



Proyecto de Ciencia de Datos

Introducción

La ciencia de datos nos ha abierto la puerta un sin número de posibilidades de mejora en casi cualquier aspecto de nuestra vida, desde facilitarnos las búsquedas en internet, hasta mejorar nuestra salud mediante el uso de herramientas digitales. Se centra en el estudio de los datos. Data Science combina la estadística, las matemáticas y la informática para interpretar datos. El objetivo es tomar decisiones

Tomando en específico, este proyecto el cual busca analizar tus hábitos alimenticios y como estos están relacionados las calorías que necesita una persona para tener una dieta saludable. Además, te muestra de una manera clara los datos nutrimentales de tus alimentos lo cual de una manera directa tiene un impacto en la conciencia de lo que se introduce en nuestro organismo

Este proyecto también tiene la finalidad de comenzar a observar los registros nutrimentales de fibra, lo cual es un dato relevante para las personas que padecen enfermedades como gastritis crónica. Si bien esta característica puede quedar en fase beta ya que faltaría más tiempo de desarrollo.

- ¿Quién es el cliente?

El cliente en este caso es el usuario que busca llevar un mejor control de sus hábitos alimenticios.

- ¿Qué problemas estás tratando de resolver?

México es un país con altos índices de sobre peso y eso viene en gran medida por la falta de conciencia en nuestros hábitos alimenticios, por lo que este programa busca crear conciencia en la alimentación de las personas

- ¿Qué solución o soluciones la Ciencia de Datos tratará de proveer?

Se buscará dar solución mediante un programa de computadora basado Python.

- ¿Qué necesitas aprender para poder desarrollar la solución o soluciones?

Se necesitarán conocimientos generales en programación (Python) y además de eso conocimientos de probabilidad y estadística para poder analizar los datos obtenidos

- ¿Qué deberás hacer para desarrollar tu solución?

Registrar los datos de alimentación diaria del cliente y desarrollar el código necesario

- ¿Qué tipos de datos se necesitaron?

Fecha	Hora	Nombre	Calorias (kcal)	Lípidos (g)	Carbohidratos (g)	Proteína (g)	Fibra(g)
21/05/2022	08:00 AM	Almuerzo	400	20	50	30	5

- ¿De dónde se obtuvieron los datos?

Los alimentos y sus fechas de consumo son obtenidas por el cliente.

Los valores nutrimentales son obtenidos mediante una pagina de nutrición llamada Fatsecret.com.

- ¿Los datos a usar son adecuados para hacer el análisis?

Si, ya que previo al proyecto se descartaron otros datos no necesarios-

- ¿Qué ajustes se tuvieron que hacer a los datos (agregar datos, integrar datos, modificar datos (remove o eliminar información)?

Se eliminaron los datos de sodio y se agregaron los datos de fibra para darle otro enfoque al proyecto.

Modelación de los datos

1. ¿Cuántas interacciones realizaste para obtener el resultado?

Se realizaron muchas iteraciones. Un numero aproximado sería más de 50, ya que se realizaron muchos cambios al código conforme fue desarrollándose.

2. ¿Qué resultados arrojo en análisis?

Ya con el apoyo del programa podemos organizar mejor la información y obtener datos estadísticos y cuál es la correlación que hay entre ciertos tipos de datos con las calorías.

```
https://DutifulClientsideApplet.jessdavid3.repl.run

(128, 8)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 128 entries, 0 to 127
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Fecha                 128 non-null   object
1   Hora                  128 non-null   object
2   Nombre                128 non-null   object
3   Calorias (kcal)       128 non-null   int64
4   Lípidos (g)           128 non-null   float64
5   Carbohidratos (g)     128 non-null   float64
6   Proteína (g)          128 non-null   float64
7   Fibra (g)             128 non-null   float64
dtypes: float64(4), int64(1), object(3)
memory usage: 8.1+ KB
None
Fecha                0
Hora                 0
Nombre               0
Calorias (kcal)      0
Lípidos (g)          0
Carbohidratos (g)    0
Proteína (g)         0
Fibra (g)            0
dtype: int64
False
Empty DataFrame
Columns: [Fecha, Hora, Nombre, Calorias (kcal), Lípidos (g), Carbohidratos (g), Proteína (g), Fibra(g)]
Index: []
Estadística descriptiva de los datos:

```

	count	mean	std	...	50%	75%	max
Calorias (kcal)	128.0	162.640625	172.571096	...	121.0	187.25	1160.0
Lípidos (g)	128.0	5.524844	9.390896	...	1.5	8.70	68.0
Carbohidratos (g)	128.0	16.891016	22.668615	...	5.5	23.25	112.0
Proteína (g)	128.0	11.520625	15.182003	...	4.2	20.00	100.0
Fibra (g)	128.0	1.405469	2.651825	...	0.1	1.50	13.0

```

[5 rows x 8 columns]
Coeficientes
Carbohidratos (g)    5.261055
Lípidos (g)          4.537442
Proteína (g)         4.186693
Fibra (g)            -2.428102
```

https://DutifulClientsideApplet.jessdavid3.repl.run

```
15    355  352.156423
16     14   28.491203
17     23   22.395002
18    204  231.920678
19    199  168.267129
```

Raíz de la desviación media al cuadrado: 26.43923039699253

Starting X

.....	Fecha	Hora	Nombre	...	Carbohidratos (g)	Proteína (g)	Fibra (g)
67	03/05/2020	13:00	Cerdo	...	0.0	40.0	0.0
86	05/05/2020	10:00	platano	...	26.9	1.2	3.0
34	03/05/2020	11:00	fresa	...	5.0	1.0	0.5
124	09/05/2020	09:00	huevo	...	1.5	20.0	0.0
89	05/05/2020	15:00	Pescado	...	0.0	34.0	0.0

[5 rows x 8 columns]

	Lípidos (g)	Calorias (kcal)
67	40.00	500
86	0.30	105
34	0.20	20
124	12.00	160
89	2.00	160
21	0.00	50
109	0.10	23
6	0.20	16
75	0.40	90
121	0.12	49

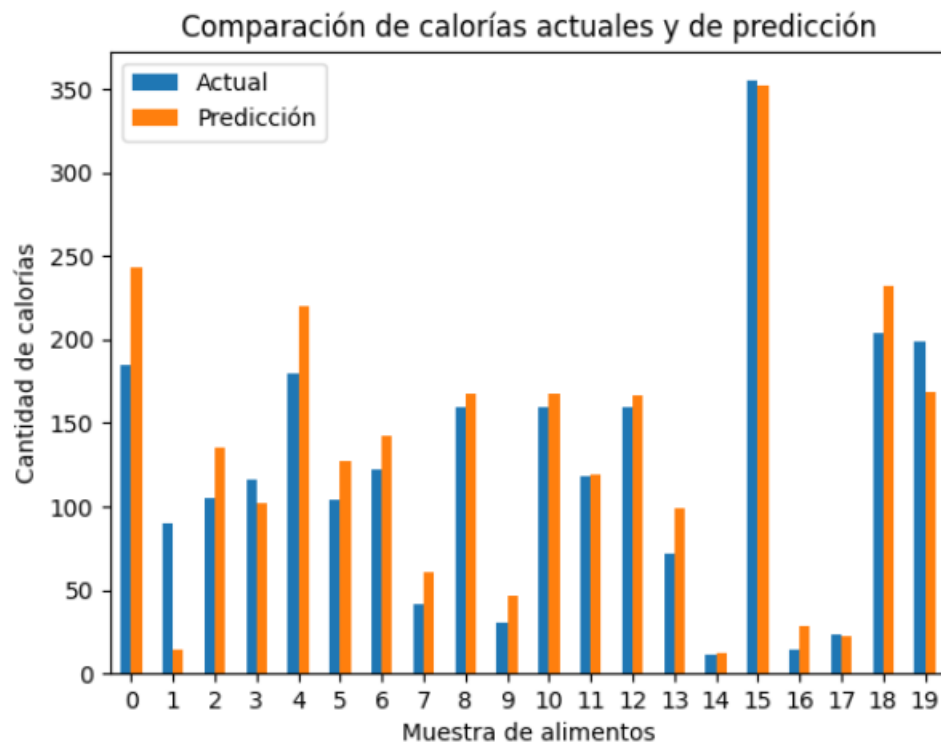
0.7825822883703811

Correlación con "Calorias (kcal)"

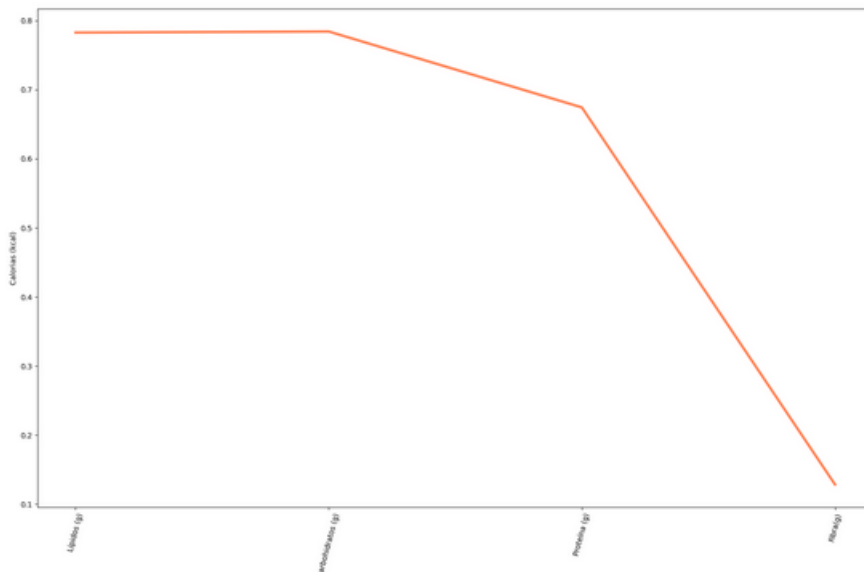
Lípidos (g): 0.782582
Carbohidratos (g): 0.784192
Proteína (g): 0.674274
Fibra(g): 0.128477

Conteo de datos en cada cluster

```
0    84
1    32
2     9
3     3
dtype: int64
```



En la figura anterior podemos observar gráficamente como mediante un data frame nos muestra nuestra alimentación actual y las predicciones futuras de como podrían ser.



Esta grafica no muestra una correlación de 0 a 1 donde 1 es 100% y podemos observar cómo los lípidos y los carbohidratos tienen una mayor correlación con las

calorías. Estos datos coinciden con opiniones nutricionales de profesionales, ya que la cantidad de grasas siempre va de la mano con las calorías que puede aportarnos un alimento.

3. ¿Qué conclusiones brindó el análisis para aceptar o rechazar la hipótesis inicial?

El análisis de información cumplió su propósito el cual era dar a conocer la importancia de los alimentos que consumimos y concientizar al cliente de los valores nutrimentales que contienen los alimentos que ingiere. Este proyecto no muestra datos como las calorías recomendadas o datos nutricionales óptimos, pues cada persona requiere diferentes niveles de estos. Sin más se advierte que este programa sirve como una herramienta para ser utilizada de la mano de un profesional de la salud que marque esos límites necesarios.

Reflexión final (Conclusiones)

Mi conclusión va de la mano con el comentario anterior. Las cantidades nutrimentales óptimas incluidas las calorías varían de persona a persona y dependen mucho también de sus objetivos, ya sean estéticos (para subir o bajar), deportivos o de salud. Por lo que siempre debe ser utilizado este programa con asesoría nutricional de un profesional el cual plantee las metas nutrimentales diarias objetivo.

Por ende, la hipótesis inicial se pone en un segundo plano, dejando como objetivo primordial la exposición de utilización de datos como método de control y medición para regular acciones futuras de nutrición.

Los modelos utilizados fueron diversos, pero el que más me gusto fue el de correlación pues muestra de una manera cuantitativa la relación entre la variable dependiente con las variables independientes. Siendo bastante interesante y más práctico verla gráficamente, lo cual facilita la explicación de datos al cliente.

El manejo de datos personales puede ser utilizado para diversos fines, y las empresas actualmente cuentan con el derecho de utilizar y guardar muchos de nuestros datos privados. Tomando un caso en específico, podemos nombrar a las compañías de teléfonos e internet. Las cuales almacenan datos de con quien hablamos, cuando, donde y cuanto hablamos con diferentes personas. Además, pueden utilizar tu ubicación para saber tus rutas y actividades diarias. Teniendo esto en cuenta creo es importante cuidar los datos que se utilizan y manejarlos como lo que son, datos privados que merecen ser salvaguardados para que no puedan ser utilizados con fines no éticos.

Anexo Liga de código de replit para revisión de código:

<https://repl.it/join/imxnpbzv-jessdaviddavid3>