

Entrega 2: Versión Avanzada del Proyecto Final de Visualización y Storytelling

Título: FitData Insights: Analizando el Impacto del Entrenamiento en la Salud y Bienestar

Texto de Contexto: Este dashboard interactivo presenta un análisis detallado de cómo diferentes tipos de entrenamiento influyen en la salud y bienestar de las personas. Basado en una base de datos con más de 10,000 registros, exploramos tendencias de actividad física, impacto en la quema de calorías, relación con la hidratación y el descanso, y cambios en el estado de ánimo antes y después del ejercicio. Esta herramienta está diseñada para ayudar a individuos y profesionales del fitness a identificar patrones clave para mejorar la calidad de vida a través del entrenamiento.

Dashboard de Visualizaciones:

- Distribución de usuarios por género y edad.
- Tendencias de tipos de entrenamiento más populares.
- Análisis de la quema de calorías según la duración del entrenamiento.
- Impacto del sueño y la hidratación en el rendimiento.
- Comparación del estado de ánimo antes y después de entrenar.
- Relación entre la TMB (Tasa Metabólica Basal) y el requerimiento de ejercicio.
- Relación entre el tipo de deporte y el porcentaje de grasa corporal.
- Relación entre el tipo de deporte y la calidad del sueño.

Presentación Web: El sitio web contará con la siguiente estructura:

1. **Introducción:** Explicación del propósito del dashboard y su utilidad.
2. **Módulo de visualización interactiva:** Gráficas con filtros dinámicos por edad, género y tipo de entrenamiento.
3. **Análisis de patrones y recomendaciones:** Insights obtenidos de los datos con recomendaciones prácticas.
4. **Conclusiones y documentación del modelado:** Explicación de las variables y transformaciones de datos realizadas.

Interactividad:

- Filtros por tipo de entrenamiento y género.
- Comparación de tendencias por rango de edad.
- Ajuste de escalas de tiempo en gráficas.

Documentación de Modelado de Datos:

- **Variables Utilizadas:**

- Edad
- Género
- Tipo de Entrenamiento
- Duración
- Calorías Quemadas
- Ritmo Cardíaco
- Intensidad
- Sueño
- Hidratación
- Estado de Ánimo.

- **Variables Modeladas:**

1. **TMB (Tasa Metabólica Basal):** Calculada con la ecuación de Mifflin-St Jeor.
2. **Rango de Edad:** Agrupación en categorías etarias.
3. **Calidad del Sueño:** Clasificación basada en horas dormidas y frecuencia cardíaca en reposo.
4. **Requerimiento de Ejercicio:** Determinación del nivel de actividad basado en la TMB y el consumo calórico.

- **Procesamiento de Datos:**

- Limpieza de datos nulos o atípicos.
- Agrupación y normalización de datos.
- Cálculo de promedios y variaciones por tipo de entrenamiento.
- Creación de indicadores clave de rendimiento.

Detalle del Modelado de Datos:

1. Variables en la base de datos original:

Variable	Tipo de variable	Descripción
Age	Número Entero (int64)	Edad de las personas

Gender	Categorica (object)	Género, tomando valores de Female, Male y Other
Height (cm)	Número Entero (int64)	Altura expresada en centímetros
Weight (kg)	Número Entero (int64)	Peso, expresado en kilogramos
Workout Type	Categorica (object)	Tipo de entrenamiento que la persona reporta que hace en sus sesiones. Las categorías son: HIIT, Strenght, Yoga, Cycling, Cardio y Running
Workout Duration (mins)	Número Entero (int64)	Duración del entrenamiento, cantidad de tiempo expresada en minutos
Calories Burned	Número Entero (int64)	Calorías quemadas por sesión de entrenamiento reportada
Heart Rate (bpm)	Número Entero (int64)	Frecuencia cardiaca media que reporta la persona en el entrenamiento
Steps Taken	Número Entero (int64)	Número de pasos diarios reportados
Distance (km)	Número Decimal (float64)	Distancia recorrida expresada en kilómetros
Workout Intensity	Categorica (object)	Intensidad del entrenamiento. Las categorías son: Low, Medium, High
Sleep Hours	Número Decimal (float64)	Total de horas dormidas
Water Intake (liters)	Número Decimal (float64)	Cantidad de agua consumida, expresada en litros
Daily Calories Intake	Número Entero (int64)	Cantidad de calorías diarias consumidas
Resting Heart Rate (bpm)	Número Entero (int64)	Frecuencia cardiaca en reposo
VO2 Max	Número Decimal (float64)	Cantidad máxima de oxígeno que una persona puede consumir por minuto y por kilogramo de peso corporal
Body Fat (%)	Número Decimal (float64)	Porcentaje de grasa corporal en relación con la masa corporal de una persona
Mood Before Workout	Categorica (object)	Estado de ánimo reportado previo al entrenamiento. Las categorías son: Tired, Happy, Neutral, Stressed
Mood After Workout	Categorica (object)	Estado de ánimo reportado posterior al entrenamiento. Las categorias son: Neutral, Energized, Fatigued

En el procesamiento de datos se optó por la creación de algunas variables tomando como entrada las variables de la base. Así, por ejemplo, se crearon las variables de Rango_Edad y Calidad_sueño las cuales buscan mostrar la información que está en otras variables, pero de una forma más condensada y en categorías más fáciles de interpretar que la variable original. Adicionalmente, se crearon variables utilizando la ecuación Mifflin-St Jeor que permite calcular la tasa metabólica basal (TMB). A partir de esta variable se calculó una segunda variable que es el requerimiento de ejercicio, el cual establece cual es el nivel de actividad de ejercicio que debería ejecutar una persona, a partir de las calorías diarias que consume, para mantener su peso.

Es importante señalar que no se realizó ningún tipo de transformación adicional a las variables originales de la base de datos ni tampoco un tratamiento a valores perdidos dado que la información para esta base se encontraba completa, como se muestra a continuación:

Variable	Cantidad de registros	Registros Nulos
Age	10000	non-null
Gender	10000	non-null
Height (cm)	10000	non-null
Weight (kg)	10000	non-null
Workout Type	10000	non-null
Workout Duration (mins)	10000	non-null
Calories Burned	10000	non-null
Heart Rate (bpm)	10000	non-null
Steps Taken	10000	non-null
Distance (km)	10000	non-null
Workout Intensity	10000	non-null
Sleep Hours	10000	non-null
Water Intake (liters)	10000	non-null
Daily Calories Intake	10000	non-null
Resting Heart Rate (bpm)	10000	non-null
VO2 Max	10000	non-null
Body Fat (%)	10000	non-null
Mood Before Workout	10000	non-null
Mood After Workout	10000	non-null
TMB	10000	non-null

Rango_Edad	10000	non-null
Calidad_sueño	10000	non-null
Requerimiento_Ejercicio	10000	non-null

A continuación, se muestran las variables modeladas y se detalla más el proceso que se realizó para la creación de estas.

2. Variables modeladas:

TMB	Número Decimal (float64)	TMB, tasa metabólica basal. Es la cantidad mínima de energía que el cuerpo necesita para sobrevivir.
Rango_Edad	Categorica (object)	Rangos de edad basados en los grupos etarios. Las categorías son: '1. Menor 18', '2. 18 - 26', '3. 26 - 40', '4. 40 - 59', '5. > 59'
Calidad_sueño	Categorica (object)	Calidad del sueño basada en las variables de "Sleep Hours" y "Resting Heart Rate (bpm)". Las categorías pueden ser: Mala, Regular, Buena, Excelente
Requerimiento_Ejercicio	Categorica (object)	Establece el requerimiento de ejercicio que debe tener la persona con base a la TMB calculada y la cantidad de calorías consumidas "Daily Calories Intake". Las categorías son: 'Sedentario', 'Actividad ligera', 'Actividad moderada', 'Actividad Intensa', 'Muy activa, Ejercicio intenso'

- **TMB, Tasa Metabólica Basal**

Para la primera variable: "TMB" que es la cantidad de calorías que se queman cuando el cuerpo está en reposo absoluto y que varía según el nivel de actividad que se realiza, se utilizó la ecuación Mifflin-St Jeor que establece los siguiente:

Cuando Genero = Male,

$$TMB_{Male} = 10 \times \text{peso (kg)} + 6,25 \times \text{altura (cm)} - 5 \times \text{edad (años)} + 5$$

Cuando Genero = Female,

$$TMB_{Female} = 10 \times \text{peso (kg)} + 6,25 \times \text{altura (cm)} - 5 \times \text{edad (años)} + 161$$

En otros casos,

$$TMB_{Other} = \frac{(TMB_{Male} + TMB_{Female})}{2}$$

Como se puede observar, en el caso de que Genero sea diferente de Male o Female entonces lo que se hace es calcular el TMB para ambos casos y promediar, esto ya que es la mejor aproximación cuando no se tiene el género declarado.

- **Requerimiento de Ejercicio**

Para la variable Requerimiento_Ejercicio se buscaba establecer cuál es el nivel de actividad que debería ejecutar una persona en sus entrenamientos a partir de la TMB calculada y el consumo de calorías diarias. Así a partir de la siguiente tabla:

Nivel de actividad	Factor multiplicador
Sedentario (poco o ningún ejercicio)	$TMB \times 1.2$
Actividad ligera (1-3 días/semana)	$TMB \times 1.375$
Actividad moderada (3-5 días/semana)	$TMB \times 1.55$
Actividad intensa (6-7 días/semana)	$TMB \times 1.725$
Muy activa (ejercicio físico extremo)	$>TMB \times 1.725$

Esta tabla establece los límites personalizados para cada persona. Con el dato de consumo de calorías diarias se puede establecer cuál es el nivel de actividad que una persona debería tener para poder mantener su peso estable. Ejemplo una persona con una TMB de 1800 calorías por día y con un consumo de 2700 calorías diarias debería realizar un nivel de actividad moderada ya que ese consumo de calorías se encuentra justo por debajo del factor multiplicador de la actividad moderada ($1800 \times 1.55 = 2790$).

- **Calidad de sueño**

La variable calidad de sueño utiliza como entradas la variable “Sleep Hours” y Resting Heart Rate (bpm) para poder establecer si el sueño tiene una calidad excelente, buena, regular o mala. Primero, se consideran 3 rangos para las horas de sueño tal que es malo tener menos de 7 horas, bueno tener entre 7 – 9 y regular más de 9 horas. En el caso de la frecuencia cardíaca en reposo se considera bueno si está por debajo de 60, regular entre 60 – 75 y mala por encima de 75. Esto nos genera 9 posibles escenarios que se establecen así:

Horas de sueño	Frecuencia Cardíaca en Reposo	Calidad del sueño
Más de 9 horas	Menor a 60 bpm	Regular
Más de 9 horas	Entre 60 - 75 bpm	Buena
Más de 9 horas	Mayor a 75 bpm	Mala
Entre 7 y 9 horas	Menor a 60 bpm	Excelente
Entre 7 y 9 horas	Entre 60 - 75 bpm	Buena
Entre 7 y 9 horas	Mayor a 75 bpm	Regular
Menos de 7 horas	Menor a 60 bpm	Buena
Menos de 7 horas	Entre 60 - 75 bpm	Regular
Menos de 7 horas	Mayor a 75 bpm	Mala

De esta manera se utilizan estas dos variables para mostrar de una manera diferente la misma información, pero en categorías más fáciles de comprender para cualquier público.

- **Rango de edad**

Para la construcción de esta variable se utilizaron los grupos etarios los cuales son agrupaciones que se suelen utilizar en múltiples estudios. Los grupos etarios son:

- Primera infancia: de 0 a 5 años
- Infancia: de 6 a 11 años
- Adolescencia: de 12 a 18 años
- Juventud: de 18 a 26 años
- Adultez: de 27 a 59 años
- Persona mayor: de 60 años o más

A estos grupos se les hizo una modificación, dado que el mínimo de la variable “Age” es 18 entonces los grupos de Primera infancia, Infancia y Adolescencia no serían necesario construirlos. Por el contrario, el grupo de adultez se podría abrir un poco más para no generar un grupo muy agrupado entonces los rangos de edad basado en estos grupos etarios que se construyeron son:

- '1. Menor 18'
- '2. 18 - 26'
- '3. 26 - 40'
- '4. 40 - 59'
- '5. > 59'

Al igual que la variable de calidad de sueño, esta variable en rangos muestra de una manera diferente la misma información, pero en categorías más fáciles de comprender para cualquier público

3. Para replicar la base obtenida se deben seguir los siguientes pasos:

Cargue de librerías e información


```
[1]: import kagglehub
import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.linear_model import LogisticRegressionCV
from sklearn.metrics import classification_report
from sklearn.preprocessing import StandardScaler, LabelEncoder, OneHotEncoder

pd.set_option('display.max_columns', None)

/Library/Frameworks/Python.framework/Versions/3.12/lib/python3.12/site-packages/tqdm/auto.py:21: TqdmWarning: IPProgress not found. Please update jupyter and ipywidgets. See https://ipywidgets.readthedocs.io/en/stable/user_install.html
from .autonotebook import tqdm as notebook_tqdm

[2]: # Download latest version
path = kagglehub.dataset_download("adilshamim8/workout-and-fitness-tracker-data")

print("Path to dataset files:", path)
print(os.listdir(path))

Path to dataset files: /Users/danielgaravito/.cache/kagglehub/datasets/adilshamim8/workout-and-fitness-tracker-data/versions/1
['workout_fitness_tracker_data.csv']

[3]: archivo = os.listdir(path)
df_WOF = pd.read_csv(os.path.join(path, archivo[0]), sep=',')
df_WOF
```

Creación de variables

```
## Creación de variables ###

## Rangos de edad
conditions_age = [
    df_WOF['Age'] <= 18,
    (df_WOF['Age'] > 18) & (df_WOF['Age'] <= 26),
    (df_WOF['Age'] > 26) & (df_WOF['Age'] <= 40),
    (df_WOF['Age'] > 40) & (df_WOF['Age'] <= 59),
    df_WOF['Age'] > 59
]

values_age = [
    '1. Menor 18',
    '2. 18 - 26',
    '3. 26 - 40',
    '4. 40 - 59',
    '5. > 59'
]

## Calidad del sueño
conditions_sueno = [
    (df_WOF['Sleep Hours'] > 9) & (df_WOF['Resting Heart Rate (bpm)'] <= 60),
    (df_WOF['Sleep Hours'] > 9) & (df_WOF['Resting Heart Rate (bpm)'] > 60) & (df_WOF['Resting Heart Rate (bpm)'] <= 75),
    (df_WOF['Sleep Hours'] > 9) & (df_WOF['Resting Heart Rate (bpm)'] > 75),
    (df_WOF['Sleep Hours'] >= 7) & (df_WOF['Sleep Hours'] <= 9) & (df_WOF['Resting Heart Rate (bpm)'] <= 60),
    (df_WOF['Sleep Hours'] >= 7) & (df_WOF['Sleep Hours'] <= 9) & (df_WOF['Resting Heart Rate (bpm)'] > 60) & (df_WOF['Resting Heart Rate (bpm)'] <= 75),
    (df_WOF['Sleep Hours'] >= 7) & (df_WOF['Sleep Hours'] <= 9) & (df_WOF['Resting Heart Rate (bpm)'] > 75),
    (df_WOF['Sleep Hours'] < 7) & (df_WOF['Resting Heart Rate (bpm)'] <= 60),
    (df_WOF['Sleep Hours'] < 7) & (df_WOF['Resting Heart Rate (bpm)'] > 60) & (df_WOF['Resting Heart Rate (bpm)'] <= 75),
    (df_WOF['Sleep Hours'] < 7) & (df_WOF['Resting Heart Rate (bpm)'] > 75)
]
```

```

values_sueno = [
    'Regular',
    'Buena',
    'Mala',
    'Excelente',
    'Buena',
    'Regular',
    'Buena',
    'Regular',
    'Mala'
]

# Fórmula de Mifflin-St Jeor y la TMB tasa metabolica basal

def calcular_TMB(df):
    if df['Gender'] == 'Male':
        return (10 * df['Weight (kg)'] + (6.25 * df['Height (cm)']) - (5 * df['Age'])) + 5
    elif df['Gender'] == 'Female':
        return (10 * df['Weight (kg)'] + (6.25 * df['Height (cm)']) - (5 * df['Age']) - 161
    else:
        # Promedio en caso de género no especificado
        tmb_male = (10 * df['Weight (kg)'] + (6.25 * df['Height (cm)']) - (5 * df['Age'])) + 5
        tmb_female = (10 * df['Weight (kg)'] + (6.25 * df['Height (cm)']) - (5 * df['Age']) - 161
        return (tmb_male + tmb_female) / 2

df_WOF['TMB'] = df_WOF.apply(calcular_TMB, axis=1)

### Requerimiento de Ejercicio con base a la TMB calculada

conditions_RE = [
    df_WOF['Daily Calories Intake'] <= df_WOF['TMB'] * 1.2,
    (df_WOF['Daily Calories Intake'] > df_WOF['TMB'] * 1.2) & (df_WOF['Daily Calories Intake'] <= df_WOF['TMB'] * 1.375),
    (df_WOF['Daily Calories Intake'] > df_WOF['TMB'] * 1.375) & (df_WOF['Daily Calories Intake'] <= df_WOF['TMB'] * 1.55),
    (df_WOF['Daily Calories Intake'] > df_WOF['TMB'] * 1.55) & (df_WOF['Daily Calories Intake'] <= df_WOF['TMB'] * 1.725),
    (df_WOF['Daily Calories Intake'] > df_WOF['TMB'] * 1.725)
]

values_RE = [
    'Sedentario',
    'Actividad ligera',
    'Actividad moderada',
    'Actividad intensa',
    'Muy activa, Ejercicio intenso'
]

#TMB = (10 x peso en kg) + (6,25 x altura en cm) - (5 x edad en años) + 5 para hombres y TMB = (10 x peso en kg) + (6,25 x altura en cm) - (5 x edad en años) - 161 para mujeres

df_WOF['Rango_Edad'] = np.select(conditions_age, values_age)
df_WOF['Calidad_sueno'] = np.select(conditions_sueno, values_sueno)
df_WOF['Requerimiento_Ejercicio'] = np.select(conditions_RE, values_RE)
df_WOF = df_WOF.drop('User ID', axis = 1)
df_WOF

```

Referencias

Harris, J. A., & Benedict, F. G. (1919). A biometric study of human basal metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, 4(12), 370–373. <https://doi.org/10.1073/pnas.4.12.370>

Roza, A. M., & Shizgal, H. M. (1984). The Harris Benedict equation reevaluated: Resting energy requirements and the body cell mass. *The American Journal of Clinical Nutrition*, 40(1), 168–182. <https://doi.org/10.1093/ajcn/40.1.168>

Mifflin, M. D., St Jeor, S. T., Hill, L. A., Scott, B. J., Daugherty, S. A., & Koh, Y. O. (1990). A new predictive equation for resting energy expenditure in healthy individuals. *The American Journal of Clinical Nutrition*, 51(2), 241–247. <https://doi.org/10.1093/ajcn/51.2.241>

World Health Organization. (1985). *Energy and protein requirements: Report of a joint FAO/WHO/UNU expert consultation*. WHO Technical Report Series 724. Ginebra: OMS.

Institute of Medicine. (2005). *Dietary reference intakes for energy, carbohydrate, fiber, fat, fatty acids, cholesterol, protein, and amino acids*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/10490>