## Línea horizontal



Caso práctico Bellabeat

09.Marzo.2024

**─**

Fabio Esparza Cetto

Track A Case study A   
Certificación Google Data Analyst

# Descripción general

El departamento de Marketing y Análisis, quiere explorar la posibilidad de guiar sus próximos esfuerzos de marketing en base al análisis de uso de dispositivos de monitoreo de activador para tratar de descubrir tendencias o comportamientos que permitan implementar una campaña más efectiva.

# Reportes a entregar

1. Resumen claro de la tarea.
2. Descripción de las fuentes de datos usadas.
3. Documentación de las actividades de limpieza, manipulación y transformación de los datos.
4. Resumen del análisis realizado.
5. Visualizaciones de apoyo y los descubrimientos clave.
6. Las principales recomendaciones de alto nivel basadas en el análisis.

# Resumen de la tarea

Presentar un reporte claro y conciso de los hallazgos derivados del análisis de los datos de uso de los dispositivos de monitoreo de actividad que permitan guiar el enfoque del departamento de Marketing, brindar recomendaciones de actuación.

# Los datos

Los datos utilizados para este análisis se obtuvieron de un dataset de Kaggle provenientes de usuarios de FitBit, estos datos son OpenSource por lo que no hay problema en utilizarlos, consta de 18 tablas largas en formato.csv con datos sobre actividad, pasos, sueño, calorías y frecuencia cardiaca.

Tengo algunas preocupaciones acerca de las limitaciones de los datos a usar, en primer lugar la muestra es de 30 personas por lo que hay alto riesgo que no sea representativo de la poblacion en general, por otro lado no hay datos de los usuarios por lo que no se pueden segmentar en categorias como sexo, edad, ubicacion, etc..

El periodo de recolección de datos es adecuado y proporciona buena representación de un uso típico.

Por otro lado los datos confiables, la fuente es bastante directa, aunque son algo viejos (2016).

# Preparación de los datos

Lo primero que hice para preparar el análisis es determinar tas tablas que usare y las que descartar debido a su baja relevancia para el propósito del análisis, las tablas descartadas son las que brindan información al minuto ya que es demasiada resolución para el análisis,

igualmente descarte la de peso ya que es auto reportada por el usuario y presenta inconsistencia en la frecuencia de reporte por el mismo.

El análisis lo haré utilizando hojas de cálculo, ya que la cantidad de datos no es muy grande y podré hacer el análisis y las gráficas en una misma herramienta.

# Limpieza de datos

Me centré al principio en el dataset de dailyActivity\_merged.csv ya que es el que contenía más información concentrada y que podría ser de más utilidad, comencé haciendo Limpieza:

Busqué espacios en blanco, y no encontré.

Me asegure que las fechas tuvieran el mismo formato y alineación.,

Recorte el número de decimales de las columnas a solo 3 digitos.

Elimine la columna SedentaryActiveDistance ya que carece de valor para el análisis.

Posteriormente continué con los datasets de dailyIntensities, dailiCalories y dailySteps, al inspeccionarlos descubrí que los datos de estos 3 ya están incluidos en dailyActivities por lo que decidí descartar los.

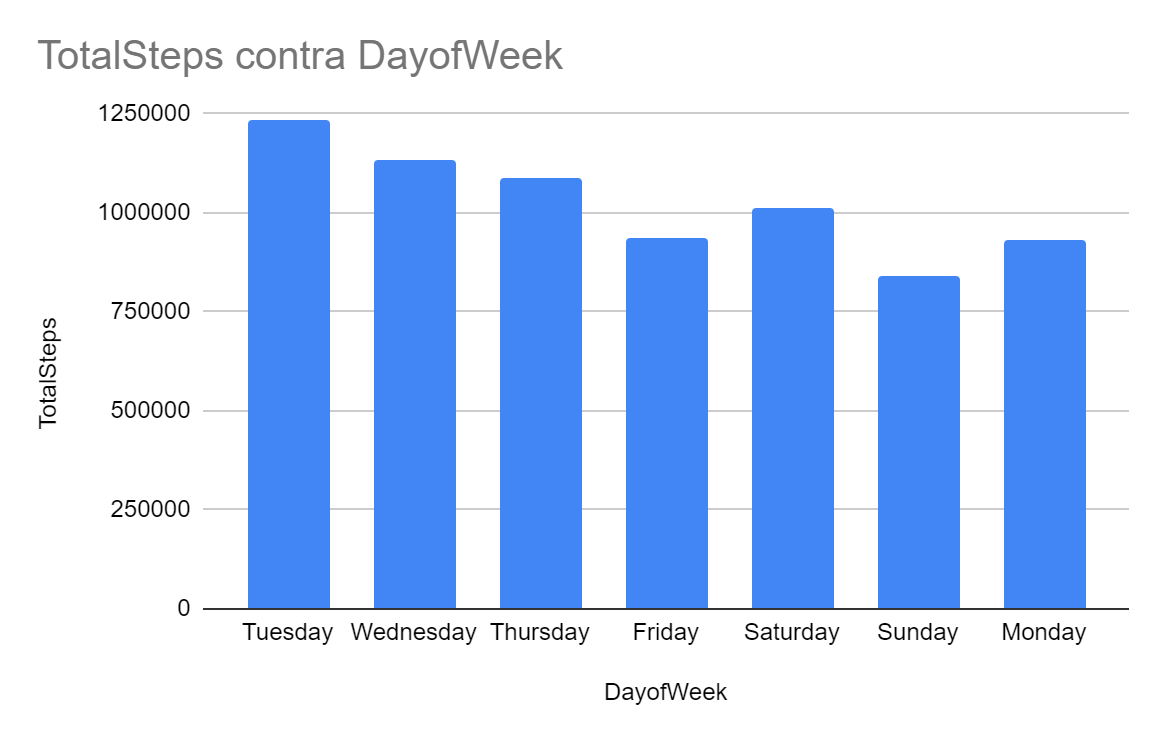
En el dataset de hourlyIntensities, lo que hice fue separar la columna de fecha y hora en 2 columnas y descartar la columna AverageIntensity.

Hice lo mismo que en la anterior con hourlySteps\_merged pero no elimine columnas.

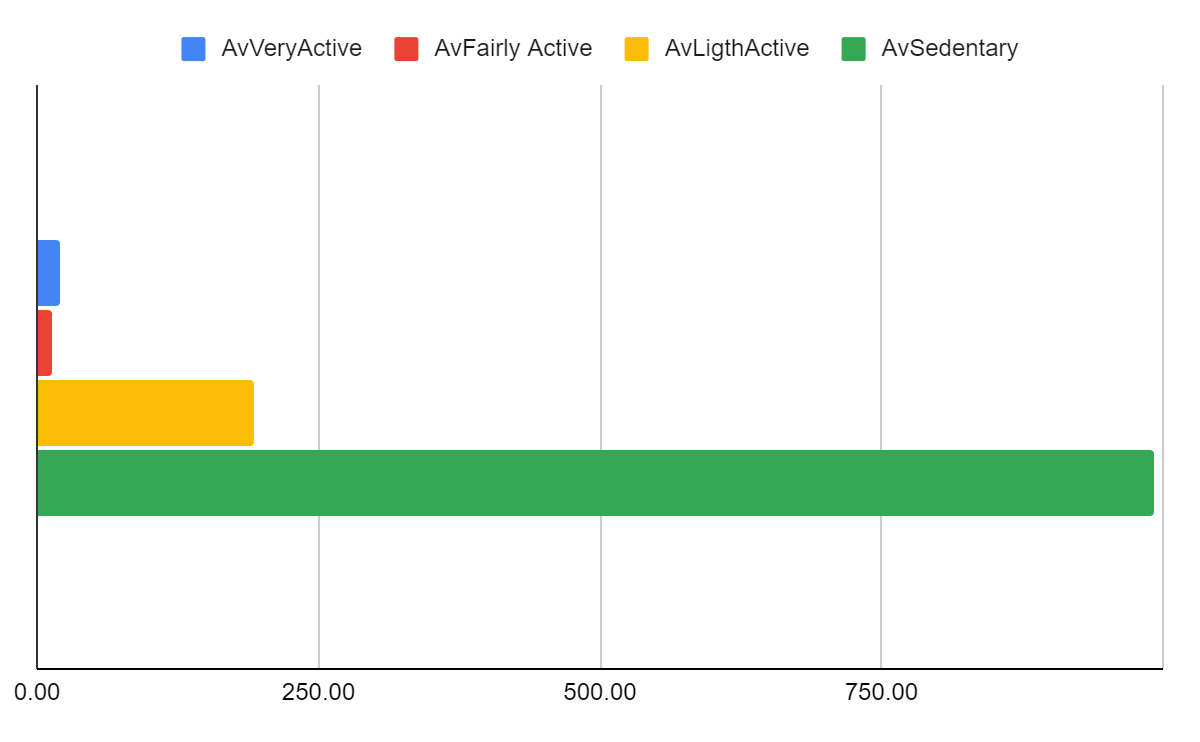
Ahora bien el dataset de sleepDay presenta un problema y es que no todos los Id presentan la misma cantidad de registros y estos varían mucho, en este caso no modifique los datos ya que aun no estoy seguro de que sean realmente útiles ya que solo cuenta con 25 de los 30 participantes, igual pienso que de aquí se puede extraer algún dato interesante.

# Análisis

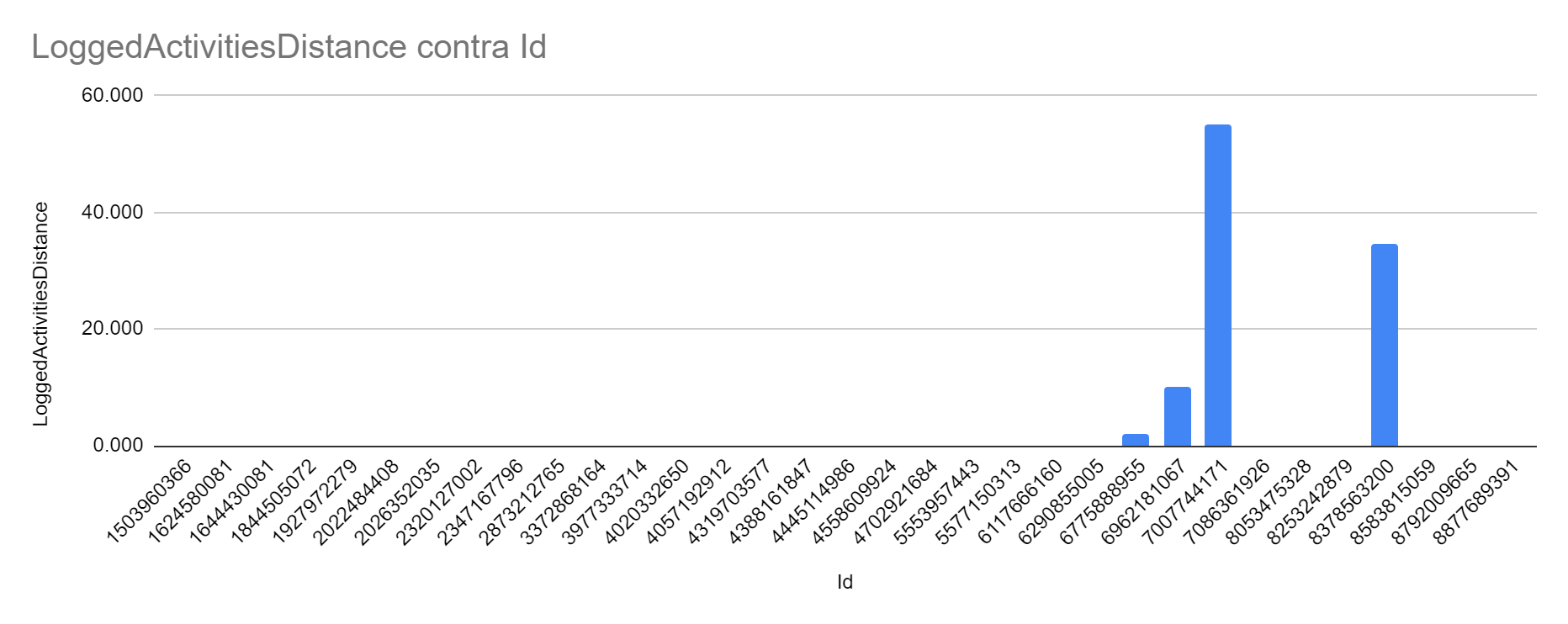
Comencé el análisis, tratando de determinar los días de la semana con mas actividad para determinar si los usuarios son activos toda la semana hay diferencias extraordinarias entre días. después del análisis estos fueron los gráficos.



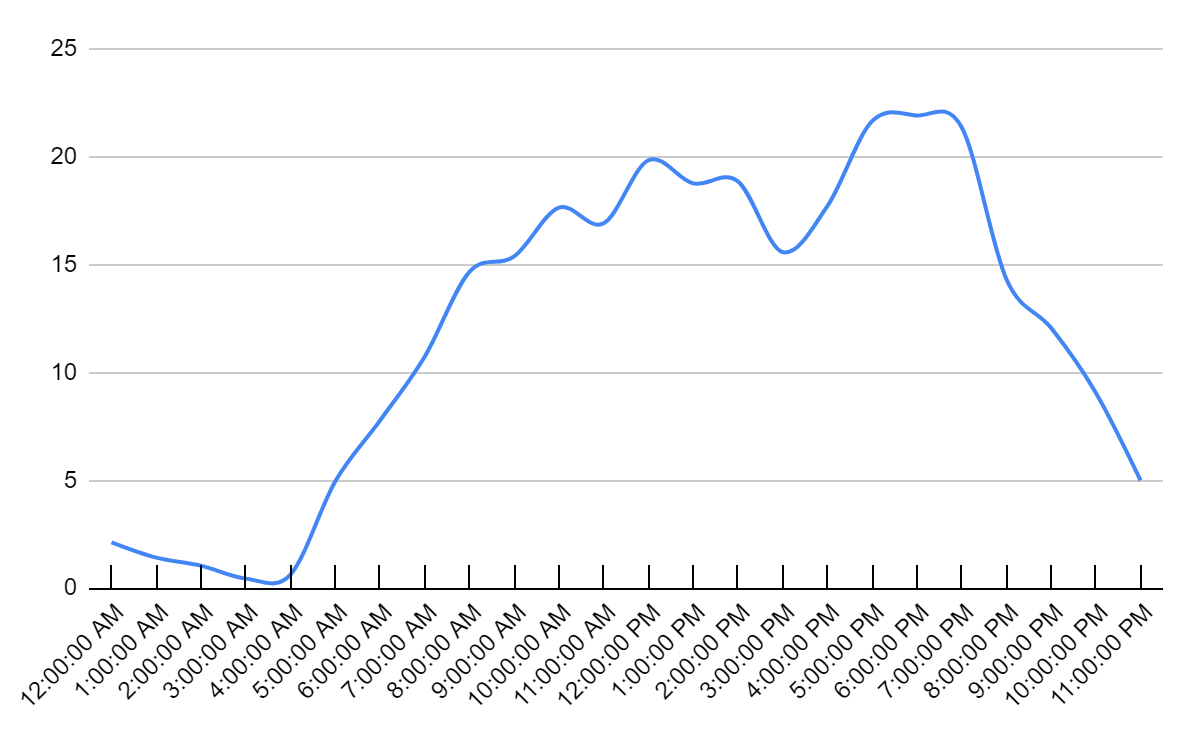
En este gráfico se puede observar que los días con menos actividad son los domingos viernes y lunes de aquí podemos ver que el lunes empieza flojo con la actividad y el martes inmediatamente alcanza el máximo para ir gradualmente disminuyendo a lo largo de la semana y vuelve a tener un pico el sábado de este gráfico se puede extraer que los usuarios suelen tener actividad regular en la semana y los sábados agregan un poco más probablemente relacionadas con actividades recreativas variadas, lo que es un punto interesante a destacar, igualmente al parecer los domingos son días de descanso y el lunes cuesta trabajo recomenzar a ritmo alto.



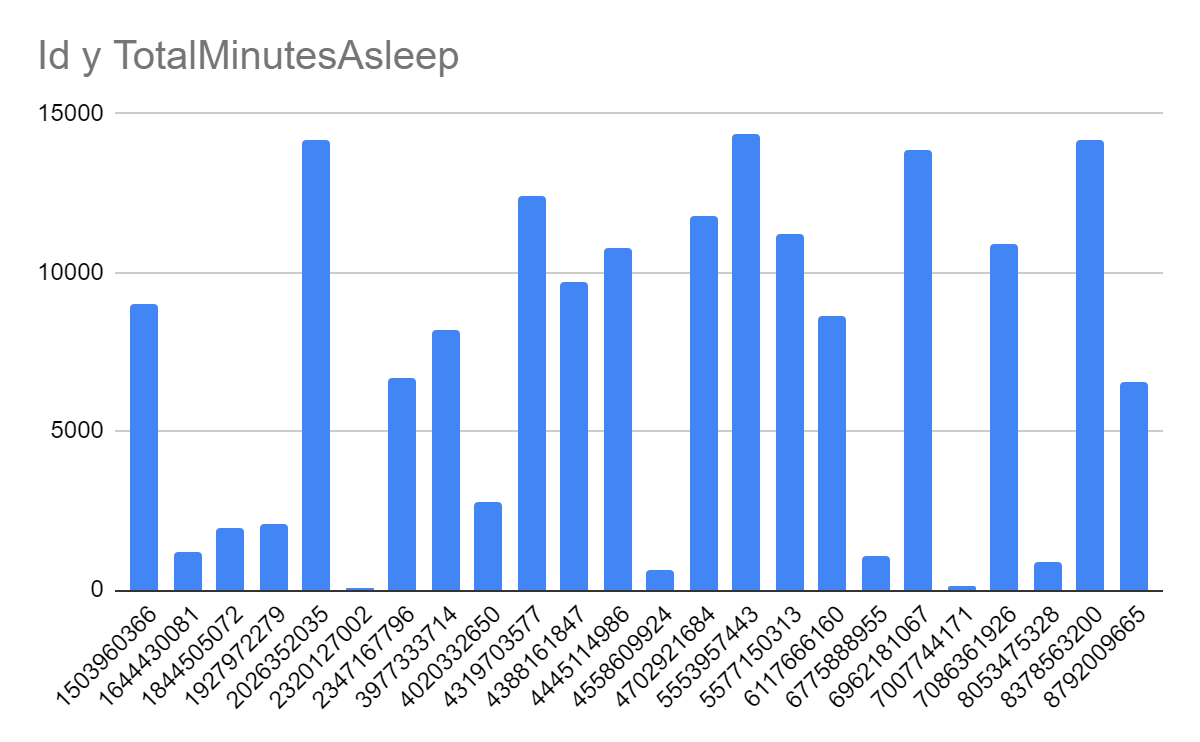
En el gráfico de arriba se puede observar claramente que la porción más grande del dia no hay actividad física lo que tiene sentido ya que las personas pasan el mayor tiempo haciendo otras cosas ademas de actividad física por lo que podemos deducir que la mayoría de los sujetos tienen trabajos sedentarios, la siguiente franja es actividad leve lo que no se puede considerar ejercicio, esto es caminar de un lado a otro en casa haciendo cosas ir al supermercado, y desplazarse de un lugar a otro, ya viendo el resto de métricas se puede concluir que los usuarios son sedentarios que dedican algunos minutos al dia a actividades físicas intensas.



En el gráfico de arriba se puede observar que solo 4 de los usuarios usan sus monitores de actividad para registrar sesiones de actividad específica por lo que es una buena oportunidad para resaltar las funciones de detección de actividad, también nos permite saber que los usuarios suelen no ser de los que se toman demasiado en serio las sesiones de actividad



En este gráfico se puede observar la distribución de las actividades durante el día se concentra más por la tarde y alrededor de mediodía con una disminución a la hora de comer.

En este gráfico queda claro que los registros de sueño no son confiables ya que hay participantes que apenas tienen registros o en definitiva no tienen y la mayoría tiene muy bajo número de registros; esto aún indica más que los usuarios son usuarios casuales ya que se quitan el dispositivo para dormir y no les importa llevar registro de sueño.

# Actuar

Después de estos análisis podemos observar que los usuarios de los dispositivos son usuarios casuales probablemente con trabajos poco activos y que solo dedican algunos cuantos minutos para hacer alguna actividad física un poco más intensa y que solo muy pocos usan el dispositivo para rastrear progreso en actividades específicas, y pudimos constatar que muchos se quitan el dispositivo para dormir.

En este caso sabemos que la mayoría de los usuarios no usan las funciones más avanzadas y tienen el dispositivo como una mera referencia de su actividad, en este caso sugiero 2 enfoques distintos, el primero es para tratar de enforcer la campaña para llegar a un tipo de población que aun no esta usando los dispositivos, es son los usuarios un poco más serios y que son más conscientes de su salud, resaltando las funciones más avanzadas y útiles que tienen los dispositivos para rastrear actividad y no solo el cuenta pasos si no las funciones de entrenamiento específico, y las funciones que incorpora la app y la utilidad que se les puede dar, de esta manera también podemos enfocar por otro lado en destacar estas mismas funciones para que los usuarios más casuales descubran y usen también algunas de estas funciones sobre todo las que ayudan a medir marcadores importantes de la salud.

Igualmente sería recomendable comenzar un programa en el que se invite a los usuarios de los dispositivos a compartir de manera anónima los datos recabados por los dispositivos propios para de esta manera tener una base de datos mucho más útil y que sea realmente representativa de los usuarios en las que se pidan también opiniones y sentires de los clientes.