LA Apparel Technical Interview - R

Joshua Susanto

2024-06-23

File titled Theoretical Lot Audits - Assessment contains data regarding a theoretical audit of fabric inventory. Since admin staff are unable to export adjustments from the fabrics system, they export the following information from our fabrics system and maintain a spreadsheet of adjustments. The following information is exported from the fabrics system:

Lot# Date Style# Desc# Color# Received Pcs Stock Pcs

Warehouse staff inform administrative staff of the number of pcs in inventory for each lot at the warehouse. When the administrative staff update the system, they update columns A, B and J.

When admin staff update each record, they update column A with the date the fabric system is updated and admin staff record their own names in column B. The admin staff then enter adjustments in column J which are the difference between initial stock pcs previously recorded in the fabric system and final count of pcs warehouse staff count in the warehouse. Received pcs is the amount of units in the lot when the company initially receives each lot. If warehouse staff discover lots which are not currently in the fabric system, then the admin staff will add the required info to the system and a new row to the bottom of the dataset.

```
# import tidyverse
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

1. What are your observations about this data set? Please include observations regarding:

```
# load dataset
fabrics <-suppressWarnings( readxl::read_excel("theoretical_lot_audits.xlsx"))

# first looks
head(fabrics)
```

```
## # A tibble: 6 x 11
##   `System Entry Date` `System User` `Lot#`  `Received Date` `Style#` `Desc#`
##   <dttm>              <chr>         <chr>              <dbl> <chr>    <chr>
## 1 2022-05-26 00:00:00 Don           Lot77              44642 Style1   Descriptio~
## 2 2022-05-31 00:00:00 Sam           Lot665             44706 Style1   Descriptio~
## 3 2022-05-23 00:00:00 Sam           Lot1171            44035 Style1   Descriptio~
## 4 2022-05-31 00:00:00 Sam           Lot1278            44232 Style1   Descriptio~
## 5 2022-05-31 00:00:00 Sam           Lot1279            44232 Style1   Descriptio~
## 6 2022-05-31 00:00:00 Dan           Lot1301            44536 Style1   Descriptio~
## # i 5 more variables: `Color#` <chr>, `Received Pcs` <dbl>,
## #   `Initial Stock Pcs` <dbl>, `Adjusted #Pcs` <dbl>, `Final Count Pcs` <dbl>
```

```
str(fabrics)
```

```
## tibble [2,694 x 11] (S3: tbl_df/tbl/data.frame)
##  $ System Entry Date: POSIXct[1:2694], format: "2022-05-26" "2022-05-31" ...
##  $ System User      : chr [1:2694] "Don" "Sam" "Sam" "Sam" ...
##  $ Lot#             : chr [1:2694] "Lot77" "Lot665" "Lot1171" "Lot1278" ...
##  $ Received Date    : num [1:2694] 44642 44706 44035 44232 44232 ...
##  $ Style#           : chr [1:2694] "Style1" "Style1" "Style1" "Style1" ...
##  $ Desc#            : chr [1:2694] "Description1" "Description1" "Description1" "Description1" ...
##  $ Color#           : chr [1:2694] "Color140" "Color188" "Color20" "Color20" ...
##  $ Received Pcs     : num [1:2694] 2 2 4 1 1 2 0 2 3 3 ...
##  $ Initial Stock Pcs: num [1:2694] 2 2 4 1 1 2 0 2 3 3 ...
##  $ Adjusted #Pcs    : num [1:2694] 0 -2 -4 -1 -1 0 0 0 0 0 ...
##  $ Final Count Pcs  : num [1:2694] 2 0 0 0 0 2 0 2 3 3 ...
```

```
summary(fabrics)
```

```
##  System Entry Date                System User            Lot#
##  Min.   :2021-05-24 00:00:00.00   Length:2694        Length:2694
##  1st Qu.:2022-05-25 00:00:00.00   Class :character   Class :character
##  Median :2022-05-27 00:00:00.00   Mode  :character   Mode  :character
##  Mean   :2022-05-26 17:47:35.77
##  3rd Qu.:2022-05-30 00:00:00.00
##  Max.   :2022-06-06 00:00:00.00
##  NA's   :227
##  Received Date      Style#             Desc#              Color#
##  Min.   :42587   Length:2694        Length:2694        Length:2694
##  1st Qu.:44341   Class :character   Class :character   Class :character
##  Median :44523   Mode  :character   Mode  :character   Mode  :character
##  Mean   :44464
##  3rd Qu.:44656
##  Max.   :44767
##  NA's   :132
##   Received Pcs     Initial Stock Pcs Adjusted #Pcs     Final Count Pcs
##  Min.   :-19.000   Min.   :-19.000   Min.   :-4.3000   Min.   :-19.000
##  1st Qu.:  2.000   1st Qu.:  2.000   1st Qu.: 0.0000   1st Qu.:  1.000
##  Median :  2.000   Median :  2.000   Median : 0.0000   Median :  2.000
##  Mean   :  2.062   Mean   :  2.166   Mean   :-0.1569   Mean   :  1.977
##  3rd Qu.:  2.000   3rd Qu.:  2.000   3rd Qu.: 0.0000   3rd Qu.:  2.000
##  Max.   : 87.000   Max.   : 87.000   Max.   : 4.0000   Max.   : 87.000
##                    NA's   :127       NA's   :245
```

```r
# missing values
missing_values <- colSums(is.na(fabrics))
print(missing_values)
```

```
## System Entry Date       System User              Lot#     Received Date
##               227               226               126               132
##            Style#             Desc#            Color#      Received Pcs
##               129               129               126                 0
## Initial Stock Pcs     Adjusted #Pcs   Final Count Pcs
##               127               245                 0
```

We seem to have many missing values, we know that the system date and user are going to be filled out for audited logs. There seems to be other missing information, for logs that are audited but have missing adjusted pieces, a fair assumption would be that there was no adjustment.

Additionally, received pieces and initial stock pieces should more or less remain consistent across logs according to the information given. Taking a closer look:

```r
# filtering data for any instances where we received or expect differently than our initial count
fabrics %>% filter(`Received Pcs` != `Initial Stock Pcs`)
```

```
## # A tibble: 5 x 11
##   `System Entry Date` `System User` `Lot#`  `Received Date` `Style#` `Desc#`
##   <dttm>              <chr>         <chr>              <dbl> <chr>    <chr>
## 1 NA                  <NA>          Lot624             44592 Style10  Descriptio~
## 2 NA                  <NA>          Lot2552            44592 Style10  Descriptio~
## 3 2022-05-31 00:00:00 Dan           Lot2240            44579 Style16  Descriptio~
## 4 2022-06-01 00:00:00 Sam           Lot1718            44361 Style87  Descriptio~
## 5 2022-06-01 00:00:00 Sam           Lot2277            44307 Style100 Descriptio~
## # i 5 more variables: `Color#` <chr>, `Received Pcs` <dbl>,
## #   `Initial Stock Pcs` <dbl>, `Adjusted #Pcs` <dbl>, `Final Count Pcs` <dbl>
```

There are only 5 cases in which this holds true, and in all cases we counted more than we received. Not sure if this is a log error, but with only 0.0018% of cases facing this discrepency, it doesn't seem like a huge issue.

Total # Unique Lots

```r
# pipeline: from fabrics, select Lot#, drop missing values, and find how many unique values
fabrics %>%
  select(`Lot#`) %>%
  drop_na() %>%
  n_distinct()
```

```
## [1] 2562
```

Using a dplyr pipeline we can easily find that there are 2562 unique lots, not accounting for missing values. Finding those that are audited, we need to find logs where a system date and user are recorded.

Unique Lots Audited (# & %)

```
# pipeline: from fabrics, filter when either date or user is not missing, select Lot#, drop missing val
fabrics %>%
  filter(!is.na(`System Entry Date`) | !is.na(`System User`)) %>%
  select(`Lot#`) %>%
  drop_na() %>%
  n_distinct()
```

```
## [1] 2462
```

By using an "or" operator we can see that there are 2462 different audited lots, or 91.389% of total unique lots.

```
# pipeline: from fabrics, filter when both date and user is missing, select Lot#, drop missing values,
fabrics %>%
  filter(!is.na(`System Entry Date`) & !is.na(`System User`)) %>%
  select(`Lot#`) %>%
  drop_na() %>%
  n_distinct()
```

```
## [1] 2461
```

However by using an "and" operator we get 2461, which means that there is exactly 1 instance where a user input their name but not the date. Since both date and user must be manually inputted, it makes more sense to consider the "or" operator as the correct scenario.

Unique Lots Remaining to be Audited (# & %)

```
fabrics %>%
  filter(is.na(`System Entry Date`) & is.na(`System User`)) %>%
  select(`Lot#`) %>%
  drop_na() %>%
  n_distinct()
```

```
## [1] 100
```

```
fabrics %>%
  filter(is.na(`System Entry Date`)) %>%
  filter(!is.na(`Lot#`))
```

```
## # A tibble: 101 x 11
##    `System Entry Date` `System User` `Lot#`  `Received Date` `Style#` `Desc#`
##    <dttm>              <chr>         <chr>              <dbl> <chr>    <chr>
##  1 NA                  <NA>          Lot742             44042 Style4   Descripti~
##  2 NA                  <NA>          Lot689             44334 Style9   Descripti~
##  3 NA                  <NA>          Lot690             44643 Style9   Descripti~
##  4 NA                  <NA>          Lot694             44702 Style9   Descripti~
##  5 NA                  <NA>          Lot2285            44569 Style9   Descripti~
##  6 NA                  <NA>          Lot611             44592 Style10  Descripti~
##  7 NA                  <NA>          Lot612             44592 Style10  Descripti~
```

```
##  8 NA                 <NA>          Lot613           44592 Style10  Descripti~
##  9 NA                 <NA>          Lot614           44592 Style10  Descripti~
## 10 NA                 <NA>          Lot615           44592 Style10  Descripti~
## # i 91 more rows
## # i 5 more variables: 'Color#' <chr>, 'Received Pcs' <dbl>,
## #   'Initial Stock Pcs' <dbl>, 'Adjusted #Pcs' <dbl>, 'Final Count Pcs' <dbl>
```

From our previous observation we know that this case must have both the system entry data as well as the system user be missing. We end up with 100 unique lots where these two entries are missing, or 3.712% of our lots needing to be audited.

Total Adjustments (pcs) (# & %) any other relevant insights

```
# first look: summary statistics of adjusted pieces
fabrics %>%
  select(`Adjusted #Pcs`) %>%
  summary()
```

```
##  Adjusted #Pcs
##  Min.   :-4.3000
##  1st Qu.: 0.0000
##  Median : 0.0000
##  Mean   :-0.1569
##  3rd Qu.: 0.0000
##  Max.   : 4.0000
##  NA's   :245
```

```
# a frequency table for distribution
adjustment_table <- table(fabrics$`Adjusted #Pcs`)
print(adjustment_table)
```

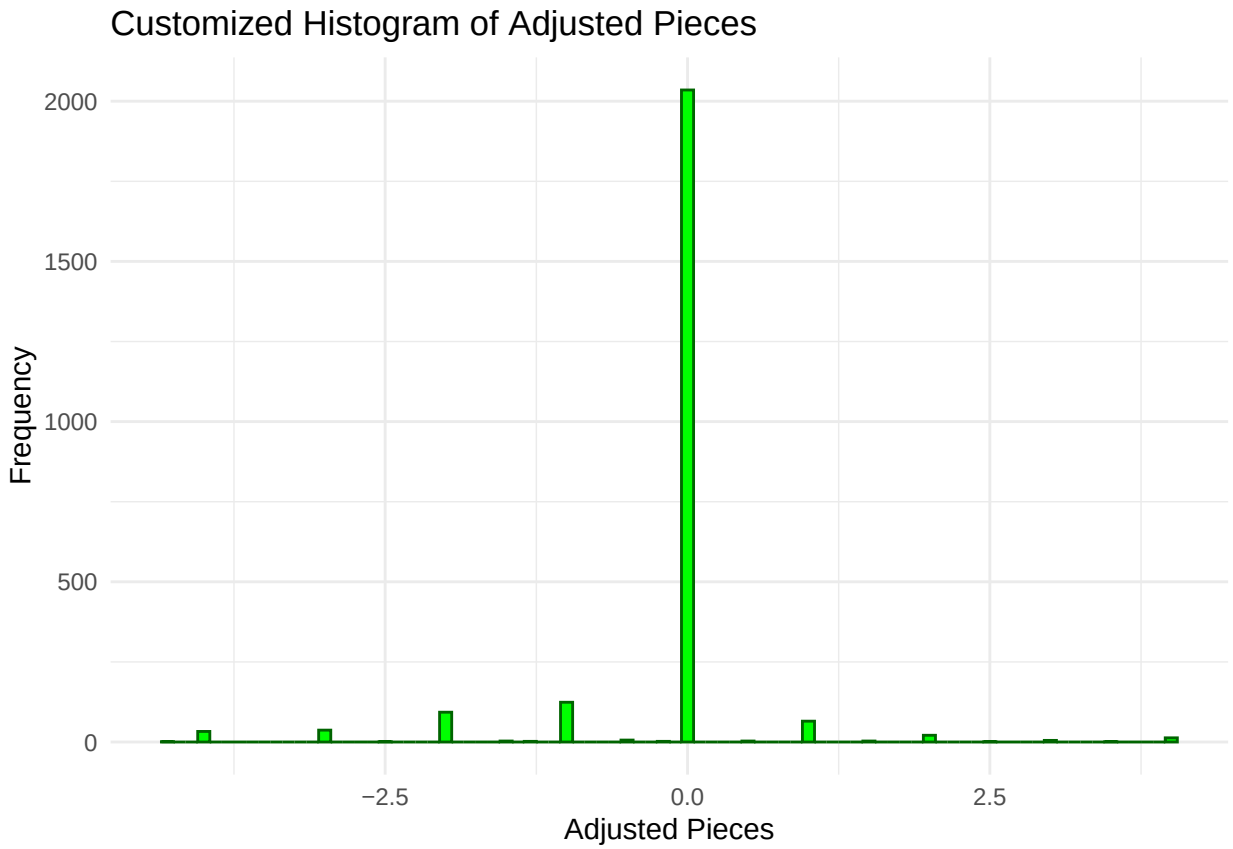
```
##
## -4.3   -4   -3 -2.5   -2 -1.5 -1.3   -1 -0.5 -0.2    0  0.5    1  1.5    2  2.5
##    1   33   37    1   93    3    2  124    6    2 2035    3   65    3   21    1
##    3  3.5    4
##    5    1   13
```

```
# gives us a quick glance on how many non-missing values are in adjusted pieces
sum(adjustment_table)
```

```
## [1] 2449
```

```
# histogram of adjusted pieces to visualize this distribution
ggplot(fabrics, aes(x = `Adjusted #Pcs`)) +
  geom_histogram(binwidth = 0.1, fill = "green", color = "darkgreen") +
  labs(title = "Customized Histogram of Adjusted Pieces", x = "Adjusted Pieces", y = "Frequency") +
  theme_minimal()
```

```
## Warning: Removed 245 rows containing non-finite values (`stat_bin()`).
```



Upon first looks we can see that a majority of adjustment logs turn out to be 0, meaning inventory has not changed. This distribution is overwhelmingly skewed toward 0, with 2035 out of 2449 non-missing adjustment logs being 0, or around 83%

Digging deeper a few questions about the dataset arise:

1. Can there be non audits with adjustments?

```
fabrics %>%
  filter(is.na(`System Entry Date`) & is.na(`System User`) & !is.na(`Adjusted #Pcs`))
```

```
## # A tibble: 1 x 11
##   `System Entry Date` `System User` `Lot#`  `Received Date` `Style#` `Desc#`
##   <dttm>              <chr>         <chr>              <dbl> <chr>    <chr>
## 1 NA                  <NA>          Lot2519            44708 Style150 Descriptio~
## # i 5 more variables: `Color#` <chr>, `Received Pcs` <dbl>,
## #   `Initial Stock Pcs` <dbl>, `Adjusted #Pcs` <dbl>, `Final Count Pcs` <dbl>
```

It seems that our assumption holds true; adjustments are made when an audit log is made. There is a singular case that differs, but this could either be a system error or a manual log error.

2. Can there be missing adjustment with an audit? What does this mean?

```
fabrics %>%
  filter((!is.na(`System Entry Date`) | !is.na(`System User`)) & is.na(`Adjusted #Pcs`))
```

```
## # A tibble: 20 x 11
##    `System Entry Date` `System User` `Lot#`  `Received Date` `Style#` `Desc#`
##    <dttm>              <chr>         <chr>              <dbl> <chr>    <chr>
##  1 2022-05-31 00:00:00 Don           Lot949             44331 Style2   Descripti~
##  2 2022-05-27 00:00:00 Dan           Lot2228            44590 Style16  Descripti~
##  3 2022-05-30 00:00:00 Dan           Lot1045            44524 Style20  Descripti~
##  4 2022-05-27 00:00:00 Dan           Lot1189            44642 Style28  Descripti~
##  5 2022-05-26 00:00:00 Don           Lot1291            44677 Style36  Descripti~
##  6 2022-05-31 00:00:00 Dan           Lot1499            44683 Style36  Descripti~
##  7 2022-05-27 00:00:00 Dan           Lot2429            44651 Style36  Descripti~
##  8 2022-05-27 00:00:00 Jill          Lot1910            44645 Style48  Descripti~
##  9 2022-05-25 00:00:00 John          Lot2406            44678 Style55  Descripti~
## 10 2022-05-27 00:00:00 Jill          Lot1795            44481 Style114 Descripti~
## 11 2022-05-27 00:00:00 Jill          Lot1949            44666 Style119 Descripti~
## 12 2022-05-27 00:00:00 Dan           Lot1441            44649 Style123 Descripti~
## 13 2022-05-27 00:00:00 John          Lot1714            44363 Style132 Descripti~
## 14 2022-05-25 00:00:00 Don           Lot2286            44586 Style142 Descripti~
## 15 2022-05-31 00:00:00 John          Lot2473            44270 Style150 Descripti~
## 16 2022-05-30 00:00:00 Dan           Lot2485            44642 Style150 Descripti~
## 17 2022-05-27 00:00:00 Dan           Lot1860            44652 Style200 Descripti~
## 18 2022-05-25 00:00:00 Don           Lot863             44390 Style252 Descripti~
## 19 2022-05-25 00:00:00 Don           Lot84              44390 Style253 Descripti~
## 20 2022-06-02 00:00:00 Don           Lot1782               NA <NA>     <NA>
## # i 5 more variables: `Color#` <chr>, `Received Pcs` <dbl>,
## #   `Initial Stock Pcs` <dbl>, `Adjusted #Pcs` <dbl>, `Final Count Pcs` <dbl>
```

It appears as though there are cases of audited logs containing missing adjustments, though few and far in between, fortunately For the sake of thoroughness and simplicity, we will assume that these lots are still considered audited and that these missing values mean no adjustments made (0). Realistically, this is also probably due to either a system or manual log error.

Now with these questions addressed we have no issues finding our adjustment totals and insights:

```
adjusted_audits <- fabrics %>%
  filter((!is.na(`System Entry Date`) | !is.na(`System User`)) & `Adjusted #Pcs` != 0 )

dim(adjusted_audits)
```

```
## [1] 414  11
```

We find that 414 of our audits have nonzero adjustments, meaning that around 8.7% of audited logs see some sort of adjustments.

```
negative_adjust <- adjusted_audits %>%
  filter(`Adjusted #Pcs` < 0) %>%
  select(`Adjusted #Pcs`)

positive_adjust <- adjusted_audits %>%
  filter(`Adjusted #Pcs` > 0) %>%
  select(`Adjusted #Pcs`)

paste('There are', dim(negative_adjust)[1], 'logs with negative adjustments.')
```

```
## [1] "There are 302 logs with negative adjustments."
```

```
paste('A total of', abs(sum(negative_adjust)), 'pieces have been removed from inventory.')
```

```
## [1] "A total of 570.3 pieces have been removed from inventory."
```

```
paste('There are', dim(positive_adjust)[1], 'logs with positive adjustments.')
```

```
## [1] "There are 112 logs with positive adjustments."
```

```
paste('A total of', sum(positive_adjust), 'pieces have been added' )
```

```
## [1] "A total of 186 pieces have been added"
```

```
paste('This means that we have a net loss of', abs(sum(negative_adjust)) - sum(positive_adjust), 'pieces
```

```
## [1] "This means that we have a net loss of 384.3 pieces of fabric"
```

This also means that of the 414 audit logs with any sort of nonzero adjustment, 72.9% are negative adjustments whilst 27.1% are positive, with inventory losing 2.7 times as much fabric as it gains.