Project 2-Inbound costs per kg by supplier country&transportation

Andy Shih

2023-07-11

Data cleaning

I use some text processing functions in R to clean the messy data entry in the original Excel sheet, extracting the real useful numbers in the inbound and outbound logistics

```
library(dplyr)
library(tidyr)
library(readxl)
project2 <- read_excel("/Users/andy/B futurist/project/data/Trade Support - Logistics Tracking.xlsx", sh

## Inbound costs data cleaning
# to get rid of dollar signs and space just in case
project2$Inbound_cost_final = gsub("\\$", "", project2$Inbound_cost_final)
project2$Inbound_cost_final = gsub("\\€", "", project2$Inbound_cost_final)
project2$Inbound_cost_final = gsub(" ", "", project2$Inbound_cost_final)

# turn character to numeric
project2$Inbound_cost_final <- as.numeric(project2$Inbound_cost_final)
```

I use some text processing functions in R to clean the messy data entry in the original Excel sheet, extracting the real useful numbers in the weight of each inbound

```
## Weight data cleaning
project2$Weight = gsub("k", "", project2$Weight)
project2$Weight = gsub("g", "", project2$Weight)
project2$Weight = gsub("K", "", project2$Weight)
project2$Weight = gsub("G", "", project2$Weight)
project2$Weight = gsub(" ", "", project2$Weight)

# turn character to numeric
project2$Weight <- as.numeric(project2$Weight)
```

Data wrangling and summarization

I use dplyr package to do aggregation, grouping the data by country of import and summarizing the transportation costs per kg and the number of cases

```
# inbound costs per kg for each transaction
project2$costPerKg <- project2$Inbound_cost_final/project2$Weight

# inbound costs per kg for every country
# aggregate(costPerKg ~ Supplier_country, FUN=mean, data = project2, na.rm = T)

project2 %>% group_by(Supplier_country) %>%
    summarize(mean_costPerKg = mean(costPerKg),
              n = n()) %>% drop_na() %>% print(n = Inf) # to print all rows of the tbl dataframe
```

```
## # A tibble: 30 x 3
##    Supplier_country mean_costPerKg     n
##    <chr>                     <dbl> <int>
##  1 Austria                   1.06      8
##  2 Belgium                   0.324     2
##  3 Bulgaria                  0.544     8
##  4 Cyprus                    0.877    10
##  5 Czech Replublic           1.65      9
##  6 France                    5.24     18
##  7 Germany                   0.858    21
##  8 Greece                    1.46      8
##  9 HK                        9.49     26
## 10 Hungary                   1.21      7
## 11 Ireland                   0.899     5
## 12 Israel                    1.61      1
## 13 Italy                     3.13      7
## 14 Korea                    53.5       3
## 15 Lebanon                   2.51      2
## 16 Lithuania                 1.69     10
## 17 Monaco                    0.644     3
## 18 Netherlands               0.626    23
## 19 Poland                    1.16     26
## 20 Portugal                  1.29      7
## 21 Saudi Arabia              6.02      1
## 22 Singapore                 4.76      1
## 23 Slovakia                  3.73      9
## 24 Spain                     2.17     25
## 25 Switzerland               4.35      2
## 26 Turkey                    4.83      1
## 27 UAE                       4.43     16
## 28 UK                        4.07     46
## 29 US                        3.05     16
## 30 Uruguay                   9.21      1
```

The company is doing international trade of high-valued fragrances and cosmetics, especially ones between Asia and Europe. Therefore, I hereby focus on Asian clients. After figuring out the mean of transportation (by sea or by air), I aggregate data by country of import and mean of transportation and summarize the transportation costs per kg and the number of cases

```
## # A tibble: 12 x 4
## # Groups:   Supplier_country [10]
##    Supplier_country transportation mean_costPerKg     n
##    <chr>            <chr>                   <dbl> <int>
##  1 HK               a                        9.49    26
##  2 Israel           a                        1.61     1
##  3 Korea            a                       10.5      1
##  4 Korea            s                       75.0      2
##  5 Lebanon          a                        1.74     1
##  6 Saudi Arabia     a                        6.02     1
##  7 Singapore        a                        4.76     1
##  8 Turkey           a                        4.83     1
##  9 UAE              a                        4.87    14
## 10 UAE              s                        1.33     2
## 11 US               a                        3.25    15
## 12 Uruguay          a                        9.21     1
```