Exercise 2. Association analysis (20% final grade)

The aim of data mining is to extract information out of data. The data available here are sales data from a retailer that sells via branches and online. Your new manager knows a bit about Association rule analysis and wants you to apply it to the shopping data.

A friendly co-worker already extracted the data of 2 weeks from the system. You can find it in the file excercise_WS2020_data.csv

The following columns are available:

channel - Online or Branch

date - date of the transaction

article_id - the ID of the article the is in the basket

basket_id - the ID of the basket (so all articles in a basket/transaction have the same basket_id)

article_group - a group that can be used to aggregate similar articles

article_name - the name of the article

Import and analyse the data. The marketing department should focus their activities on cross selling and bundling of products. Therefore they need your help.

Visualize and explain your results. The main part of the analysis should be an Association rule analysis.

How to deliver:

- Create a short power-point presentation (PDF format) with 4-6 slides of content. (Tittle and good-bye slides not counted). It should contain at least two diagrams from association analysis. The next step is still under consideration, so your short presentation should end with a recommendation on how to move forward.

- Include a text document (PDF format) of max. 2 pages, explaining your slides. Add your code fully commented as an annex at the end of your document.

- Include both files in an archive format, such as .zip, .rar, .7z, .tar… and upload this file in Moodle.

Important:

Your deliveries will be checked for plagiarism (presentation, text document, and code) and will be punished severely if this academic dishonesty is found.

Due date: 28-02-2021

Hints:

To be able to use arules exploration and plotting functions you need to convert the baskets in a different kind of object (transactions)

Visit https://cran.r-project.org/web/packages/arules/vignettes/arules.pdf for detailed information.

Option 1 use split:

```
# for split to work correctly, the column vector to group should be of type factor.
# replace the placeholders with the right variable names.
# split receives first the data you want to group as data frame or column vector
# and the second input of the function is the vector you use for grouping

shopping_basket <- split(<data_to_group >,<vector_groupping>)

#create transaction data type needed for arules.
#The input should be a basket (list of the articles)

retail_transactions <- as(shopping_basket, "transactions")
```

Option 2 use tidyverse commands: group_by and summarise:

```
# you can use group_by to define how to group your data and then
# use summarise(<your_data>,<new_varname> = list(feature_to_list)) to create the
# baskets
# replace the placeholders with the right variable names.

shopping_basket <- <your_data> %>%
  group_by(<variable_groupping>) %>%
  summarise(basket_list = list(<data_to_group>)

#create transaction data type needed for arules.
#The input should be a basket (list of the articles)

retail_transactions <- as(shopping_basket$basket_list, "transactions")
```