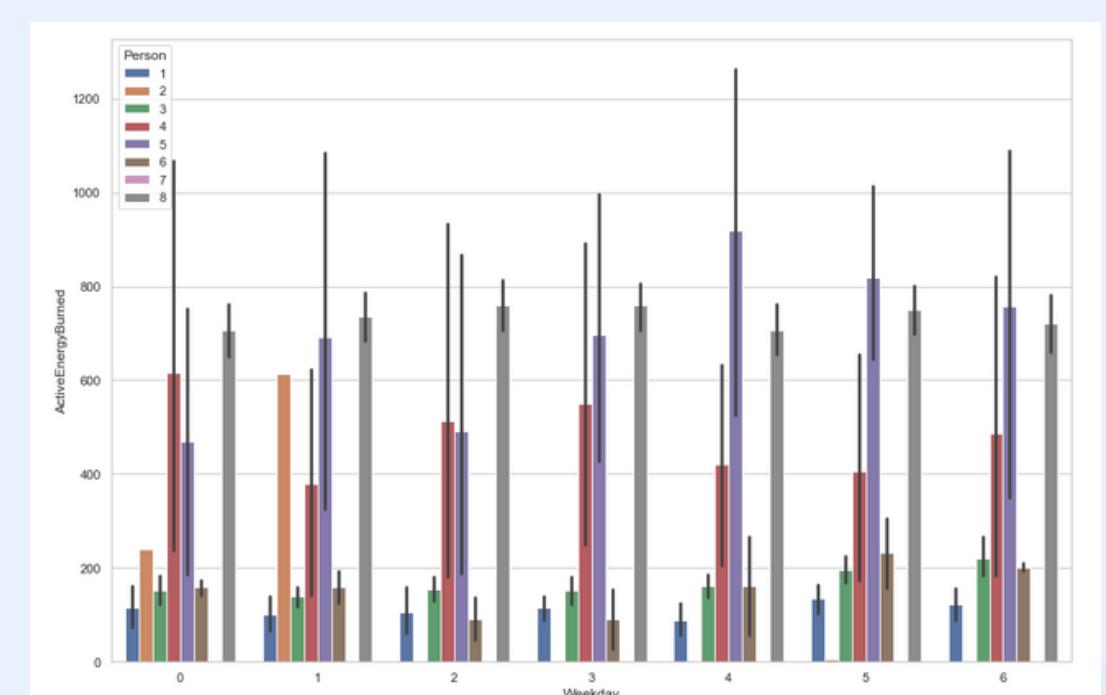
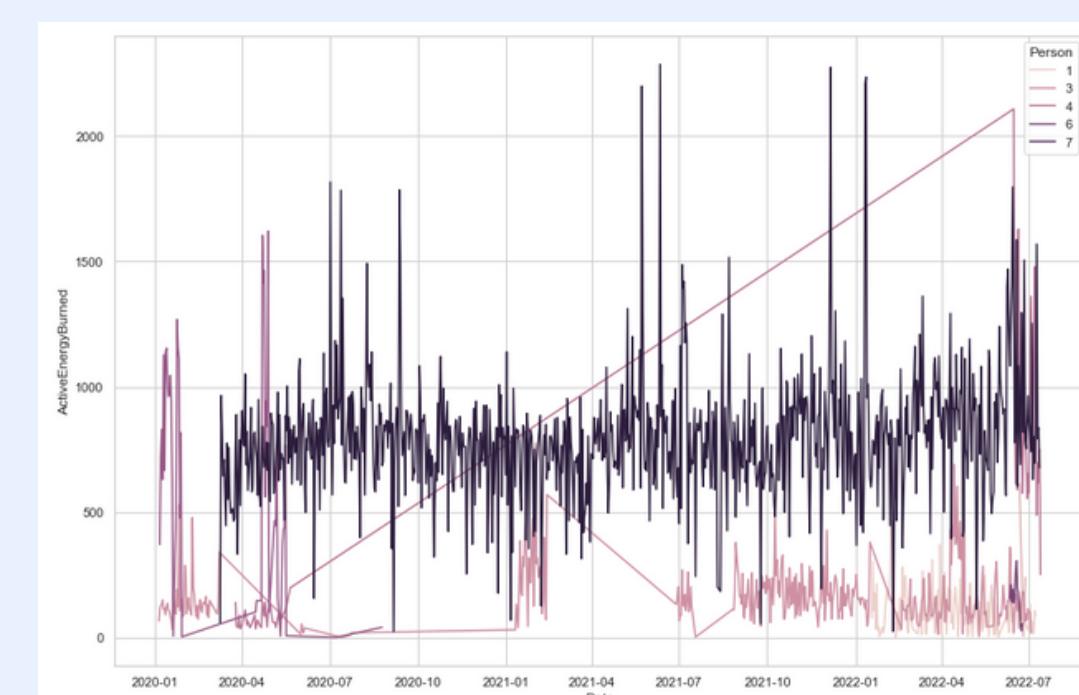
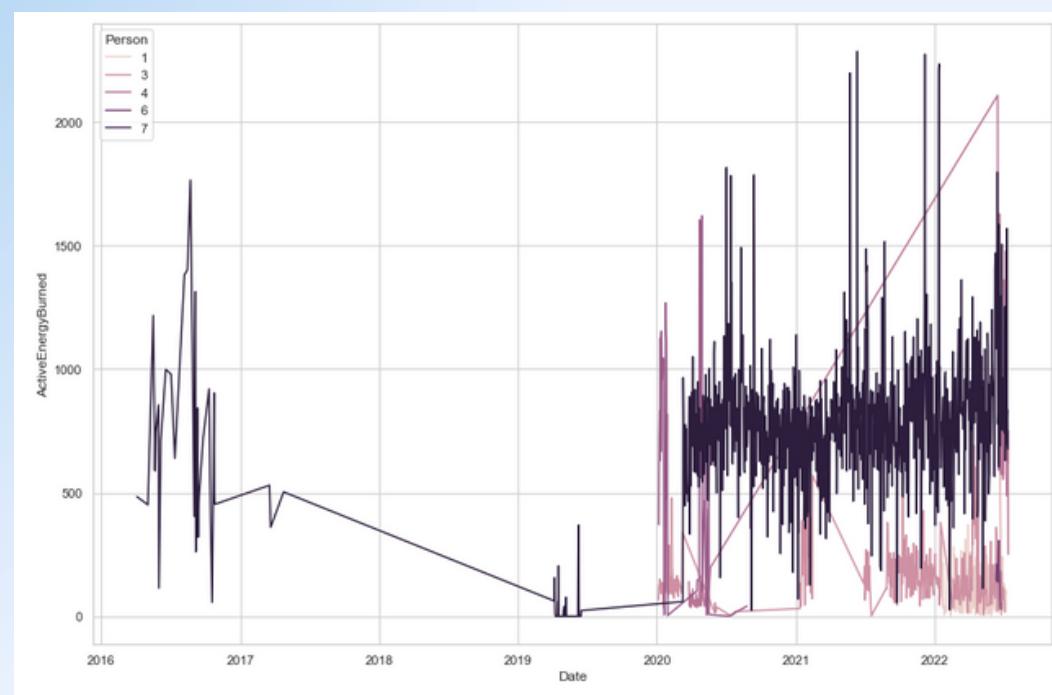
Sueño



Ritmo cardiaco



Calorias



Observaciones finales

- Los datos de salud y actividad física deben ser observados bajo los patrones individuales de cada individuo y con la perspectiva de calidad de cada una de las recolecciones de datos (ej. cantidad de días por cada variables por individuo).
- A medida que se cuentan con más datos, podemos ver que los patrones de cada persona tienden más a una tendencia moderada y no se posicionan con valores extremos.
- Es necesario realizar agrupaciones y medidas de minimización de ruido para poder identificar patrones de forma gráfica.
- Tener diferentes aproximaciones gráficas a los datos del estudio permite identificar diferentes conclusiones y poder asociar los datos de cada individuo con posibles patrones de rutina y actividad.
- Para sacar observaciones más concluyentes, sería necesario contar con muchos más individuos en el estudio que nos pudieran dar una aproximación representativa de cada variable.

Gracias

