Dominic Byerly

---

title: "Mall Customers"

author: "Shwetabh123"

date: "8/6/2021"

output: word\_document

---

# Dataset information

#My dataset is called Mall\_customers and collects data from 200 different mall customers. This dataset collects data on what gender they are, their age, annual income and a spending score that goes from 1-100. The business problem or goal that they wish to accomplish from this dataset, is to better understand the demographics of shoppers who enter the mall in order to better target them for advertising and get them to spend more. The business goal should be to increase sales.`

`#This dataset was retrieved from Kaggle.com and was uploaded 4 years ago by a user by the name of “Shwetabh123”.`

## R Markdown

I used library(rmarkdown) to open R markdown

I used render(“Mall\_Customers”) to render the file and I uploaded the datasheet through Rstudio.

Then i used the read command below to read to me the first 10 rows of the data and display it in the console terminal.

read\_csv("Mall\_Customers")

head(Mall\_Customers)

# A tibble: 6 x 5

CustomerID Genre Age `Annual Income (k$)` `Spending Score (1-100)`

<dbl> <chr> <dbl> <dbl> <dbl>

1 1 Male 19 15 39

2 2 Male 21 15 81

3 3 Female 20 16 6

4 4 Female 23 16 77

5 5 Female 31 17 40

6 6 Female 22 17 76

#My data has 200 rows of observations, meaning that 200 people where interviewed for this survey. This dataset also contains 5 rows that go from customer id, genre, age, annual income and spending score (1-100).

> colnames(Mall\_Customers)

[1] "CustomerID" "Genre" "Age"

[4] "Annual Income (k$)" "Spending Score (1-100)"

I renamed the above variables to make the dataset cleaner and more accurate.

#As you can see above I used that command to list out the names of the columns.

#When it comes to data preparation and errors I do not believe there is too much to do as this comes off as a pretty simple data set. The only possible cleaning that might be necessary would be changing the gender column from being male and female to being 0 and 1, so that it can be a binary variable.

`

> summary(Mall\_Customers)

CustomerID Genre Age Annual Income (k$)

Min. : 1.00 Length:200 Min. :18.00 Min. : 15.00

1st Qu.: 50.75 Class :character 1st Qu.:28.75 1st Qu.: 41.50

Median :100.50 Mode :character Median :36.00 Median : 61.50

Mean :100.50 Mean :38.85 Mean : 60.56

3rd Qu.:150.25 3rd Qu.:49.00 3rd Qu.: 78.00

Max. :200.00 Max. :70.00 Max. :137.00

Spending Score (1-100)

Min. : 1.00

1st Qu.:34.75

Median :50.00

Mean :50.20

3rd Qu.:73.00

Max. :99.00

```

#As we can see above from the summary command we can see a glimpse into the overall output and meaning of the dataset.



#Above we can also see a scatter plot that shows the relationship between annual income and the spending score. While we would expect to see that as spending score goes up so does annual income. But that is not completely the case as we can see a concentration around the middle that says even richer people aren’t spending proportionately more at the mall.

> summary.data.frame(Mall\_Customers)

CustomerID Genre Age Annual Income (k$)

Min. : 1.00 Length:200 Min. :18.00 Min. : 15.00

1st Qu.: 50.75 Class :character 1st Qu.:28.75 1st Qu.: 41.50

Median :100.50 Mode :character Median :36.00 Median : 61.50

Mean :100.50 Mean :38.85 Mean : 60.56

3rd Qu.:150.25 3rd Qu.:49.00 3rd Qu.: 78.00

Max. :200.00 Max. :70.00 Max. :137.00

Spending Score (1-100)

Min. : 1.00

1st Qu.:34.75

Median :50.00

Mean :50.20

3rd Qu.:73.00

Max. :99.00



#However, when we measure the age vs spending score we see a correlation as the older someone is, the less they are likely to spend during their trip to the mall. I found this result to be quite surprising as I would of thought that the older people are the more they spend but that is not the case.

#Perhaps the best strategy for the Mall to employ would be to spend more time advertising promotional campaigns with the younger generation specifically in mind, as it seems the older generation is more wary to spend a great deal at the mall.