

# LBB Kafani Irawan

Kafani Irawan

2022-04-27

## Contents

R Markdown . . . . .	1
Input Data . . . . .	1
Data Inspection . . . . .	1
Data Cleansing and Coertions . . . . .	3

```
tinytex::install_tinytex()  
# to uninstall TinyTeX, run  
# tinytex::uninstall_tinytex()
```

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

Hi! Welcome to my Rmd :) in this LBB i will use previous data which is Retail.csv I hope you enjoy it !

## Input Data

```
retail <- read.csv("retail.csv")
```

## Data Inspection

```
head(retail)
```

##	Row.ID	Order.ID	Order.Date	Ship.Date	Ship.Mode	Customer.ID
## 1	1	CA-2016-152156	11/8/16	11/11/16	Second Class	CG-12520
## 2	2	CA-2016-152156	11/8/16	11/11/16	Second Class	CG-12520
## 3	3	CA-2016-138688	6/12/16	6/16/16	Second Class	DV-13045
## 4	4	US-2015-108966	10/11/15	10/18/15	Standard Class	S0-20335
## 5	5	US-2015-108966	10/11/15	10/18/15	Standard Class	S0-20335
## 6	6	CA-2014-115812	6/9/14	6/14/14	Standard Class	BH-11710
##	Segment	Product.ID	Category	Sub.Category		

## 1	Consumer	FUR-BO-10001798	Furniture	Bookcases
## 2	Consumer	FUR-CH-10000454	Furniture	Chairs
## 3	Corporate	OFF-LA-10000240	Office Supplies	Labels
## 4	Consumer	FUR-TA-10000577	Furniture	Tables
## 5	Consumer	OFF-ST-10000760	Office Supplies	Storage
## 6	Consumer	FUR-FU-10001487	Furniture	Furnishings
##			Product.Name	Sales
## 1			Bush Somerset Collection Bookcase	261.9600
## 2			Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back	731.9400
## 3			Self-Adhesive Address Labels for Typewriters by Universal	14.6200
## 4			Bretford CR4500 Series Slim Rectangular Table	957.5775
## 5			Eldon Fold 'N Roll Cart System	22.3680
## 6			Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood	48.8600
##	Quantity	Discount	Profit	
## 1	2	0.00	41.9136	
## 2	3	0.00	219.5820	
## 3	2	0.00	6.8714	
## 4	5	0.45	-383.0310	
## 5	2	0.20	2.5164	
## 6	7	0.00	14.1694	

tail(etail)

##	Row.ID	Order.ID	Order.Date	Ship.Date	Ship.Mode	Customer.ID
## 9989	9989	CA-2017-163629	11/17/17	11/21/17	Standard Class	RA-19885
## 9990	9990	CA-2014-110422	1/21/14	1/23/14	Second Class	TB-21400
## 9991	9991	CA-2017-121258	2/26/17	3/3/17	Standard Class	DB-13060
## 9992	9992	CA-2017-121258	2/26/17	3/3/17	Standard Class	DB-13060
## 9993	9993	CA-2017-121258	2/26/17	3/3/17	Standard Class	DB-13060
## 9994	9994	CA-2017-119914	5/4/17	5/9/17	Second Class	CC-12220
##	Segment	Product.ID	Category	Sub.Category		
## 9989	Corporate	TEC-PH-10004006	Technology	Phones		
## 9990	Consumer	FUR-FU-10001889	Furniture	Furnishings		
## 9991	Consumer	FUR-FU-10000747	Furniture	Furnishings		
## 9992	Consumer	TEC-PH-10003645	Technology	Phones		
## 9993	Consumer	OFF-PA-10004041	Office Supplies	Paper		
## 9994	Consumer	OFF-AP-10002684	Office Supplies	Appliances		
##					Product.Name	
## 9989					Panasonic KX - TS880B Telephone	
## 9990					Ultra Door Pull Handle	
## 9991					Tenex B1-RE Series Chair Mats for Low Pile Carpets	
## 9992					Aastra 57i VoIP phone	
## 9993					It's Hot Message Books with Stickers, 2 3/4" x 5"	
## 9994					Acco 7-Outlet Masterpiece Power Center, Wihtout Fax/Phone Line Protection	
##	Sales	Quantity	Discount	Profit		
## 9989	206.100	5	0.0	55.6470		
## 9990	25.248	3	0.2	4.1028		
## 9991	91.960	2	0.0	15.6332		
## 9992	258.576	2	0.2	19.3932		
## 9993	29.600	4	0.0	13.3200		
## 9994	243.160	2	0.0	72.9480		

```
dim(retail)
```

```
## [1] 9994 15
```

```
names(retail)
```

```
## [1] "Row.ID"      "Order.ID"    "Order.Date"  "Ship.Date"   "Ship.Mode"
## [6] "Customer.ID" "Segment"     "Product.ID"  "Category"    "Sub.Category"
## [11] "Product.Name" "Sales"       "Quantity"    "Discount"    "Profit"
```

kesimpulan: 1. data retail terdiri dari 9994 baris dan 15 kolom 2. nama dari masing-masing kolom yaitu [1] "Row.ID" "Order.ID" "Order.Date" "Ship.Date" [5] "Ship.Mode" "Customer.ID" "Segment" "Product.ID" [9] "Category" "Sub.Category" "Product.Name" "Sales" [13] "Quantity" "Discount" "Profit"

## Data Cleansing and Coertions

### Check for Missing Value

```
colSums(is.na(retail))
```

```
##      Row.ID      Order.ID      Order.Date      Ship.Date      Ship.Mode      Customer.ID
##          0          0          0          0          0          0
##      Segment      Product.ID      Category      Sub.Category      Product.Name      Sales
##          0          0          0          0          0          0
##      Quantity      Discount      Profit
##          0          0          0
```

```
anyNA(retail)
```

```
## [1] FALSE
```

Kesimpulan: 1. Tidak ada missing value pada data retail.csv dengan pengecekan menggunakan colsums 2. Tidak ada missing value pada data retail.csv dengan pengecekan menggunakan anyNA

### Manipulating Data

```
head(retail)
```

```
##      Row.ID      Order.ID      Order.Date      Ship.Date      Ship.Mode      Customer.ID
## 1      1 CA-2016-152156      11/8/16      11/11/16      Second Class      CG-12520
## 2      2 CA-2016-152156      11/8/16      11/11/16      Second Class      CG-12520
## 3      3 CA-2016-138688      6/12/16      6/16/16      Second Class      DV-13045
## 4      4 US-2015-108966      10/11/15      10/18/15      Standard Class      S0-20335
## 5      5 US-2015-108966      10/11/15      10/18/15      Standard Class      S0-20335
```

```
## 6      6 CA-2014-115812      6/9/14      6/14/14 Standard Class      BH-11710
##      Segment      Product.ID      Category Sub.Category
## 1 Consumer FUR-BO-10001798      Furniture      Bookcases
## 2 Consumer FUR-CH-10000454      Furniture      Chairs
## 3 Corporate OFF-LA-10000240 Office Supplies      Labels
## 4 Consumer FUR-TA-10000577      Furniture      Tables
## 5 Consumer OFF-ST-10000760 Office Supplies      Storage
## 6 Consumer FUR-FU-10001487      Furniture      Furnishings
##
##                                     Product.Name      Sales
## 1                                     Bush Somerset Collection Bookcase 261.9600
## 2      Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back 731.9400
## 3      Self-Adhesive Address Labels for Typewriters by Universal 14.6200
## 4                                     Bretford CR4500 Series Slim Rectangular Table 957.5775
## 5                                     Eldon Fold 'N Roll Cart System 22.3680
## 6 Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood 48.8600
##      Quantity Discount      Profit
## 1      2      0.00      41.9136
## 2      3      0.00      219.5820
## 3      2      0.00      6.8714
## 4      5      0.45 -383.0310
## 5      2      0.20      2.5164
## 6      7      0.00      14.1694
```

```
retail$Row.ID <- as.character(retail$Row.ID)
retail$Ship.Date <- as.Date(retail$Ship.Date, format = "%m/%d/%y")
retail$Order.Date <- as.Date(retail$Ship.Date, format = "%m/%d/%y")
retail$Ship.Mode <- as.factor(retail$Ship.Mode)
retail$Segment <- as.factor(retail$Segment)
retail$Category <- as.factor(retail$Category)
retail$Sub.Category <- as.factor(retail$Sub.Category)
retail$Product.Name <- as.factor(retail$Product.Name)

str(retail)
```

```
## 'data.frame':      9994 obs. of      15 variables:
## $ Row.ID      : chr      "1" "2" "3" "4" ...
## $ Order.ID      : chr      "CA-2016-152156" "CA-2016-152156" "CA-2016-138688" "US-2015-108966" ...
## $ Order.Date      : Date, format: "2016-11-11" "2016-11-11" ...
## $ Ship.Date      : Date, format: "2016-11-11" "2016-11-11" ...
## $ Ship.Mode      : Factor w/ 4 levels "First Class",...: 3 3 3 4 4 4 4 4 4 4 ...
## $ Customer.ID      : chr      "CG-12520" "CG-12520" "DV-13045" "SO-20335" ...
## $ Segment      : Factor w/ 3 levels "Consumer","Corporate",...: 1 1 2 1 1 1 1 1 1 1 ...
## $ Product.ID      : chr      "FUR-BO-10001798" "FUR-CH-10000454" "OFF-LA-10000240" "FUR-TA-10000577" ...
## $ Category      : Factor w/ 3 levels "Furniture","Office Supplies",...: 1 1 2 1 2 1 2 3 2 2 ...
## $ Sub.Category: Factor w/ 17 levels "Accessories",...: 5 6 11 17 15 10 3 14 4 2 ...
## $ Product.Name: Factor w/ 1850 levels "\"While you Were Out\" Message Book, One Form per Page",...: 3
## $ Sales      : num      262 731.9 14.6 957.6 22.4 ...
## $ Quantity      : int      2 3 2 5 2 7 4 6 3 5 ...
## $ Discount      : num      0 0 0 0.45 0.2 0 0 0.2 0.2 0 ...
## $ Profit      : num      41.91 219.58 6.87 -383.03 2.52 ...
```

```
head(retail)
```

```
##      Row.ID      Order.ID Order.Date      Ship.Date      Ship.Mode Customer.ID
```

```
## 1      1 CA-2016-152156 2016-11-11 2016-11-11    Second Class    CG-12520
## 2      2 CA-2016-152156 2016-11-11 2016-11-11    Second Class    CG-12520
## 3      3 CA-2016-138688 2016-06-16 2016-06-16    Second Class    DV-13045
## 4      4 US-2015-108966 2015-10-18 2015-10-18    Standard Class   SO-20335
## 5      5 US-2015-108966 2015-10-18 2015-10-18    Standard Class   SO-20335
## 6      6 CA-2014-115812 2014-06-14 2014-06-14    Standard Class   BH-11710
##      Segment      Product.ID      Category Sub.Category
## 1 Consumer FUR-BO-10001798      Furniture    Bookcases
## 2 Consumer FUR-CH-10000454      Furniture    Chairs
## 3 Corporate OFF-LA-10000240 Office Supplies    Labels
## 4 Consumer FUR-TA-10000577      Furniture    Tables
## 5 Consumer OFF-ST-10000760 Office Supplies    Storage
## 6 Consumer FUR-FU-10001487      Furniture    Furnishings
##
##      Product.Name      Sales
## 1      Bush Somerset Collection Bookcase 261.9600
## 2      Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back 731.9400
## 3      Self-Adhesive Address Labels for Typewriters by Universal 14.6200
## 4      Bretford CR4500 Series Slim Rectangular Table 957.5775
## 5      Eldon Fold 'N Roll Cart System 22.3680
## 6 Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood 48.8600
##      Quantity Discount      Profit
## 1      2      0.00      41.9136
## 2      3      0.00      219.5820
## 3      2      0.00      6.8714
## 4      5      0.45     -383.0310
## 5      2      0.20      2.5164
## 6      7      0.00      14.1694
```

```
retail <- retail[,c(3:15)]
retail$selisih <- difftime(retail$Ship.Date, retail$Order.Date, units = "day")
head(retail)
```

```
##      Order.Date Ship.Date      Ship.Mode Customer.ID Segment      Product.ID
## 1 2016-11-11 2016-11-11    Second Class    CG-12520 Consumer FUR-BO-10001798
## 2 2016-11-11 2016-11-11    Second Class    CG-12520 Consumer FUR-CH-10000454
## 3 2016-06-16 2016-06-16    Second Class    DV-13045 Corporate OFF-LA-10000240
## 4 2015-10-18 2015-10-18    Standard Class    SO-20335 Consumer FUR-TA-10000577
## 5 2015-10-18 2015-10-18    Standard Class    SO-20335 Consumer OFF-ST-10000760
## 6 2014-06-14 2014-06-14    Standard Class    BH-11710 Consumer FUR-FU-10001487
##      Category Sub.Category
## 1      Furniture    Bookcases
## 2      Furniture    Chairs
## 3 Office Supplies    Labels
## 4      Furniture    Tables
## 5 Office Supplies    Storage
## 6      Furniture    Furnishings
##
##      Product.Name      Sales
## 1      Bush Somerset Collection Bookcase 261.9600
## 2      Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back 731.9400
## 3      Self-Adhesive Address Labels for Typewriters by Universal 14.6200
## 4      Bretford CR4500 Series Slim Rectangular Table 957.5775
## 5      Eldon Fold 'N Roll Cart System 22.3680
## 6 Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood 48.8600
##      Quantity Discount      Profit selisih
```

```
## 1      2      0.00  41.9136  0 days
## 2      3      0.00 219.5820  0 days
## 3      2      0.00   6.8714  0 days
## 4      5      0.45 -383.0310  0 days
## 5      2      0.20   2.5164  0 days
## 6      7      0.00  14.1694  0 days
```

```
retail$selisih <- as.integer(retail$selisih)
```

## Data Explanation

```
summary(retail)
```

```
##      Order.Date      Ship.Date      Ship.Mode
## Min.   :2014-01-07  Min.   :2014-01-07  First Class  :1538
## 1st Qu.:2015-05-27  1st Qu.:2015-05-27  Same Day    : 543
## Median :2016-06-29  Median :2016-06-29  Second Class :1945
## Mean   :2016-05-03  Mean   :2016-05-03  Standard Class:5968
## 3rd Qu.:2017-05-18  3rd Qu.:2017-05-18
## Max.   :2018-01-05  Max.   :2018-01-05
##
## Customer.ID      Segment      Product.ID
## Length:9994      Consumer   :5191  Length:9994
## Class :character  Corporate  :3020  Class :character
## Mode  :character  Home Office:1783  Mode  :character
##
##
##
##      Category      Sub.Category      Product.Name
## Furniture   :2121  Binders    :1523  Staple envelope      : 48
## Office Supplies:6026  Paper      :1370  Easy-staple paper    : 46
## Technology   :1847  Furnishings: 957  Staples               : 46
##              Phones    : 889  Avery Non-Stick Binders : 20
##              Storage   : 846  Staples in misc. colors : 19
##              Art       : 796  KI Adjustable-Height Table: 18
##              (Other)   :3613  (Other)                :9797
##
##      Sales      Quantity      Discount      Profit
## Min.   : 0.444  Min.   : 1.00  Min.   :0.0000  Min.   : -6599.978
## 1st Qu.: 17.280  1st Qu.: 2.00  1st Qu.:0.0000  1st Qu.:   1.729
## Median : 54.490  Median : 3.00  Median :0.2000  Median :   8.666
## Mean   : 229.858  Mean   : 3.79  Mean   :0.1562  Mean   :  28.657
## 3rd Qu.: 209.940  3rd Qu.: 5.00  3rd Qu.:0.2000  3rd Qu.:  29.364
## Max.   :22638.480  Max.   :14.00  Max.   :0.8000  Max.   : 8399.976
##
##      selisih
## Min.   :0
## 1st Qu.:0
## Median :0
## Mean   :0
## 3rd Qu.:0
```

```
## Max. :0
##
```

```
View(retail)
```

```
#Untuk melihat rincian masing-masing Category berdasarkan Sub Category
```

```
table_category <- table(retail$Category, retail$Sub.Category)
table_category
```

```
##
##      Accessories Appliances  Art Binders Bookcases Chairs Copiers
## Furniture           0         0    0      0      228    617      0
## Office Supplies     0        466   796    1523        0      0      0
## Technology          775         0    0      0        0      0     68
##
##      Envelopes Fasteners Furnishings Labels Machines Paper Phones
## Furniture           0         0      957      0        0      0      0
## Office Supplies    254        217        0    364        0   1370      0
## Technology          0         0        0      0     115      0   889
##
##      Storage Supplies Tables
## Furniture           0         0    319
## Office Supplies    846        190      0
## Technology          0         0      0
```

```
#Untuk melihat Segment berdasarkan mode pengiriman
```

```
contoh_table <- table(retail$Segment, retail$Ship.Mode)
contoh_table
```

```
##
##      First Class Same Day Second Class Standard Class
## Consumer          769        317        1020        3085
## Corporate          485        114         609        1812
## Home Office        284        112         316        1071
```

```
#untuk melihat profit yang dihasilkan dari masing-masing Segment
```

```
jumlah <- aggregate(formula = Profit~Segment, data = retail, FUN = sum)
jumlah
```

```
##      Segment      Profit
## 1 Consumer 134119.21
## 2 Corporate  91979.13
## 3 Home Office 60298.68
```

```
#Untuk melihat rata-rata profit dari masing-masing segment
```

```
avg_cons <- aggregate(formula = Profit~Segment, data = retail, FUN = mean)
avg_cons
```

```
##      Segment      Profit
## 1 Consumer 25.83687
## 2 Corporate 30.45667
## 3 Home Office 33.81866
```

Kesimpulan 1. Pesanan pertama kali pada 2014 2. Mode pengiriman yang paling populer adalah standard class, sementara yang paling tidak populer ada pengiriman dengan mode Same Day 3. 3 Sub Category yang paling banyak dibeli dari category office supplies adalah Binders, Paper dan Storage 4. 3 Sub Category yang paling banyak dibeli dari category Furniture adalah Furnishings, chairs dan tables 5. Segment yang paling banyak menghasilkan profit secara total adalah Segment Consumer, walaupun secara rata-rata, profit yang dihasilkan dari Segment Consumer lebih rendah dari Segment Corporate dan Segment Home Office.