EXAM ASSIGNMENT

Study Programme and level	MSc Logistics and Supply Chain Management + elective						
Term	V21-22o						
Course name and exam code(s)	Tools for Analytics					460202E016	
Exam form and duration	WOAI					6 hours	
Date and time	20 December 2021					09.00 – 15.00	
Supplementary material/aids	All	X	Specified		No		
Other relevant information	**Avoid being suspected of exam cheating** Remember to state references and use quotation marks, if you copy text from other sources or re-use parts of a previously submitted exam paper (plagiarism and self-plagiarism). Students must answer the exam assignment **individually**. All submitted exam papers are checked for plagiarism, so cheating and collaboration between students will be detected.						
Hand-in of hand-written material allowed	Yes		No	X			
Anonymous exam?	Yes	X	No		Please do not write your name or student ID number anywhere. Use your flow-id number (find this on the cover sheet in WISEflow).		
Number of pages (incl. front page)	9						

Practical information

- This exam is open book, open internet, closed other people. You may use any on-line 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 an appendix available for download from WISEflow with 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 that 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 version 4.1.1 (2021-08-10).
- 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 yourFlowId.html'.
  - 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(fitdistrplus)
    library(skimr)
    library(conflicted)
    ```
- 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.
  - Be very careful to save your file as an `.xlsm` file (and not a `.xlsx` file).
  - An Excel template file is given in the appendix that you may use as a starting point. This file also contains the data used in Assignments 4-6.

R - Assignment 1 (14%)

Question 1

Make functions:

- `SecToMin` which takes an input argument `sec` in seconds and return the number converted to minutes.
- `SecToHours` which takes an input argument `sec` in seconds and return the number converted to hours.
- `MinToSec` which takes an input argument `min` in minutes and return the number converted to seconds.
- `MinToHours` which takes an input argument `min` in minutes and return the number converted to hours.
- `HoursToMin` which takes an input argument `hours` in hours and return the number converted to minutes.
- `HoursToSec` which takes an input argument `hours` in hours and return the number converted to seconds.

All numbers may be decimal numbers, e.g. 90 seconds is 1.5 minutes and 1.5 hours is 90 minutes.

Question 2

Make a function `ConvertTime` which takes two input arguments:

- `val` A number.
- `unit` A string that can take values "sec", "min" and "hours".

The function should return `val` converted to seconds, minutes and hours with features:

- works for all possible values for `unit`,
- uses the functions in Question 1,
- returns a vector with 3 numbers (seconds, minutes and hours) or `NA` if `unit` does not equals "sec", "min" or "hours".

For example:

```
ConvertTime(val = 1.5, unit = "min")
```

should return

```
##    sec    min  hours
## 90.000  1.500  0.025
```

and

```
ConvertTime(val = 1.5, unit = "hours")
```

should return

```
##    sec    min  hours
## 5400.0   90.0    1.5
```

R - Assignment 2 (18%)

The dataset in the file `election.csv` lists election votes for municipal elections in Denmark. The column/variables are:

- `area`: municipality,
- `party`: political party,
- `year`: election year,
- `validVotes`: Number of (valid) votes,
- `personalVotes`: total number of personal votes,
- `listedMen`: men listed,
- `listedWomen`: women listed,
- `electedMen`: elected men,
- `electedWomen`: elected women.

You can read the dataset file into the dataset `dat` using

```
library(tidyverse)
dat <- read_csv("elections.csv")
```

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

Question 1

How many different municipalities were there in Denmark in 2017?

Question 2

In which election year was the total number of votes highest?

Question 3

Consider the 2017 election. How many votes and elected candidates did the different parties get in total (sort your result descending by votes)?

Question 4

Calculate the percentage of females listed for each year. Are there on average more women than men listed? Is there any tendency over the years?

Question 5

Consider the 2017 election. In which municipality did the Liberal Democratic Party get the highest percentage of votes?

R - Assignment 3 (18%)

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 from Assignment 2 which can be read using:

```
library(tidyverse)
dat <- read_csv("elections.csv")
```

Question 1

Create a visualization showing the total number of votes for each election year with the following features:

- Number of votes is represented using columns.
- Columns are filled with colours for each party.
- Informative figure title and axis titles are given.

Which year had the lowest number of votes?

Question 2

Create a visualization showing the relative number of elected women in each municipality with the following features:

- The relative number of elected women is shown using columns.
- Municipalities are rotated 90 degrees on the x-axis.
- The columns are sorted increasing.
- Informative figure title and axis titles are given.

Which municipality has the lowest/highest percentage of women elected?

Question 3

Create a visualization showing the elected number of candidates as a function of election year with the following features:

- We consider only municipalities Herning, Copenhagen and Aarhus.
- We consider only the Social Democratic Party, Conservative Peoples Party and Liberal Democratic Party.
- The plot is divided using `area` (facet).
- Different colours are used for each party.
- Informative figure title and axis titles are given.
- Data points are given in the plot.
- A line for each party is added.

Consider the Social Democratic Party in Copenhagen. Based on the plot do the elected number of candidates increase or decrease over the election years?

Question 4

Create a visualization showing the relative number of women elected compared to listed (e.g. if listed is 12 and elected is 4 then the number is 4/12) with the following features:

- Columns are used for each party.
- Colours are used to identify the party.
- A sub-plot is given for each year.
- Informative figure title and axis titles are given.

Which party seems to get the highest number of the listed women elected?

Question 5

Create a visualization showing the variance in the relative number of personal votes compared to valid votes (i.e. divide the two numbers) with the following features:

- We consider only municipalities Herning, Copenhagen and Aarhus.
- Variation is shown using a box-plot for each municipality.
- Informative figure title and axis titles are given.

Which municipality has the highest median?

VBA - Assignment 4 (10%)

Save all your VBA code in the module named `modA4`.

In this exercise you are *not* allowed to use any procedures made during the TFA course. You have to code your procedure using plain VBA.

Create a procedure that reads the strings in worksheet *DataA4* into a 2D array with the following features:

- Indexing must start from 1 in both dimensions.
- The procedure should work for other datasets with a different number of columns and rows. You may assume that the data starts in cell A1.
- You must use for loop(s) to assign values to the array.
- A message box should be given after the data has been read with the value of array entry (1,3). If the entry does not exists a warning should be given instead.
- A button in worksheet *A4* should run the procedure.

VBA - Assignment 5 (20%)

Save all your VBA code in the module named `modA5`.

Given an investment of x thousand dollars you have the option of investing in two projects:

- Project 1: The profit function is

$$\pi_1(x) = \begin{cases} A_1 x, & \text{if } 0 \leq x < 20 \\ A_2 x - 40, & \text{if } 20 \leq x < 50 \\ A_3 x - 140 & \text{if } x \geq 50 \end{cases}$$

where A_i is normal distributed with mean $2i$ and standard deviation $0.1i$.

- Project 2: The profit function is

$$\pi_2(x) = \begin{cases} 1.25x, & \text{if } 0 \leq x < 40 \\ (x - A_1)^2 + 50, & \text{if } 40 \leq x < 60 \\ A_2 x + 510 & \text{if } x \geq 60 \end{cases}$$

where A_1 is a continuous random number between 38 and 42 (uniform distributed) and A_2 is a continuous random number between -2 and -0.5 (uniform distributed).

Question 1

Create two functions that return the profit of Project 1 and 2 given investment x as input argument, i.e. the random samples A_i is calculated inside the functions.

Question 2

Make a sub that do a simulation with the following features:

- Considers investment sizes $x = 10, 20, 30, ..., 100$.
- Uses 50 samples for each investment x.
- Calculates the average profit and standard deviation of the two projects for each x.
- Stores the results in a worksheet named *A5*.

VBA - Assignment 6 (20%)

Save all your VBA code in the module named `modA6`*.*

Consider worksheet *DataA6-1* which contains a table with three columns. You want to search the table and output the matching rows to a new sheet.

Question 1

Create a procedure that searches the table with the following features:

- First, a message box is used to ask if a name should be searched. If the answer is *yes*, then use an input box to type the name.
- Next, use a message box to ask if an amount should be searched. If the answer is *yes*, then use an input box to type the amount (search for amounts greater than or equal the typed amount).
- Scan the table and output all rows that match the specified criteria, and output the resulting table with matching rows in worksheet *A6*. If no criteria are used, then return all rows.
- Add a button to worksheet *A6* that run the procedure.
- The procedure should work for other datasets with different number of rows too (such as *DataA6-2*).

Question 2

Create a procedure that searches the table with the following features:

- Copy the whole table to worksheet *A6*.
- Sort the table non-increasing order of the column **Sales Amount**, then next based on the **Title** column.
- Use an input box to ask for an amount x.
- Highlight the cells in the rows where the **Sales Amount** is less than x.
- Add a button to worksheet *A6* that run the procedure.