DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

EXAM ASSIGNMENT

Study programme and level	**MSc Logistics and Supply Chain Management + elective**						
Term	V20-21r						
Course name and exam code(s)	Tools for Analytics						460202E016
Exam form and duration	WHAI (changed from WOAI due to Covid restrictions)						6 hours
Date and time	15 February, 2021						9.00 – 15.00
Supplementary material/aids	All	X	Specified		No		
Hand-in of hand-written material allowed	Yes		No	X	Comments:		
Anonymous exam?	Yes	X	No		Comments: Please do not write your name or student ID number anywhere.		
Number of pages (incl. front page)	9						

How to hand in your exam paper
Start preparing the hand in well in advance of the exam deadline.
Your exam paper should be handed in as a set of files (.Rmd and .xlsm) under **'appendix material'**. The total maximum permitted file size here is 5 GB. Due to the system, you must also upload an empty pdf-document named after yourFlowId.pdf (max 200 MB).

If you experience problems uploading and handing in your exam paper in WISEflow, you can send the paper to the following email address: bss.exam@au.dk. You need to ask for permission to hand your paper for final assessment in WISEflow. Use the formula "Exemption" under "Applications to Study Councils" in the Student Self-Service. You need to apply as soon as possible after sending your paper to the email address.
If you need technical assistance during the exam, you can contact BSS IT-support, phone: 8715 0933. Contact the invigilator, if the exam is on-site.

Be aware that exam papers are as a rule only permitted for final assessment, if handed in in the right format/size and within the exam deadline.

Avoid being suspected of exam cheating

Remember to use quotation marks and to insert references if you copy text from other sources, incl. indicative exam solutions (plagiarism) or if you re-use parts of a previously submitted (passed) exam paper (self-plagiarism). Do not share your exam paper with others, or communicate about the assignment during the exam. Students must answer the exam assignment individually.
All submitted exam papers will be checked for plagiarism, so exam cheating (incl. collaboration between students) will be detected.

Questions during the first hour of the exam:

In case you have any *clarification* questions during the first hour of the exam, please email them to Lars Relund at larsrn@econ.au.dk or Sanne Wøhlk at sanw@econ.au.dk. Do not expect instant answers. Answers to questions of a general concern will be posted to Blackboard.

Practical information:

- This exam is open book, open internet, closed other people. You may use any online or book-based resource you would like, but you must include citations for any code that you use (directly or indirectly). You **may not** consult with anyone else during this exam. You cannot ask direct questions on the internet, or consult with each other, not even hypothetical questions.
- This assignment has one appendix available for download from WISEflow. The file is a zip file containing the files you may need during the exam.
- The exam has a VBA and R part with approx. equal weight.
- Please note that the weights on each assignment are only guideline weights, and they only provide information regarding the relative weight of the assignments. The final evaluation will be given based on the total material handed in.
- If you find that some information is missing in the assignments, you may make the necessary assumptions and clearly specify these.
- Handing in: You must hand in a set of files (.Rmd, .xlsm and .html) as "Appendix material". Due to the system, you must also upload an empty pdf-document named `yourFlowId.pdf`.
- Your VBA code will be tested using Excel 2016 and your R code will be tested using R 4.0.3. As operating system Windows will be used.
- **About assignments 1-3: R**
    - Your R code must be written up in an R Markdown (Rmd) file named `yourFlowId.Rmd`. Moreover, also hand in the rendered/knitted html file.
    - Your file must include your code and a (brief) comment for each question. For example, "The three companies with smallest profit are …" or "The plot shows that …".
    - An R markdown template file is given in the appendix that you may use as a starting point.
    - You may load and use the following packages:
        - `library(tidyverse)`
        - `library(skimr)`
- **About assignments 4-6: VBA**
    - Your VBA code should be contained in a single Excel file named `yourFlowId.xlsm`.
    - Do not protect your code with password or turn it into an Add-In.

Assignment 1 - R (10%)

Consider the vector:

```
v <- c(9, 19, 2, 8, NA, 12, 9, 23, NA, 34)
v

##  [1]  9 19  2  8 NA 12  9 23 NA 34
```

Use R to answer the following questions:

Question 1

Is any of the entries in `v` below or equal to 2?

Question 2

Are all of the entries in `v` above or equal to 2?

Question 3

Does `v` have missing values?

Question 4

Which entries in `v` are above 10? You must return the indices, e.g. the index of `v[3]` is 3.

Question 5

Create a vector `res` where `res[i]` is equal to `v[i]` if `v[i]` is less than 10 and otherwise zero (also if `v[i]` is `NA`).

Assignment 2 - R (20%)

The dataset `titanic`, given in the appendix, lists approx. 1300 passengers on Titanic. The column/variables are:

- `pclass`: Passenger class (1 = 1st; 2 = 2nd; 3 = 3rd).
- `survived`: Survival (0 = No; 1 = Yes).
- `name`: Name.
- `sex`: Sex.
- `age`: Age.
- `fare`: Passenger Fare.

- `cabin`: Cabin number.
- `embarked`: Port of embarkation (C = Cherbourg; Q = Queenstown; S = Southampton).
- `boat`: Lifeboat number.

You can read the dataset file `titanic.csv` into the dataset `dat` using

```
dat <- read_csv("titanic.csv")
```

Use the *dplyr* package in *tidyverse* to calculate relevant summary tables (data frames) and answer the following questions:

Question 1

Create a new column named `male`, which is true, if the person is a male.

Question 2

How many persons are we considering? How many survived? How many men (in percentage)?

Question 3

How many of the females survived in percent (and how many males)?

Question 4

Define children as people with age below 19. How many children survived?

Question 5

Did relatively more people survive at first class compared to third class?

Question 6

How many persons that entered a lifeboat died (in percent)?

Question 7

How many persons with `Hansen` in their name survived?

Assignment 3 – R (20%)

Answer this assignment using the *ggplot2* package in *tidyverse* (you might need *dplyr* for preparing the datasets you want to plot). We work with the dataset `titanic` from Assignment 2, which can be read using:

```
dat <- read_csv("titanic.csv")
```

Question 1

Create a visualization showing the number of persons embarking the different harbors with the following features:

- Number of persons is represented using bars.
- Informative figure title and axis titles are given.
- The labels on the x-axis are the harbor names (not abbreviations).

What harbor was the main departure harbor?

Question 2

Create a visualization showing the age distribution of the persons with the following features:

- Number of persons is represented using a histogram.
- Informative figure title and axis titles are given.

Where there most people on board in their twenties or thirties?

Question 3

Create a visualization showing the fare as a function of age with the following features:

- Different colors are used for each sex.
- Informative figure title and axis titles are given.
- A trend line for each sex is added using `geom_smooth`.

Based on the trend lines, do females in general pay more for a ticket?

Question 4

Create a visualization showing the survival rate for each passenger class with the following features:

- Bars are used for each passenger class.
- All bars have same height (100%).
- Colors are used to identify who survived and who did not survived.
- Informative figure title and axis titles are given.

Is the survival rate different on first and third class?

Question 5

Let column `level` denote the first letter in `cabin`. Create a visualization showing the variance in fare prices with the following features:

- Ignore rows with missing `level`.

- Variation is shown using a boxplot.
- Informative figure title and axis titles are given.

Is the fare price different at the B and C level?

Assignment 4 - VBA (13%)

Please use ”Sheet4” in the Excel file “TFA reeks 20-21 assignment4-6.xlsx” that can be downloaded from WISEflow, and put the statement “Worksheets("Sheet4").Activate” in the top of your code to ensure that the code is using the correct sheet.

Question A. Write a sub “Assignment4a” that reads two integer numbers, a and b (assume that $a < b$) from cells E1 and E2, and generates two random integer numbers, r_1 and r_2, uniformly between a and b such that $r_1 \neq r_2$,. The sub should write r_1 and r_2 in cells B1 and B2 with the smaller of the two numbers in B1.

Question B. Write two subs “Assignment4b” and “Assignment4bmaster”.

The sub “Assignment4b” should take two integer numbers, a and b, as input and must generate two random numbers r_1 and r_2 following the same rules as in question A. However, r_1 and r_2 may not be written to the spreadsheet. Instead, they should be given to the sub calling “Assignment4b”.

The sub “Assignment4bmaster” should have four integer variables: g, h, j1, and j2. Assign some values to g and h such that g < h. For instance, g = 2 and h = 17. The sub should then call “Assignment4b” with g and h as input. After the call, j_1 and j_2 should have the values of the two random numbers generated in “Assignment4b”, with j_1 having the smallest value. You may not write anything to the spreadsheet, but “Assignment4b” may take as many input variables as you would like it to. Hint: input can be given by value or by reference.

Assignment 5 - VBA (17%)

Please consult ”Sheet5” in the Excel file “TFA reeks 20-21 assignment4-6.xlsx” that can be downloaded from WISEflow, and put the statement “Worksheets("Sheet5").Activate” in the top of your code to ensure that the code is using the correct sheet.

In column 1, rows 4 to 23, you see 20 names. In cells B1 and B2, you see two numbers between 1 and 20. You may assume that the number in B2 is strictly larger than the one in B1. In the following, let r_1 and r_2 represent these two numbers.

Sometimes, we need to change the order of the items in such lists, and in this assignment, you will be asked to do that in three different ways.

Question A. Write a sub "Assignment5a" that, given values of r_1 and r_2, (to be read from B1 and B2), swaps the r_1'th and the r_2'th name in the list.

Figure 1 illustrates this for an example, where *Bente* and *Ellen* are swapped.

	A	B	C
1	Index 1	2	
2	Index 2	5	
3			
4	Anders		
5	Bente		
6	Chris		
7	Dorthe		
8	Ellen		
9	Frederik		
10	George		

	A	B	C
1	Index 1	2	
2	Index 2	5	
3			
4	Anders		
5	Ellen		
6	Chris		
7	Dorthe		
8	Bente		
9	Frederik		
10	George		

Figure 1. Swapping two names. Left: before swap. Right: after swap

Question B. Write a sub "Assignment5b" that, given values of r_1 and r_2 (to be read from B1 and B2), reverses the sequence of from the r_1'th to the r_2'th name.

See Figure 2 for an example. For illustration purpose, the sequence is marked by green.

	A	B	C
1	Index 1	2	
2	Index 2	8	
3			
4	Anders		
5	Bente		
6	Chris		
7	Dorthe		
8	Ellen		
9	Frederik		
10	George		
11	Hans		
12	Ida		
13	Jacob		
14	Kent		
15	Lasse		

	A	B	C
1	Index 1	2	
2	Index 2	8	
3			
4	Anders		
5	Hans		
6	George		
7	Frederik		
8	Ellen		
9	Dorthe		
10	Chris		
11	Bente		
12	Ida		
13	Jacob		
14	Kent		
15	Lasse		

Figure 2. Reversing the sequence from r_1 to r_2. Lefts: before. Right: after.

Question C. Write a sub “Assignment5c” that changes the order of the names so they appear in random order.

Assignment 6 - VBA (20%)

Please consult ”Sheet6” in the Excel file “TFA reeks 20-21 assignment4-6.xlsx” that can be downloaded from WISEflow, and put the statement “Worksheets("Sheet6").Activate” in the top of your code to ensure that the code is using the correct sheet.

In the sheet, you see a list of tasks. The number of tasks is stated in cell C1. For each task, you see the task ID and a descriptive text, along with the following information:

- Column D: The number indicates any task that must start before the current task. For instance, task 6 (“Check social media”) must be started before task 8 (“Filter false news”) can be started. Note that the task just have to start, not finish. You may assume that at most one task is required to start in each case.
- Column E: A number indicating how attractive the task is to perform. Higher values indicate higher attractiveness.
- Column F: A “1” is stated, if the task is already started. (the first three tasks are already started.)

Given a set of tasks, where some are already started, please write a sub, “Assignment6” that can be used to select the next task to start. Among the tasks not started, your sub should select the most attractive task that respects the predecessor requirement.

Your code should state the selected task in cell C2 and indicate in column F that the task is now selected.