DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

EXAM ASSIGNMENT

Study Programme and level	**MSc Logistics and Supply Chain Management + elective**						
Term	V22-23r						
Course name and exam code(s)	Tools for Analytics						460202E016
Exam form and duration	Written exam, WOAI						3 hours
Date and time	21 February 2023						9.00 – 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)	6 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 (50%)

Consider the dataset in the file `orders.csv` with purchase orders for a group of ships.

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

- `ship`: The ship considered.
- `order_id`: Order id. An order is a group of items purchased in one batch from a single supplier.
- `item_id`: Item id.
- `item_desc`: Item description.
- `quantity`: Number of items ordered.
- `price`: Price per unit.
- `order_date`: Order date.
- `delivery_date`: Expected delivery date when order is made.
- `delivery_place`: Delivery place.
- `recieved_date`: Actual date the order is recieved.
- `supplier`: Supplier for the order.
- `delivery_cost`: Delivery cost.
- `order_year`: Year the order was placed.

The dataset can be read using:

```
library(tidyverse)
dat <- read_csv("orders.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 ships, suppliers, and different items are considered?

Question 2

Do all orders use a single supplier and delivery place?

Question 3

For each order, calculate/find the number of different items, number of items, total price, total delivery cost, and order year (missing values are assumed zero).

Which order has the highest delivery cost?

Question 4

For each order, find the supplier and delivery place. Hint: The `first` function may be used to select the first item within a group.

Which supplier and delivery place is used most?

Question 5

Add a column to the dataset equal the *value* of the items in a row calculated as the price times the quantity. Next, create a summary table named `res1` that for each item id calculates the aggregated value and arrange them in descending order.

Which item is the most costly one with respect to the calculated value in `res1`?

Given the dataset `res1`, add two new columns:

- The relative value for each item as the ratio between the item's value and the total value of all items.
- The cumulative relative value. Hint: You can use the `cumsum` function here.

Consider the most costly item with respect to the calculated value in `res1`. At which ships is the item used and how many times?

R - Assignment 2 (50%)

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("orders.csv")
```

Question 1

Create a visualization showing the number of purchase orders 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 orders?

Question 2

Create a visualization showing the total number of items ordered 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

For each ship, 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 items with id:

```
items <- c("601.003.004" ,"601.004.006", "601.026.128", "601.026.052")
```

For each row in the dataset, calculate an index variable as price divided with the maximum item id price. Hint: Group by item id and use `mutate` to add the column. Create a visualization showing the index as a function of order year with the following features:

- The numbers are shown using points.
- The plot is divided using item id (facet).
- A trend line is added.

- Informative figure title and axis titles are given.

Take a closer look at the item with most price fluctuations. Are the prices reasonable for a singe product?

Question 5

Make a summary table in the following way:

- Add a column `val` with the value of an item in a row equaling the price times quantity.
- Group each item id and set the total value `total_value` to the sum of `val` and calculate the total quantity purchased.
- Arrange the total value in descending order.
- Add columns:
  - Relative total value equals `total_val/sum(total_val)`.
  - Relative item number equals `row_number()/n()`.
  - Cumulative relative total value (`cum_pct_val`). Hint: You can use the `cumsum` function here.
  - Class equals `class = case_when(cum_pct_val <= 0.8 ~ "A", cum_pct_val <= 0.95 ~ "B", TRUE ~ "C")` classifying items in A (high value items), B (middle), and C (low value items).

Create a visualization showing the cumulative relative total value as a function of relative item number with the following features:

- Data points are given in the plot.
- Different colors are used for each class.
- Different sizes are used for each quantity.
- A line is added with black color and fixed size 0.3.
- Informative figure title and axis titles are given.

How many items contribute to 80% of the total value?