Shopify_Challenge

Ankita Sarkar

10/21/2021

Reading the data set:

```
file <- read.csv('Sheet1.csv')
```

Creating a data frame to work with it

```
df <- data.frame(file)
```

Getting the summary of the data

```
summary(df)
```

```
##     order_id        shop_id          user_id        order_amount
## Min.   :  1    Min.   :  1.00   Min.   :607.0    Min.   :    90
## 1st Qu.:1251   1st Qu.: 24.00   1st Qu.:775.0    1st Qu.:   163
## Median :2500   Median : 50.00   Median :849.0    Median :   284
## Mean   :2500   Mean   : 50.08   Mean   :849.1    Mean   :  3145
## 3rd Qu.:3750   3rd Qu.: 75.00   3rd Qu.:925.0    3rd Qu.:   390
## Max.   :5000   Max.   :100.00   Max.   :999.0    Max.   :704000
##   total_items        payment_method       created_at
## Min.   :   1.000   Length:5000         Length:5000
## 1st Qu.:   1.000   Class :character    Class :character
## Median :   2.000   Mode  :character    Mode  :character
## Mean   :   8.787
## 3rd Qu.:   3.000
## Max.   :2000.000
```

Checking the minimum and maximum value for order_amount

```
max(df$order_amount)
```
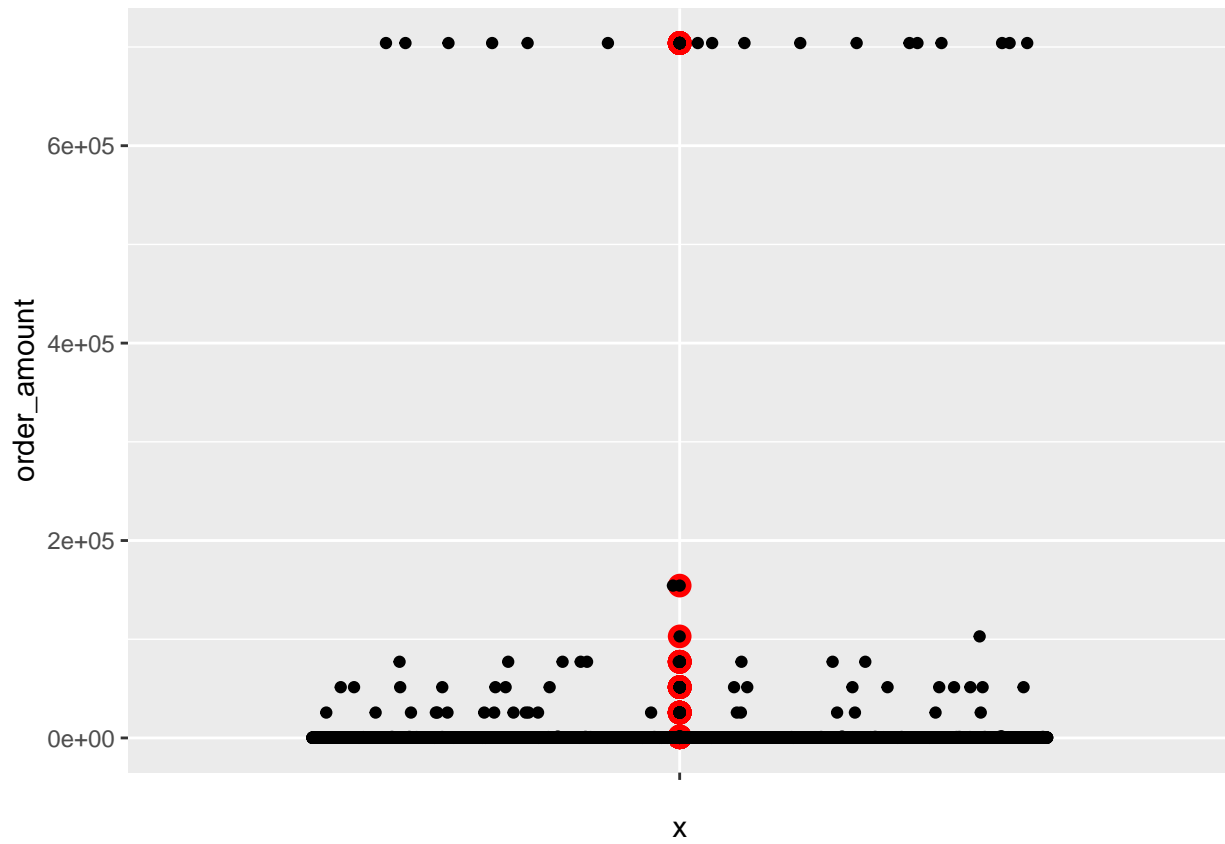
```
## [1] 704000
```

```
min(df$order_amount)
```

```
## [1] 90
```

Plotting a boxplot for the data

```
library(ggplot2)

ggplot(data = df) +
  aes(x = '', y = order_amount) +
  geom_boxplot(outlier.colour = 'red', outlier.size = 3.5, colour = 'green') +
  geom_point() +
  geom_jitter()
```


We have shown the boxplot as well as the all the orders of sneakers being sold. Clearly outliers present in the order_amount(shown in red).

Printing out the entries containing outliers

```
out <- boxplot.stats(df$order_amount)$out
out_ind <- which(df$order_amount %in% c(out))
#out_ind

out_df <- df[out_ind, ]
out_df
```

```
##        order_id shop_id user_id order_amount total_items payment_method
## 16          16      42     607       704000        2000    credit_card
## 61          61      42     607       704000        2000    credit_card
## 100        100      18     752          780           5           cash
## 137        137      15     961          765           5    credit_card
## 161        161      78     990        25725           1    credit_card
```

```
## 220    220    23   884     780        5  credit_card
## 223    223    96   829     765        5        debit
## 260    260    99   963     780        4  credit_card
## 265    265    16   950     780        5  credit_card
## 491    491    78   936   51450        2        debit
## 494    494    78   983   51450        2         cash
## 512    512    78   967   51450        2         cash
## 521    521    42   607  704000     2000  credit_card
## 523    523    46   761     830        5  credit_card
## 618    618    78   760   51450        2         cash
## 652    652    51   987     748        4  credit_card
## 692    692    78   878  154350        6        debit
## 738    738    50   873     772        4        debit
## 743    743    12   727     804        4         cash
## 772    772    19   818     815        5        debit
## 880    880    60   870     885        5        debit
## 939    939    42   808    1056        3  credit_card
## 995    995    89   723     784        4         cash
## 1057  1057    78   800   25725        1        debit
## 1105  1105    42   607  704000     2000  credit_card
## 1124  1124    29   911     815        5  credit_card
## 1151  1151    82   853     885        5        debit
## 1194  1194    78   944   25725        1        debit
## 1205  1205    78   970   25725        1  credit_card
## 1257  1257     6   942     935        5  credit_card
## 1260  1260    78   775   77175        3  credit_card
## 1363  1363    42   607  704000     2000  credit_card
## 1365  1365    42   797    1760        5         cash
## 1368  1368    42   926    1408        4         cash
## 1385  1385    78   867   25725        1         cash
## 1420  1420    78   912   25725        1         cash
## 1437  1437    42   607  704000     2000  credit_card
## 1453  1453    78   812   25725        1  credit_card
## 1472  1472    42   907    1408        4        debit
## 1485  1485    96   702     765        5  credit_card
## 1504  1504    11   904     736        4        debit
## 1530  1530    78   810   51450        2         cash
## 1563  1563    42   607  704000     2000  credit_card
## 1564  1564    91   934     960        6        debit
## 1603  1603    42   607  704000     2000  credit_card
## 1629  1629    91   982     800        5        debit
## 1765  1765    12   789     804        4        debit
## 1924  1924    91   880     800        5  credit_card
## 1947  1947    33   866     865        5         cash
## 1949  1949    87   801     745        5  credit_card
## 1963  1963    46   879     830        5        debit
## 2033  2033    88   798     880        5         cash
## 2040  2040    11   756     920        5  credit_card
## 2044  2044    84   721     765        5  credit_card
## 2128  2128    83   745     774        6         cash
## 2137  2137    61   931     790        5        debit
## 2141  2141    89   767     784        4         cash
## 2154  2154    42   607  704000     2000  credit_card
## 2271  2271    78   855   25725        1  credit_card
```

```
## 2298   2298   42   607   704000   2000   credit_card
## 2308   2308   61   723      948      6   credit_card
## 2354   2354   27   811      845      5          cash
## 2387   2387   38   847      760      4          cash
## 2390   2390   87   754      745      5         debit
## 2453   2453   78   709    51450      2          cash
## 2493   2493   78   834   102900      4         debit
## 2495   2495   50   757      965      5         debit
## 2496   2496   78   707    51450      2          cash
## 2513   2513   78   935    51450      2         debit
## 2549   2549   78   861    25725      1          cash
## 2561   2561    6   845      935      5   credit_card
## 2565   2565   78   915    77175      3         debit
## 2671   2671   18   740      780      5         debit
## 2691   2691   78   962    77175      3         debit
## 2758   2758   66   772      805      5   credit_card
## 2774   2774   78   890    25725      1          cash
## 2819   2819   78   869    51450      2         debit
## 2822   2822   78   814    51450      2          cash
## 2836   2836   42   607   704000   2000   credit_card
## 2907   2907   78   817    77175      3         debit
## 2923   2923   78   740    25725      1         debit
## 2968   2968   46   774      830      5         debit
## 2970   2970   42   607   704000   2000   credit_card
## 2988   2988   42   819     1056      3          cash
## 3074   3074   90   877      890      5         debit
## 3078   3078   89   754      980      5         debit
## 3086   3086   78   910    25725      1          cash
## 3102   3102   78   855    51450      2   credit_card
## 3118   3118   38   794      760      4   credit_card
## 3152   3152   78   745    25725      1   credit_card
## 3168   3168   78   927    51450      2          cash
## 3203   3203    6   841      748      4   credit_card
## 3253   3253   67   706      786      6   credit_card
## 3333   3333   42   607   704000   2000   credit_card
## 3404   3404   78   928    77175      3         debit
## 3429   3429   11   922      736      4          cash
## 3439   3439   66   842      805      5   credit_card
## 3441   3441   78   982    25725      1         debit
## 3514   3514   42   726     1056      3         debit
## 3518   3518   11   732      736      4         debit
## 3533   3533   51   828      935      5          cash
## 3539   3539   43   830     1086      6         debit
## 3610   3610   11   901      736      4         debit
## 3706   3706   78   828    51450      2   credit_card
## 3725   3725   78   766    77175      3   credit_card
## 3781   3781   78   889    25725      1          cash
## 3866   3866   68   815      816      6         debit
## 3928   3928   97   979      810      5   credit_card
## 3967   3967   10   976      740      5         debit
## 4041   4041   78   852    25725      1          cash
## 4057   4057   42   607   704000   2000   credit_card
## 4080   4080   78   946    51450      2          cash
## 4142   4142   54   733     1064      8         debit
```

```
## 4193   4193   78   787   77175      3   credit_card
## 4296   4296   99   760     780      4          cash
## 4312   4312   78   960   51450      2         debit
## 4413   4413   78   756   51450      2         debit
## 4421   4421   78   969   77175      3         debit
## 4491   4491   57   941     735      5   credit_card
## 4506   4506   78   866   25725      1         debit
## 4513   4513   38   848     760      4         debit
## 4524   4524   26   995     880      5   credit_card
## 4555   4555   18   833     780      5         debit
## 4575   4575    6   922     748      4         debit
## 4581   4581   51   799     748      4          cash
## 4585   4585   78   997   25725      1          cash
## 4597   4597    6   846     748      4          cash
## 4620   4620   91   926     800      5   credit_card
## 4647   4647   42   607  704000   2000   credit_card
## 4712   4712   86   883     780      6          cash
## 4716   4716   78   818   77175      3         debit
## 4848   4848   13   993     960      6          cash
## 4869   4869   42   607  704000   2000   credit_card
## 4871   4871   61   958     790      5   credit_card
## 4883   4883   42   607  704000   2000   credit_card
## 4906   4906   38   748     760      4         debit
## 4919   4919   78   823   25725      1          cash
## 4928   4928   74   849     765      5          cash
## 4953   4953   26   786     880      5          cash
## 4959   4959   70   711     865      5   credit_card
## 4981   4981   50   942     772      4   credit_card
##                  created_at
## 16    2017-03-07 4:00:00
## 61    2017-03-04 4:00:00
## 100   2017-03-06 23:41:16
## 137   2017-03-26 5:06:46
## 161   2017-03-12 5:56:57
## 220   2017-03-03 20:11:05
## 223   2017-03-13 15:16:25
## 260   2017-03-25 11:41:01
## 265   2017-03-24 7:32:42
## 491   2017-03-26 17:08:19
## 494   2017-03-16 21:39:35
## 512   2017-03-09 7:23:14
## 521   2017-03-02 4:00:00
## 523   2017-03-26 19:07:51
## 618   2017-03-18 11:18:42
## 652   2017-03-17 10:33:12
## 692   2017-03-27 22:51:43
## 738   2017-03-02 21:26:04
## 743   2017-03-14 16:38:01
## 772   2017-03-07 8:48:16
## 880   2017-03-27 20:15:11
## 939   2017-03-13 23:43:45
## 995   2017-03-05 9:05:23
## 1057 2017-03-15 10:16:45
## 1105  2017-03-24 4:00:00
```

```
## 1124  2017-03-26 0:53:49
## 1151 2017-03-24 20:47:47
## 1194 2017-03-16 16:38:26
## 1205 2017-03-17 22:32:21
## 1257 2017-03-12 19:49:08
## 1260  2017-03-27 9:27:20
## 1363  2017-03-15 4:00:00
## 1365  2017-03-10 6:28:21
## 1368  2017-03-13 2:38:34
## 1385 2017-03-17 16:38:06
## 1420 2017-03-30 12:23:43
## 1437  2017-03-11 4:00:00
## 1453 2017-03-17 18:09:54
## 1472 2017-03-12 23:00:22
## 1485 2017-03-03 20:53:28
## 1504 2017-03-06 23:08:25
## 1530  2017-03-29 7:12:01
## 1563  2017-03-19 4:00:00
## 1564  2017-03-23 8:25:49
## 1603  2017-03-17 4:00:00
## 1629 2017-03-10 20:05:33
## 1765  2017-03-03 3:10:50
## 1924 2017-03-08 10:58:01
## 1947  2017-03-14 5:05:37
## 1949  2017-03-12 4:35:06
## 1963 2017-03-14 17:11:01
## 2033  2017-03-18 4:24:14
## 2040 2017-03-04 10:51:41
## 2044  2017-03-13 2:18:47
## 2128  2017-03-27 6:59:46
## 2137  2017-03-07 8:12:31
## 2141 2017-03-05 23:52:45
## 2154  2017-03-12 4:00:00
## 2271 2017-03-14 23:58:22
## 2298  2017-03-07 4:00:00
## 2308 2017-03-26 11:29:37
## 2354  2017-03-13 7:07:39
## 2387  2017-03-25 9:48:36
## 2390 2017-03-10 21:07:31
## 2453 2017-03-27 11:04:04
## 2493  2017-03-04 4:37:34
## 2495  2017-03-04 7:32:45
## 2496  2017-03-26 4:38:52
## 2513 2017-03-18 18:57:13
## 2549 2017-03-17 19:36:00
## 2561 2017-03-16 22:24:30
## 2565  2017-03-25 1:19:35
## 2671 2017-03-12 22:44:22
## 2691  2017-03-22 7:33:25
## 2758  2017-03-14 8:43:29
## 2774 2017-03-26 10:36:43
## 2819  2017-03-17 6:25:51
## 2822 2017-03-02 17:13:25
## 2836  2017-03-28 4:00:00
```

```
## 2907  2017-03-16  3:45:46
## 2923 2017-03-12 20:10:58
## 2968  2017-03-23  9:22:12
## 2970  2017-03-28  4:00:00
## 2988  2017-03-03  9:09:25
## 3074  2017-03-26  8:08:27
## 3078  2017-03-13  5:27:58
## 3086  2017-03-26  1:59:27
## 3102  2017-03-21  5:10:34
## 3118  2017-03-04  7:58:59
## 3152 2017-03-18 13:13:07
## 3168 2017-03-12 12:23:08
## 3203  2017-03-26  8:24:52
## 3253 2017-03-29 16:05:41
## 3333  2017-03-24  4:00:00
## 3404  2017-03-16  9:45:05
## 3429 2017-03-04 21:14:40
## 3439 2017-03-22 17:58:37
## 3441 2017-03-19 19:02:54
## 3514 2017-03-24 17:51:05
## 3518  2017-03-20  8:36:02
## 3533 2017-03-17 16:05:35
## 3539 2017-03-17 19:56:29
## 3610  2017-03-08  2:21:55
## 3706 2017-03-14 20:43:15
## 3725 2017-03-16 14:13:26
## 3781 2017-03-11 21:14:50
## 3866  2017-03-11  9:31:50
## 3928  2017-03-11  7:37:13
## 3967 2017-03-13 17:09:41
## 4041 2017-03-02 14:31:12
## 4057  2017-03-28  4:00:00
## 4080 2017-03-20 21:14:00
## 4142 2017-03-07 17:05:18
## 4193  2017-03-18  9:25:32
## 4296 2017-03-16 21:22:47
## 4312  2017-03-01  3:02:10
## 4413  2017-03-02  4:13:39
## 4421 2017-03-09 15:21:35
## 4491  2017-03-20  8:55:46
## 4506 2017-03-22 22:06:01
## 4513  2017-03-27  7:18:04
## 4524  2017-03-09  8:28:31
## 4555  2017-03-22  8:16:22
## 4575 2017-03-10 11:17:39
## 4581  2017-03-02  5:45:50
## 4585 2017-03-25 21:48:44
## 4597  2017-03-04  5:57:05
## 4620  2017-03-01  5:10:03
## 4647  2017-03-02  4:00:00
## 4712 2017-03-18 14:18:19
## 4716  2017-03-05  5:10:44
## 4848 2017-03-27 11:00:45
## 4869  2017-03-22  4:00:00
```

```
## 4871 2017-03-02 23:46:58
## 4883  2017-03-25 4:00:00
## 4906 2017-03-30 17:21:52
## 4919 2017-03-15 13:26:46
## 4928  2017-03-28 6:55:57
## 4953  2017-03-17 1:50:18
## 4959 2017-03-08 17:22:51
## 4981  2017-03-23 5:15:19
```

We can use median of the order_amount to correct this problem since the median does not get affected by the presense of outliers in data

```
median(df$order_amount)
```

```
## [1] 284
```

We see the median for order_amount is 284. One possible solution could be to change all the rows holding excessive large amount with this median and process again.

```
df[out_ind, ] <- median(df$order_amount)
```

If we check the summary of the data now, it seems to solve our problem

```
summary(df)
```

```
##      order_id        shop_id          user_id        order_amount
##   Min.   :   1   Min.   :  1.00   Min.   :284.0   Min.   :  90.0
##   1st Qu.:1136   1st Qu.: 25.00   1st Qu.:769.0   1st Qu.:163.0
##   Median :2424   Median : 52.00   Median :847.0   Median :284.0
##   Mean   :2435   Mean   : 56.46   Mean   :833.9   Mean   :293.4
##   3rd Qu.:3713   3rd Qu.: 77.00   3rd Qu.:923.0   3rd Qu.:374.0
##   Max.   :5000   Max.   :284.00   Max.   :999.0   Max.   :730.0
##    total_items     payment_method       created_at
##   Min.   :  1.000   Length:5000         Length:5000
##   1st Qu.:  1.000   Class :character    Class :character
##   Median :  2.000   Mode  :character    Mode  :character
##   Mean   :  9.904
##   3rd Qu.:  3.000
##   Max.   :284.000
```