EJERCICIOS CON DATOS



VANESA SIERRA SÁNCHEZ

#understand 2ed 2023

Pruebas obligatorias

El objetivo del ejercicio es obtener una tabla (Excel, csv, html,...) donde se puedan ver los 15 cereales que más proteína tienen, atendiendo a las siguiente reglas:

- Queremos los productos que, además de tener el máximo de proteínas, estén ordenados de forma inversa por el número de calorías. Es decir, los que menos calorías tengan primero.
- Establecer una comparativa entre los MFR (fabricante) K y G de manera que obtengamos una columna adicional indicando la diferencia en porcentaje de GRASAS (FAT) donde sea visualmente evidente si la divergencia es mayor o menor. Recomendable calcular en porcentaje (una idea, seguid el camino que prefiráis).

He resuelto la prueba siguiendo los pasos siguientes:

 He accedido a la web de Kaggle a través del link proporcionado por la plataforma, a continuación he leído los campos del dataset o tabla, que son los nombres que se van poniendo en cada una de las columnas de la tabla y que hacen referencia a la distinta información que tenemos sobre los cereales, como el nombre del producto, el fabricante, las calorías que tiene por ración, los gramos de proteína, etc.

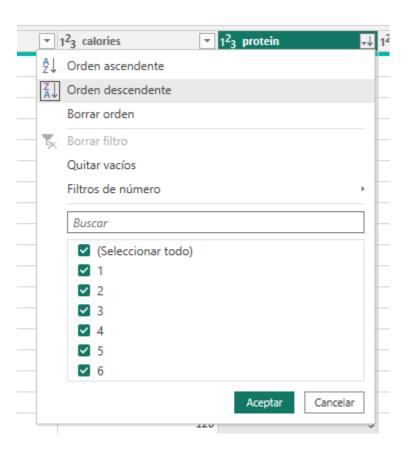
Fields in the dataset:

- · Name: Name of cereal
- · mfr: Manufacturer of cereal
 - o A = American Home Food Products;
 - o G = General Mills
 - K = Kelloggs
 - o N = Nabisco
 - o P = Post
 - Q = Quaker Oats
 - R = Ralston Purina
- type:
 - o cold
 - hot
- · calories: calories per serving
- · protein: grams of protein
- · fat: grams of fat
- · sodium: milligrams of sodium
- fiber: grams of dietary fiber
- · carbo: grams of complex carbohydrates
- · sugars: grams of sugars
- · potass: milligrams of potassium

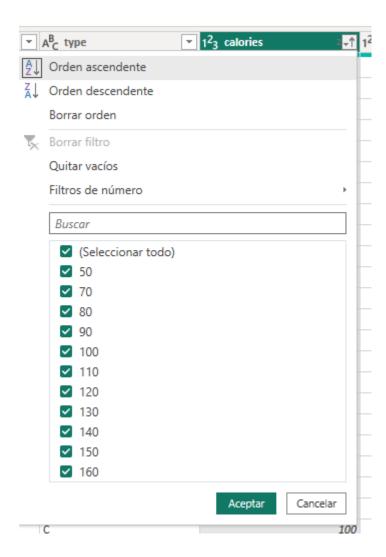
 Me he descargado el dataset y lo he importado a Power BI, que es una herramienta que me permite poder filtrar los datos a mi antojo para poder ver la relación entre los distintos valores de los datos.



• En el ejercicio nos piden obtener una tabla donde se puedan ver los 15 cereales que tengan más proteína. Como queremos que los productos, es decir, los distintos cereales, estén ordenados de mayor a menor por el número de proteínas debemos ir a la columna de proteínas seleccionar en el margen superior derecho el botón y al abrir el desplegable seleccionar "orden descendente". De esta manera como se puede observar la columna queda ordenada de mayor a menor número de proteínas.

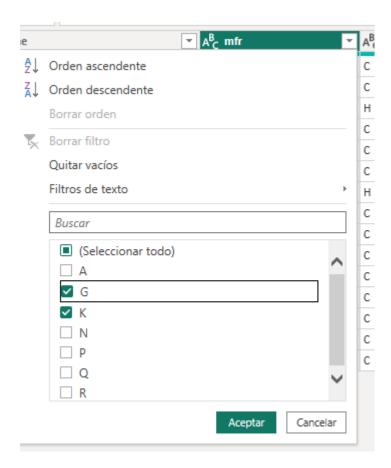


 A continuación queremos que la columna de las calorías quede ordenada a la inversa, es decir, los que menos calorías tengan primero, así que aplicamos el "orden ascendente".

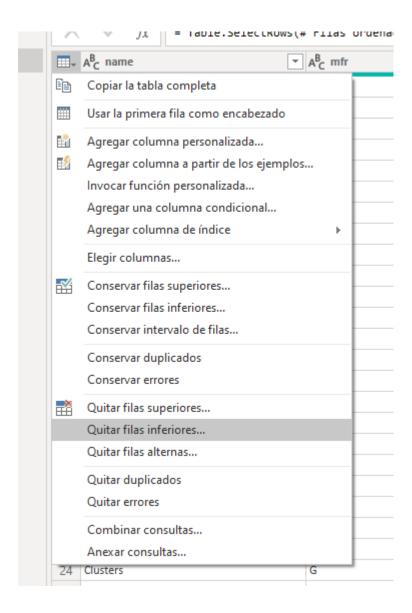


Establecemos una comparativa entre dos fabricantes: Kelloggs, representado en la tabla por la letra "K" y General Mills, representado por la letra "G". Primero filtramos la segunda columna, llamada mfr, que hace referencia al fabricante, en ella hay un triángulo con el vértice hacia abajo en el margen superior derecho. Pues bien, si se hace click con el ratón a ése triángulo sale un desplegable con varios filtros de búsqueda. En la parte inferior de éste desplegable podemos observar que están todos los datos posibles que pueden aparecer en ésta columna según el fabricante de los cereales (A,G,N,P,Q,R y K). Seleccionamos únicamente K y G.

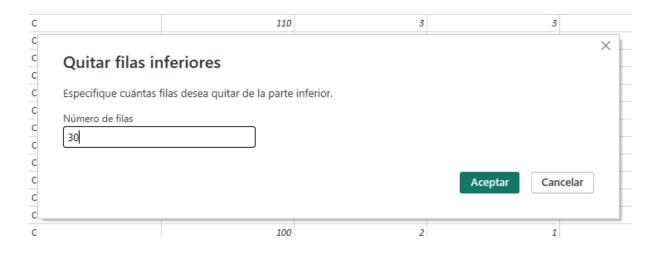
🐝 alumno 105@programasimpact.org



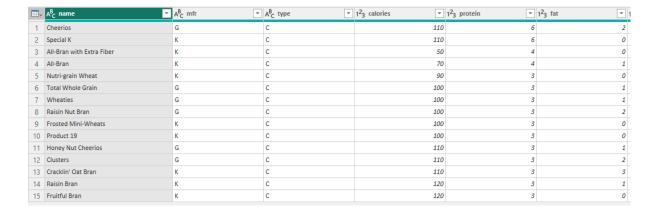
El resultado es una tabla en el que el producto estaría ordenado de más a menos proteínas y a su vez dentro de que tengan la misma cantidad de gramos de proteínas se ordene de menor a mayor calorías por ración. Como queremos tener una tabla solamente con los datos de los 15 primeros cereales con más proteína quitamos las filas que no nos interesan y haciendo una nueva tabla con ellos, pues son los que nos pide el ejercicio. Primero seleccionamos "Quitar filas inferiores" y a continuación nos pide el número de filas que queremos eliminar.



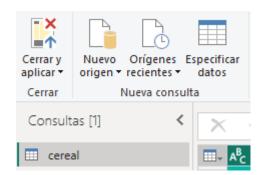
 Como la tabla contiene 45 cereales y nosotros queremos 15 eliminamos 30 filas y pulsamos el botón aceptar.



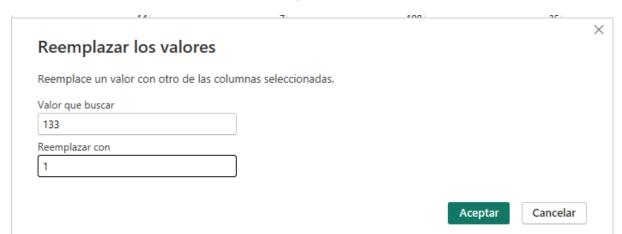
• La tabla quedaría de la siguiente manera:



 Cargamos los datos para visualizarlos con Power BI Desktop dando al botón cerrar y aplicar, en el margen izquierdo superior.

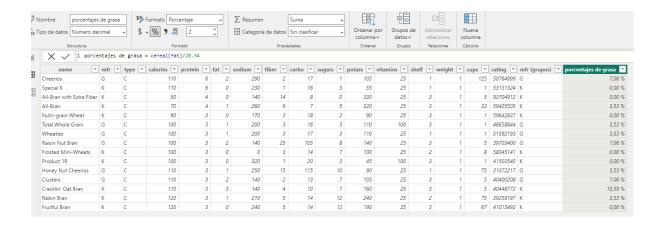


Como podemos ver en la leyenda el peso "weight" es por onzas por una ración de cereales. Que en todos los productos es igual a 1 onza excepto en Raising Bran y en Fruitful Bran que es igual a 133 onzas. En mi opinión éstos datos deben ser una errata pues una onza equivale a 28, 34 gramos y 133 onzas serían poco menos de 3 kilos. Así que doy por hecho que han tomado el resto de valores en relación a el mismo peso de producto, es decir una onza y paso a modificar la tabla.

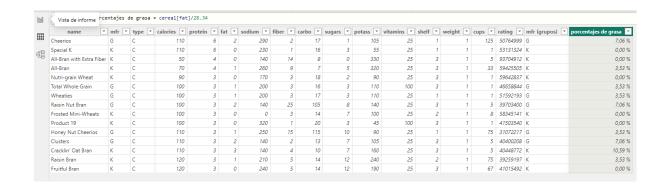


En el ejercicio nos piden obtener una columna adicional indicando la diferencia en porcentaje de grasas "fat" donde sea visualmente evidente si la divergencia entre los dos fabricantes es mayor o menor. En la vista de de datos creo una nueva columna que va a ser igual a el valor en cada fila de "fat" entre los gramos a los que equivale una onza, es decir 28, 34 gramos. Seguidamente doy al símbolo de "%" situado en la parte superior de la interfaz y cambio el nombre de la columna por el de "porcentajes de grasa".

Como resultado tendríamos la tabla como nos la pide el ejercicio

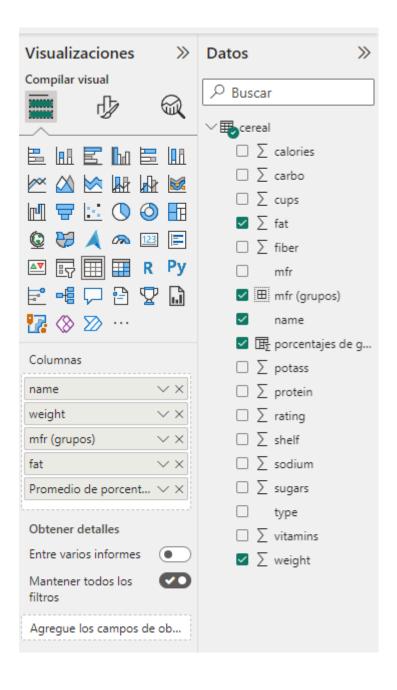


• En el margen superior izquierdo podemos ver un icono con columnas en el que tenemos que pulsar para tener la vista del informe.



🐝 alumno 105@programasimpact.org

Accedemos y podemos ver en el margen derecho una columna llamada visualizaciones y a continuación otra llamada datos. En la columna "Datos" podemos ver el nombre de nuestra tabla "cereal" y clicando en el símbolo mayor que ">" se despliegan todos los campos de la tabla. Vamos a establecer la comparativa de los fabricantes K y G con respecto a su diferencia de porcentaje en grasas, por lo que tomamos los campos necesarios para hacerlo: name, weight, fat, mfr y promedio de porcentajes de grasa. Otra opción para hacer lo mismo es coger cada uno de los campos y transportarlos a la sección "Columnas" en la columna anteriormente nombrada llamada Visualizaciones. Para ello colocamos el ratón encima del campo deseado, pulsamos el click derecho y sin soltar lo arrastramos hasta "columnas". Y clicamos en la opción "mantener todos los filtros" de la tabla.

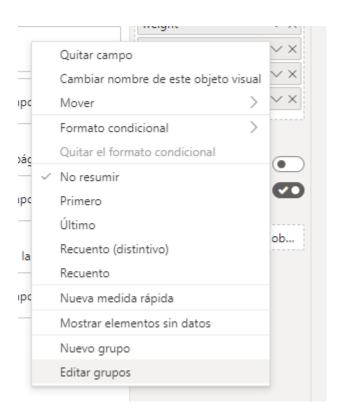


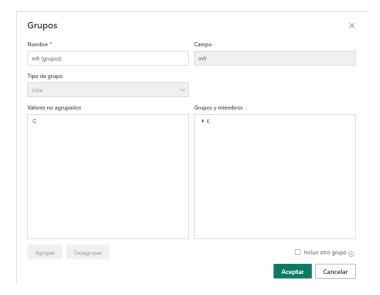




🕉 alumno 105@ programas impact.org

 Como quiero separar en dos tablas a el fabricante K y al fabricante G para que se pueda visualizar más la diferencia creo grupos haciendo clic en "mft" y seleccionando nuevo grupo.





• Una vez tenemos todos los campos en columnas hacemos una tabla con todos los productos de General Mills y otra con los productos de Kellogg's .

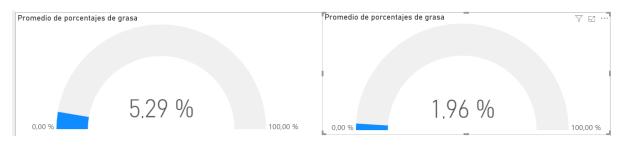
General Mills

name	weight	mfr (grupos)	fat	Promedio de porcentajes de grasa
Cheerios	1	G	2	7,06 %
Clusters	1	G	2	7,06 %
Honey Nut Cheerios	1	G	1	3,53 %
Raisin Nut Bran	1	G	2	7,06 %
Total Whole Grain	1	G	1	3,53 %
Wheaties	1	G	1	3,53 %
Total				5,29 %

Kelloggs

name	weight	mfr (grupos)	fat	Promedio de porcentajes de grasa
All-Bran	1	K	1	3,53 %
All-Bran with Extra Fiber	1	K	0	0,00 %
Cracklin' Oat Bran	1	K	3	10,59 %
Frosted Mini-Wheats	1	K	0	0,00 %
Fruitful Bran	1	K	0	0,00 %
Nutri-grain Wheat	1	K	0	0,00 %
Product 19	1	K	0	0,00 %
Raisin Bran	1	K	1	3,53 %
Special K	1	K	0	0,00 %
Total				1,96 %

• Como podemos observar el porcentaje de grasa promedio de los productos de K que es 1,96% y el de los productos de G que es 5,29%.



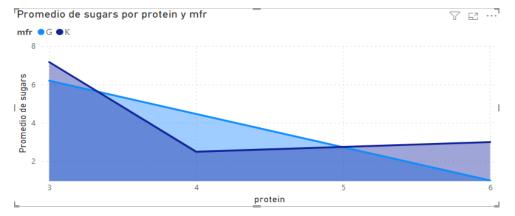
- los productos de Kelloggs (k) tienen un promedio mucho menor en grasas que General Mills, luego, podemos decir en comparativa que los productos fabricados por Kellogg's en general son menos grasos y por lo tanto más saludables que los de General Mills.
- Como podemos visualizar en la tabla, Special K (Kelloggs) y Cheerios (General Mills) teniendo el mismo número y máximo de proteínas (6) e idénticas calorías (110) Special K tiene 0% de materia grasa mientras que Cheerios tiene 7,06 % de materia grasa.



 Pero si tenemos en consideración la cantidad de calorías por ración el mejor producto para una persona que quiera llevar una dieta rica en proteínas y lo más baja posible en calorías y grasas sería All- Bran with Extra Fiber de Kelloggs.

name	mfr	fat	porcentajes de grasa	Suma de calories	Suma de protein
All-Bran with Extra Fiber	K	0	0,00 %	50	4
All-Bran	K	1	3,53 %	70	4

 Hago una comparación de promedio de azúcares del fabricante K y el G donde se pueda apreciar que mientras que en G va bajando el porcentaje de azúcar según aumentan las proteínas en K hay un descenso de azúcares significativo en 4 proteínas, pero luego asciende ligeramente en el valor máximo de proteínas, aunque si analizamos por producto All- Bran with Extra Fiber de Kelloggs sigue siendo el mejor por contener 0 % azúcares.

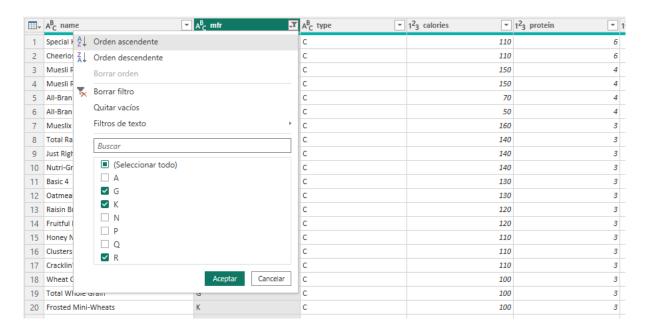


Pruebas voluntarias

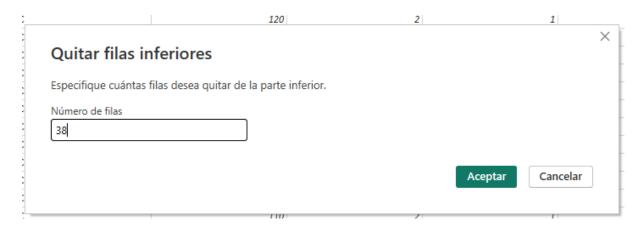
(2) 80 cereales: https://www.kaggle.com/datasets/crawford/80-cereals

El objetivo del ejercicio es obtener una tabla (Excel, csv, html,...) donde se puedan ver los 15 cereales que más proteína tienen, atendiendo a las siguiente reglas:

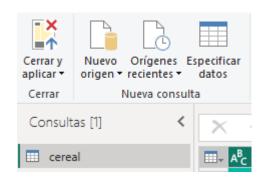
- Queremos los productos que, además de tener el máximo de proteínas, estén ordenados por el número de calorías.
- Establecer una comparativa entre los MFR (fabricante) K, G y R de manera que obtengamos una columna adicional indicando la diferencia en porcentaje de GRASAS (FAT) donde sea visualmente evidente si la divergencia es mayor o menor.
- Escribir una recomendación sobre qué fabricante es el ideal para construir una dieta, atendiendo exclusivamente a las proteínas y las grasas. Justificar la explicación.
 - Como hemos realizado en el primer ejercicio, subimos la tabla "cereales" a PoweBI.
 Primero filtramos las proteínas de mayor a menor y luego las calorías tambien de mayor a menor para ordenar el número de calorías. A continuación filtramos los fabricantes seleccionando únicamente Kellogs(K), General Mills (g) y Ralston Purina(R).



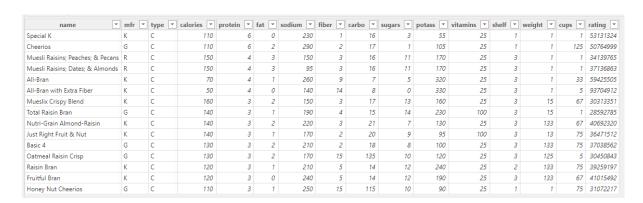
 Como queremos tener una tabla solamente con los datos de los 15 primeros cereales con más proteína quitamos las filas que no nos interesan y haciendo una nueva tabla con ellos, pues son los que nos pide el ejercicio. Primero seleccionamos "Quitar filas inferiores" y a continuación nos pide el número de filas que queremos eliminar.



 Cargamos los datos para visualizarlos con Power BI Desktop dando al botón cerrar y aplicar, en el margen izquierdo superior.



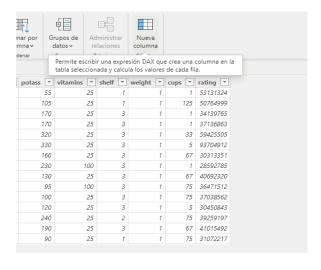
La tabla quedaría de la siguiente manera.



- Como podemos ver en la leyenda el peso "weight" es por onzas por una ración de cereales. Que en todos los productos es igual a 1 onza excepto en Raising Bran, Fruitful Bran, Basic 4 y Nutri-Grain Almond-Raisin que es igual a 133 onzas. En Oatmeal Raisin Crisp que es igual a 125 onzas. En Muesli Crispy Blend y Total Raisin Bran que es igual a 15 onzas. Y por último Just Right Fruit & Nut que es igual a 13 onzas.
- En mi opinión éstos datos deben ser una errata pues una onza equivale a 28, 34 gramos y 133 onzas serían poco menos de 3 kilos. 15 serían casi medio kilo. Así que doy por hecho que han tomado el resto de valores en relación a el mismo peso de producto, es decir una onza y paso a modificar la tabla.

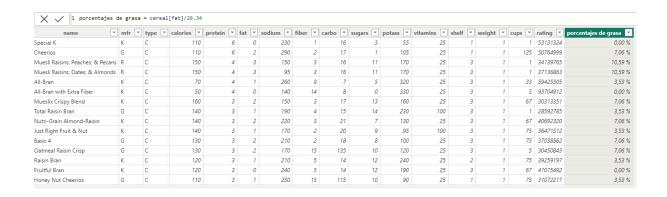


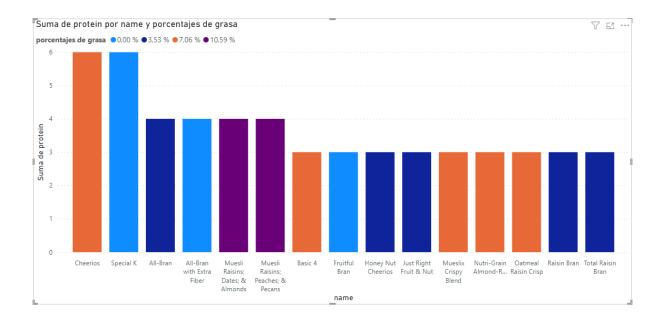
En el ejercicio nos piden obtener una columna adicional indicando la diferencia en porcentaje de grasas "fat" donde sea visualmente evidente si la divergencia entre los dos fabricantes es mayor o menor. En la vista de de datos creo una nueva columna que va a ser igual a el valor en cada fila de "fat" entre los gramos a los que equivale una onza, es decir 28, 34 gramos. Seguidamente doy al símbolo de "%" situado en la parte superior de la interfaz y cambio el nombre de la columna por el de "porcentajes de grasa".



Como resultado tendríamos la tabla como nos la pide el ejercicio

0





Mi recomendación sobre qué fabricante es el ideal para construir una dieta saludable es Kelloggs, ya que no es indispensable consumir todos los productos de un fabricante, sino simplemente elegir uno. El mejor producto para una persona que quiera llevar una dieta rica en proteínas y lo más baja posible en grasas (por ejemplo un deportista) sería Special K. Ya que posee el valor máximo de calorías y 0% de materia grasa.

En el caso de que además quisiera consumir la mitad de calorías que el anterior producto y 0% en azúcares, mi recomendación sería All- Bran with Extra Fiber de Kelloggs teniendo en cuenta que bajaría el contenido proteico de 6 a 4 gramos.

name	- T	nfr 🖸	type	▼ calories	protein '	fat	sodium 💌	fiber 💌	carbo 💌	sugars 💌	potass 💌	vitamins 💌	shelf 💌	weight 💌	cups 💌	rating 💌	porcentajes de grasa 🗾
Fruitful Bran	K	(C	12	0 .	3	240	5	14	12	190	25	3	1	67	41015492	0,00 %
All-Bran with Extra Fiber	K	(С	5	0 .	4	140	14	8	0	330	25	3	1	5	93704912	0,00 %
Special K	K	(С	11	0	5	230	1	16	3	55	25	1	1	1	53131324	0.00 %