# Eltecon Data Science Course by Emarsys

## Sample project for a continuous prediction problem

András Bérczi

October 13, 2021

# About me

- Background in Economics
- Works as Data Scientist @ Emarsys

# Today's topics

- How to set up a project?
- Doing a sample project together

# Use git, if you know how (not necessary)

- You can submit homeworks there, using your own repository.

# Make sure your code is easy to run for others!

- Most of the time others will have to run your code. Make it easy for them!
- "Caveat: only you can decide how much you care about this." [1]

[1] https://www.tidyverse.org/blog/2017/12/workflow-vs-script/

[2] https://martinctc.github.io/blog/rstudio-projects-and-working-directories-a-beginner's-guide/

# Some suggestions

- Use R projects
- Use the `here` package
- Best practice depends on the situation - we don't use R projects in the class either
  - Definatelly recommended for final project
  - It helps in the homeworks as well

# What should be part of a prediction project?

- Understand your data
- Data cleaning
- EDA (for selecting features for your model)
- Modelling
- Evaluating the model
- How the output should look like

# Let's do it!

You will do every step by yourself first or you can do it with your partner!

After that I will show you my solution!

Use `data/laptop_price.csv`

# Homework

- Homework:
  1. Choose a data, which you will use for your final project!
  2. Explore the data, what kind of features/variables you have. Is it good/usable for your final project?
- You have 3 weeks for this homework (until 3rd of November)!
- Since this is a bigger homework, you will receive more points for it!

The aim is to help you with your final project!

- Based on this you will have to decide on your research question for your final project! (This will be a homework later.)
  - Eg.: I will be predicting XY, this and that feature will probably be helpful as features in my model (proved by plots or some statistics).
- (Probably there will be another homework for that week as well, so make sure you start in time!)