

625 Trend

Longfei Zhang

12/5/2021

```
mydata <- read_csv('C:/Users/ZLF/Documents/Iowa_Data.csv', show_col_types = FALSE)
```

```
## New names:  
## * ' -> ...1
```

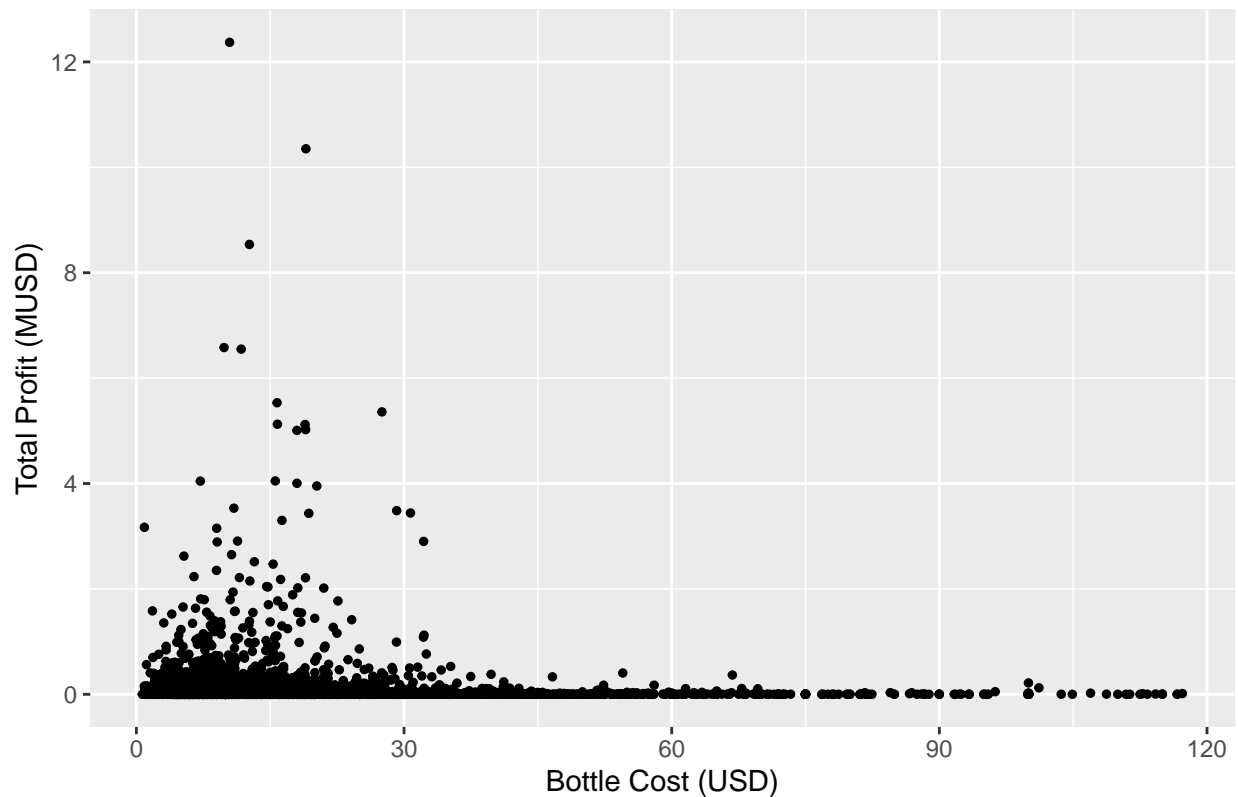
```
mydata$mydate <- unclass(as.Date(mydata$Date, format = '%m/%d/%Y'))  
mydata$mydate <- mydata$mydate-min(mydata$mydate)+1
```

```
mydata <- mydata%>%  
  group_by(Item.Number) %>%  
  mutate(State.Bottle.Retail=mean(State.Bottle.Retail)) %>%  
  mutate(State.Bottle.Cost=mean(State.Bottle.Cost)) %>%  
  mutate(Profit=(State.Bottle.Retail-State.Bottle.Cost)*Bottles.Sold)
```

```
total_profit <- mydata %>% group_by(Item.Number) %>%  
  mutate(tol_profit=sum(Profit)/1e6) %>%  
  mutate(botl_profit=State.Bottle.Retail-State.Bottle.Cost) %>%  
  distinct(Item.Number, tol_profit, botl_profit, State.Bottle.Cost) %>%  
  filter(botl_profit>0) %>%  
  filter(botl_profit<60) %>%  
  arrange(desc(tol_profit), desc(botl_profit))
```

```
p_plot <- ggplot(total_profit, aes(State.Bottle.Cost, tol_profit)) + geom_point(size = 1) + labs(title =  
p_plot
```

Total Profit Plot



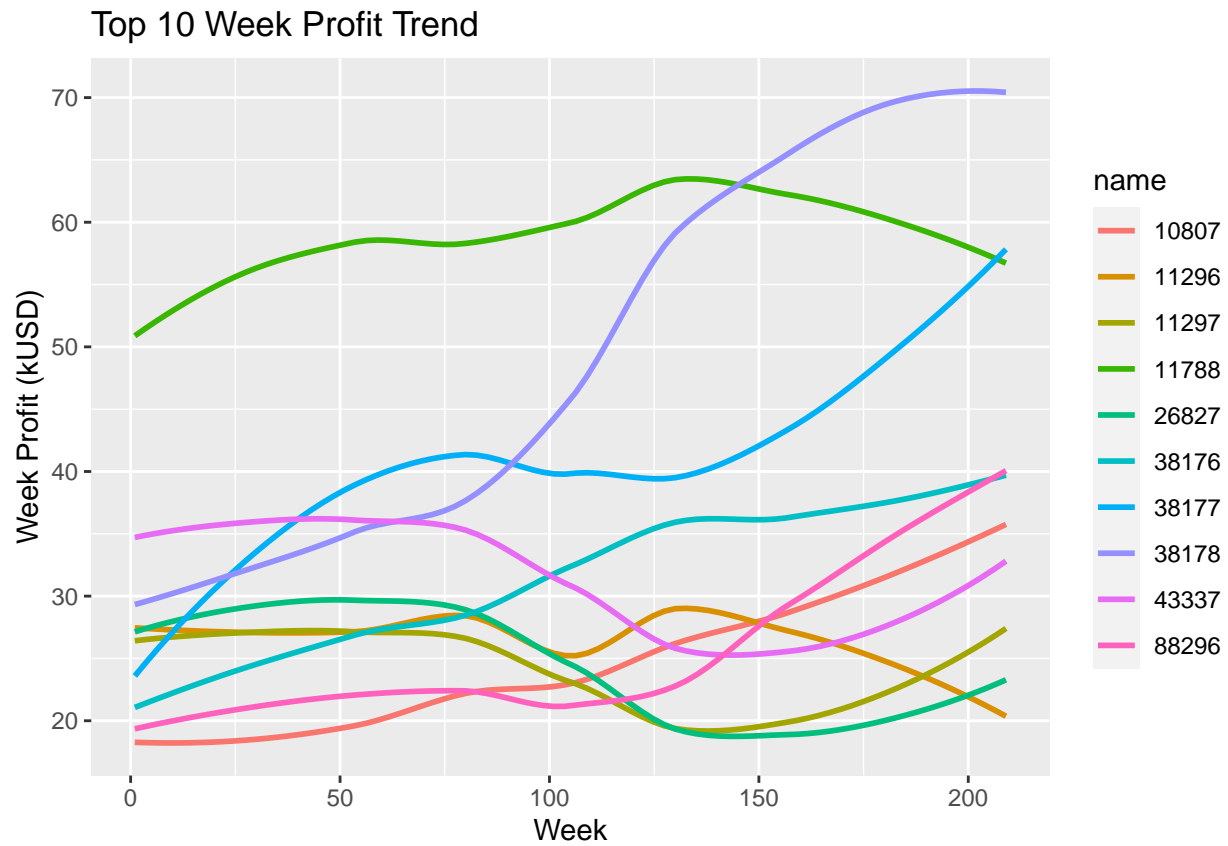
```
top10 <- total_profit[1:10, ]
```

```
profit_date <- mydata %>% group_by(Item.Number, mydate) %>%
  mutate(p=sum(Profit)) %>%
  distinct(Item.Number, mydate, p)
```

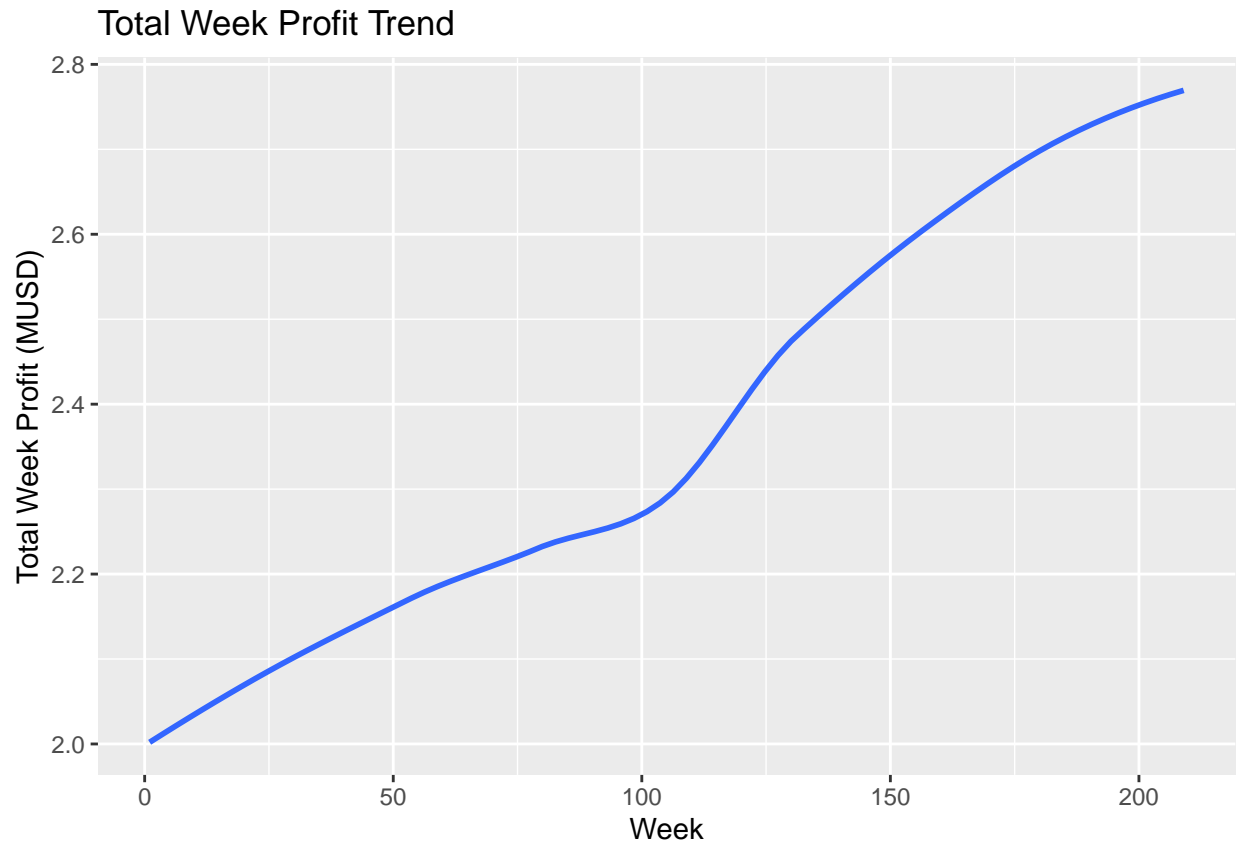
```
trend <- matrix(0, nrow = 209, ncol = 11)
trend[, 1] <- 1:209
colnames(trend) <- c('week', top10$Item.Number)
for(i in 1:10){
  tmp <- profit_date %>%
    filter(Item.Number==top10$Item.Number[i]) %>%
    ungroup %>%
    mutate(week=rep(1:209, each=7)[mydate]) %>%
    add_row(week = 1:209, p = 0) %>%
    group_by(week) %>%
    mutate(week_p = sum(p)/1e3) %>%
    distinct(week, week_p) %>%
    arrange(week)
  trend[, i+1] <- as.vector(tmp$week_p)
}
```

```
trend.df <- as.data.frame(trend)
trend.df <- trend.df %>% pivot_longer(-week)
```

```
trend_plot <- ggplot(trend.df) + geom_smooth(aes(week,value,colour=name), method = 'loess', se = FALSE,
trend_plot
```



```
tot.trend <- profit_date %>%
  ungroup %>%
  mutate(week=rep(1:209, each=7)[mydate]) %>%
  add_row(week = 1:209, p = 0) %>%
  group_by(week) %>%
  mutate(week_p = sum(p)/1e6) %>%
  distinct(week, week_p) %>%
  arrange(week)
tot.trend_plot <- ggplot(tot.trend) + geom_smooth(aes(week, week_p), method = 'loess', se = FALSE, form
tot.trend_plot
```



```
tmp <- mydata %>%
  distinct(Category, Item.Number) %>%
  group_by(Category) %>%
  count(Category)
tmp <- tmp[order(tmp$n, decreasing = T), ]
tmp
```

```
## # A tibble: 62 x 2
## # Groups:   Category [62]
##   Category      n
##   <dbl> <int>
## 1 1901200    732
## 2 1701100    615
## 3 1011200    499
## 4 1022200    380
## 5 1031200    367
## 6 1031100    307
## 7 1011100    267
## 8 1012300    235
## 9 1071100    225
## 10 1081300    214
## # ... with 52 more rows
```

```
sales_week <- mydata %>%
  group_by(mydate) %>%
```

```

mutate(p=sum(Sale..Dollars.)) %>%
distinct(mydate, p) %>%
ungroup %>%
mutate(week=rep(1:209, each=7)[mydate]) %>%
add_row(week = 1:209, p = 0) %>%
group_by(week) %>%
mutate(week_p = sum(p)/1e6) %>%
distinct(week, week_p) %>%
arrange(week)

sales_month <- mydata %>%
  group_by(mydate) %>%
  mutate(p=sum(Sale..Dollars.)) %>%
  distinct(mydate, p) %>%
  ungroup %>%
  mutate(month=rep(1:49, each=30)[mydate]) %>%
  add_row(month = 1:49, p = 0) %>%
  group_by(month) %>%
  mutate(month_p = sum(p)/1e6) %>%
  distinct(month, month_p) %>%
  arrange(month)

```

```

myweek <- 1:104
meanX <- mean(myweek)
SSX <- sum((myweek-meanX)^2)

mylm <- function(sale, date){
  sale1 <- sale[date<105]
  date1 <- date[date<105]
  b1 <- sum(sale1*(date1-meanX))/SSX

  sale2 <- sale[date>104]
  date2 <- date[date>104]
  b2 <- sum(sale2*(date2-104-meanX))/SSX

  return(b2-b1)
}

```

```

Covid_index <- mydata %>%
  group_by(Item.Number) %>%
  mutate(tot_sale = sum(Bottles.Sold)) %>%
  filter(tot_sale>=730) %>%
  mutate(week = rep(1:209, each=7)[mydate]) %>%
  filter(week < 209) %>%
  ungroup %>%
  group_by(Item.Number, week) %>%
  mutate(week_s = sum(Bottles.Sold)) %>%
  ungroup %>%
  distinct(Item.Number, week, week_s) %>%
  group_by(Item.Number) %>%
  mutate(Covid_index = mylm(week_s, week)) %>%
  distinct(Item.Number, Covid_index)

```

```
cl <- makeCluster(20)
stopCluster(cl)
```