

EXAM ASSIGNMENT

Study Programme and level	MSc Logistics and Supply Chain Management + elective						
Term	V22-23o						
Course name and exam code(s)	Tools for Analytics					460202E016	
Exam form and duration	Written exam, WOAI					3 hours	
Date and time	11 January 2023					09.00 – 12.00	
Supplementary material/aids	All	X	Specified			No	
Hand-in of extra material (appendix) in WISEflow allowed	Yes	X					
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).		
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.						
Number of pages (incl. front page)	7 pages						

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.
- 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 and .html) as “Appendix material”. Due to the system, you must also upload an empty PDF document named `yourFlowId.pdf`.
- Your R code will be tested using R version 4.2.2 (2022-10-31).
- 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 ...”.
- You may load and use the following packages:

```
```r
library(tidyverse)
library(skimr)
library(conflicted)
```
```

- An R markdown template file is given in the appendix that you may use as a starting point.

R - Assignment 1 (40%)

Consider the dataset in the file `jobs.csv` with engine maintenance jobs for a group of ships.

The dataset contains a row for each item used. The columns/variables are:

- `ship`: The ship considered.
- `job_id`: Maintenance job id. A job is a collection of items replaced.
- `job_desc`: Job description.
- `item_id`: Item id.
- `item_name`: Item name.
- `item_quantity`: Number of items used.
- `item_manufacturer`: Item manufacturer.
- `component_id`: Engine component id.
- `component_desc`: Engine component description.
- `done_date`: Date the job finished.
- `year`: Year of done date.
- `days`: Days since the item was last used for maintenance on the ship.

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

Question 1

Find the delimiter used in the csv file and load the file into a tibble/data frame called `dat`.

Question 2

Provide a short overview of the data. How many rows are there in the data, what is the number of different ships, and what is the range of dates?

Question 3

What is the total number of different jobs, number of different jobs per ship, and average number of jobs per ship? Hint: The function `n_distinct` may be used to find distinct values within a group.

Question 4

What is the minimum, average, and maximum number of *different* items considered at each job? What is the minimum, average, and maximum number of items used for maintenance at each job?

Question 5

It seems that some items considered in a job are not used anyway. Which ship has most jobs where an item should have been used, but has not (quantity is zero)?

Question 6

Which item is the most used one for maintenance?

R - Assignment 2 (40%)

Answer this assignment using the *ggplot2* package in *tidyverse* (you may need *dplyr* for preparing the datasets you want to plot). We work with the dataset from Assignment 1, which can be read using:

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

Question 1

For each year and job id, identify the ship the job was done on. Hint: The `first` function may be used to select the first item within a group. Use this to create a visualization showing the number of maintenance jobs for each year with the following features:

- Bars are used for each year.
- Fill colors are used to identify the ship.
- Legends are put at the bottom of the plot.
- Informative figure title and axis titles are given.

Which year had the most jobs?

Question 2

Create a visualization showing the number of different items used for each ship with the following features:

- The numbers are shown using columns.
- Reorder the columns so they increase along the x-axis.
- Informative figure title and axis titles are given.

Which ship uses most items?

Question 3

Create a visualization showing the number of items used as a function of done date with the following features:

- The numbers are shown using columns with a fixed line width of 1 and a blue color.
- The plot is divided using `ship` (facet). Hint: You may use `scales = "free_y"`.
- Informative figure title and axis titles are given.

Which ship has most items used at a specific date?

Question 4

Consider two items with id:

```
items <- c("601.004.006", "601.026.128")
```

Create a visualization showing the variation of demand (item quantity) of each item with the following features:

- A density is plotted for each item with a fixed bandwidth (`bw`) of 1 and transparency (`alpha`) of 0.5.
- Item name is used as fill.
- The plot is divided using `year` (facet).
- Informative figure title and axis titles are given.

Over the years, is it consistent that on average the demand of one of the items is higher than the other?

Question 5

Consider two items with id:

```
items <- c("601.004.006", "601.026.128")
```

Create a visualization showing the variation of days since last used for maintenance given an item with the following features:

- A violin is used to plot days since last used given an item.
- Item name is used as fill.
- Informative figure title and axis titles are given.

Comment on the plot.

R - Assignment 3 (20%)

Create a function `process` with the following features:

- The arguments are:
 - `v`: A vector with numbers (may contain missing values, NAs).
 - `find`: A number to search for. Default value is `NULL` indicating not to search for a number.
 - `replace`: A number. If `find` is a number, then replace all occurrences of `find` in the vector with `replace` (if any). Default value is `NULL` indicating not to replace anything.
- The function returns a list with
 - the original vector `v`,
 - the length of `v`,
 - the number of missing values,
 - if `find` is a number, the number of times the number is found, and
 - if `replace` is a number, the new vector with the replaced numbers.
- Document your function.

You may test your function using:

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

```
## $original
## [1] 8 19 -2 8 NA 12 9 23 NA -34
##
## $length
## [1] 10
##
## $missing
## [1] 2
```

```
process(v, find = 1)
```

```
## $original
## [1] 8 19 -2 8 NA 12 9 23 NA -34
##
## $length
## [1] 10
##
## $missing
## [1] 2
##
## $found
## [1] 0
```

```
process(v, find = 8)
```

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

```
## $length
## [1] 10
##
## $missing
## [1] 2
##
## $found
## [1] 2

process(v, find = 8, replace = 5)

## $original
## [1] 8 19 -2 8 NA 12 9 23 NA -34
##
## $length
## [1] 10
##
## $missing
## [1] 2
##
## $found
## [1] 2
##
## $new
## [1] 5 19 -2 5 NA 12 9 23 NA -34
```