

RECETAS, INGREDIENTES y SALUD

Introducción

Todos hemos tenido algún contacto con la cocina directa o indirectamente, ya sea calentar, preparar comidas, ayudando a alguien a preparar una deliciosa receta que encontró en redes. Algunos han pasado buenos momentos, como aprender a cocinar y preparar la comida por sí solo, pero también no muy buenos momentos como la primera quemadura, cortada o que se haya quemado la comida. Pero a pesar de todo esto, lo más esperado después de cocinar es cuando llega el momento de probar y disfrutar ese delicioso plato.

Por otro lado, la alimentación hace parte fundamental en nuestro vivir. Nos limita o nos activa, y condiciona en gran medida nuestra salud. Quien no ha escuchado por la televisión, redes, en una visita al médico acerca de la necesidad de equilibrar nuestra dieta, con un 50% de vegetales, 25% hidratos y 25% de proteína, y complementarlos con agua, frutas y lácteos. También de controlar el consumo de productos azucarados, y por supuesto complementarlo con actividad física.

Por eso he escogido algún tema relacionado con la cocina ya que la cocina es uno de los hobbies que he encontrado como favoritos últimamente, pero lo mejor de ellos es que fue una actividad que llego a mí, por el momento. Me entusiasma conocer más de otras recetas provenientes de diferentes países; pero teniendo en cuenta de controlar el consumo de sodio, grasas y azúcar que en grandes medidas causan grandes consecuencias a nuestro organismo.

Objetivo

Como tema principal de dicho proyecto, se trata de conocer los ingredientes de las diferentes cocinas del mundo, y tal vez conocer y empezar a usar ingredientes que no sabían que existían en mis nuevas preparaciones. Y como la salud toma protagonismo en la alimentación. Todo lo anterior se llevará a cabo con la ayuda de técnicas de análisis de datos como, Python la principal herramienta, y distintas librerías de Python como Pandas, Seaborn y BeautifulSoup, por mencionar las principales.

Así, el principal objetivo de dicho análisis era corroborar las siguientes hipótesis:

- El arroz es el ingrediente más usado en las preparaciones de las diferentes cocinas en estudio con respecto al resto de los ingredientes.
- Las cocinas de los países hispanos tienen una alta relación. Es decir, comparten gran porcentaje de ingredientes en sus preparaciones, en que el tomate y la cebolla son los más usados.
- Estados Unidos es el país con mayor media de IMC (Índice de Masa Corporal) y, tiene una correlación positiva con los componentes del valor nutricional de las recetas (Grasas, Colesterol, Sodio, Azúcar, etc.)

Descripción del proyecto

Se obtendrá información a partir del cual se entenderá los diferentes valores nutricionales de las recetas de un país y como estos pueden estar relacionados, si positiva o negativamente con condiciones o enfermedades que son causadas por la mala alimentación, como por ejemplo el sobrepeso, hipertensión y diabetes.

Desarrollo del proyecto

Durante el proyecto hubo momentos de bastante trabajo y dedicación que fueron encontrar la información correcta y por otro lado la exploración, estandarización, y la agrupación de todos los datos.

Cabe resaltar que encontrar el tema del proyecto no fue fácil, no sabia de que se iba a tratar el análisis al principio, para ello realice un listado de todas aquellos temas o materias me gustaban y me encantaría investigar, aprender más y presentarlo a los demás.

Después de haber definido el proyecto, se procede con adquirir cierto conocimiento para conocer el análisis que se abordará. Para ello, se investiga documentación, páginas relacionadas con el tema. En mi caso, empecé con definir las fuentes de donde podría encontrar la información pertinente, por lo que estuve revisando y verificando diferentes datasets tipo CSV y API's; pero allí fue donde me encontré con el primer reto, limitadas variables por dataset, dado que solo se especificaba en la información los ingredientes y el país al cual pertenece la receta, y no más. Por ello, decidí realizar un Web Scrapping por medio de Selenium y BeautifulSoup de la INDRAPRASTHA Instituto de información tecnológica Delhi. Esta obtención en específico requirió de bastante tiempo, más de 8 horas, debido a que el repositorio de recetas es inmenso. Pero logré obtener los ingredientes de las distintas recetas con su respectivo valor nutricional.

Otras fuentes de información utilizados fueron NCD Risk Factor Collaboration, para obtener datos acerca de los ratios y métricas de salud por país como, por ejemplo, el Índice de Masa Corporal (), Cholesterol tanto HDL y LDL, entre otros.

Con la información obtenida, el paso a seguir fue agrupar y limpiar un poco los datasets, que fue la actividad que más toma tiempo en realizar, por razones como, la no estandarización de los datos, por ejemplo, en algunas páginas mencionadas anteriormente, se refieren a Estados Unidos como US, otros como United States of America, lo que era indispensable verificar con los demás tipos de variables para que cuando vaya a unir los Dataframes no tenga conflicto y se me generen valores de tipo NaN.

En mi caso yo realicé dos limpiezas cuando obtuve los datos por medio de Webscrapping y después, antes de realizar el análisis de las variables. La primera limpieza la realice con el motivo de definir si los datos obtenidos funcionarían sin que pudiera tener mayores complicaciones.

Con los Dataframes obtenidos, se continua con la exploración de estos, en el que conoce un poca más de las variables que lo componen, los outliers que existen en ellos, para luego continuar con la limpieza de missing values, de datos duplicados, eliminación de columnas que no se vaya utilizar y por su puesto la agrupación de los distintos datasets para poder hallar o encontrar si hay o no correlación entre las variables.

Con los Dataframes completamente limpios y preparados para analizarlos. Se tienen en cuenta los diferentes estadísticos como media, moda, desviación estándar, cuartiles percentiles, y por supuesto gráficos con el de caja, histograma, mapa de calor. Con ello pude confirmar o no las hipótesis planteadas.

Conclusiones

- El cereal más usado es la harina y no el arroz. En general, la sal, el aceite de oliva, la mantequilla y la cebolla son los 5 ingredientes más usados en los 75 países estudiados
- Además de la sal, que es algo comprensible. La cebolla se comparte en la mayoría de las cocinas latinas.
- Estados Unidos si es el país con mayor IMC en promedio en los hombres, mientras que el país con mayor IMC es Egipto. Por otro lado, existe una correlación negativa entre la fibra y el colesterol Non DHL, y una correlación positiva entre la grasa saturada y el colesterol DHL o bueno. Ya que según, la Organización Mayo Clinic, que la fibra que se encuentra en cereales como la avena ayuda a reducir el colesterol de lipoproteína de baja densidad (LDL), o como lo conocemos comúnmente el colesterol “malo”.
- Una conclusión en general del proyecto, es que he aprendido bastante en el manejo de datos, limpieza y análisis, aunque haya momentos de estancamiento, lo mejor que pude hacer en esos momentos fue tomarme 5 minutos, descansar y volver con todo el entusiasmo.

Fuentes:

Colesterol: Los principales alimentos que mejorarán tus niveles. (2022). Recuperado 14 de marzo de 2022, de Mayo Clinic website: <https://www.mayoclinic.org/es-es/diseases-conditions/high-blood-cholesterol/in-depth/cholesterol/art-20045192#:~:text=La%20fibra%20soluble%20puede%20reducir,a%204%20gramos%20de%20fibra.>