# Exercise description

For the first group exercise, you should choose one og the four cases to work with. The cases will be similar for exercise 2, but with other requirements, of cause.

Try to work together and learn from each other. We will give feedback for the full report, and that is more useful, if you actually did try all tasks.

There can be more solutions to each task. But if I ask for e.g. a loop, please create a loop!

**Formalities for all four cases**

* **Format**: A pdf-file of the report and code – it can be output from Quarto or RMarkdown.
* **Size**: 5 pages.
* One page is defined as 2275 symbols (including spaces). Font minimum 11 pixls (Quarto default is fine).
* Include all code chunks. If you work in word, make a screen dump and if you work in Quarto you can ensure the code is visible. Messages and warnings should not be shown.

# Case 1: Economics - Analyzing Inflation Data

### Scenario:

You are an economist working for a government agency tasked with monitoring inflation rates. You have been given a dataset containing monthly inflation rates for the past five years. Your goal is to analyze the data and create a summary report.

### Tasks:

* Load the dataset into R and inspect its structure.
* Calculate the average inflation rate over the five-year period and the standard deviation of the inflation rates.
* Subset the data.frame for each year and calculate the average core inflation.
* Identify and print the month with the highest inflation rate.
* Create – based on quantiles – a variable categorize inflation as "High," "Medium," or "Low" for different goods based on 2023-2017. Use a conditional statement for this.
* Regex: Replace the “M” with a “\_” in Row “Time\_code”
* Use a loop to iterate through each category of inflation (from column 5) and calculate and print the average inflation rate for each category.
* Use the results and your business knowledge to describe the situation. If you do not know a lot on inflation, you can read up on it.

# Case 2: Logistics/Sustainability - Carbon Emission Analysis

### Scenario:

You work for a logistics company interested in reducing its carbon emissions. You have access to a dataset containing monthly carbon emission levels for your company's delivery trucks over the past year. Your task is to analyze this data to identify areas for improvement.

### Tasks:

* Load the carbon emission dataset into R.
* Calculate the total carbon emissions for the year.
* Calculate the average emissions per month.
* Subset the data.frame for each year and calculate the average carbon emission.
* Identify and print the month with the lowest emissions and suggest possible reasons.
* Use a conditional statement to determine if emissions exceeded a specified threshold in any month.
* Regex: Use regex to shorten the month to 3 letters.
* Create a new column. Use a loop to go through each row and add the accumulated carbon emission in the column.
* Use the results and your business knowledge to describe the situation.

# Case 3: Market Segmentation - Customer Behavior Analysis

### Scenario:

You work for a marketing company, and your client wants to understand their customer base better. You have a dataset containing customer information, including purchase history. Your task is to segment the customers based on their behavior.

### Tasks:

* Import the customer data into R.
* Calculate and print summary statistics of customer purchases.
* Use a conditional statement to create a new variable to categorize customers as "High," "Medium," or "Low" spenders.
* Calculate the average purchase amount for each spending category.
* Use the apply() function to find the total purchases made by each customer.
* Regex: Replace Other with Unknown in the gender column.
* Create a new column (segmentation). Decide boarders for segments, e.g. Gender, Age, Income.
* Use a loop to go through all individuals and create a segment for each (e.g. 1,2,3,4 or give the segments a good name). Can you do it smarter, too?
* Use the results and your business knowledge to describe the situation.

## Case 4: Governmental Analytics - Integration of Data Sources

### Scenario:

You work for a government agency responsible for public health. You have multiple datasets from different sources related to healthcare, and your task is to integrate them to gain a comprehensive view of the current health situation.

In this first assignment, however, we only look at the health metrics dataset.

### Tasks:

* Load the health\_metrics datasets into R.
* Calculate summary statistics for key health indicators (blood pressure, BMI and Cholesterol Level).
* Identify smoking status with the highest and lowest values for a chosen health metric.
* Use a conditional statement to create a new variable that identifies if blood pressure is "High," "Medium," or "Low". Use either quartiles or other documentation for your cuts.
* Regex: remove space and “-” in the Smoking\_status column
* Create a new column (high\_risk). Use a loop to go through each patient and if both BMI, Blood Pressure and cholesterol is above average, categorize patient as “High risk”, and else “Not high risk” in the new column.
* Use the results and your business knowledge to describe the situation.