

# Data Description

*Author: Rajat Jain*

*Last Updated: 2018-04-23*

## Contents

### Schema

We have first 7-day product usage data from Adobe photography plan users. Data is a CSV file containing 2650 rows. It has following columns:

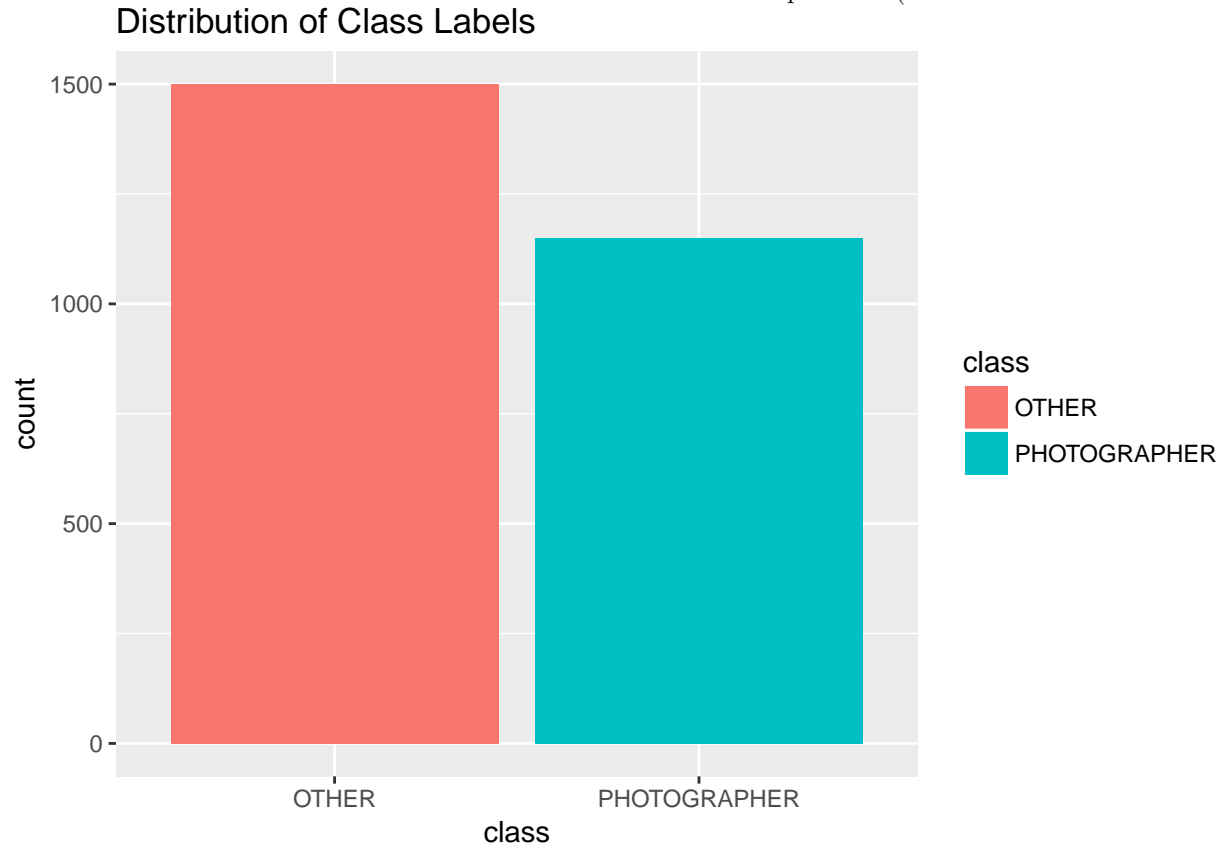
- **member\_guid:** Customer Identifier.
- **class:** Output Class - PHOTOGRAPHER or OTHER.
- **lr\_cc\_usage:** # of times customer used Lightroom CC product in first 7-days.
- **lr\_cl\_usage:** # of times customer used Lightroom Classic product in first 7-days.
- **lr\_mo\_usage:** # of times customer used Lightroom Mobile product in first 7-days.
- **storage\_usage:** # of times customer accessed Cloud Storage in first 7-days.
- **ps\_usage:** # of times customer used Photoshop product in first 7-days.
- **stock\_usage:** # of times customer searched for a Stock Image in first 7-days.

### Data Summary

##	class	lr_cc_usage	lr_cl_usage	lr_mo_usage
##	OTHER :1500	Min. : 0.0000	Min. : 0.000	Min. : 0.00
##	PHOTOGRAPHER:1150	1st Qu.: 0.0000	1st Qu.: 0.000	1st Qu.: 0.00
##		Median : 0.0000	Median : 2.000	Median : 0.00
##		Mean : 0.3475	Mean : 4.217	Mean : 0.86
##		3rd Qu.: 0.0000	3rd Qu.: 6.000	3rd Qu.: 0.00
##		Max. :22.0000	Max. :185.000	Max. :24.00
##	storage_usage	ps_usage	stock_usage	
##	Min. : 0.0	Min. : 0.000	Min. : 0.000	
##	1st Qu.: 0.0	1st Qu.: 0.000	1st Qu.: 0.000	
##	Median : 0.0	Median : 3.000	Median : 0.000	
##	Mean : 300.5	Mean : 4.649	Mean : 1.002	
##	3rd Qu.: 1.0	3rd Qu.: 6.000	3rd Qu.: 0.000	
##	Max. :107556.0	Max. :182.000	Max. :246.000	

## Distribution of Class Labels

The bar chart below shows the distribution of class labels in the input data (in absolute number of records).



It contains **43.4%** records classified as “PHOTOGRAPHER” class and 56.6% records in “OTHER” class.

## Binomial Test

Now we run Binomial Test on this data.

```
binom.test(nrow(usage.data[usage.data[,2] == "PHOTOGRAPHER",]), nrow(usage.data))

##
## Exact binomial test
##
## data:  nrow(usage.data[usage.data[, 2] == "PHOTOGRAPHER", ]) and nrow(usage.data)
## number of successes = 1150, number of trials = 2650, p-value =
## 1.127e-11
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.4149853 0.4530853
## sample estimates:
## probability of success
##           0.4339623
```