Statistical Analysis of Professional Photographer's Estimated Return on Investment (ROI) Turnaround Time

Collin Guarino, crguarin@uncg.edu
STA-301: Statistical Methods - Fall 2017
The University of North Carolina at Greensboro

Abstract

Professional photography relies heavily on engaging audiences and delivering to a diverse range of clients on a schedule. The business requires an initial investment of time to be on-site, shoot photographs, and conduct image processing before delivering the final product. The time required in this workflow is difficult for humans to gauge, meaning an unknown duration before a return on investment (ROI) between the time that money is spent to take the photograph and the time the photographs can be licensed or sold. Statistical techniques can be used to extract, analyze, and visualize how a professional[1] photographer can estimate how many days, on average, pass between when a photograph is taken and when the photograph gets uploaded. Since each client requires a specific image category[2], such as "Landscapes" or "Journalism", a more accurate estimate can be calculated on a categorical basis.

Introduction

500px.com is a large social networking site for photography professionals to share and sell images. Unlike most social networking sites, a large majority of images are posted to 500px after an extensive process of transferring large image files from a DSLR to a computer and performing time intensive image touchups. Additionally, professional photography requires a great deal of travel therefore the time between taking a photo and uploading it is significantly longer than other websites. 500px.com/popular is a section that lists the top photographs from the website's population of professional photographers.

[1] A "professional photographer" in this study is defined as an individual who has a sufficiently popular following of fans, whose photographs make it to the "popular" section of 500px.com

[2] An "image category" is defined as the primary subject of the photograph

Data Collection

There is no existing public database that sufficiently represents the target population. 500px.com is an outlet for the photographer community and the "popular" section best represents the population that is classified as "professional" workers, subset of the entire professional photographer population. The site has a public developer application interface program (API) hosted at developers.500px.com which can be accessed using an API token[3]. A python script[4] was written to collect all the image metadata and add it to a database so it can be analyzed using R.

Table 1: Excerpt from sample data set[5]

URL	Category	Days To Upload
500px.com/photo/234815327	Landscapes	358
500px.com/photo/234793089	Travel	1070
500px.com/photo/234774847	Landscapes	262
500px.com/photo/234850045	Landscapes	53
500px.com/photo/234840743	Landscapes	33
...	...	...

The "URL" is a direct link to the photograph and the "Category" is parsed from that image's metadata. There are thirty categories on the site, of which the following are collected: Sport, Journalism, Landscapes, and Travel. The "Days To Upload" column is calculated by subtracting the uploaded date from the date that the photograph was taken, both of which are found in the metadata.

[3] Instructions on how to obtain an API token is available at https://github.com/500px/api-documentation

[4] The python script which collected the data set is available for download at https://github.com/Collinux/photo-roi-turnaround/blob/master/collect.py

[5] Table 1: Raw data collection available for download at https://github.com/Collinux/photo-roi-turnaround/blob/master/photos.dat

Table 2: Summary of the collected dataset[6]

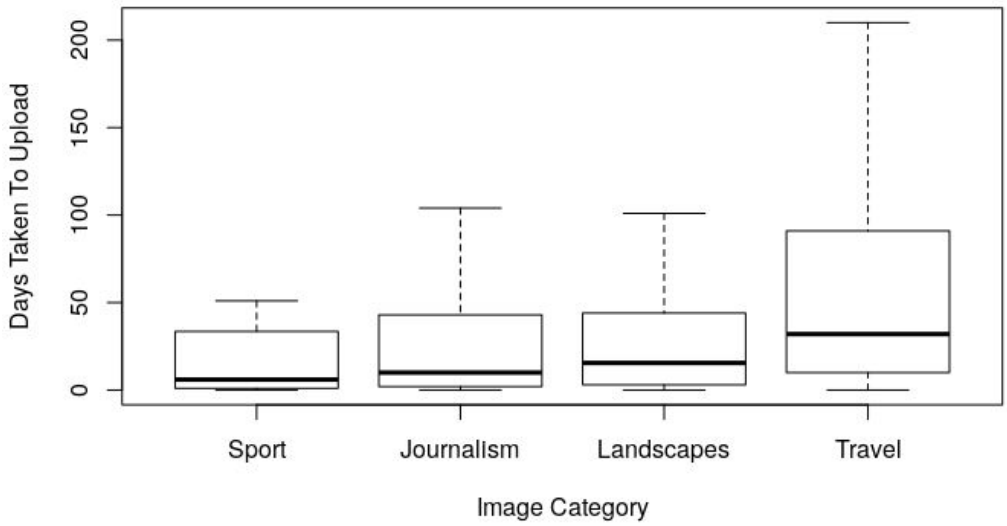
Category	Sport	Journalism	Landscapes	Travel	**TOTAL**
Images Collected	119	51	6216	765	**7151**

Image metadata was for the above categories and all images where the "Days To Upload" were greater than or equal to 365 (days) were omitted from the study. If the difference between the date taken and the date uploaded is over a year then it does not fit into the return on investment for the same fiscal or taxable year.

Analysis

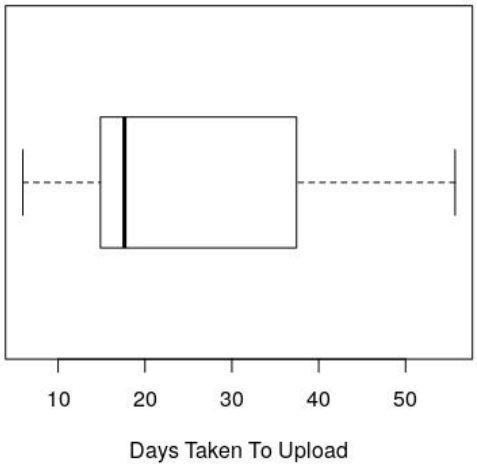
The data collected by the python script was imported into R and analyzed to determine the effect of the image category on the days taken to upload. The independent images appear to be normally distributed with a sufficiently large sample size and the variances appear roughly equal, with the exception that the Travel category has a slightly higher variance than the rest.

Figure 3: Metadata from 500px.com/popular (outliers removed)



[6] Table 2: Summary of dataset available for download at https://github.com/Collinux/photo-roi-turnaround/blob/master/status.txt

Figure 4: Elapsed Days Across All Categories



Across all the sampled image categories, on average, professional photographers take on average between approximately 20 and 22 days to start earning a return on investment for a photograph, 95% confidence .

Given that the image category is “Sport” then professional photographers take on average between approximately 17 and 27 days to start earning a return on investment for the photograph, with 95% confidence.

Given that the image category is “Journalism” then professional photographers take on average between approximately 6 and 13 days to start earning a return on investment for the photograph, with 95% confidence.

Given that the image category is “Landscapes” then professional photographers take on average between approximately 17 and 18 days to start earning a return on investment for the photograph, with 95% confidence.

Given that the image category is “Travel” then professional photographers take on average between approximately 48 and 56 days to start earning a return on investment for the photograph, with 95% confidence.

The above data shows the average upload time ordered as:

μ Journalism < μ Landscapes < μ Sport < μ Travel

To determine how the image category effects the elapsed days a t-test[7] is used to test the following:

H_0: μ Journalism = μ Landscapes
H_a: μ Journalism < μ Landscapes

Since the p-value is nearly 0 there is strong evidence that journalism photographs take professional photographers less time to upload than landscape photographs.

```
> t.test(journalism, landscapes, alternative =
"less")

       Welch Two Sample t-test

data:  journalism and landscapes
t = -4.4014, df = 52.926, p-value = 2.623e-05
alternative hypothesis: true difference in
means is less than 0
95 percent confidence interval:
      -Inf -4.866829
sample estimates:
mean of x mean of y
 9.470588 17.324968
```

H_0: μ Landscapes = μ Sport
H_a: μ Landscapes < μ Sport

Since the p-value is approximately 0.025 there is strong evidence that landscape photographs take professional photographers less time to upload than sports photographs.

```
> t.test(landscapes, sport, alternative =
"less")

       Welch Two Sample t-test

data:  landscapes and sport
t = -1.9823, df = 121.47, p-value = 0.02485
alternative hypothesis: true difference in
means is less than 0
95 percent confidence interval:
       -Inf -0.8087302
sample estimates:
mean of x mean of y
 17.32497  22.26050
```

H_0: μ Sport = μ Travel
H_a: μ Sport < μ Travel

Since the p-value is nearly 0 there is strong evidence that sports photographs take professional photographers less time to upload than travel photographs.

```
> t.test(sport, travel, alternative = "less")

       Welch Two Sample t-test

data:  sport and travel
t = -9.2147, df = 308.79, p-value < 2.2e-16
alternative hypothesis: true difference in
means is less than 0
95 percent confidence interval:
      -Inf -24.18851
sample estimates:
mean of x mean of y
 22.26050  51.72418
```

[7] Data analysis written in R is available at https://github.com/Collinux/photo-roi-turnaround/analysis.r