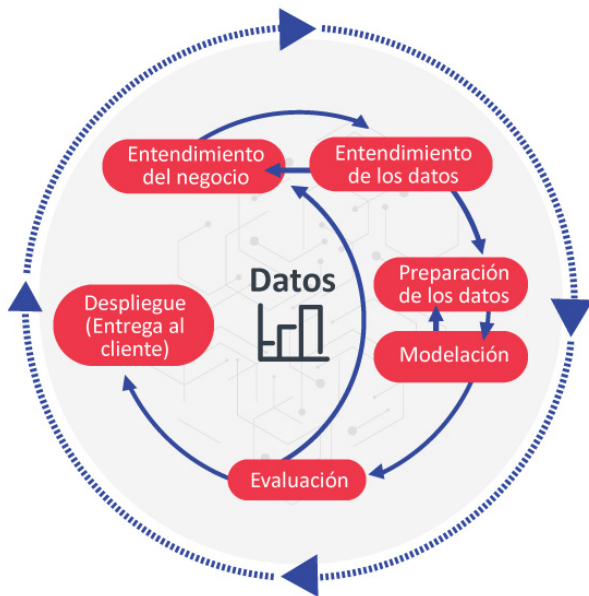


La ciencia de datos combina múltiples campos que incluyen estadísticas, métodos científicos y análisis de datos para extraer el valor de los datos. Así mismo, un proyecto de ciencia de datos se basa en la recolección, preparación, análisis e interpretación de datos para llegar a un resultado en particular. Desde mi punto de vista, la ciencia de datos puede traer muchos beneficios al desarrollo de personal, educativo, industrial e incluso del país en múltiples aspectos.

Tanto desde el área donde las empresas utilizan estas herramientas para crear mejores estrategias y aumentar sus ingresos por medio de mejoras en sus servicios o productos o en estrategias de mercadotecnia.

Como lo sería en el uso del sector salud, como es el ejemplo de este proyecto de ciencia de datos.

Para la realización del mismo proyecto se utilizará una metodología como la siguiente:



El cliente para este proyecto de ciencia de datos será mi propia persona, ya que yo elegí tomar la ruta de un proyecto personal, donde analizaré mi ingesta de alimentos para un estudio sobre mi propia salud.

Lo que intento lograr con este proyecto es realizar un estudio para mejorar mi estado físico. Para esto realicé un conteo de las calorías que consumí durante todo el día. Esto

se realizó utilizando una aplicación llamada *MyFitnessPal*, la cual también sirvió para tener un conteo de los macronutrientes (Proteínas, carbohidratos y grasas) que consumí, así como de las calorías de cada alimento y su contenido en miligramos de sodio.

Estos datos fueron vaciados en una tabla de excel que servirá como base de datos y principal fuente de información para este proyecto de ciencia de datos.

Esta también la uní con los datos que me proporcionaron otras dos aplicaciones, las cuales son *Wyze* y la aplicación *Strong*. La primera está unida a una báscula digital que me proporciona datos sobre mi peso, mi índice de grasa corporal y mi porcentaje de músculo en cuerpo.

La segunda me ayudará a llevar una bitácora de los entrenamientos que realice, esto me ayudará a saber si el peso que levanto ha aumentado o si la intensidad en mis entrenamientos ha variado. Los datos de estas también se encuentran en hojas de cálculo para analizarlos con respecto al tiempo.

La razón por la que realizo este proyecto con la ayuda de la ciencia de datos es porque debido a la misma naturaleza de los objetivos que se buscan, se requiere de una extensa cantidad de días, y en cada día se generan una gran cantidad de datos. Lo cual volvería el analizar todos estos datos de manera sería una tarea casi imposible de realizar. Debido a esto busco que con la ayuda de la ciencia de datos pueda tener un acercamiento mayor al entendimiento de que alimentos son más útiles para mi caso personal en el objetivo que busco lograr, el cual reitero es tener un mejor aspecto y condición física.

Para esto necesito aprender varias cosas, como el efecto que tienen diferentes macronutrientes en el cuerpo, cuáles comidas son más amigables para este objetivos. Así mismo como una manera eficiente de analizar estos datos por medio de la programación en python.

De esta manera, este proyecto se realizó siguiendo la metodología SMART, definiendo así metas a corto, mediano y largo plazo que me permitieron el poder tener en tiempo y forma los datos necesarios para el análisis deseado.

Considero que las columnas de la base de datos que más información me proporcionaron sobre mi salud e ingesta fueron: “Nombre”, “Fecha”, “Calorías” y “Proteína”.

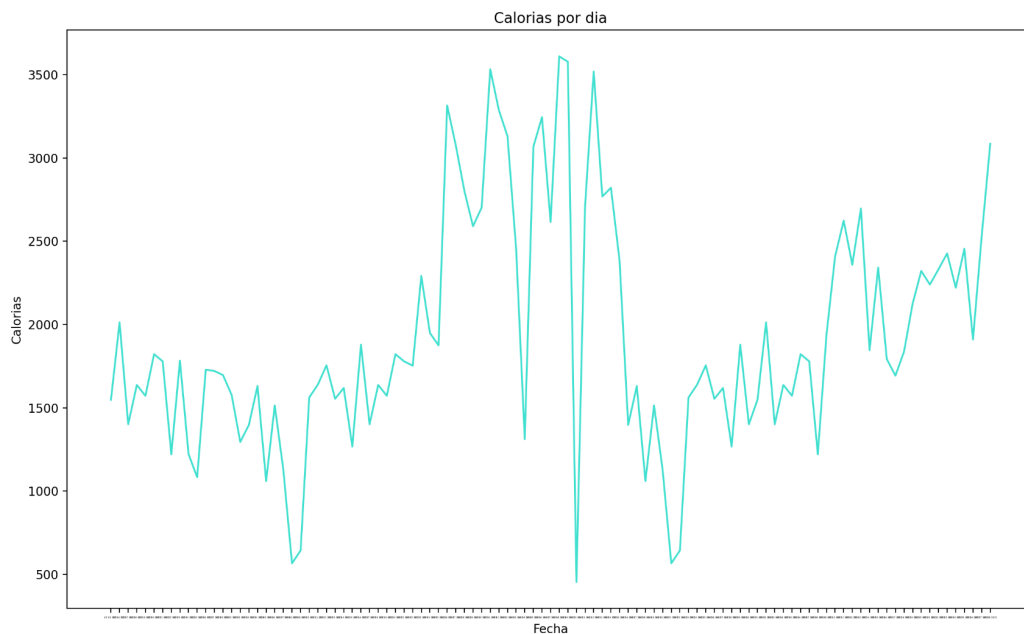
Esto es debido a que estos datos, en combinación con las otras bases de datos ya mencionadas anteriormente ayudaron a poder crear un análisis mucho más preciso y certero sobre cómo estos alimentos contribuyen a los cambios fisiológicos en mi persona.

Como ya se ha mencionado, se busca encontrar la correlación entre cuestiones como peso, grasa corporal y masa muscular con las calorías consumidas y los diferentes macronutrientes. Es debido a que el sodio no contiene calorías que esta columna no parece aportar información pertinente hacía este proyecto de ciencia de datos. Los apartados de datos que se establecieron fueron suficientes para realizar un modelo fácil de interpretar. Esto se presentará más adelante en este trabajo.

Se contó con 955 entradas de alimentos, 38 entradas de entrenamiento y 38 entradas de mediciones de peso corporal a lo largo de estas 15 semanas de recolección de datos. No fueron necesarias tantas iteraciones para lograr los objetivos esperados debido a que con anticipación se planeó como sería el código, utilizando el llamado “pseudocódigo” lo cual ayudó a poder crear el código de una manera más fluida y sin tantos errores en el proceso.

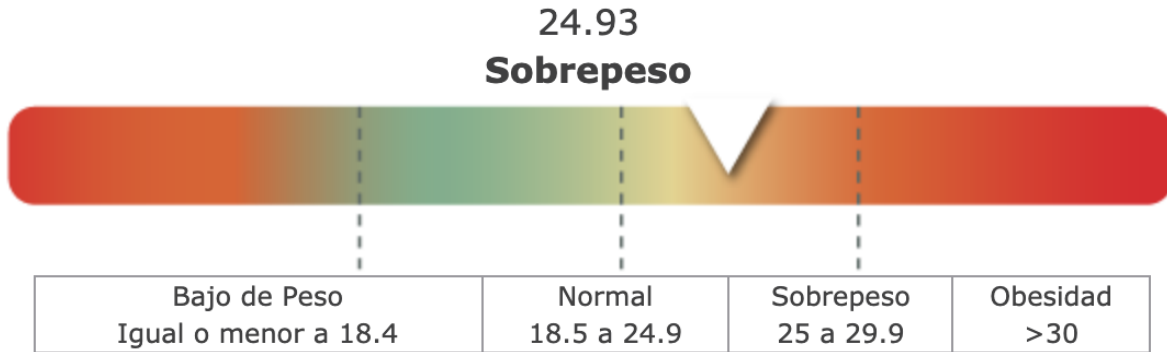
Por la misma razón de que se están utilizando múltiples bases de datos, fue necesario el realizar una fusión de los mismos. Esto es debido a que se recolectó información acerca de diferentes elementos que son complementarios para el análisis completo del proyecto. Debido a la naturaleza de estos datos, la fusión de estos será de naturaleza cooperativa.

Además de estos ajustes no resultó necesario el realizar mayores modificaciones a los datos, ya que se intentó que se registraran de la mejor manera posible para evitar así tener que lidiar con la limpieza de los datos, eliminación de registros o que se presentaran complicaciones. De la misma manera, como es común en un proyecto de ciencia de datos, sería bastante útil contar con una mayor cantidad de registros, pero debido a las limitaciones de tiempo esto no es posible, aún así se pudo realizar un proyecto bastante confiable con los datos que se recolectaron, como se puede observar a continuación:



Como se observa en la gráfica arriba, el consumo diario de calorías ha variado a lo largo del tiempo, siendo unos días muy alto y otros manteniéndose muy bajo. Según el Instituto Mexicano del Seguro Social, una persona con mis características físicas (Tamaño, peso, edad y sexo) así mismo con el nivel de actividad promedio que yo presento, necesita de 2,900 kcal diarias para mantener su peso.

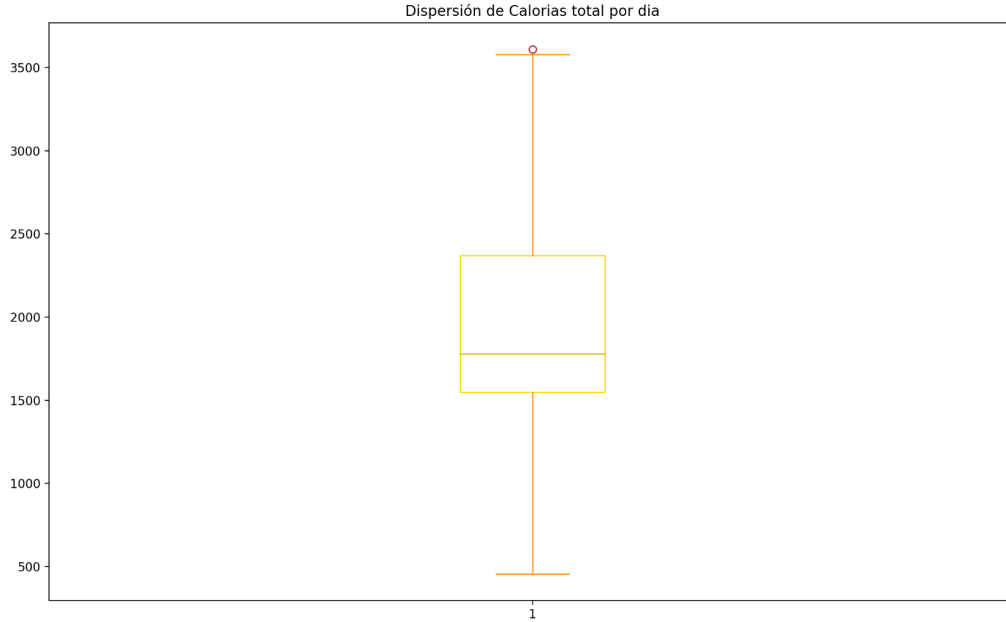
Así mismo, gracias a la tabla de IMC presentada a continuación, podemos observar que al inicio de estas 15 semanas me encontraba en la categoría de “sobrepeso”.



De acuerdo al Pub Médico Central de Europa (McMurray R, 1985). El mantener déficits calóricos elevados por un tiempo prolongado puede tener efectos negativos directos en las aptitudes físicas de los individuos, no solo frenando su mejora deportiva, sino que incluso empeorarla. Así mismo esto puede tener efectos negativos en la salud física y emocional y puede conducir a lo conocido como “dietas yo-yo” donde el individuo vuelve a subir de peso después de una dieta muy restrictiva.

Es debido a esto que también se encontraba dentro de mi interés el conocer si estaba en un déficit muy elevado. Esto fue posible gracias a la utilización de python y librerías como matplotlib, numpy y pandas. Pude obtener la siguiente información.

	Calorias (kcal)	Carbohidratos (g)	Lípidos (g)	Proteína (g)	Sodio (mg)
count	103.000000	103.000000	103.000000	103.000000	103.000000
mean	1923.779612	152.236893	76.171845	145.705049	1964.044660
std	700.937139	66.688303	32.268668	80.006741	1219.303511
min	454.000000	51.700000	9.900000	14.800000	250.000000
25%	1548.000000	108.300000	52.250000	89.650000	1196.500000
50%	1779.000000	149.400000	69.900000	121.300000	1895.100000
75%	2371.450000	199.500000	98.100000	181.450000	2491.000000
max	3610.000000	340.100000	168.000000	381.700000	7046.400000



Gracias a estos datos pude determinar que en promedio me encontraba en un déficit calórico de 977 Kcal. De acuerdo con diario “Clarín”, 3500Kcal son equivalentes a 1lb de grasa. Tomando esto en cuenta, a la semana me encontraba en un déficit de 6839Kcal o lo equivalente a 1.95lb semanales. En mi hipótesis de predicciones esto significaba que al final de mis registros de 15 semanas debí haber perdido 29.3lb.



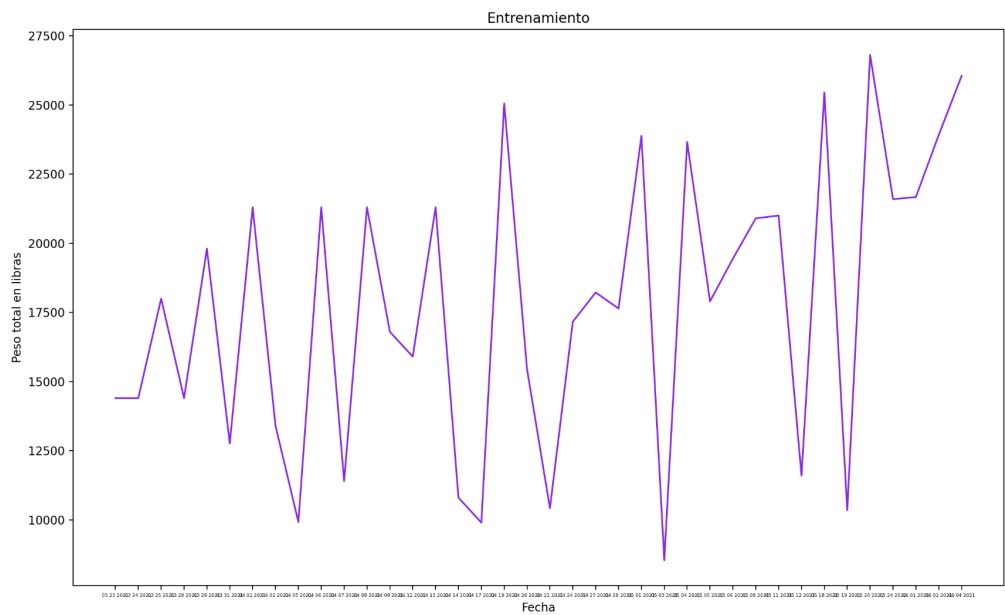
Como podemos observar, la gráfica de mi peso corporal muestra picos en medio de la misma, esto lo atribuyo a que en este mismo periodo de tiempo, mi ingesta calórica era en superávit de mi nivel de mantención de peso, lo cual provocó que en lugar de perder grasa corporal, aumentara.

Así mismo se puede observar que mi pico en peso fue de 93 kg, teniendo mi estado más bajo en la última medición con 86 kg. Esto se traduce a una pérdida de 7kg o 15.5lb, siendo esta casi la mitad de lo predicho en mi hipótesis inicial. Esto se lo atribuyo a tres cosas principalmente, la primera de estas es a los días en los que me encontré en un superávit calórico, los cuales me hacían aumentar grasa corporal, y la segunda a los días donde era más sedentario y no tenía un gasto calórico tan elevado, la tercera es debido a los cambios en mi masa muscular la cual discutiré más adelante. Con esto concluyo que si me mantengo en un déficit calórico, y realizo suficiente ejercicio para conseguir este desbalance en energía, podré seguir bajando de peso. Aún busco realizar esto ya que aún no me encuentro en el estado físico que propuse como meta.

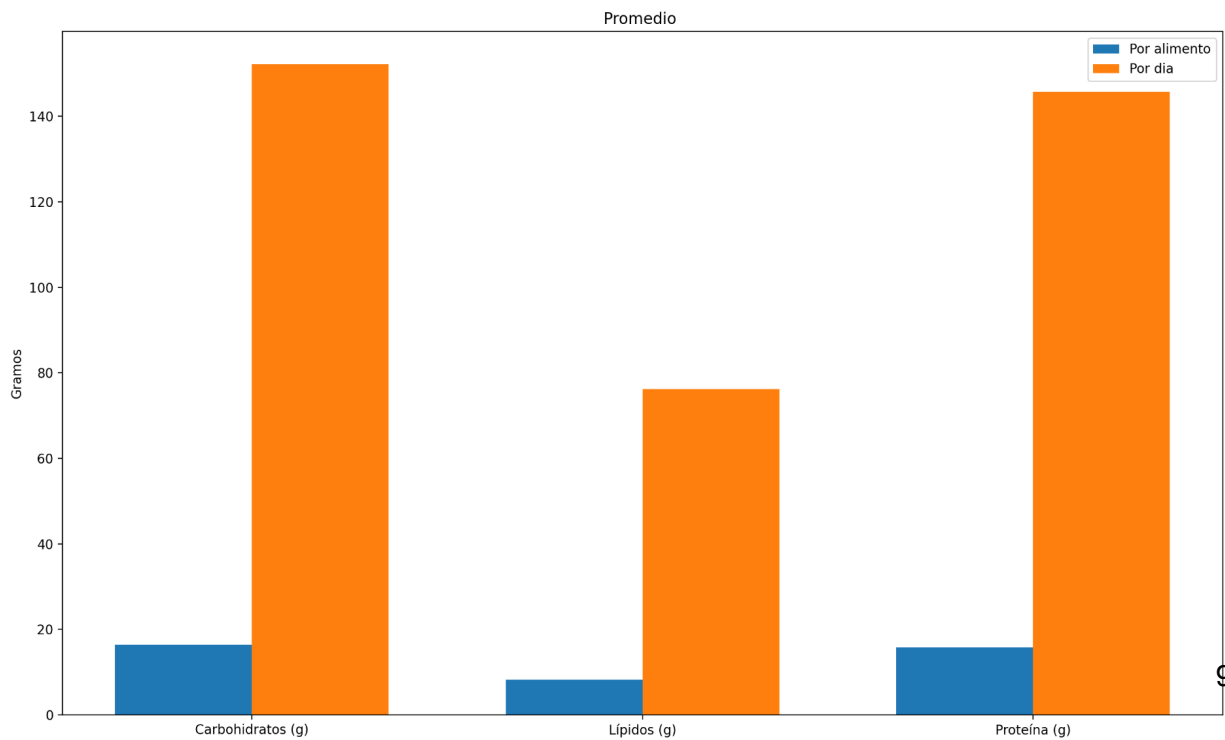
Como promedio podría decirse que logré perder 0.47kg (7kg/15 semanas) por semana, lo cual se encuentra saludable ya que es menor a un 0.52% de pérdida de peso por semana, el cual de acuerdo al Instituto Mexicano del Seguro Social se encuentra en rangos saludables. (Esto es considerando el promedio de todos los registros de peso total, siendo este de 89.25kg)

Por otro lado, también consideraba la hipótesis que al bajar mi ingesta calórica y bajar de peso, mis entrenamientos se verían afectados, causando que levantara menos peso total en mis entrenamientos. Esto debido a una menor energía para realizar los entrenamientos, un déficit muy elevado que no permitiría el realizar mucha actividad física para preservar mejor calorías o incluso que una menor masa corporal lo cual se traducir a menor fuerza siguiente principios de física $F=M*A$, donde F es fuerza, M es masa y A es aceleración.

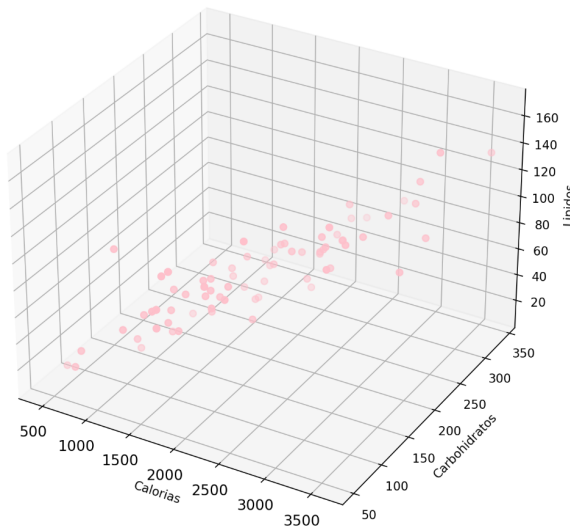
Evidencia 2. Proyecto de Ciencia de Datos



Como se puede observar, esto no sucedió, incluso teniendo mi entrenamiento con mayor peso dentro de esta semana #15. Esto se lo atribuyo a que de acuerdo con The Royal Society of Chemistry, se recomienda un consumo de proteína mínimo de 0.8 gramos por kilogramo de peso o de 1 gramo de proteína por kg para hacer posible el crecimiento muscular.



Como se puede observar, mi consumo diario de proteína se encuentra cerca de 146 gramos diarios, lo cual equivale a casi 1.5 gramos por kilogramo de peso en mi periodo de mayor peso con 93kg. Con esto puedo concluir que mi masa muscular aumentó, permitiendo tener mejores entrenamientos a lo largo del tiempo, y así mismo, frenando de cierta medida mi pérdida de peso total representado en la báscula. Esto fue lo que pudo permitir que mi peso de entrenamiento pudiera escalar hasta 26,810 lb de peso total levantado durante el entrenamiento, siendo el peso más bajo de 8,540 lb.



Así mismo como se puede observar en esta gráfica y en la pasada, mi consumo de carbohidratos en promedio no es bajo, estando cerca de 152 gramos por día. De acuerdo al Centro Farmacéutico Otsuka, los carbohidratos son la fuente de energía más fácilmente procesable por el cuerpo, siendo utilizados estos casi al instante, teniendo únicamente 4 kcal/gr. Al contrario de las grasas que tardan más en digerirse, tienen un efecto térmico menor y proporcionan 9 Kcal/gr.

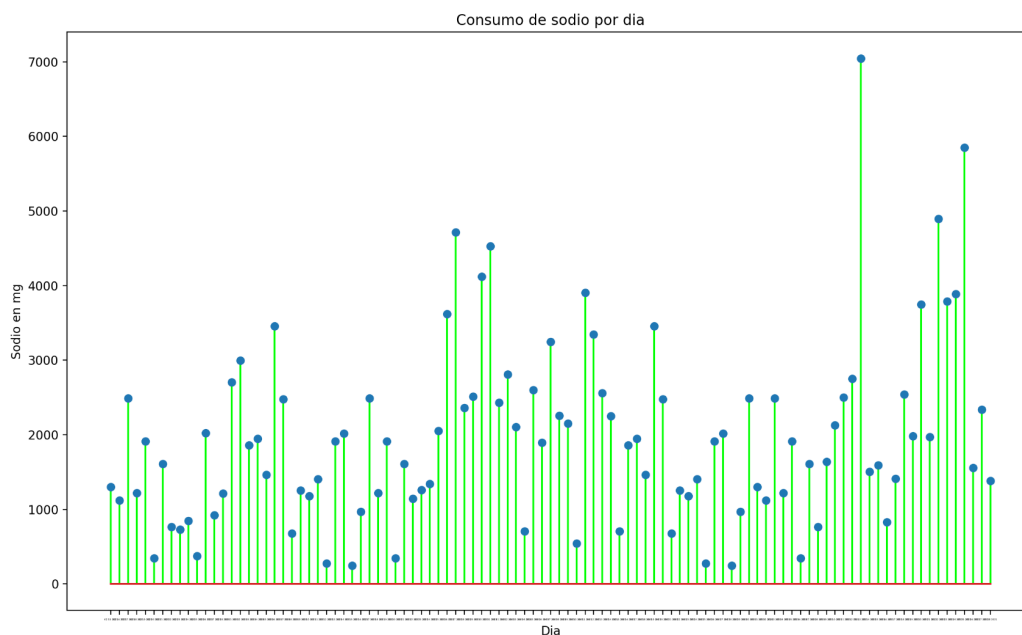
Es debido a esto que la energía necesaria para realizar mis actividades nunca faltó y que pude mantener el consumo calórico en un nivel bajo, evitando en gran medida las grasas, que son caloricamente densas y se está mejor utilizando más carbohidratos.

Gracias a estos análisis, puede determinar aspectos que son más perceptibles visualmente, como el peso levantado o la grasa corporal en una persona. Pero dentro de este proyecto también quise enfocarme en el consumo de micronutrientes y efectos no tan fácilmente notorios como los ya mencionados.

Me enfoqué específicamente en el sodio, debido a que es un elemento muy común en los alimentos pero que consumido en altas concentraciones puede tener efectos severos en la salud. La Organización Mundial de la Salud recomienda mantener un consumo diario de sodio por debajo de los 2 gramos.

Se estima que en 2017 murieron 17,8 millones de personas afectadas por enfermedades cardiovasculares, lo que representa el 32% de todas las muertes registradas en el mundo. Los datos indican que la reducción de la ingesta de sodio reduce significativamente la tensión arterial en los adultos.

El consumo alto de sodio también puede causar problemas de retención de líquidos, artritis, problemas cerebro vasculares, cerebrales, entre otros.



Como se puede observar en la tabla, mi consumo de sodio se encuentra en un promedio diario de 1963 mg (1.97gr) de sodio, existiendo días donde solo han sido 250mg (0.25gr). El problema radica en esos días donde el consumo de sodio supera los límites recomendados, habiendo un pico de incluso 7000 mg (7 gr) de sodio, estando 3.5 veces arriba de los niveles recomendados. Con esto debo detectar que ciertos alimentos pueden contener una alta cantidad de sodio. Debido a que el promedio se encontró por debajo del límite recomendado, puedo concluir que mi consumo del mismo no se encuentra en niveles peligrosos para mi salud, ni que esto pueda suponer indicios de enfermedades relacionadas al alto consumo de sodio.

Es debido a estos análisis, que después de estas semanas de trabajar en un proyecto de ciencia de datos, puedo decir a manera de conclusión que la ciencia de datos es una disciplina extremadamente útil. La ciencia de datos sirve para identificar de forma masiva y rápida todos los sets de datos, desde los más sencillos hasta los más robustos y complicados. En el caso de este proyecto me sirvió ampliamente para conocer en qué medida he progresado con mis objetivos, que ha logrado que mis cambios sean positivos o negativos y así mismo el cómo se encuentra mi salud. Pero como ya se mencionó anteriormente, esta también puede ser muy útil en un infinidad de industrias como lo es la médica, legal, mercadotecnia, aeronáutica, por mencionar algunas.

Además, gracias a los sistemas de análisis de datos que utilizan lo que es machine learning, es posible implementar modelos de analítica predictiva capaces de analizar la información de gran cantidad. En un proyecto más avanzado de ciencia de datos estas son demasiado útiles para intentar conocer datos del futuro basándonos en datos del pasado. Está dentro de mis intereses el continuar con este proyecto e implementar esta capacidad predictiva para conocer de mejor manera cuando podría estimar llegar a cierto resultado.

Así mismo considero que debido a la gran cantidad de datos que un analista de datos puede llegar a manipular, se deberían implementar ciertas regulaciones para evitar el mal uso de los datos.

El manejo de una gran cantidad de datos ya sean estructurados, semiestructurados y no estructurados, tienen el potencial de ser extraídos para obtener información de cualquier tipo ya sea del sector financiero, económico, social, geoespacial, meteorológico, la medicina, la planificación de una ciudad, la publicidad, entre otros.

Desde mi punto de vista, la recopilación y utilización de cantidades masivas de datos tiene que llevarse a cabo con mucha cautela y siempre bajo el consentimiento de los usuarios de quienes se extrae la misma.

Así, a manera de conclusión puedo decir que el uso del análisis de datos no debería ser prohibido. Pero los profesionales en esta área, así como las empresas y los gobiernos deberían implementar mejores políticas para su uso. Tanto una mayor transparencia y comunicación con el usuario, como una mayor seguridad en el resguardo de los mismos

Referencias:

McMurray R. (1985). *Responses of endurance-trained subjects to caloric deficits induced by diet or exercise*. Recuperado de: <https://europepmc.org/article/med/4068964>

Clarín. (2019). *Ejercicios para bajar de peso: cuántas calorías quemas en una hora*. Recuperado de: https://www.clarin.com/buena-vida/ejercicios-bajar-peso-cuantas-calorias-quemas-hora_0_ry3Mi99qM.html

Guoyao Wu. (2016). *Dietary protein intake and human health*. Recuperado de: 15 <https://pubs.rsc.org/en/content/articlehtml/2016/fo/c5fo01530h>

Centro Farmacéutico Otsuka. (2018). *Carbohydrates: an efficient energy source*. Recuperado de: <https://www.otsuka.co.jp/en/nutraceutical/about/nutrition/sports-nutrition/essential-nutrients/carbohydrates.html>

OMS. (2019). *Reducir la ingesta de sodio para reducir la tensión arterial y el riesgo de enfermedades cardiovasculares en adultos*. Recuperado de: https://www.who.int/elena/titles/sodium_cvd_adults/es/

Tecnologías. (s.f). Técnicas de Fusión de Datos. Recuperado de: <https://www.tecnologias-informacion.com/fusion.html>

Khan Academy. (2018). La tasa metabólica. Recuperado de: <https://es.khanacademy.org/science/ap-biology/ecology-ap/energy-flow-through-ecosystems/a/metabolic-rate>

BBC. (2017). Por qué dormir más de 7 horas puede ayudarte a perder peso. Recuperado de: <https://www.bbc.com/mundo/noticias-38919178>

Microsoft. (2020). Fase de descripción de negocio del ciclo de vida del Proceso de ciencia de datos en equipo. Recuperado de: <https://docs.microsoft.com/es-es/azure/machine-learning/team-data-science-process/lifecycle-business-understanding>

Lipschutz, Seymour (1991). Teoría de conjuntos y temas afines. McGraw-Hill.

Microsoft. (2020). Fase de modelado del ciclo de vida del proceso de ciencia de datos en equipos. Recuperado de: <https://docs.microsoft.com/es-es/azure/machine-learning/team-data-science-process/lifecycle-modeling>

Delgado, F. (2019). *Los Desafíos Éticos de la Ciencia de Datos*. Recuperado de: <https://medium.com/saturdays-ai/los-desaf%C3%ADos-%C3%A9ticos-de-la-ciencia-de-datos-25ce771d892e>

Colmenarejo, R. (2017). *Ética aplicada a la gestión de datos masivos*. Recuperado de: file:///Users/justin/Downloads/6553-Texto%20del%20art%C3%ADculo-16236-1-10-20171115.pdf

Linkedin. (2021). *Ciencia De Datos Para Todos*. Recuperado de:
<https://www.linkedin.com/in/ciencia-de-datos-para-todos-0b91a9190/>