Explore and Transform REWE Products

Preparation

Load the REWE products data set with the following lines:

```
library(tidyverse)
rewe <- read_csv("data/rewe_products.csv")
```

Exercise 1: Explore the data set

1. How many columns and rows are in the data set?
2. Output all column names to the console!
3. Which and how many columns have a numeric data type?
4. Display the first 20 product names! How can you see all products?
5. Output the first 30 brands! What do you notice? How could you solve this?
6. Summarize the value ranges of the columns 'vegan' and 'vegetarian'. What do you think about the data quality of the two columns?
7. Create a summary of all columns that contain a value in grams. What different ways do you find to achieve this?

Exercise 2: Select columns with `select`

Find solutions for the following tasks, in which you have to select a subset of the columns:

1. Create a new data frame containing the product name, the product category and the sales price!
2. Create a new data frame with all nutritional information as well as the product name and the product category!

3. Create a new data frame that contains only numeric columns. Check the columns and write down what you notice!

Exercise 3: Reduce rows with `filter`

Find a solution to restrict the rows of the data set as described:

1. Filter the data so that only products from Germany are included! Keep only the product name and the country of origin in the result!
2. Find all vegan organic products!
3. Which types of beer are offered in the REWE online store?
4. Find all red wines under 2 EUR!
5. Find all products which contain the allergen soy! Take a look at the function `str_detect()` for this.

Exercise 4: Modify or create new columns with `mutate` or `transmute`

1. Change the data type of the two columns `productId` and `gtin` to strings (`chr`).
2. Create a new column `sum_nutrition` in which you form the sum of all nutritional values. Leave only the new sum column and the individual nutrition columns in the result.
3. Calculate if a product contains more than 90% fat and store this information on a new column `high_fat`! For checking purposes, displays only rows where the value is `TRUE`. The new column should be inserted before the `productDescription` column.
4. Create a column `imported_bio` which should contain `TRUE` if the product is an organic product and at the same time is not from Germany.

Exercise 5: Summarize data with `group_by` and `summarize`.

1. How many products are in the data set?
2. How many products does each product category have?
3. In addition to the product category and the number of products, now include the average selling price!
4. List all product categories according to the average fat content of their products! What problem do you encounter and how can you solve it?

5. Use the previous result and keep only the top 5 categories with the highest average fat content of their products!
6. Which brands (`brand`) have the products with the highest protein content in their assortment? List the top 10!