

EXAM ASSIGNMENT

Study Programme and level	MSc Operations and Supply Chain Analytics + elective					
Term	Winter 2024/25 – ordinary exam					
Course name and exam code(s)	Tools for Analytics				460202E016	
Exam form and duration	Written onsite exam, internet allowed				3 hours	
Date and time	13 January 2025				15:00-18:00	
Supplementary material/aids	All	X	Specified		No	
Anonymous exam	Yes	X	No		Please do not write your name or student ID number anywhere.	
Use of generative AI (GAI) allowed	Yes		No	X		
Hand-in of handwritten material allowed	Yes		No	X		
Hand-in of extra material (appendix) in WISEflow allowed	Yes	X	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.					
Number of pages (incl. front page)	6 pages					

Other instructions:

It is important that you start uploading your exam paper well in advance - at least 10 min. before end of exam.

Practical information

- This exam is with internet. 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. That is, you cannot ask direct questions on the internet, use answers given on the internet after the start of this exam, or consult with each other, not even for hypothetical questions.
- The use of tools based on Generative AI, such as ChatGPT or Copilot, is not allowed.
- This assignment has an appendix available for download from WISEflow.
- 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 <participant index>.pdf.
- Your R code will be tested using R version 4.4.2 (2024-10-31).
- Your R code must be written in an R Markdown (Rmd) file named <participant index>.Rmd. Moreover, hand in the rendered/knitted HTML file <participant index>.html.
- Your file must include your code and a (brief) comment for each question. For example, “The three companies with the 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 on WISEflow that you may use as a starting point.

R - Assignment 1 (55%)

Consider the dataset in the file `bom.csv` containing a Bill of Materials (BOM) for a group of brewery products and materials. The dataset contains a row for each material and product, i.e. a row gives information about how much of the material is needed to produce the product. A product is produced in batches, and the BOM provides information about the quantities of the materials needed to produce one batch of each product.

The columns/variables are:

- `product_id`: Product id.
- `product_desc`: Product description.
- `product_type`: Product type.
- `product_group_id`: Product group id.
- `product_group_desc`: Product group description.
- `material_id`: Material id.
- `material_quantity`: Amount needed for producing a batch of the product.
- `material_unit`: Unit used for the material (kilogram (KG), meters (M), items (I)).
- `material_unit_price`: Unit price/cost in Euros.
- `material_valid_from`: First date the material is valid to use.
- `material_valid_to`: Last date the material is valid to use.

The dataset can be read using:

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

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

Question 1

How many different products, product groups and materials are considered in the dataset?

Question 2

Find all materials that are valid on date January 1st 2022 (2022-01-01). Is the material with id M34491 used?

Question 3

Create a function `get_valid` with the following features

- Takes a BOM dataset and a date as arguments.
- Removes all materials that are not valid any more at the date and returns the filtered dataset.

That is, do the same as in the previous question with an arbitrary date. Note: `get_valid` returns a data frame which can be used in a pipe. Test the function by running:

```
get_valid(dat, "2023-02-15")
```

Question 4

How many rows have negative material quantity? Modify your data in the following way:

- If `material_quantity` is negative, then assume this is a typo and set the quantity to the positive number instead.
- Add a column `material_cost` equal to the material quantity times the material unit price.
- Remove all rows where material quantity is zero.

Question 5

Consider the modified dataset from the previous question (if you did not succeed in modifying the data, then consider the data in the file `bom-modified.csv`).

For each product group, calculate/find the number of different products and materials. Moreover, for each product group, calculate the average number of different materials used per product. Hint: The function `n_distinct` can be used to find the number of different items within a group.

Which product group contains the most products, and which product group has the highest average number of different materials?

Question 6

A dataset with material info is given in the file `materials.csv`.

The columns/variables are:

- `material_id`: Material id.
- `material_desc`: Material description.
- `material_group_id`: Material group id.
- `material_group_desc`: Material group description

The dataset can be read using:

```
dat_mat <- read_csv("materials.csv")
```

Answer the following questions:

- How many different material groups are considered?
- Join the dataset with the BOM dataset. Which material group is used in most products?
- Consider material groups MG2130201 and MG2130202. For each material group and product type, calculate the minimum, average and maximum quantity used.

R - Assignment 2 (45%)

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

```
library(tidyverse)
dat <- read_csv("bom-m-groups.csv")
```

The columns/variables are as in Assignment 1 plus the following columns:

- `material_cost`: The cost of using the material in the production of a batch of a product.
- `material_desc`: Material description.
- `material_group_id`: Material group id.
- `material_group_desc`: Material group description.

Question 1

Create a visualization showing the variation in material costs with the following features:

- A histogram is used (with argument `binwidth = 100`).
- Fill colors are used to identify the product type.
- A plot is made for each product type (facet).
- Legends are added at the bottom of the plot.
- Informative figure title and axis titles are given.

In which interval does the material cost mostly lie? What product type has high material cost?

Question 2

Create a visualization showing the number of products for each product type with the following features:

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

Which product type contains most products?

Question 3

Create a visualization showing the total costs of selected products with the following features:

- The total cost of a product is the sum of the material costs.
- Only 6 products for each product type are plotted. Pick the 6 with the highest total cost (most expensive).
- Bars are used for each product ordered such that the height of the bars decreases over the x-axis.
- Fill colors are used to identify the product type.
- Legends are added at the bottom of the plot.
- Informative figure title and axis titles are given.

What product type category has the lowest total costs among the products visualized?

Question 4

Create a visualization showing the proportion of material costs used in each material group given a product with the following features:

- Consider products with id P10393, P64433 and P100160.
- Bars are used for each product such that the height is the same since we want to compare proportions (`position = "fill"`).
- Fill colors are used to identify the material group.
- Legends are added at the bottom of the plot.
- Informative figure title and axis titles are given.

Which material group has the highest relative cost in the product with id P100160.

Question 5

Create a visualization showing the variation in material costs given the product type with the following features:

- Only consider the material group with id MG2130201.
- Variation is visualized using a density.
- Fill colors are used to identify the product type.
- Legends are added at the bottom of the plot.
- Informative figure title and axis titles are given.

Which product type has the highest mean value in this product group?