

Foundation of Data Science-UCS548

Dashboard Project

By Sourav Choubey

102016075

3CS11

DATASET USED: GLOBAL SUPERSTORE

LINK FOR THE SAME : <https://www.kaggle.com/datasets/tahir1413/global-superstore-2016>

Initially I was having dataset of dataset of dimension 51290 X 24

```
> dim(data)
[1] 51290    24
```

After Data refining and filtering I am reduced to 48690 X 26

```
> dim(df)
[1] 48690    26
```

COLUMN NAMES BEFORE FILTERING AND REFINING

```
> colnames(data)
[1] "Row.ID"      "Order.ID"      "Order.Date"      "Ship.Date"
[5] "Ship.Mode"   "Customer.ID"   "Customer.Name"   "Segment"
[9] "Postal.Code" "City"          "State"           "Country"
[13] "Region"      "Market"        "Product.ID"      "Category"
[17] "Sub.Category" "Product.Name"  "Sales"           "Quantity"
[21] "Discount"    "Profit"        "Shipping.Cost"   "Order.Priority"
```

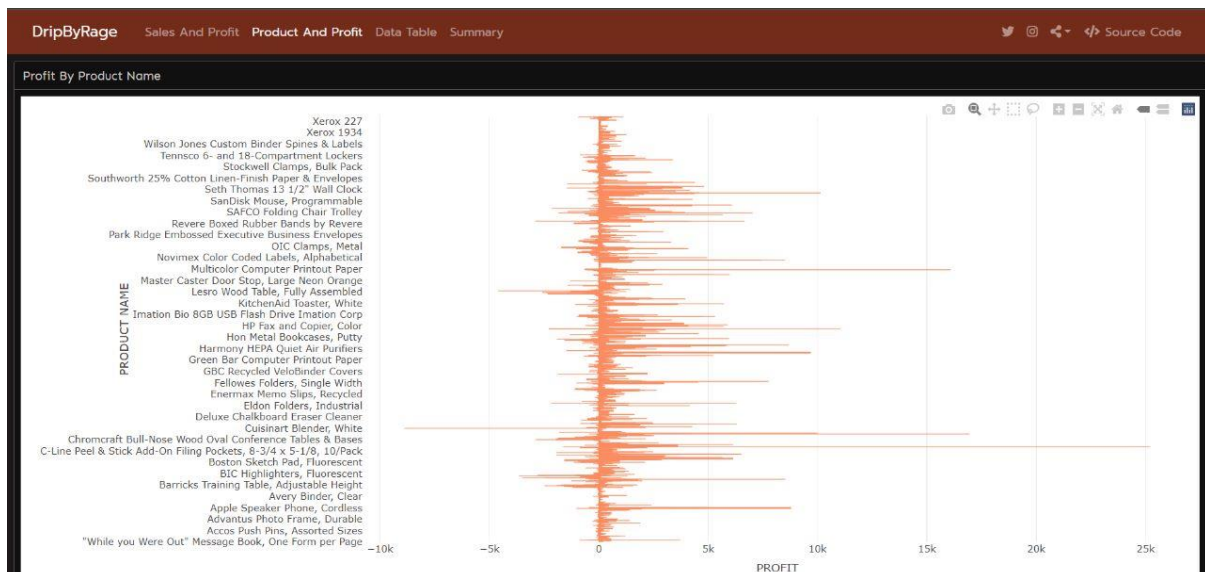
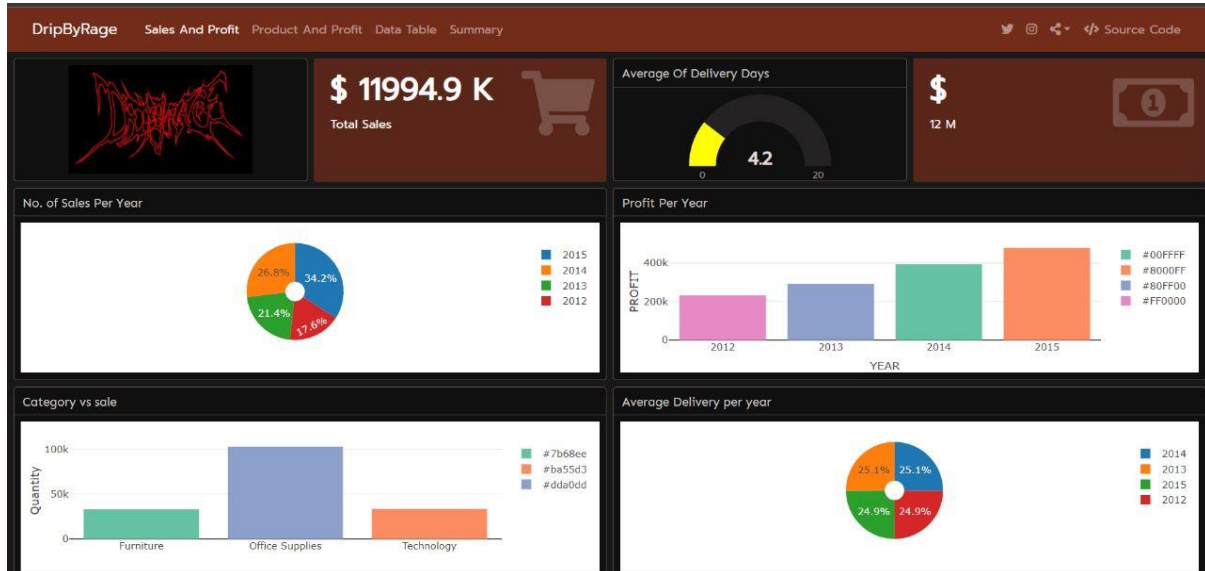
COLUMN NAMES AFTER FILTERING AND REFINING

```
> colnames(df)
[1] "X"           "Row.ID"      "Order.ID"      "Order.Date"
[5] "Ship.Date"   "Ship.Mode"   "Delivery"       "year"
[9] "State"       "Country"     "Region"        "Market"
[13] "Product.ID"  "Customer.ID" "Customer.Name"  "Segment"
[17] "City"        "Category"    "Sub.Category"   "Product.Name"
[21] "Sales"       "Quantity"    "Discount"       "Profit"
[25] "Shipping.Cost" "Order.Priority"
```

DASHBOARD

Inside Dashboard I made 4 pages with source code embedded

FOLLOWING ARE THE SCREENSHOTS OF THE DASHBOARD



DripByRage

Sales And ProfitProduct And ProfitData TableSummary

<

DripByRage

Sales And Profit

Product And Profit

Data Table

Summary

48690

No. Of Rows

3780

Diff. Type of Products

165

Diff. Countries for Sales

169458

Quantity Of Sales

Summary

This Project Uses following Library

1.flexdashboard

2.knitr

3.DT

4.ggplot2

5.dplyr

6.plotly

7.openintro

Basic Work Flow is Data Fetching-> Data Cleaning->Data Visualisation

Here we Plotted Different Graphs and charts using ploty,ggplot etc.

Designed by: Sourav Choubey

Enroll. No:102016075

Group: 3CS11

Project: DashBoard On General SuperStore

[Source Code](#)

Twitter
Facebook
LinkedIn
Pinterest

DATA CLEANING

I refined my dataset, removed unwanted column and added sum new columns

FOLLOWING ARE THE QUERIES OF DATA CLEANING

```
library("dplyr")
# getwd()
# setwd("C:/Users/sourav/OneDrive/Desktop")
#Reading the main data in the form of data.Frame.....
data<- data.frame(read.csv("C:/Users/sourav/OneDrive/Desktop/global_superstore_2016.csv"))
colnames(data)
dim(data)

#Analyzing our columns.
str(data)
#Since profit and sales looks like it must be numeric but after analyzing the structure of data its character type hence some pollution in our data

#Dividing rows into different data frame for some cleaning.....
df1<- data.frame(data[,c(1:5)])

df2<-data.frame(data[,c(1,6:10)])

df3<-data.frame(data[,c(1,11:15)])

df4<- data.frame(data[,c(1,16:20)])

df5<- data.frame(data[,c(1,21:24)])

#removing the column with null values.....
df1[sapply(df1,function(x) !all(sapply(x,is.null)))]
df2[sapply(df2,function(x) !all(sapply(x,is.null)))]
df3[sapply(df3,function(x) !all(sapply(x,is.null)))]
df4[sapply(df4,function(x) !all(sapply(x,is.null)))]
df5[sapply(df5,function(x) !all(sapply(x,is.null)))]

#Removing the column with na value.....
df1<-df1[, colSums(is.na(df1))==0]
df2<-df2[, colSums(is.na(df2))==0]
df3<-df3[, colSums(is.na(df3))==0]
df4<-df4[, colSums(is.na(df4))==0]
df5<-df5[, colSums(is.na(df5))==0]

dim(df1)
dim(df2)
dim(df3)
dim(df4)
dim(df5)

#creating a new column with delivery days.....
df1$Order.Date<-as.Date(df1$Order.Date,format = "%m/%d/%Y")
df1$Ship.Date<- as.Date(df1$Ship.Date,format = "%m/%d/%Y")
df1$Deliverydays <-as.numeric(difftime(df1$Ship.Date,df1$Order.Date,units=c("days"))))

#Removing the rows of delivery column containing 0 as the day we are ordering and shipping are the same hence will not contribute in our analysis
df1<-df1[df1$Deliverydays != 0,]

dim(df1)
#Removing the $ sign from the column containing the $ sign.....
df4$Sales = as.numeric(gsub("[\\$,]", "", df4$Sales))
df5$Profit = as.numeric(gsub("[\\$,]", "", df5$Profit))

#Creating a new column year so as to make easy access.....
df1$year <- format(as.Date(df1$Order.Date, format="%m/%d/%Y"),"%Y")

#Performing Joins.....
df6<-right_join(df3,df2,by="Row.ID")
df7<-left_join(df1,df6,by="Row.ID")
df8<-inner_join(df7,df4,by="Row.ID")
df9<-inner_join(df8,df5,by="Row.ID")
data<-df9

#Creating a new csv.....
write.csv(data,file = "Sourav.csv")

df<-read.csv("Sourav.csv")
dim(df)

OrderDetails<- data.frame(df[,c(1:6,22)])
finance<- data.frame(df[,c(2,20:25)])
Region_details<- data.frame(df[,c(2,9:12)])
Id<-data.frame(df[,c(2,3,13,14)])
Time<-data.frame(df[,c(2,4,5,7,8)])
Product<-data.frame(df[,c(2,13,18:20)])
```

```
write.csv(OrderDetails,file="Order.csv")
write.csv(finance,file="Finance.csv")
write.csv(Region_details,file="Geo.csv")
write.csv(Id,file="Id.csv")
write.csv(Time,file="Time.csv")
write.csv(Product,file="Product.csv")

data1<-data.frame(df[1:5000,])

write.csv(data1,file="table.csv")
```

QUERIES

Since I used r for my dashboard,

SO FOLLOWING ARE THE QUERIES:

title: "DripByRage"

output:

flexdashboard::flex_dashboard:

theme:

version: 4

bg: "#101010"

fg: "#FDF7F7"

primary: "#732C1A"

navbar-bg: "#3ADAC6"

base_font:

google: Prompt

heading_font:

google: Sen

code_font:

google:

family: JetBrains Mono

local: false

```
orientation: rows
vertical_layout: fill
social: ["twitter","instagram","menu"]
source_code: embed
```

```
{r setup, include=FALSE}
```

```
library(flexdashboard)
```

```
library(knitr)
```

```
library(DT)
```

```
library(rpivotTable)
```

```
library(ggplot2)
```

```
library(dplyr)
```

```
library(plotly)
```

```
library(openintro)
```

```
library(highcharter)
```

```
{r}
```

```
mycolors<-c("#4D4DFF", "#FFAD00", "#44D62C", "#D22730")
```

```
{r read}
```

```
data<- read.csv("Sourav.csv")
```

```
data1<- read.csv("table.csv")
```

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

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

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

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

```
Time<-read.csv("Time.csv")
Product1<-read.csv("Product.csv")
```

Sales And Profit

=====

Row

###

{r}

```
include_graphics("DripByRage.jpg",dpi=.5)
```

Total Sales

{r}

```
valueBox(paste("$",format(round(sum(Finance$Sales) / 1e3, 1), trim = TRUE),"K"),
  icon = "fa-shopping-cart")
```

Average Of Delivery Days

{r}

```
gauge(round(mean(Time$Deliverydays),
  digits =1),
  min= 0,
  max = 20,
  gaugeSectors(success = c(0,3),
```

```
warning = c(3,10),  
danger = c(10,20),  
colors = c('red','yellow','green'))))
```

Total Profit by global store

```
{r}  
valueBox("$",paste(format(round(sum(Finance$Sales) / 1e6, 1)), "M"),  
            icon='fa-money')
```

Row

No. of Sales Per Year

```
{r}  
p1<- Time %>%  
  group_by(year) %>%  
  summarise(z = n()) %>%  
  plot_ly(  
    labels = ~year,  
    values = ~z,  
    marker = list(color = mycolors)) %>%  
  add_pie(hole=0.2) %>%  
  layout(xaxis=list(zeroline = F,  
                    showline = F,  
                    showticklabels = F,  
                    showgrid = F),  
         yaxis = list(zeroline = F,
```



```

        showline = F,
        showticklabels = F,
        showgrid = F)
    )
p1

```

```

### Profit Per Year
{r}
da<-right_join(Finance,Time,by="Row.ID")
p2 <- da %>%
  group_by(year) %>%
  summarise(m = sum(Profit)) %>%
  plot_ly(
    x = ~year,
    y = ~m,
    color = rainbow(4),
    type = 'bar') %>%
  layout(xaxis = list(title = "YEAR"),
         yaxis = list(title = "PROFIT"))
p2

```

Row

```

### Category vs sale
{r}
da1<-right_join(Product1,Order,by="Row.ID")

```

```

p2 <- da1 %>%
  group_by(Category) %>%
  summarise(m = sum(Quantity)) %>%
  plot_ly(
    x = ~Category,
    y = ~m,
    color = c("#7b68ee", "#ddaodd", "#ba55d3"),
    type = 'bar') %>%
  layout(xaxis = list(title = ""),
         yaxis = list(title = "Quantity"))
p2

```

```

### Average Delivery per year
{r}
p4 <- Time %>%
  group_by(year) %>%
  summarise(z = mean(Deliverydays))%>%
  plot_ly(
    labels = ~year,
    values = ~z,
    marker = list(color = rainbow(4))) %>%
  add_pie(hole=0.2) %>%
  layout(xaxis=list(zeroline = F,
                    showline = F,
                    showticklabels = F,
                    showgrid = F),
         yaxis = list(zeroline = F,
                      showline = F,

```

```
showticklabels = F,  
showgrid = F))
```

p4

Product And Profit

=====

```
### Profit By Product Name
```

```
{r}
```

```
p3 <- data %>%
```

```
  group_by(Product.Name) %>%
```

```
  summarise(m = sum(Profit)) %>%
```

```
  plot_ly(  
    x = ~m,
```

```
    y = ~Product.Name,
```

```
    color = '99FF1D',
```

```
    type = 'bar') %>%
```

```
  layout(xaxis = list(title = "PROFIT"),
```

```
         yaxis = list(title = "PRODUCT NAME"))
```

p3

Data Table

=====

```
{r}
```

```
datatable(data1,
```

```
  caption = "Global SuperStore",
```

```
  rownames = T,
```

```
filter = "top",
options = list(pageLength=50))
```

```
Summary{data-orientation=columns}
```

```
=====
```

```
Column {data-width = 100}
```

```
-----
```

```
### No. Of Rows
```

```
{r}
```

```
valueBox(nrow(data),
  icon='fa-store')
```

```
### Diff. Type of Products
```

```
{r}
```

```
valueBox(n_distinct(data$Product.Name),
  icon='fa-product-hunt')
```

```
### Diff. Countries for Sales
```

```
{r}
```

```
valueBox(n_distinct(data$Country),
  icon='fa-flag')
```

```
### Quantity Of Sales
```

```
{r}  
valueBox(sum(data$Quantity),  
         icon = 'fa-box')
```

column

Summary

* This Project Uses following Library

\n
1.flexdashboard

\n
2.knitr

\n
3.DT

\n
4.ggplot2

\n
5.dplyr

\n
6.plotly

\n

7.openintro

* Basic Work Flow is Data Fetching-> Data Cleaning->Data Visualisation

* Here we Plotted Different Graphs and charts using ploty,ggplot etc.

Designed by: Sourav Choubey

Enroll. No.:102016075

Group: 3CS11

Project: DashBoard On General SuperStore