


Data Processing in RStudio

MB4030 Business Intelligence and Analytics


Riefka Azka Aulia **Hasna** (19022036)


Data gathered from Kaggle.com


ROHIT SAHOO · UPDATED 4 YEARS AGO

▲ **542**

New Notebook


 **Download**





Superstore Sales Dataset

Predict Sales using Time Series



[Data Card](#) [Code \(184\)](#) [Discussion \(1\)](#) [Suggestions \(0\)](#)

About Dataset

Context

Retail dataset of a global superstore for 4 years.

Usability ⓘ

10.00

License

GPL 2

Olah Data 1

```
data <- read.csv("C:/Users/Riefka Hasna/Downloads/archive (3)/train.csv")  
head(data)  
summary(data)
```

Data										
data	9800 obs. of 18 variables									
\$ Row.ID	: int	1	2	3	4	5	6	7	8	9 10 ...
\$ Order.ID	: chr	"CA-2017-152156"	"CA-2017-152156"	"CA-2017-152156"	"CA-2017-152156"	"CA-2017-152156"	"CA-2017-152156"	"CA-2017-152156"	"CA-2017-152156"	"CA-2017-152156"
\$ Order.Date	: Date, format: NA NA NA ...									
\$ Ship.Date	: chr	"11/11/2017"	"11/11/2017"	"11/11/2017"	"11/11/2017"	"11/11/2017"	"11/11/2017"	"11/11/2017"	"11/11/2017"	"11/11/2017"
\$ Ship.Mode	: chr	"Second Class"	"Second Class"	"Second Class"	"Second Class"	"Second Class"	"Second Class"	"Second Class"	"Second Class"	"Second Class"
\$ Customer.ID	: chr	"CG-12520"	"CG-12520"	"CG-12520"	"CG-12520"	"CG-12520"	"CG-12520"	"CG-12520"	"CG-12520"	"CG-12520"
\$ Customer.Name	: chr	"Claire Gute"	"Claire Gute"	"Claire Gute"	"Claire Gute"	"Claire Gute"	"Claire Gute"	"Claire Gute"	"Claire Gute"	"Claire Gute"
\$ Segment	: chr	"Consumer"	"Consumer"	"Consumer"	"Consumer"	"Consumer"	"Consumer"	"Consumer"	"Consumer"	"Consumer"
\$ Country	: chr	"United States"	"United States"	"United States"	"United States"	"United States"	"United States"	"United States"	"United States"	"United States"
\$ City	: chr	"Henderson"	"Henderson"	"Henderson"	"Henderson"	"Henderson"	"Henderson"	"Henderson"	"Henderson"	"Henderson"
\$ State	: chr	"Kentucky"	"Kentucky"	"Kentucky"	"Kentucky"	"Kentucky"	"Kentucky"	"Kentucky"	"Kentucky"	"Kentucky"
\$ Postal.Code	: int	42420	42420	90036	33311	33311	33311	33311	33311	33311
\$ Region	: chr	"South"	"South"	"West"	"South"	"South"	"South"	"South"	"South"	"South"
\$ Product.ID	: chr	"FUR-BO-10001798"	"FUR-BO-10001798"	"FUR-BO-10001798"	"FUR-BO-10001798"	"FUR-BO-10001798"	"FUR-BO-10001798"	"FUR-BO-10001798"	"FUR-BO-10001798"	"FUR-BO-10001798"
\$ Category	: chr	"Furniture"	"Furniture"	"Furniture"	"Furniture"	"Furniture"	"Furniture"	"Furniture"	"Furniture"	"Furniture"
\$ Sub.Category	: chr	"Bookcases"	"Chairs"	"Chairs"	"Chairs"	"Chairs"	"Chairs"	"Chairs"	"Chairs"	"Chairs"
\$ Product.Name	: chr	"Bush Somerset Collection ..."	"Bush Somerset Collection ..."	"Bush Somerset Collection ..."	"Bush Somerset Collection ..."	"Bush Somerset Collection ..."	"Bush Somerset Collection ..."	"Bush Somerset Collection ..."	"Bush Somerset Collection ..."	"Bush Somerset Collection ..."
\$ Sales	: num	262	731.9	14.6	957.6	22.4	957.6	22.4	957.6	22.4

```
> summary(data)  
      Row.ID      Order.ID      Order.Date      Ship.Date      Ship.Mode  
Min.   :    1      Length:9800      Length:9800      Length:9800      Length:9800  
1st Qu.: 2451      Class :character      Class :character      Class :character      Class :character  
Median : 4900      Mode  :character      Mode  :character      Mode  :character      Mode  :character  
Mean   : 4900  
3rd Qu.: 7350  
Max.   : 9800  
  
Customer.ID      Customer.Name      Segment      Country      City  
Length:9800      Length:9800      Length:9800      Length:9800      Length:9800  
Class :character      Class :character      Class :character      Class :character      Class :character  
Mode  :character      Mode  :character      Mode  :character      Mode  :character      Mode  :character  
  
State      Postal.Code      Region      Product.ID      Category  
Length:9800      Min.   : 1040      Length:9800      Length:9800      Length:9800  
Class :character      1st Qu.:23223      Class :character      Class :character      Class :character  
Mode  :character      Median :58103      Mode  :character      Mode  :character      Mode  :character  
Mean   :55273  
3rd Qu.:90008  
Max.   :99301  
NA's   :11  
Sub.Category      Product.Name      Sales  
Length:9800      Length:9800      Min.   : 0.444  
Class :character      Class :character      1st Qu.: 17.248  
Mode  :character      Mode  :character      Median : 54.490  
Mean   : 230.769  
3rd Qu.: 210.605  
Max.   :22638.480
```

Olah Data 2

```
data <- read.csv("C:/Users/Riefka Hasna/Downloads/archive (3)/train.csv")
head(data)
data$Order.Date <- as.Date(data$Order.Date, format = "%Y-%m-%d")
sorted_data <- data[order(data$Order.Date), ]
head(data)
```

Data	
data	9800 obs. of 18 variables
sorted_data	9800 obs. of 18 variables
\$ Row.ID	: int 1 2 3 4 5 6 7 8 9 10 ...
\$ Order.ID	: chr "CA-2017-152156" "CA-2017-...
\$ Order.Date	: Date, format: NA NA NA ...
\$ Ship.Date	: chr "11/11/2017" "11/11/2017" ...
\$ Ship.Mode	: chr "Second Class" "Second Cla...
\$ Customer.ID	: chr "CG-12520" "CG-12520" "DV-...
\$ Customer.Name	: chr "Claire Gute" "Claire Gute...
\$ Segment	: chr "Consumer" "Consumer" "Cor...
\$ Country	: chr "United States" "United St...
\$ City	: chr "Henderson" "Henderson" "L...
\$ State	: chr "Kentucky" "Kentucky" "Cal...
\$ Postal.Code	: int 42420 42420 90036 33311 33...
\$ Region	: chr "South" "South" "West" "So...
\$ Product.ID	: chr "FUR-BO-10001798" "FUR-CH-...
\$ Category	: chr "Furniture" "Furniture" "O...
\$ Sub.Category	: chr "Bookcases" "Chairs" "Labe...
\$ Product.Name	: chr "Bush Somerset Collection ...
\$ Sales	: num 262 731.9 14.6 957.6 22.4 ...

```
> data$Order.Date <- as.Date(data$Order.Date, format = "%Y-%m-%d")
> sorted_data <- data[order(data$Order.Date), ]
> head(data)
```

Row.ID	Order.ID	Order.Date	Ship.Date	Ship.Mode	Customer.ID	Customer.Name
1	1 CA-2017-152156	<NA>	11/11/2017	Second Class	CG-12520	Claire Gute
2	2 CA-2017-152156	<NA>	11/11/2017	Second Class	CG-12520	Claire Gute
3	3 CA-2017-138688	<NA>	16/06/2017	Second Class	DV-13045	Darrin Van Huff
4	4 US-2016-108966	<NA>	18/10/2016	Standard Class	SO-20335	Sean O'Donnell
5	5 US-2016-108966	<NA>	18/10/2016	Standard Class	SO-20335	Sean O'Donnell
6	6 CA-2015-115812	<NA>	14/06/2015	Standard Class	BH-11710	Brosina Hoffman

Segment	Country	City	State	Postal.Code	Region	Product.ID
1 Consumer	United States	Henderson	Kentucky	42420	South	FUR-BO-10001798
2 Consumer	United States	Henderson	Kentucky	42420	South	FUR-CH-10000454
3 Corporate	United States	Los Angeles	California	90036	West	OFF-LA-10000240
4 Consumer	United States	Fort Lauderdale	Florida	33311	South	FUR-TA-10000577
5 Consumer	United States	Fort Lauderdale	Florida	33311	South	OFF-ST-10000760
6 Consumer	United States	Los Angeles	California	90032	West	FUR-FU-10001487

	Category	Sub.Category	Product.Name	Sales
1	Furniture	Bookcases	Bush Somerset Collection Bookcase	261.9600
2	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back	731.9400
3	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters by Universal	14.6200
4	Furniture	Tables	Bretford CR4500 Series Slim Rectangular Table	957.5775
5	Office Supplies	Storage	Eldon Fold 'N Roll Cart System	22.3680
6	Furniture	Furnishings	Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood	48.8600