



R Lab

Graded Assignment #1

Introduction

This assignment is to be done on an individual basis. The scoring will be 0-10, and it will have a weight of 50% of the overall R Labs grade (which counts for 10% of the overall course grade).

The data to be used for this assignment can be found in the file **Logistics.txt**. It includes data provided by the Trialto Latvia, the third-party logistics operator. The dataset consists 2279 observations with 7 features (variables). The selected features include only numerical data. All the features have an undeniable impact on the inventory management and constitute two core groups: handling-related and turnover-related. Such features as expire date, pallet weight, pallet height and number of units per pallet determine the speed and subtlety of handling. On the other hand, total outbound and number of outbound orders indicate how tradable a particular unit is.

The variables in the dataset are:

Variable name	Description
id	ID of the unit
unit_price	Unit price in Euro
exprie_date	Unit shelf-life
outbound_number	Times that a product was ordered from 2017-02-06 to 2018-02-13
total_outbound	Number of pallets sold from 2017-02-06 to 2018-02-13
pal_weight	Weight of a fully-loaded pallet
pal_height	Height of a fully-loaded pallet
units_per_pal	Number of units in a pallet

Please complete the tasks below. The use of additional packages is not allowed. Make sure you name data as indicated below so the script is reproducible. You may submit an R **script** or an R **Markdown** document (preferably HTML or PDF) as a solution.

Tasks

1. Load the data in an object named **dataLogistics**. (0.5 point)
2. Check for any missing values in **dataLogistics**. If any are found, then delete those observations. How many observations have been deleted? (0.5 point)

3. For which unit the corresponding fully-loaded pallet is the heaviest one? For how many units the corresponding fully-loaded pallets are the shortest ones? Hint: function `which()`. (0.5 point)
4. Produce a boxplot of the `pal_weight` variable, for (faceted by) each value of the variable `expire_data` observed in the dataset (in the same plot window). Draw a horizontal line at `y=200`. (1 point)
5. Calculate the total amount of money earned (in Euro) from the units sold at prices greater than 0. (1 point)
6. Find the lightest unit among those for which both the number of the units in the pallet and the weight of the pallet are greater than 0. Hint: function `which()`. (1 point)
7. Some values in `expire_date` are equal to 0. Create a new data frame, named `dataLogistics2`, which does not include these observations. In `dataLogistics2` data frame, for which unit the corresponding fully-loaded pallet is the heaviest one? (0.5 points)
8. Work with the `dataLogistics2` data frame. Create a matrix with two columns, one with the `outbound_number` variable and one with the `total_outbound` one. Calculate a vector that includes the average of the two, by using the `apply()` function. (1 point)
9. Work with the `dataLogistics2` data frame. Create a scatterplot with `pal_weight` on the x-axis and `pal_height` on the y-axis. The main title should be “Weight and Height of Pallets” and the axes should be named accordingly. Colour with black the dots for which `pal_height>1`; otherwise, colour them with red. Add a horizontal dashed line for `pal_height=1`. Make the y-axis labels to be perpendicular to the axis. (2 points)
10. Create a new variable, named `quality`, assigned it to the `dataLogistics2` data frame. The new variable describes the quality of units. If the price of a unit is greater than 15, then the quality of the unit is **Very Good**. Otherwise, the quality of the unit is **Good**. Then, create a grouped barplot of the new variable and the `expire_date`. The main title should be “Quality and Expire Date”. Also, the axes should be named, accordingly providing an appropriate legend at the topright of the grouped barplot. (2 points)