NDH802 - Assignment 1

Group no.

- The assignment includes 2 questions, 6 points each.
- Submit your assignment via Canvas before 10:00 CET, April 12, 2021.
- Your submission should be an RMarkdown file with your solutions in words and/or R code. If you handwrite parts of your assignment, insert it as an image near the corresponding question(s). Name the files NDH802_Assignment1_GroupNumber.
- You should work in groups and contribute equally.
- You can copy my code, but make sure you understand it.
- You should not have the exact solutions and/or results with other groups.
- Results without code/justifications will not be graded.

Set things up

Set your working directory and fill in your group number. For example, if you are group 3, make it `our_group <- 3`. If you don't fill in your group number or fill in the wrong number, your assignment will **not** be graded.

```
#setwd("")
our_group <- 27
```

Run this code chunk to load data into your R Environment. The command will randomly select 1,000,000 rows of data from the original data set. Hereby each and every group should have a unique `df`. Accordingly, your results should be different from other groups' and you should not be comparing them.

Data description

Variable	Description
cust.id	Unique customer id
age	Customer age in year 2021
email	If there is an email of the customer in the system
member.since	Year from which the customer become a member. They can only register at the physical stores.
distance.to.store	Distance (in km) from customer's address to the physical store they register their membership
store.trans	Total number of offline transactions the customer made in year 2021
store.spend	Total amount the customer spend from offline transaction in year 2021 (in SEK)
online.visits	Total number of time the customer visit (does not necessarily mean purchase) the online store in year 2021
online.trans	Total number of online transactions the customer made in year 2021
online.spend	Total amount the customer spend from online transaction in year 2021 (in SEK)
points	Total loyalty points the customer accumulates since they become a member deducted by points they have used
main.store	The format in which the customer made the most transactions in year 2021

Question 1. Mean and variance

(a) Plot the histogram of customer's `value`. Imagine you will present this to your boss at work. Make it readable and self-explanatory (e.g., add the title for the chart and labels for the axes where needed). (1p)

(b) Make a box plot for `visits` for 2 groups, loyal and not loyal customers. Refer to the code provided and modify it (1p). Imagine you are the customer relationship manager. What would you say about this figure? (1p)

```
# boxplot(
#   your_variable_of_interest ~ the_group,
#   data = your_df,
#   ylim = c(-10, 100) # adjust the ylim that better illustrates your data
# )
```

(c) Compute the mean and variance of `value`, `deals`, `points`. Be careful, variance is **different from** standard deviation. (1p)

(d) Compute the mean and variance of `value`, `deals`, `points` of the *loyal* customers. (1p)

(e) Compute the mean and variance of `value`, `deals`, `points` of the *loyal* customers who *made at least one offline purchase*. (1p)

Question 2. Probability theory

(a) How many loyal customers and not loyal customer do you have in your `df`? Formally, compute N_{loyal} and $N_{\overline{loyal}}$. (1p)

For Q2b-f, consider the following events in your `df`:

(E1) Being loyal

(E2) Not being loyal

(E3) Made at least one offline purchase

(E4) Made at least one online purchase

(b) Are E1 and E2 mutually exclusive? Why/why not? (1p)

(c) Are E1 and E2 collectively exhaustive? Why/why not? (1p)

(d) Are E3 and E4 mutually exclusive? Why/why not? (1p)

(e) Are E3 and E4 collectively exhaustive? Why/why not? (1p)

(f) Are E1 and E4 mutually exclusive? Why/why not? (1p)

For Q2b-f, you can write the solutions using formulas, words, Venn diagrams, code, numbers or the combination of them, whichever expresses your rationales the best. If you find handwriting more convenient, feel free to do so and attach a photo of it in the submission.

Have fun and good luck!
Huong and Emelie