

# Week 7

Musrat Jahan

2025-03-17

## 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>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
df <- read.table("https://raw.githubusercontent.com/musratjahan1/DATA607/refs/heads/main/dataweek7.txt")
head(df)
```

```
##           Category  Item.Name Item.ID      Brand   Price Variation.ID
## 1    Electronics  Smartphone     101  TechBrand  699.99        101-A
## 2    Electronics  Smartphone     101  TechBrand  699.99        101-B
## 3    Electronics    Laptop     102 CompuBrand 1099.99        102-A
## 4    Electronics    Laptop     102 CompuBrand 1099.99        102-B
## 5 Home Appliances Refrigerator    201   HomeCool  899.99        201-A
## 6 Home Appliances Refrigerator    201   HomeCool  899.99        201-B
##           Variation.Details
## 1           Color: Black, Storage: 64GB
## 2           Color: White, Storage: 128GB
## 3           Color: Silver, Storage: 256GB
## 4           Color: Space Gray, Storage: 512GB
## 5 Color: Stainless Steel, Capacity:20 cu ft
## 6           Color: White, Capacity: 18 cu ft
```

## JSON

```
library(jsonlite)

#convert to json
json_data <- toJSON(df, pretty=TRUE)

#save json to a file
write(json_data, "data.json")

#print output
cat(json_data)

## [
##   {
##     "Category": "Electronics",
##     "Item.Name": "Smartphone",
##     "Item.ID": 101,
```

```

##      "Brand": "TechBrand",
##      "Price": 699.99,
##      "Variation.ID": "101-A",
##      "Variation.Details": "Color: Black, Storage: 64GB"
##    },
##    {
##      "Category": "Electronics",
##      "Item.Name": "Smartphone",
##      "Item.ID": 101,
##      "Brand": "TechBrand",
##      "Price": 699.99,
##      "Variation.ID": "101-B",
##      "Variation.Details": "Color: White, Storage: 128GB"
##    },
##    {
##      "Category": "Electronics",
##      "Item.Name": "Laptop",
##      "Item.ID": 102,
##      "Brand": "CompuBrand",
##      "Price": 1099.99,
##      "Variation.ID": "102-A",
##      "Variation.Details": "Color: Silver, Storage: 256GB"
##    },
##    {
##      "Category": "Electronics",
##      "Item.Name": "Laptop",
##      "Item.ID": 102,
##      "Brand": "CompuBrand",
##      "Price": 1099.99,
##      "Variation.ID": "102-B",
##      "Variation.Details": "Color: Space Gray, Storage: 512GB"
##    },
##    {
##      "Category": "Home Appliances",
##      "Item.Name": "Refrigerator",
##      "Item.ID": 201,
##      "Brand": "HomeCool",
##      "Price": 899.99,
##      "Variation.ID": "201-A",
##      "Variation.Details": "Color: Stainless Steel, Capacity:20 cu ft"
##    },
##    {
##      "Category": "Home Appliances",
##      "Item.Name": "Refrigerator",
##      "Item.ID": 201,
##      "Brand": "HomeCool",
##      "Price": 899.99,
##      "Variation.ID": "201-B",
##      "Variation.Details": "Color: White, Capacity: 18 cu ft"
##    },
##    {
##      "Category": "Home Appliances",
##      "Item.Name": "Washing Machine",
##      "Item.ID": 202,

```

```

##      "Brand": "CleanTech",
##      "Price": 499.99,
##      "Variation.ID": "202-A",
##      "Variation.Details": "Type: Front Load, Capacity:4.5 cu ft"
##  },
##  {
##      "Category": "Home Appliances",
##      "Item.Name": "Washing Machine",
##      "Item.ID": 202,
##      "Brand": "CleanTech",
##      "Price": 499.99,
##      "Variation.ID": "202-B",
##      "Variation.Details": "Type: Top Load, Capacity:5.0 cu ft"
##  },
##  {
##      "Category": "Clothing",
##      "Item.Name": "T-Shirt",
##      "Item.ID": 301,
##      "Brand": "FashionCo",
##      "Price": 19.99,
##      "Variation.ID": "301-A",
##      "Variation.Details": "Color: Blue, Size: S"
##  },
##  {
##      "Category": "Clothing",
##      "Item.Name": "T-Shirt",
##      "Item.ID": 301,
##      "Brand": "FashionCo",
##      "Price": 19.99,
##      "Variation.ID": "301-B",
##      "Variation.Details": "Color: Red, Size: M"
##  },
##  {
##      "Category": "Clothing",
##      "Item.Name": "T-Shirt",
##      "Item.ID": 301,
##      "Brand": "FashionCo",
##      "Price": 19.99,
##      "Variation.ID": "301-C",
##      "Variation.Details": "Color: Green, Size: L"
##  },
##  {
##      "Category": "Clothing",
##      "Item.Name": "Jeans",
##      "Item.ID": 302,
##      "Brand": "DenimWorks",
##      "Price": 49.99,
##      "Variation.ID": "302-A",
##      "Variation.Details": "Color: Dark Blue, Size: 32"
##  },
##  {
##      "Category": "Clothing",
##      "Item.Name": "Jeans",
##      "Item.ID": 302,

```

```

##      "Brand": "DenimWorks",
##      "Price": 49.99,
##      "Variation.ID": "302-B",
##      "Variation.Details": "Color: Light Blue, Size: 34"
##  },
##  {
##      "Category": "Books",
##      "Item.Name": "Fiction Novel",
##      "Item.ID": 401,
##      "Brand": "-",
##      "Price": 14.99,
##      "Variation.ID": "401-A",
##      "Variation.Details": "Format: Hardcover, Language: English"
##  },
##  {
##      "Category": "Books",
##      "Item.Name": "Fiction Novel",
##      "Item.ID": 401,
##      "Brand": "-",
##      "Price": 14.99,
##      "Variation.ID": "401-B",
##      "Variation.Details": "Format: Paperback, Language: Spanish"
##  },
##  {
##      "Category": "Books",
##      "Item.Name": "Non-Fiction Guide",
##      "Item.ID": 402,
##      "Brand": "-",
##      "Price": 24.99,
##      "Variation.ID": "402-A",
##      "Variation.Details": "Format: eBook, Language: English"
##  },
##  {
##      "Category": "Books",
##      "Item.Name": "Non-Fiction Guide",
##      "Item.ID": 402,
##      "Brand": "-",
##      "Price": 24.99,
##      "Variation.ID": "402-B",
##      "Variation.Details": "Format: Paperback, Language: French"
##  },
##  {
##      "Category": "Sports Equipment",
##      "Item.Name": "Basketball",
##      "Item.ID": 501,
##      "Brand": "SportsGear",
##      "Price": 29.99,
##      "Variation.ID": "501-A",
##      "Variation.Details": "Size: Size 7, Color: Orange"
##  },
##  {
##      "Category": "Sports Equipment",
##      "Item.Name": "Tennis Racket",
##      "Item.ID": 502,

```

```
##      "Brand": "RacketPro",
##      "Price": 89.99,
##      "Variation.ID": "502-A",
##      "Variation.Details": "Material: Graphite, Color: Black"
##    },
##    {
##      "Category": "Sports Equipment",
##      "Item.Name": "Tennis Racket",
##      "Item.ID": 502,
##      "Brand": "RacketPro",
##      "Price": 89.99,
##      "Variation.ID": "502-B",
##      "Variation.Details": "Material: Aluminum, Color: Silver"
##    }
##  ]
```

```
#read json file as a data frame
json_data <- fromJSON("data.json")
print(json_data)
```

|       | Category                                  | Item.Name         | Item.ID | Brand      | Price   | Variation.ID |
|-------|---|-------------------|---------|------------|---------|--------------|
| ## 1  | Electronics                               | Smartphone        | 101     | TechBrand  | 699.99  | 101-A        |
| ## 2  | Electronics                               | Smartphone        | 101     | TechBrand  | 699.99  | 101-B        |
| ## 3  | Electronics                               | Laptop            | 102     | CompuBrand | 1099.99 | 102-A        |
| ## 4  | Electronics                               | Laptop            | 102     | CompuBrand | 1099.99 | 102-B        |
| ## 5  | Home Appliances                           | Refrigerator      | 201     | HomeCool   | 899.99  | 201-A        |
| ## 6  | Home Appliances                           | Refrigerator      | 201     | HomeCool   | 899.99  | 201-B        |
| ## 7  | Home Appliances                           | Washing Machine   | 202     | CleanTech  | 499.99  | 202-A        |
| ## 8  | Home Appliances                           | Washing Machine   | 202     | CleanTech  | 499.99  | 202-B        |
| ## 9  | Clothing                                  | T-Shirt           | 301     | FashionCo  | 19.99   | 301-A        |
| ## 10 | Clothing                                  | T-Shirt           | 301     | FashionCo  | 19.99   | 301-B        |
| ## 11 | Clothing                                  | T-Shirt           | 301     | FashionCo  | 19.99   | 301-C        |
| ## 12 | Clothing                                  | Jeans             | 302     | DenimWorks | 49.99   | 302-A        |
| ## 13 | Clothing                                  | Jeans             | 302     | DenimWorks | 49.99   | 302-B        |
| ## 14 | Books                                     | Fiction Novel     | 401     | -          | 14.99   | 401-A        |
| ## 15 | Books                                     | Fiction Novel     | 401     | -          | 14.99   | 401-B        |
| ## 16 | Books                                     | Non-Fiction Guide | 402     | -          | 24.99   | 402-A        |
| ## 17 | Books                                     | Non-Fiction Guide | 402     | -          | 24.99   | 402-B        |
| ## 18 | Sports Equipment                          | Basketball        | 501     | SportsGear | 29.99   | 501-A        |
| ## 19 | Sports Equipment                          | Tennis Racket     | 502     | RacketPro  | 89.99   | 502-A        |
| ## 20 | Sports Equipment                          | Tennis Racket     | 502     | RacketPro  | 89.99   | 502-B        |
| ##    | Variation.Details                         |                   |         |            |         |              |
| ## 1  | Color: Black, Storage: 64GB               |                   |         |            |         |              |
| ## 2  | Color: White, Storage: 128GB              |                   |         |            |         |              |
| ## 3  | Color: Silver, Storage: 256GB             |                   |         |            |         |              |
| ## 4  | Color: Space Gray, Storage: 512GB         |                   |         |            |         |              |
| ## 5  | Color: Stainless Steel, Capacity:20 cu ft |                   |         |            |         |              |
| ## 6  | Color: White, Capacity: 18 cu ft          |                   |         |            |         |              |
| ## 7  | Type: Front Load, Capacity:4.5 cu ft      |                   |         |            |         |              |
| ## 8  | Type: Top Load, Capacity:5.0 cu ft        |                   |         |            |         |              |
| ## 9  | Color: Blue, Size: S                      |                   |         |            |         |              |
| ## 10 | Color: Red, Size: M                       |                   |         |            |         |              |
| ## 11 | Color: Green, Size: L                     |                   |         |            |         |              |
| ## 12 | Color: Dark Blue, Size: 32                |                   |         |            |         |              |
| ## 13 | Color: Light Blue, Size: 34               |                   |         |            |         |              |

```
## 14      Format: Hardcover, Language: English
## 15      Format: Paperback, Language: Spanish
## 16      Format: eBook, Language: English
## 17      Format: Paperback, Language: French
## 18      Size: Size 7, Color: Orange
## 19      Material: Graphite, Color: Black
## 20      Material: Aluminum, Color: Silver
```

## HTML

```
library(htmlTable)

#create html table
html_data <- htmlTable(df)

#save html table to a file
writeLines(html_data, "data.html")

#print output
cat(html_data)
```

```
## <table class='gmisc_table' style='border-collapse: collapse; margin-top: 1em; margin-bottom: 1em;' >
## <thead>
## <tr><th style='border-bottom: 1px solid grey; border-top: 2px solid grey;'></th>
## <th style='font-weight: 900; border-bottom: 1px solid grey; border-top: 2px solid grey; text-align: center;'>1</th>
## <th style='font-weight: 900; border-bottom: 1px solid grey; border-top: 2px solid grey; text-align: center;'>Electronics</th>
## <th style='font-weight: 900; border-bottom: 1px solid grey; border-top: 2px solid grey; text-align: center;'>Smartphone</th>
## <th style='font-weight: 900; border-bottom: 1px solid grey; border-top: 2px solid grey; text-align: center;'>101</th>
## <th style='font-weight: 900; border-bottom: 1px solid grey; border-top: 2px solid grey; text-align: center;'>TechBrand</th>
## <th style='font-weight: 900; border-bottom: 1px solid grey; border-top: 2px solid grey; text-align: center;'>699.99</th>
## <th style='font-weight: 900; border-bottom: 1px solid grey; border-top: 2px solid grey; text-align: center;'>101-A</th>
## <th style='font-weight: 900; border-bottom: 1px solid grey; border-top: 2px solid grey; text-align: center;'>Color: Black, Storage: 64GB</th>
## </tr>
## </thead>
## <tbody>
## <tr>
## <td style='text-align: left;'>1</td>
## <td style='text-align: center;'>Electronics</td>
## <td style='text-align: center;'>Smartphone</td>
## <td style='text-align: center;'>101</td>
## <td style='text-align: center;'>TechBrand</td>
## <td style='text-align: center;'>699.99</td>
## <td style='text-align: center;'>101-A</td>
## <td style='text-align: center;'>Color: Black, Storage: 64GB</td>
## </tr>
## <tr>
## <td style='text-align: left;'>2</td>
## <td style='text-align: center;'>Electronics</td>
## <td style='text-align: center;'>Smartphone</td>
## <td style='text-align: center;'>101</td>
## <td style='text-align: center;'>TechBrand</td>
## <td style='text-align: center;'>699.99</td>
## <td style='text-align: center;'>101-B</td>
## <td style='text-align: center;'>Color: White, Storage: 128GB</td>
## </tr>
```

```

## <tr>
## <td style='text-align: left;'>3</td>
## <td style='text-align: center;'>Electronics</td>
## <td style='text-align: center;'>Laptop</td>
## <td style='text-align: center;'>102</td>
## <td style='text-align: center;'>CompuBrand</td>
## <td style='text-align: center;'>1099.99</td>
## <td style='text-align: center;'>102-A</td>
## <td style='text-align: center;'>Color: Silver, Storage: 256GB</td>
## </tr>
## <tr>
## <td style='text-align: left;'>4</td>
## <td style='text-align: center;'>Electronics</td>
## <td style='text-align: center;'>Laptop</td>
## <td style='text-align: center;'>102</td>
## <td style='text-align: center;'>CompuBrand</td>
## <td style='text-align: center;'>1099.99</td>
## <td style='text-align: center;'>102-B</td>
## <td style='text-align: center;'>Color: Space Gray, Storage: 512GB</td>
## </tr>
## <tr>
## <td style='text-align: left;'>5</td>
## <td style='text-align: center;'>Home Appliances</td>
## <td style='text-align: center;'>Refrigerator</td>
## <td style='text-align: center;'>201</td>
## <td style='text-align: center;'>HomeCool</td>
## <td style='text-align: center;'>899.99</td>
## <td style='text-align: center;'>201-A</td>
## <td style='text-align: center;'>Color: Stainless Steel, Capacity:20 cu ft</td>
## </tr>
## <tr>
## <td style='text-align: left;'>6</td>
## <td style='text-align: center;'>Home Appliances</td>
## <td style='text-align: center;'>Refrigerator</td>
## <td style='text-align: center;'>201</td>
## <td style='text-align: center;'>HomeCool</td>
## <td style='text-align: center;'>899.99</td>
## <td style='text-align: center;'>201-B</td>
## <td style='text-align: center;'>Color: White, Capacity: 18 cu ft</td>
## </tr>
## <tr>
## <td style='text-align: left;'>7</td>
## <td style='text-align: center;'>Home Appliances</td>
## <td style='text-align: center;'>Washing Machine</td>
## <td style='text-align: center;'>202</td>
## <td style='text-align: center;'>CleanTech</td>
## <td style='text-align: center;'>499.99</td>
## <td style='text-align: center;'>202-A</td>
## <td style='text-align: center;'>Type: Front Load, Capacity:4.5 cu ft</td>
## </tr>
## <tr>
## <td style='text-align: left;'>8</td>
## <td style='text-align: center;'>Home Appliances</td>
## <td style='text-align: center;'>Washing Machine</td>

```

```

## <td style='text-align: center;'>202</td>
## <td style='text-align: center;'>CleanTech</td>
## <td style='text-align: center;'>499.99</td>
## <td style='text-align: center;'>202-B</td>
## <td style='text-align: center;'>Type: Top Load, Capacity:5.0 cu ft</td>
## </tr>
## <tr>
## <td style='text-align: left;'>9</td>
## <td style='text-align: center;'>Clothing</td>
## <td style='text-align: center;'>T-Shirt</td>
## <td style='text-align: center;'>301</td>
## <td style='text-align: center;'>FashionCo</td>
## <td style='text-align: center;'>19.99</td>
## <td style='text-align: center;'>301-A</td>
## <td style='text-align: center;'>Color: Blue, Size: S</td>
## </tr>
## <tr>
## <td style='text-align: left;'>10</td>
## <td style='text-align: center;'>Clothing</td>
## <td style='text-align: center;'>T-Shirt</td>
## <td style='text-align: center;'>301</td>
## <td style='text-align: center;'>FashionCo</td>
## <td style='text-align: center;'>19.99</td>
## <td style='text-align: center;'>301-B</td>
## <td style='text-align: center;'>Color: Red, Size: M</td>
## </tr>
## <tr>
## <td style='text-align: left;'>11</td>
## <td style='text-align: center;'>Clothing</td>
## <td style='text-align: center;'>T-Shirt</td>
## <td style='text-align: center;'>301</td>
## <td style='text-align: center;'>FashionCo</td>
## <td style='text-align: center;'>19.99</td>
## <td style='text-align: center;'>301-C</td>
## <td style='text-align: center;'>Color: Green, Size: L</td>
## </tr>
## <tr>
## <td style='text-align: left;'>12</td>
## <td style='text-align: center;'>Clothing</td>
## <td style='text-align: center;'>Jeans</td>
## <td style='text-align: center;'>302</td>
## <td style='text-align: center;'>DenimWorks</td>
## <td style='text-align: center;'>49.99</td>
## <td style='text-align: center;'>302-A</td>
## <td style='text-align: center;'>Color: Dark Blue, Size: 32</td>
## </tr>
## <tr>
## <td style='text-align: left;'>13</td>
## <td style='text-align: center;'>Clothing</td>
## <td style='text-align: center;'>Jeans</td>
## <td style='text-align: center;'>302</td>
## <td style='text-align: center;'>DenimWorks</td>
## <td style='text-align: center;'>49.99</td>
## <td style='text-align: center;'>302-B</td>

```



```

## <td style='text-align: center;'>Color: Light Blue, Size: 34</td>
## </tr>
## <tr>
## <td style='text-align: left;'>14</td>
## <td style='text-align: center;'>Books</td>
## <td style='text-align: center;'>Fiction Novel</td>
## <td style='text-align: center;'>401</td>
## <td style='text-align: center;'>--</td>
## <td style='text-align: center;'>14.99</td>
## <td style='text-align: center;'>401-A</td>
## <td style='text-align: center;'>Format: Hardcover, Language: English</td>
## </tr>
## <tr>
## <td style='text-align: left;'>15</td>
## <td style='text-align: center;'>Books</td>
## <td style='text-align: center;'>Fiction Novel</td>
## <td style='text-align: center;'>401</td>
## <td style='text-align: center;'>--</td>
## <td style='text-align: center;'>14.99</td>
## <td style='text-align: center;'>401-B</td>
## <td style='text-align: center;'>Format: Paperback, Language: Spanish</td>
## </tr>
## <tr>
## <td style='text-align: left;'>16</td>
## <td style='text-align: center;'>Books</td>
## <td style='text-align: center;'>Non-Fiction Guide</td>
## <td style='text-align: center;'>402</td>
## <td style='text-align: center;'>--</td>
## <td style='text-align: center;'>24.99</td>
## <td style='text-align: center;'>402-A</td>
## <td style='text-align: center;'>Format: eBook, Language: English</td>
## </tr>
## <tr>
## <td style='text-align: left;'>17</td>
## <td style='text-align: center;'>Books</td>
## <td style='text-align: center;'>Non-Fiction Guide</td>
## <td style='text-align: center;'>402</td>
## <td style='text-align: center;'>--</td>
## <td style='text-align: center;'>24.99</td>
## <td style='text-align: center;'>402-B</td>
## <td style='text-align: center;'>Format: Paperback, Language: French</td>
## </tr>
## <tr>
## <td style='text-align: left;'>18</td>
## <td style='text-align: center;'>Sports Equipment</td>
## <td style='text-align: center;'>Basketball</td>
## <td style='text-align: center;'>501</td>
## <td style='text-align: center;'>SportsGear</td>
## <td style='text-align: center;'>29.99</td>
## <td style='text-align: center;'>501-A</td>
## <td style='text-align: center;'>Size: Size 7, Color: Orange</td>
## </tr>
## <tr>
## <td style='text-align: left;'>19</td>

```

```
## <td style='text-align: center;'>Sports Equipment</td>
## <td style='text-align: center;'>Tennis Racket</td>
## <td style='text-align: center;'>502</td>
## <td style='text-align: center;'>RacketPro</td>
## <td style='text-align: center;'>89.99</td>
## <td style='text-align: center;'>502-A</td>
## <td style='text-align: center;'>Material: Graphite, Color: Black</td>
## </tr>
## <tr>
## <td style='border-bottom: 2px solid grey; text-align: left;'>20</td>
## <td style='border-bottom: 2px solid grey; text-align: center;'>Sports Equipment</td>
## <td style='border-bottom: 2px solid grey; text-align: center;'>Tennis Racket</td>
## <td style='border-bottom: 2px solid grey; text-align: center;'>502</td>
## <td style='border-bottom: 2px solid grey; text-align: center;'>RacketPro</td>
## <td style='border-bottom: 2px solid grey; text-align: center;'>89.99</td>
## <td style='border-bottom: 2px solid grey; text-align: center;'>502-B</td>
## <td style='border-bottom: 2px solid grey; text-align: center;'>Material: Aluminum, Color: Silver</td>
## </tr>
## </tbody>
## </table>
```

## XML

```
library(XML)
```

```
#convert to xml
```

```
xml_data <- xmlTree()
```

```
xml_data$addTag("root",close=FALSE)
```

```
## Warning in xmlRoot(XMLInternalDocument(currentNodes[[1]]): empty XML document
```

```
for(i in 1:nrow(df)) {
  xml_data$addTag("record", close=FALSE)
  xml_data$addTag("Category", df$Category[i])
  xml_data$addTag("Item.Name", df$Item.Name[i])
  xml_data$addTag("Item.ID", df$Item.ID[i])
  xml_data$addTag("Brand", df$Brand[i])
  xml_data$addTag("Price", df$Price[i])
  xml_data$addTag("Variation.ID", df$Variation.ID[i])
  xml_data$addTag("Variation.Details", df$Variation.Details[i])
  xml_data$closeTag()      #close <record>
}
```

```
xml_data$closeTag()      #close root
```

```
#save xml to file
```

```
saveXML(xml_data, file="data.xml")
```

```
## [1] "data.xml"
```

```
#print xml output
```

```
cat(saveXML(xml_data))
```

```
## <?xml version="1.0"?>
```

```
##
```

```

## <root>
##   <record>
##     <Category>Electronics</Category>
##     <IItem.Name>Smartphone</IItem.Name>
##     <Item.ID>101</Item.ID>
##     <Brand>TechBrand</Brand>
##     <Price>699.99</Price>
##     <Variation.ID>101-A</Variation.ID>
##     <Variation.Details>Color: Black, Storage: 64GB</Variation.Details>
##   </record>
##   <record>
##     <Category>Electronics</Category>
##     <IItem.Name>Smartphone</IItem.Name>
##     <Item.ID>101</Item.ID>
##     <Brand>TechBrand</Brand>
##     <Price>699.99</Price>
##     <Variation.ID>101-B</Variation.ID>
##     <Variation.Details>Color: White, Storage: 128GB</Variation.Details>
##   </record>
##   <record>
##     <Category>Electronics</Category>
##     <IItem.Name>Laptop</IItem.Name>
##     <Item.ID>102</Item.ID>
##     <Brand>CompuBrand</Brand>
##     <Price>1099.99</Price>
##     <Variation.ID>102-A</Variation.ID>
##     <Variation.Details>Color: Silver, Storage: 256GB</Variation.Details>
##   </record>
##   <record>
##     <Category>Electronics</Category>
##     <IItem.Name>Laptop</IItem.Name>
##     <Item.ID>102</Item.ID>
##     <Brand>CompuBrand</Brand>
##     <Price>1099.99</Price>
##     <Variation.ID>102-B</Variation.ID>
##     <Variation.Details>Color: Space Gray, Storage: 512GB</Variation.Details>
##   </record>
##   <record>
##     <Category>Home Appliances</Category>
##     <IItem.Name>Refrigerator</IItem.Name>
##     <Item.ID>201</Item.ID>
##     <Brand>HomeCool</Brand>
##     <Price>899.99</Price>
##     <Variation.ID>201-A</Variation.ID>
##     <Variation.Details>Color: Stainless Steel, Capacity:20 cu ft</Variation.Details>
##   </record>
##   <record>
##     <Category>Home Appliances</Category>
##     <IItem.Name>Refrigerator</IItem.Name>
##     <Item.ID>201</Item.ID>
##     <Brand>HomeCool</Brand>
##     <Price>899.99</Price>
##     <Variation.ID>201-B</Variation.ID>
##     <Variation.Details>Color: White, Capacity: 18 cu ft</Variation.Details>

```

```

## </record>
## <record>
##   <Category>Home Appliances</Category>
##   <IItem.Name>Washing Machine</IItem.Name>
##   <Item.ID>202</Item.ID>
##   <Brand>CleanTech</Brand>
##   <Price>499.99</Price>
##   <Variation.ID>202-A</Variation.ID>
##   <Variation.Details>Type: Front Load, Capacity:4.5 cu ft</Variation.Details>
## </record>
## <record>
##   <Category>Home Appliances</Category>
##   <IItem.Name>Washing Machine</IItem.Name>
##   <Item.ID>202</Item.ID>
##   <Brand>CleanTech</Brand>
##   <Price>499.99</Price>
##   <Variation.ID>202-B</Variation.ID>
##   <Variation.Details>Type: Top Load, Capacity:5.0 cu ft</Variation.Details>
## </record>
## <record>
##   <Category>Clothing</Category>
##   <IItem.Name>T-Shirt</IItem.Name>
##   <Item.ID>301</Item.ID>
##   <Brand>FashionCo</Brand>
##   <Price>19.99</Price>
##   <Variation.ID>301-A</Variation.ID>
##   <Variation.Details>Color: Blue, Size: S</Variation.Details>
## </record>
## <record>
##   <Category>Clothing</Category>
##   <IItem.Name>T-Shirt</IItem.Name>
##   <Item.ID>301</Item.ID>
##   <Brand>FashionCo</Brand>
##   <Price>19.99</Price>
##   <Variation.ID>301-B</Variation.ID>
##   <Variation.Details>Color: Red, Size: M</Variation.Details>
## </record>
## <record>
##   <Category>Clothing</Category>
##   <IItem.Name>T-Shirt</IItem.Name>
##   <Item.ID>301</Item.ID>
##   <Brand>FashionCo</Brand>
##   <Price>19.99</Price>
##   <Variation.ID>301-C</Variation.ID>
##   <Variation.Details>Color: Green, Size: L</Variation.Details>
## </record>
## <record>
##   <Category>Clothing</Category>
##   <IItem.Name>Jeans</IItem.Name>
##   <Item.ID>302</Item.ID>
##   <Brand>DenimWorks</Brand>
##   <Price>49.99</Price>
##   <Variation.ID>302-A</Variation.ID>
##   <Variation.Details>Color: Dark Blue, Size: 32</Variation.Details>

```

```

## </record>
## <record>
##   <Category>Clothing</Category>
##   <IItem.Name>Jeans</IItem.Name>
##   <Item.ID>302</Item.ID>
##   <Brand>DenimWorks</Brand>
##   <Price>49.99</Price>
##   <Variation.ID>302-B</Variation.ID>
##   <Variation.Details>Color: Light Blue, Size: 34</Variation.Details>
## </record>
## <record>
##   <Category>Books</Category>
##   <IItem.Name>Fiction Novel</IItem.Name>
##   <Item.ID>401</Item.ID>
##   <Brand>-</Brand>
##   <Price>14.99</Price>
##   <Variation.ID>401-A</Variation.ID>
##   <Variation.Details>Format: Hardcover, Language: English</Variation.Details>
## </record>
## <record>
##   <Category>Books</Category>
##   <IItem.Name>Fiction Novel</IItem.Name>
##   <Item.ID>401</Item.ID>
##   <Brand>-</Brand>
##   <Price>14.99</Price>
##   <Variation.ID>401-B</Variation.ID>
##   <Variation.Details>Format: Paperback, Language: Spanish</Variation.Details>
## </record>
## <record>
##   <Category>Books</Category>
##   <IItem.Name>Non-Fiction Guide</IItem.Name>
##   <Item.ID>402</Item.ID>
##   <Brand>-</Brand>
##   <Price>24.99</Price>
##   <Variation.ID>402-A</Variation.ID>
##   <Variation.Details>Format: eBook, Language: English</Variation.Details>
## </record>
## <record>
##   <Category>Books</Category>
##   <IItem.Name>Non-Fiction Guide</IItem.Name>
##   <Item.ID>402</Item.ID>
##   <Brand>-</Brand>
##   <Price>24.99</Price>
##   <Variation.ID>402-B</Variation.ID>
##   <Variation.Details>Format: Paperback, Language: French</Variation.Details>
## </record>
## <record>
##   <Category>Sports Equipment</Category>
##   <IItem.Name>Basketball</IItem.Name>
##   <Item.ID>501</Item.ID>
##   <Brand>SportsGear</Brand>
##   <Price>29.99</Price>
##   <Variation.ID>501-A</Variation.ID>
##   <Variation.Details>Size: Size 7, Color: Orange</Variation.Details>

```

```

## </record>
## <record>
##   <Category>Sports Equipment</Category>
##   <IItem.Name>Tennis Racket</IItem.Name>
##   <Item.ID>502</Item.ID>
##   <Brand>RacketPro</Brand>
##   <Price>89.99</Price>
##   <Variation.ID>502-A</Variation.ID>
##   <Variation.Details>Material: Graphite, Color: Black</Variation.Details>
## </record>
## <record>
##   <Category>Sports Equipment</Category>
##   <IItem.Name>Tennis Racket</IItem.Name>
##   <Item.ID>502</Item.ID>
##   <Brand>RacketPro</Brand>
##   <Price>89.99</Price>
##   <Variation.ID>502-B</Variation.ID>
##   <Variation.Details>Material: Aluminum, Color: Silver</Variation.Details>
## </record>
## </root>

```

```

#read xml file
xml_parsed<- xmlParse("data.xml")
df_from_xml<- xmlToDataFrame(nodes = getNodeSet(xml_parsed, "//record"))
print(df_from_xml)

```

|       | Category                                  | IItem.Name        | Item.ID | Brand      | Price   | Variation.ID |
|-------|---|-------------------|---------|------------|---------|--------------|
| ## 1  | Electronics                               | Smartphone        | 101     | TechBrand  | 699.99  | 101-A        |
| ## 2  | Electronics                               | Smartphone        | 101     | TechBrand  | 699.99  | 101-B        |
| ## 3  | Electronics                               | Laptop            | 102     | CompuBrand | 1099.99 | 102-A        |
| ## 4  | Electronics                               | Laptop            | 102     | CompuBrand | 1099.99 | 102-B        |
| ## 5  | Home Appliances                           | Refrigerator      | 201     | HomeCool   | 899.99  | 201-A        |
| ## 6  | Home Appliances                           | Refrigerator      | 201     | HomeCool   | 899.99  | 201-B        |
| ## 7  | Home Appliances                           | Washing Machine   | 202     | CleanTech  | 499.99  | 202-A        |
| ## 8  | Home Appliances                           | Washing Machine   | 202     | CleanTech  | 499.99  | 202-B        |
| ## 9  | Clothing                                  | T-Shirt           | 301     | FashionCo  | 19.99   | 301-A        |
| ## 10 | Clothing                                  | T-Shirt           | 301     | FashionCo  | 19.99   | 301-B        |
| ## 11 | Clothing                                  | T-Shirt           | 301     | FashionCo  | 19.99   | 301-C        |
| ## 12 | Clothing                                  | Jeans             | 302     | DenimWorks | 49.99   | 302-A        |
| ## 13 | Clothing                                  | Jeans             | 302     | DenimWorks | 49.99   | 302-B        |
| ## 14 | Books                                     | Fiction Novel     | 401     | -          | 14.99   | 401-A        |
| ## 15 | Books                                     | Fiction Novel     | 401     | -          | 14.99   | 401-B        |
| ## 16 | Books                                     | Non-Fiction Guide | 402     | -          | 24.99   | 402-A        |
| ## 17 | Books                                     | Non-Fiction Guide | 402     | -          | 24.99   | 402-B        |
| ## 18 | Sports Equipment                          | Basketball        | 501     | SportsGear | 29.99   | 501-A        |
| ## 19 | Sports Equipment                          | Tennis Racket     | 502     | RacketPro  | 89.99   | 502-A        |
| ## 20 | Sports Equipment                          | Tennis Racket     | 502     | RacketPro  | 89.99   | 502-B        |
| ##    | Variation.Details                         |                   |         |            |         |              |
| ## 1  | Color: Black, Storage: 64GB               |                   |         |            |         |              |
| ## 2  | Color: White, Storage: 128GB              |                   |         |            |         |              |
| ## 3  | Color: Silver, Storage: 256GB             |                   |         |            |         |              |
| ## 4  | Color: Space Gray, Storage: 512GB         |                   |         |            |         |              |
| ## 5  | Color: Stainless Steel, Capacity:20 cu ft |                   |         |            |         |              |
| ## 6  | Color: White, Capacity: 18 cu ft          |                   |         |            |         |              |
| ## 7  | Type: Front Load, Capacity:4.5 cu ft      |                   |         |            |         |              |

```
## 8      Type: Top Load, Capacity:5.0 cu ft
## 9      Color: Blue, Size: S
## 10     Color: Red, Size: M
## 11     Color: Green, Size: L
## 12     Color: Dark Blue, Size: 32
## 13     Color: Light Blue, Size: 34
## 14     Format: Hardcover, Language: English
## 15     Format: Paperback, Language: Spanish
## 16     Format: eBook, Language: English
## 17     Format: Paperback, Language: French
## 18     Size: Size 7, Color: Orange
## 19     Material: Graphite, Color: Black
## 20     Material: Aluminum, Color: Silver
```

## Parquet

```
library(arrow)
```

```
##
## Attaching package: 'arrow'

## The following object is masked from 'package:utils':
##
##      timestamp
```

```
#save dataframe as parquet
write_parquet(df, "data.parquet")

#read parquet file
df_parquet<- read_parquet("data.parquet")
print(df_parquet)
```

```
##      Category      Item.Name Item.ID      Brand      Price Variation.ID
## 1      Electronics      Smartphone      101      TechBrand      699.99      101-A
## 2      Electronics      Smartphone      101      TechBrand      699.99      101-B
## 3      Electronics      Laptop      102      CompuBrand      1099.99      102-A
## 4      Electronics      Laptop      102      CompuBrand      1099.99      102-B
## 5      Home Appliances      Refrigerator      201      HomeCool      899.99      201-A
## 6      Home Appliances      Refrigerator      201      HomeCool      899.99      201-B
## 7      Home Appliances      Washing Machine      202      CleanTech      499.99      202-A
## 8      Home Appliances      Washing Machine      202      CleanTech      499.99      202-B
## 9      Clothing      T-Shirt      301      FashionCo      19.99      301-A
## 10     Clothing      T-Shirt      301      FashionCo      19.99      301-B
## 11     Clothing      T-Shirt      301      FashionCo      19.99      301-C
## 12     Clothing      Jeans      302      DenimWorks      49.99      302-A
## 13     Clothing      Jeans      302      DenimWorks      49.99      302-B
## 14     Books      Fiction Novel      401      -      14.99      401-A
## 15     Books      Fiction Novel      401      -      14.99      401-B
## 16     Books      Non-Fiction Guide      402      -      24.99      402-A
## 17     Books      Non-Fiction Guide      402      -      24.99      402-B
## 18     Sports Equipment      Basketball      501      SportsGear      29.99      501-A
## 19     Sports Equipment      Tennis Racket      502      RacketPro      89.99      502-A
## 20     Sports Equipment      Tennis Racket      502      RacketPro      89.99      502-B
##
##      Variation.Details
## 1      Color: Black, Storage: 64GB
```

```

## 2          Color: White, Storage: 128GB
## 3          Color: Silver, Storage: 256GB
## 4          Color: Space Gray, Storage: 512GB
## 5 Color: Stainless Steel, Capacity:20 cu ft
## 6          Color: White, Capacity: 18 cu ft
## 7          Type: Front Load, Capacity:4.5 cu ft
## 8          Type: Top Load, Capacity:5.0 cu ft
## 9          Color: Blue, Size: S
## 10         Color: Red, Size: M
## 11         Color: Green, Size: L
## 12         Color: Dark Blue, Size: 32
## 13         Color: Light Blue, Size: 34
## 14         Format: Hardcover, Language: English
## 15         Format: Paperback, Language: Spanish
## 16         Format: eBook, Language: English
## 17         Format: Paperback, Language: French
## 18         Size: Size 7, Color: Orange
## 19         Material: Graphite, Color: Black
## 20         Material: Aluminum, Color: Silver

```

## Pros and Cons

JSON is best for web based applications and APIs and for loading into JavaScript. However it does not have schema support and namespace support. HTML is good for web visualizations and reports and easy sharing. It is simple to edit. However it needs CSS and JavaScript for dynamic content and styling. XML is best for structured data exchange and has schema support and namespace support. However it is more complicated to convert a dataframe to xml. Parquet is best for large datasets because it is very fast and uses low storage. It is good for column aggregate queries so if dataframe has many columns it is better. However for a data frame with many records and not many columns there are better options.