Report of Facebook metrics

Brief description of the data set and a summary of its attributes.

Title: Facebook performance metrics

Sources: Created by: Sérgio Moro, Paulo Rita and Bernardo Vala (ISCTE-IUL) @ 2016

Past Usage: The full dataset was described and analyzed in S. Moro, P. Rita and B. Vala. Predicting social media performance metrics and evaluation of the impact on brand building a data mining approach. Journal of Business Research, Elsevier, In press, Available online since 28 February 2016.

Relevant Information: The data is related to posts' published during the year of 2014 on the Facebook's page of a renowned cosmetics brand. This dataset contains 500 of the 790 rows and part of the features analyzed by Moro et al. (2016). The remaining were omitted due to confidentiality issues.

Number of Instances: 500

Number of Attributes: 19

I got this dataset from UCI public dataset.

Summary of attribute:

Page total likes : This attribute shows the total number of likes pages have received

Category: To identify the category of page

Post Month: Month of the post

Post Weekday: Day on which the post was uploaded

Post Hour: Identifies the post time

Paid: Whether paid for promotion or not

Lifetime Post Total Reach: Total reach of each post in life time.

Lifetime Post Total Impressions: Total number of people who saw the post.

Lifetime Engaged Users: Total number of people who clicked on post

Lifetime Post Consumers: total number of people who saw your post and bought something.

Lifetime Post Consumptions: Total number of people who visit your post

Lifetime Post Impressions by people who have liked your Page

Lifetime Post reach by people who like your Page: reached to other people by people who liked the post.

Lifetime People who have liked your Page and engaged with your post: Regular visitors of the post.

Comment: Total number of comments for each post

Like: number of likes for each post

Share: number of share(forwards) of the post.

Total Interactions: Total number of people who have interacted with the post.

Initial plan for data exploration

The plan for data exploration was to first identifying the data types of attributes.

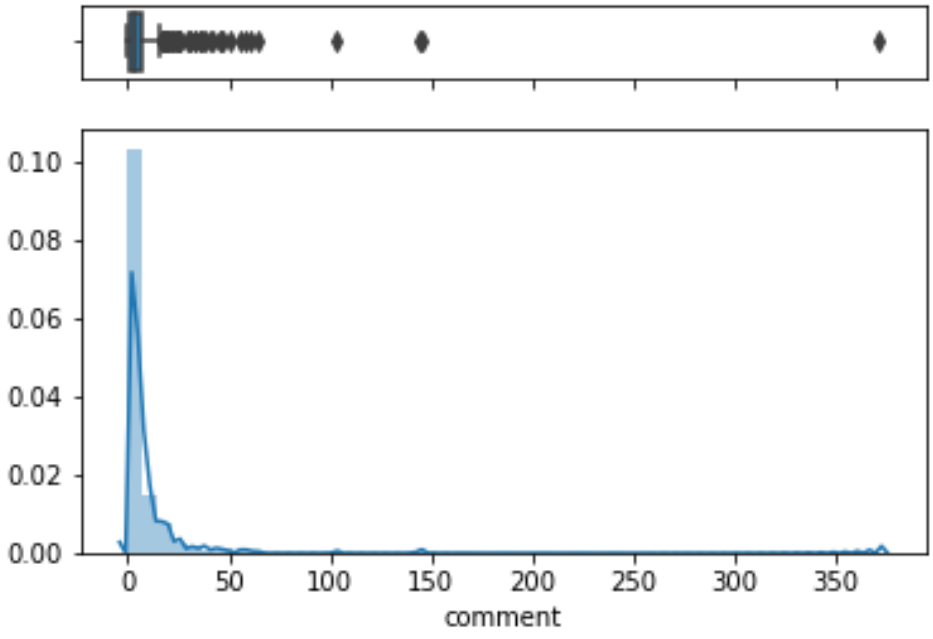
Then identifying the mean, median, minimum, maximum and the IQR of each attribute.

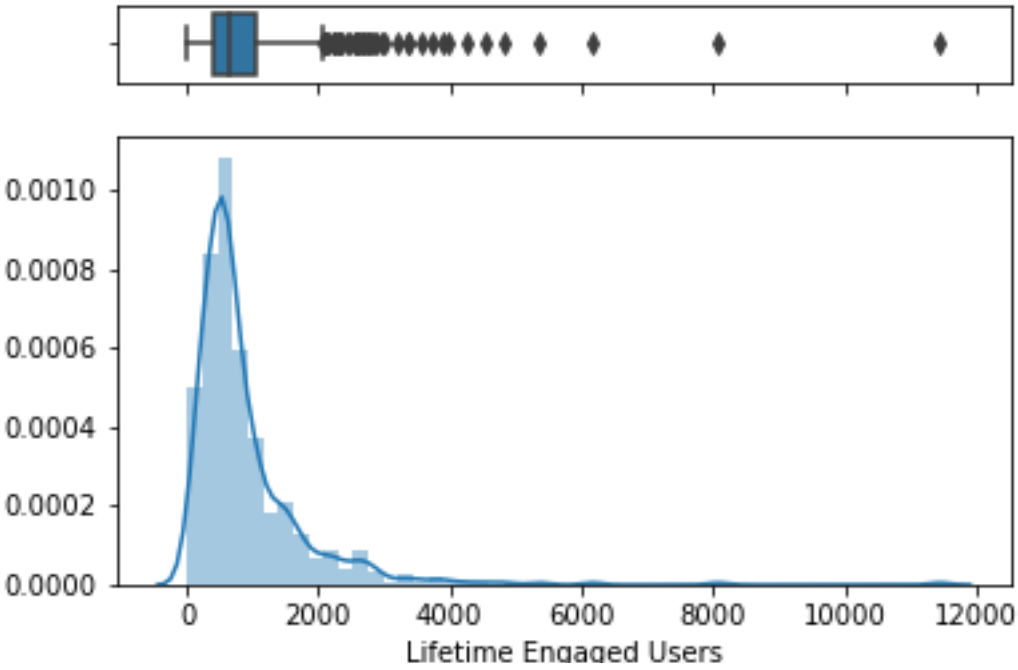
Late the attributes where visualised using distribution plot, box plot(for identifying outliers) and count plots according to data types of the attributes.

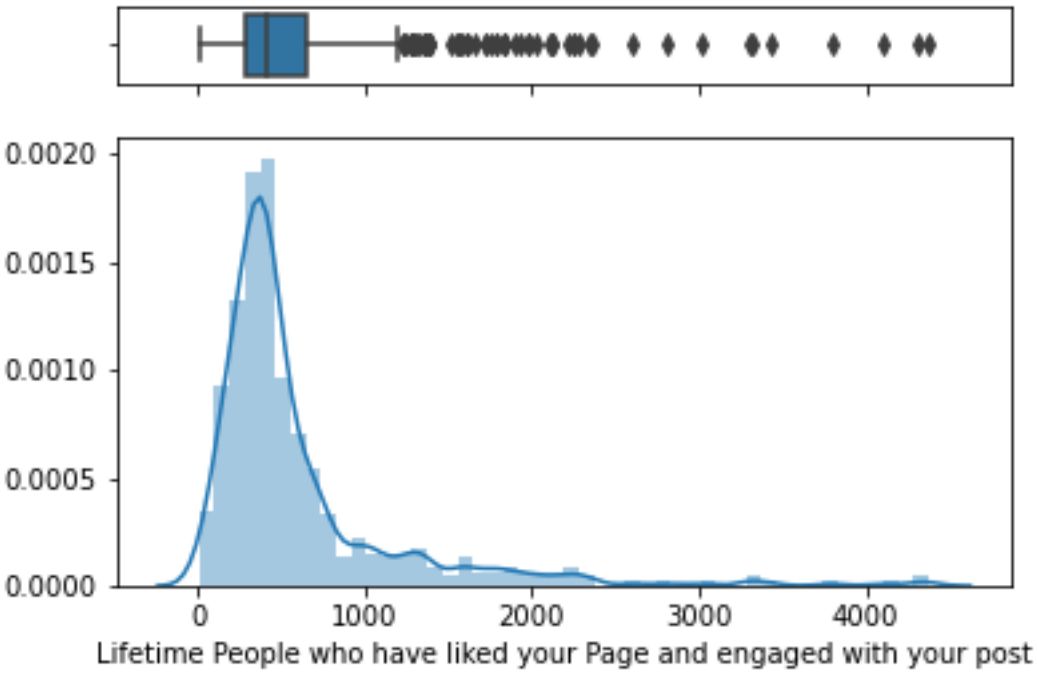
Actions taken for data cleaning and feature engineering

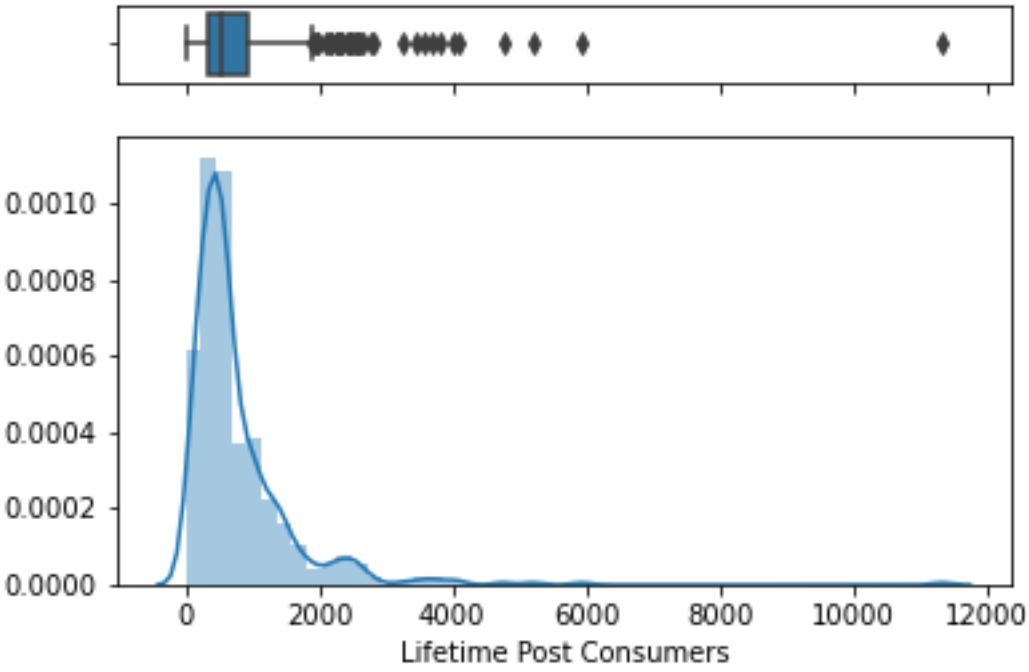
The data was cleaning by filling the missing values with the mean of the columns and the outliers where bought to 3rd inter quartile values of the columns

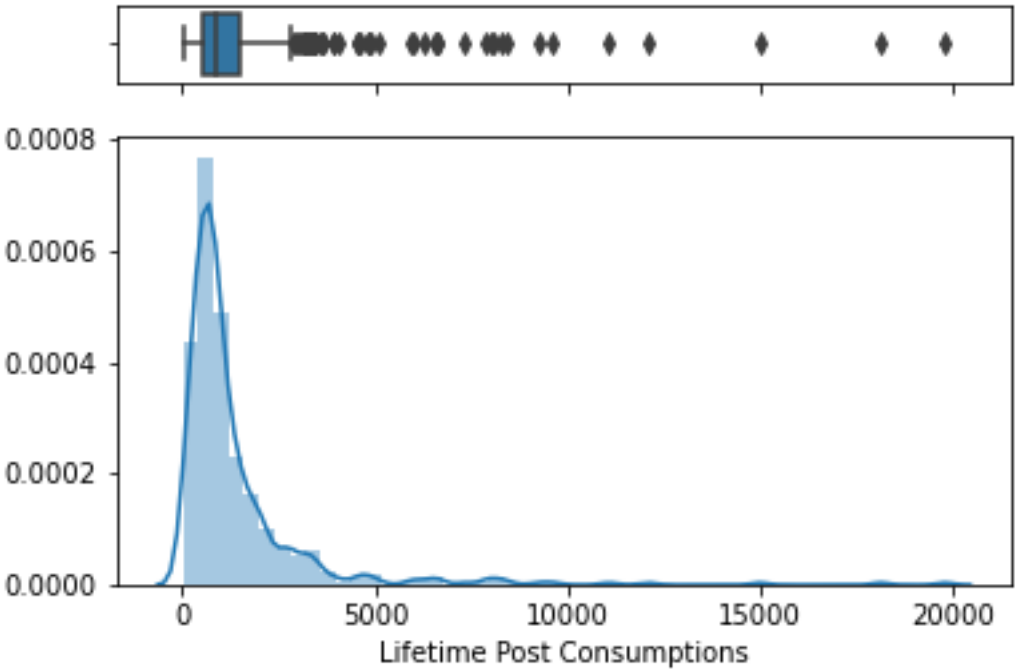
The following were the data visualized before cleaning the data.

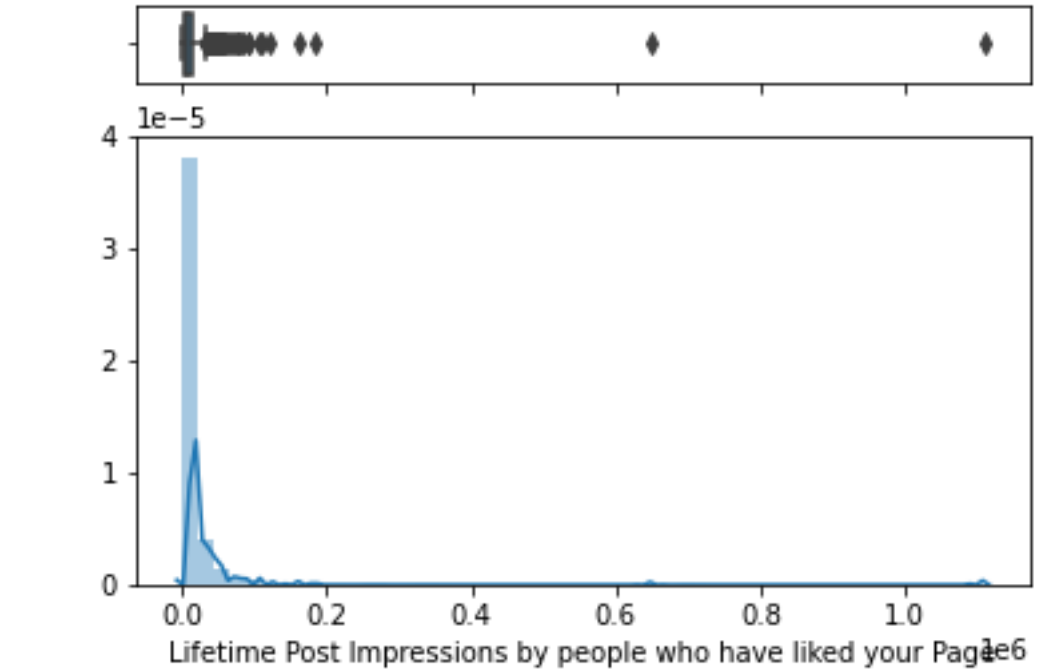


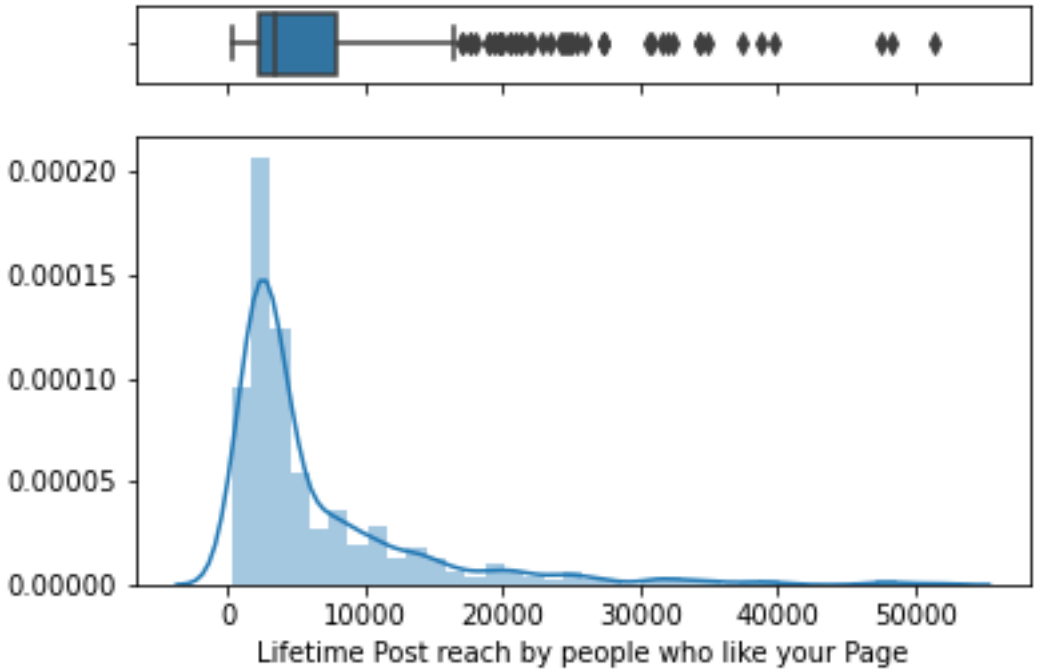

0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
0
2000
4000
6000
8000
10000
12000
Lifetime Engaged Users

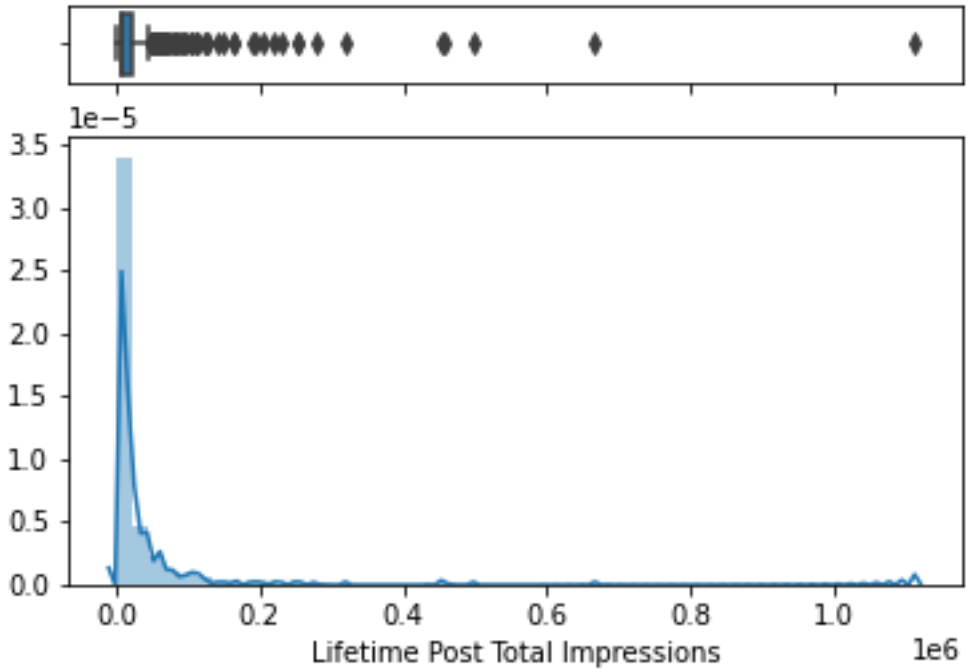

0.0020
0.0015
0.0010
0.0005
0.0000
0
1000
2000
3000
4000
Lifetime People who have liked your Page and engaged with your post

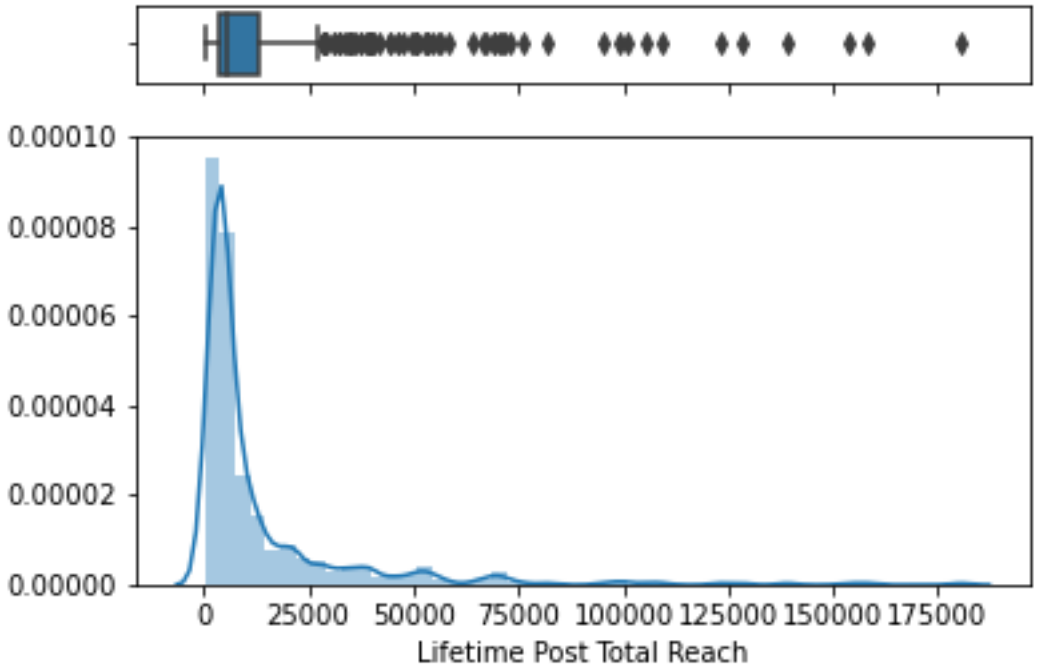

0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
0
2000
4000
6000
8000
10000
12000
Lifetime Post Consumers

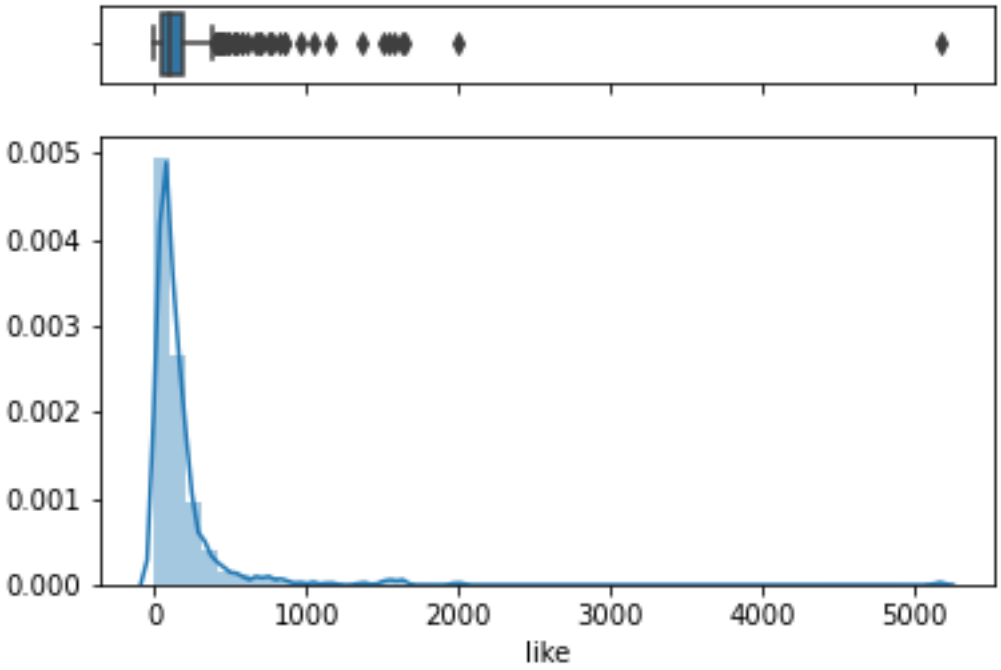

0.0008
0.0006
0.0004
0.0002
0.0000
0
5000
10000
15000
20000
Lifetime Post Consumptions

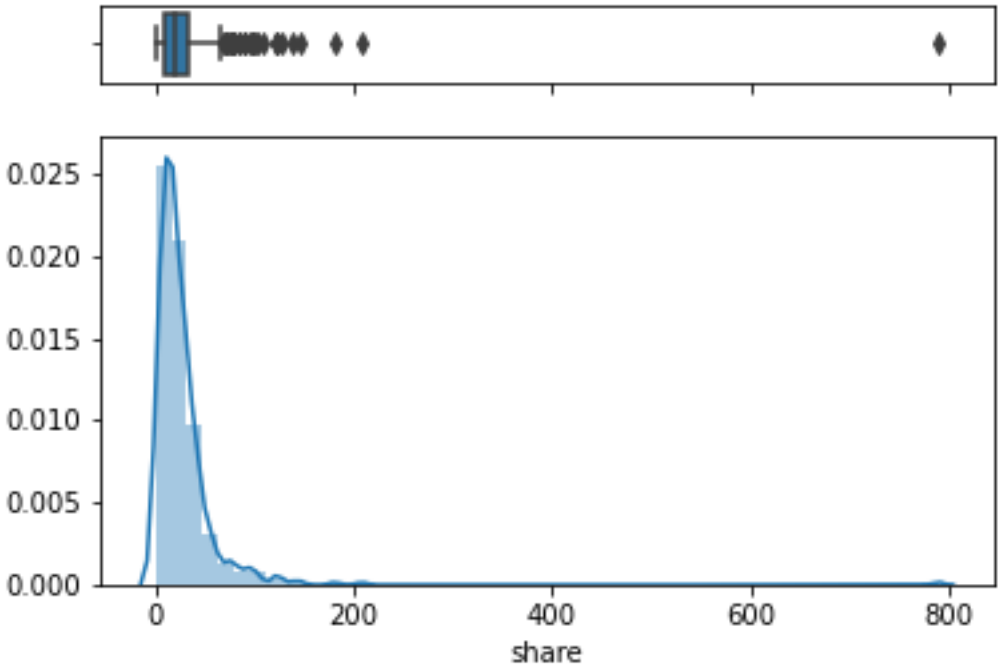

1e−5
4
3
2
1
0
0.0
0.2
0.4
0.6
0.8
1.0
1e6
Lifetime Post Impressions by people who have liked your Page

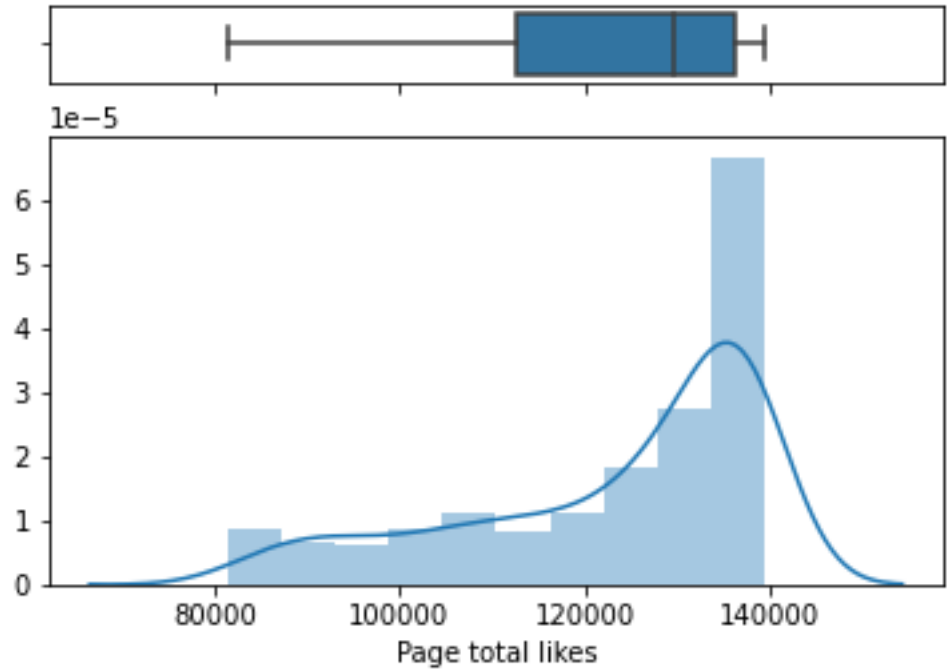

0.00020
0.00015
0.00010
0.00005
0.00000
0
10000
20000
30000
40000
50000
Lifetime Post reach by people who like your Page


1e−5
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0
1e6
Lifetime Post Total Impressions

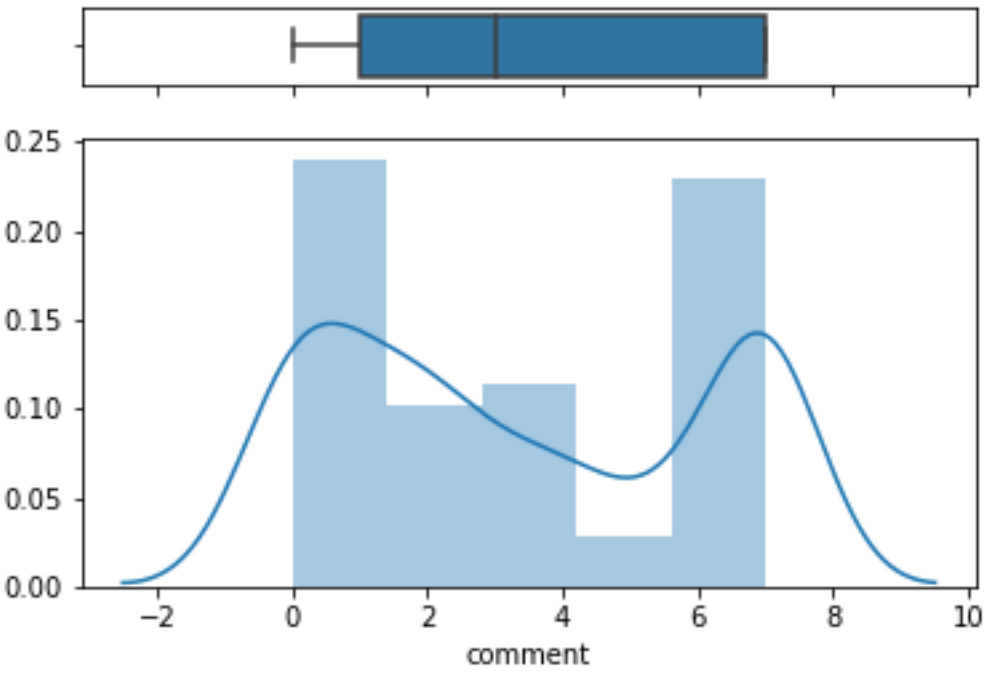

0.00010
0.00008
0.00006
0.00004
0.00002
0.00000
0
25000
50000
75000
100000
125000
150000
175000
Lifetime Post Total Reach

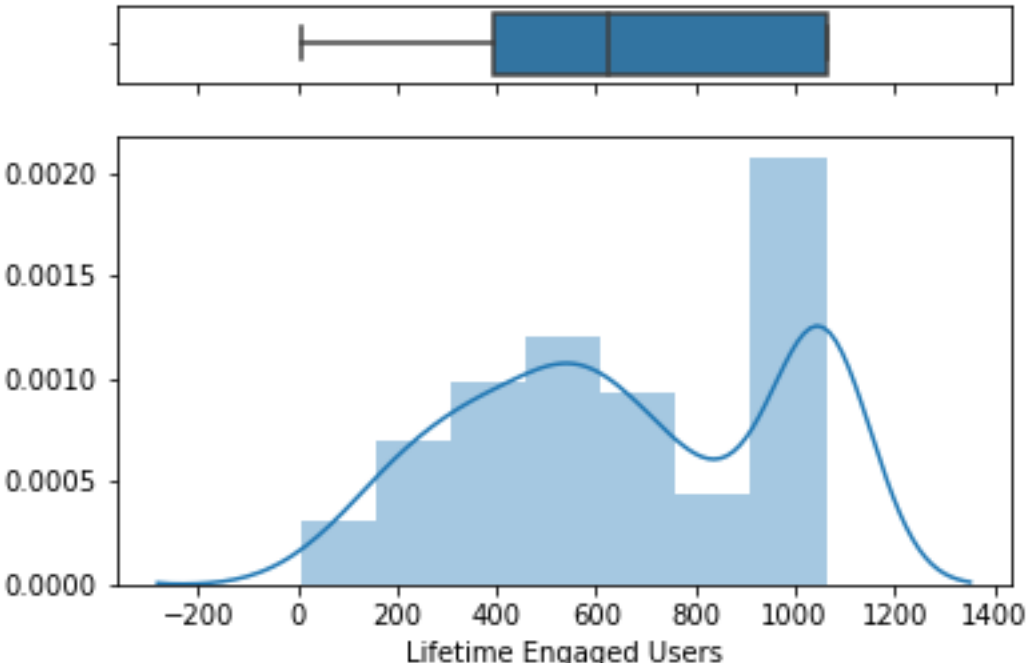

0.005
0.004
0.003
0.002
0.001
0.000
0
1000
2000
3000
4000
5000
like

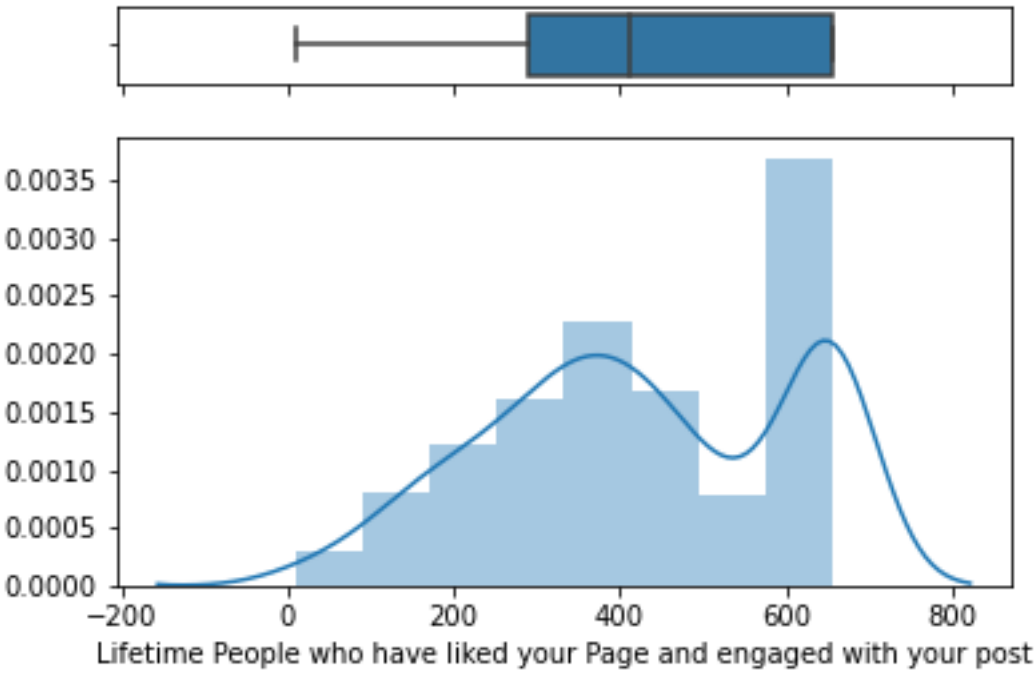

0.025
0.020
0.015
0.010
0.005
0.000
0
200
400
600
800
share

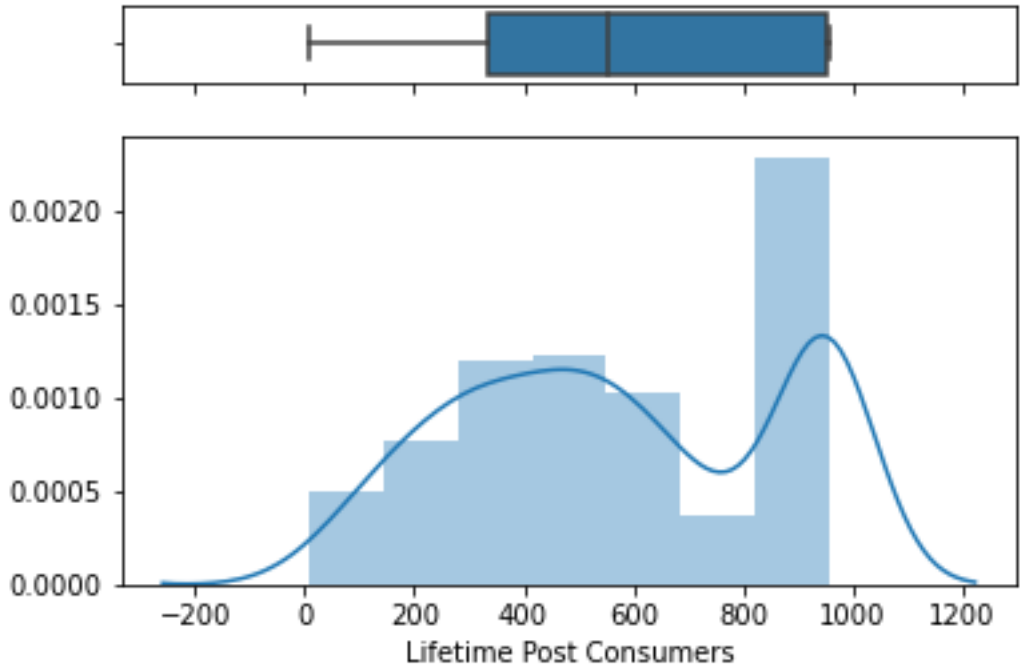


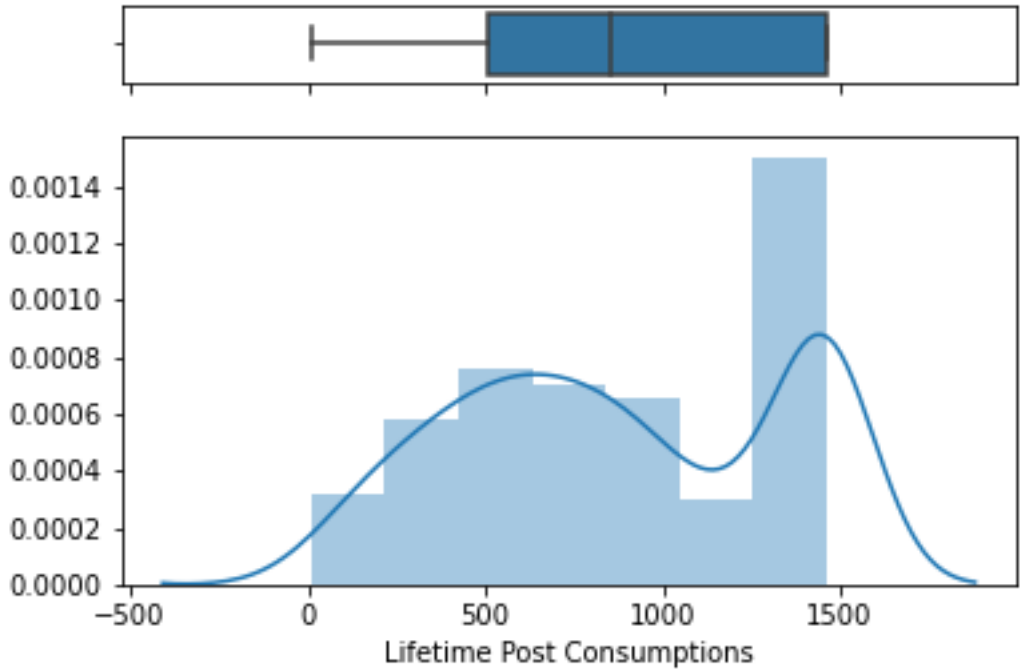
After cleaning the data.

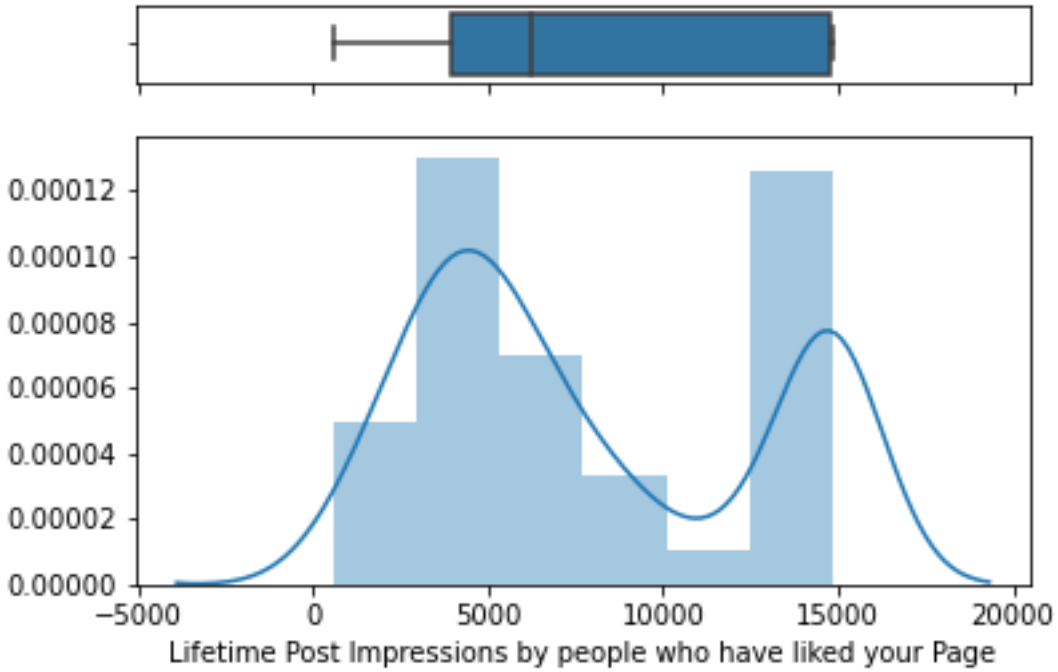


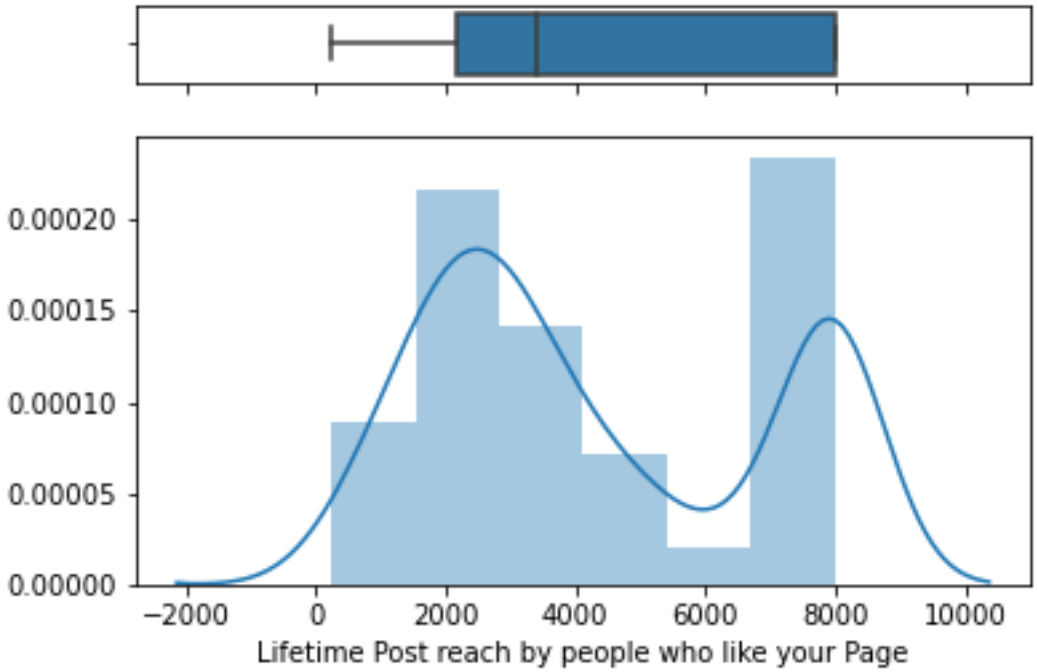

0.0020
0.0015
0.0010
0.0005
0.0000
−200
0
200
400
600
800
1000
1200
1400
Lifetime Engaged Users

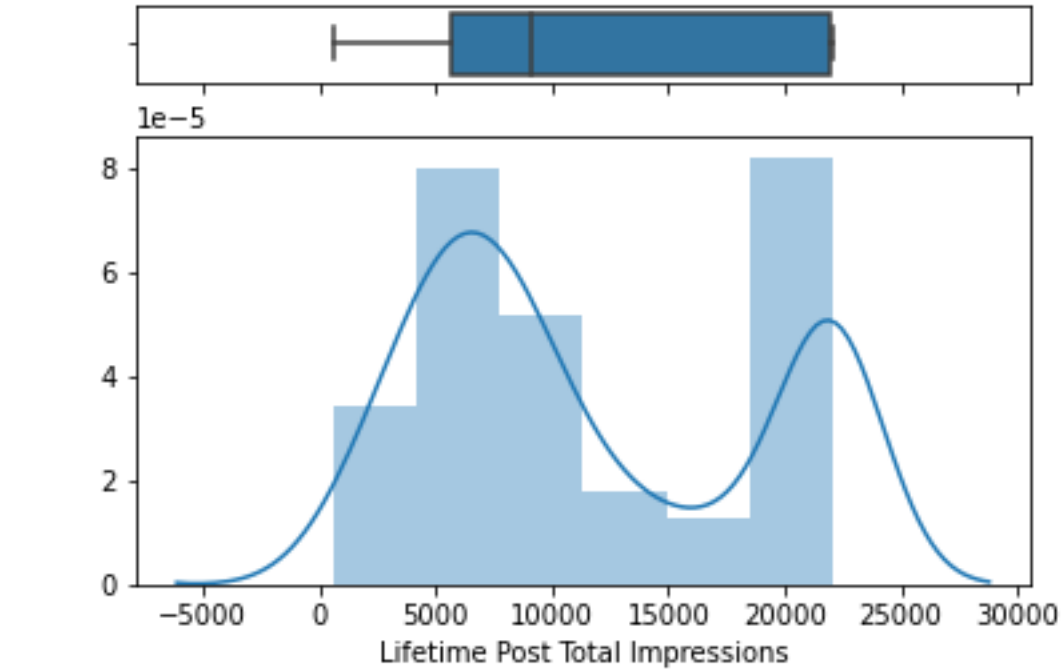

0.0035
0.0030
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
−200
0
200
400
600
800
Lifetime People who have liked your Page and engaged with your post

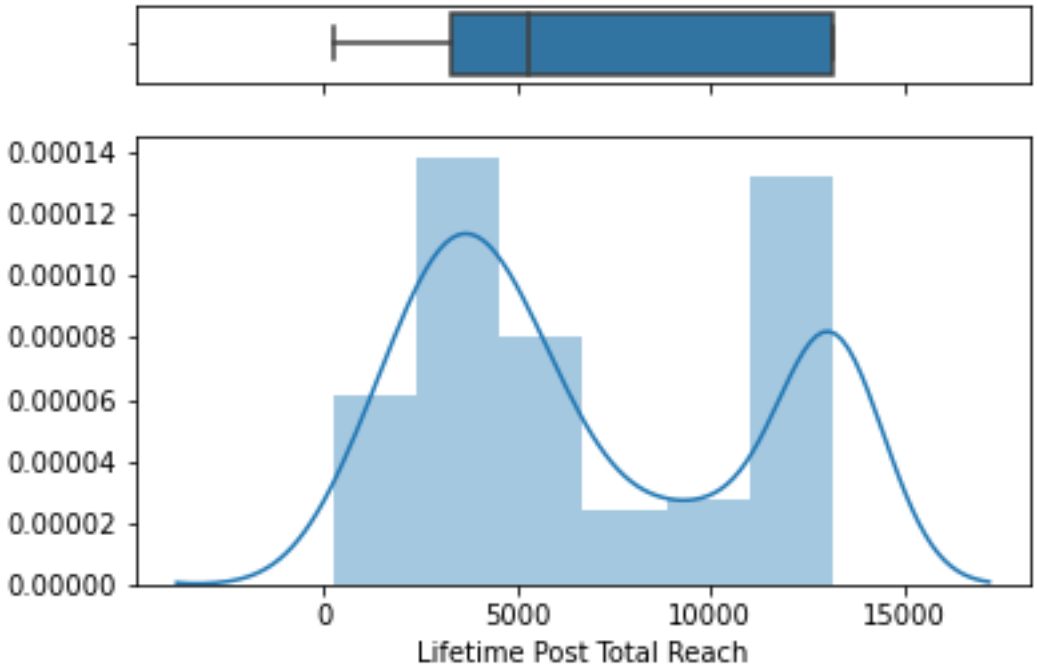

0.0020
0.0015
0.0010
0.0005
0.0000
−200
0
200
400
600
800
1000
1200
Lifetime Post Consumers

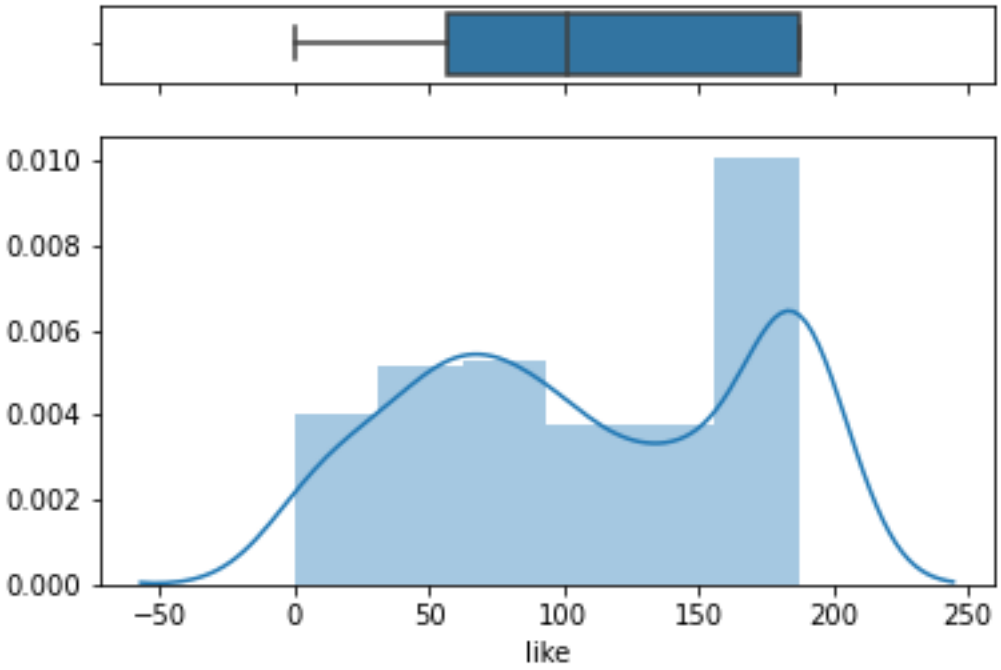

0.0014
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
−500
0
500
1000
1500
Lifetime Post Consumptions

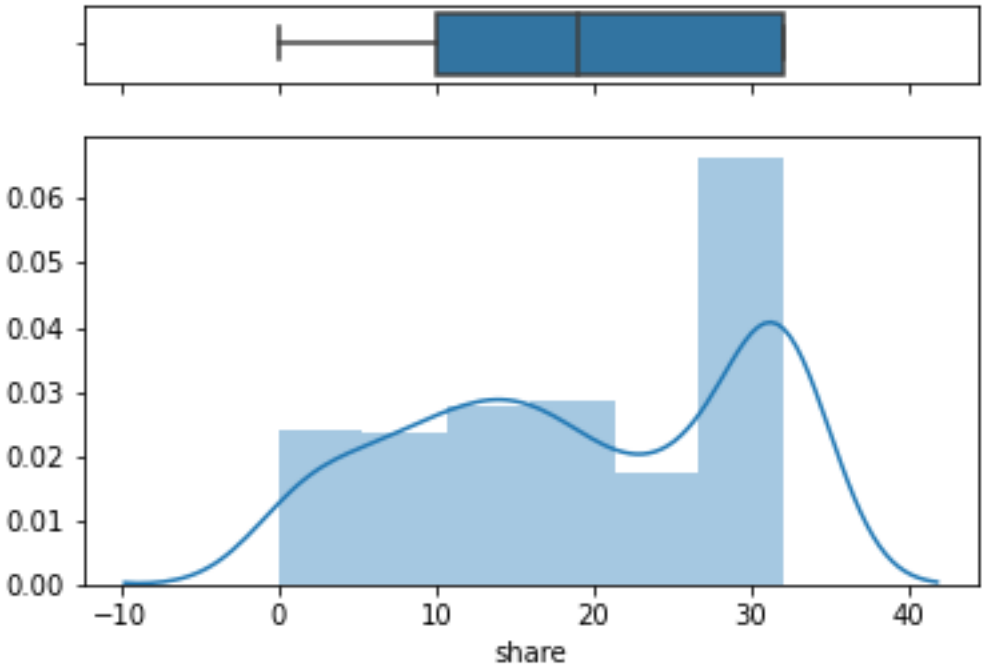

0.00012
0.00010
0.00008
0.00006
0.00004
0.00002
0.00000
−5000
0
5000
10000
15000
20000
Lifetime Post Impressions by people who have liked your Page


0.00020
0.00015
0.00010
0.00005
0.00000
−2000
0
2000
4000
6000
8000
10000
Lifetime Post reach by people who like your Page

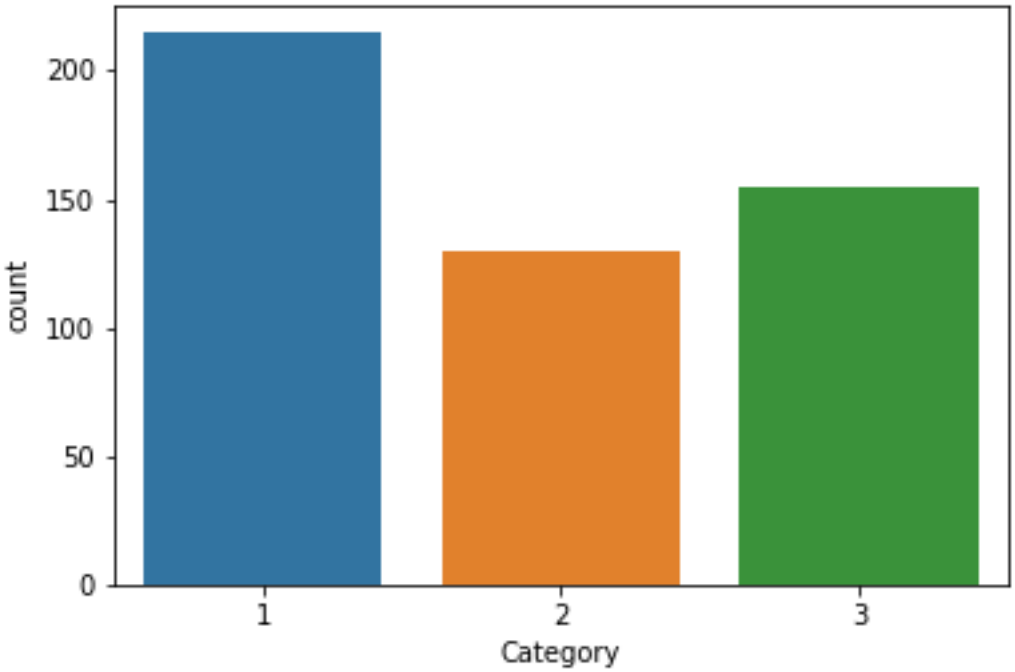

1e−5
8
6
4
2
0
−5000
0
5000
10000
15000
20000
25000
30000
Lifetime Post Total Impressions

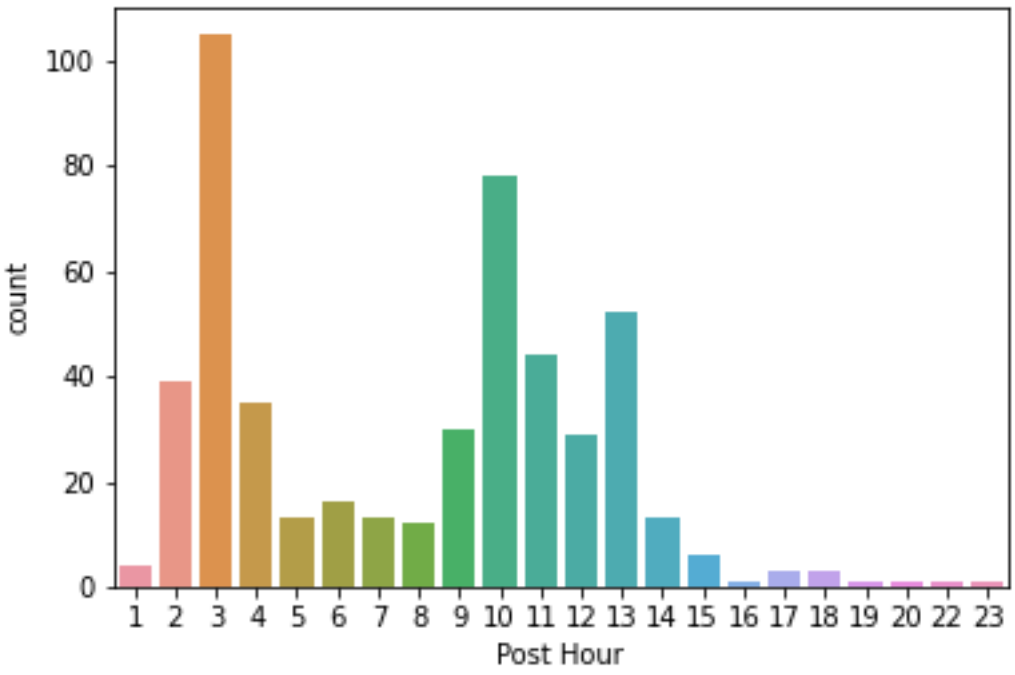

0.00014
0.00012
0.00010
0.00008
0.00006
0.00004
0.00002
0.00000
0
5000
10000
15000
Lifetime Post Total Reach

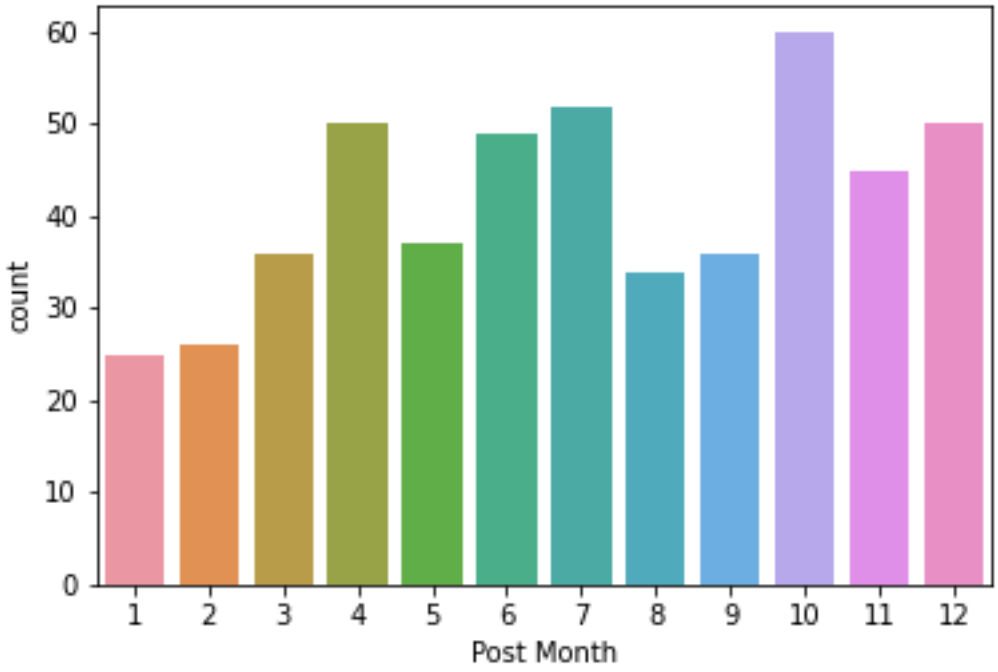

0.010
0.008
0.006
0.004
0.002
0.000
−50
0
50
100
150
200
250
like

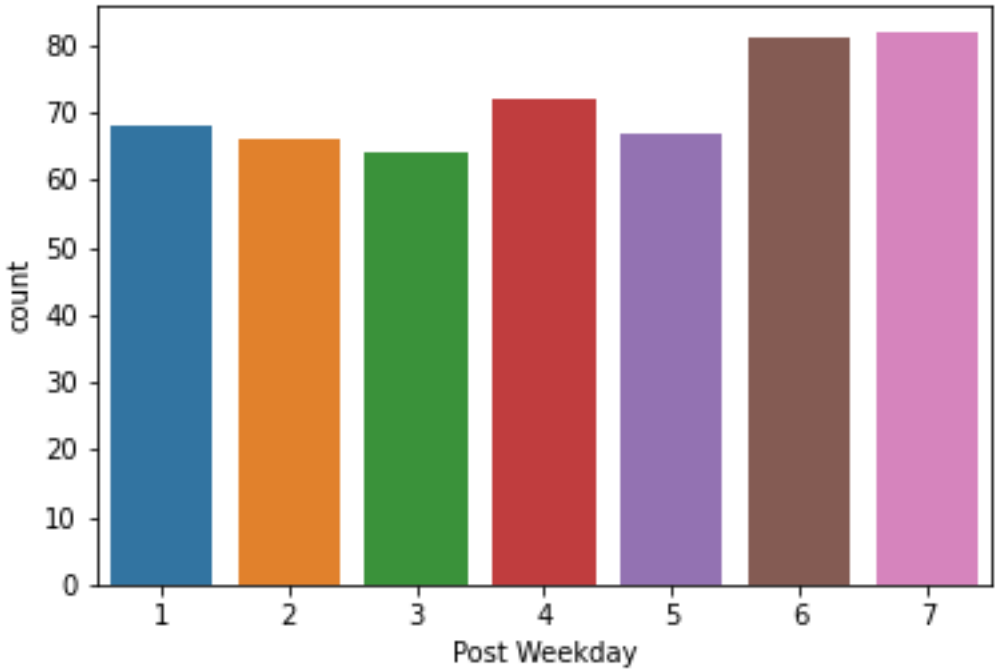

0.06
0.05
0.04
0.03
0.02
0.01
0.00
−10
0
10
20
30
40
share

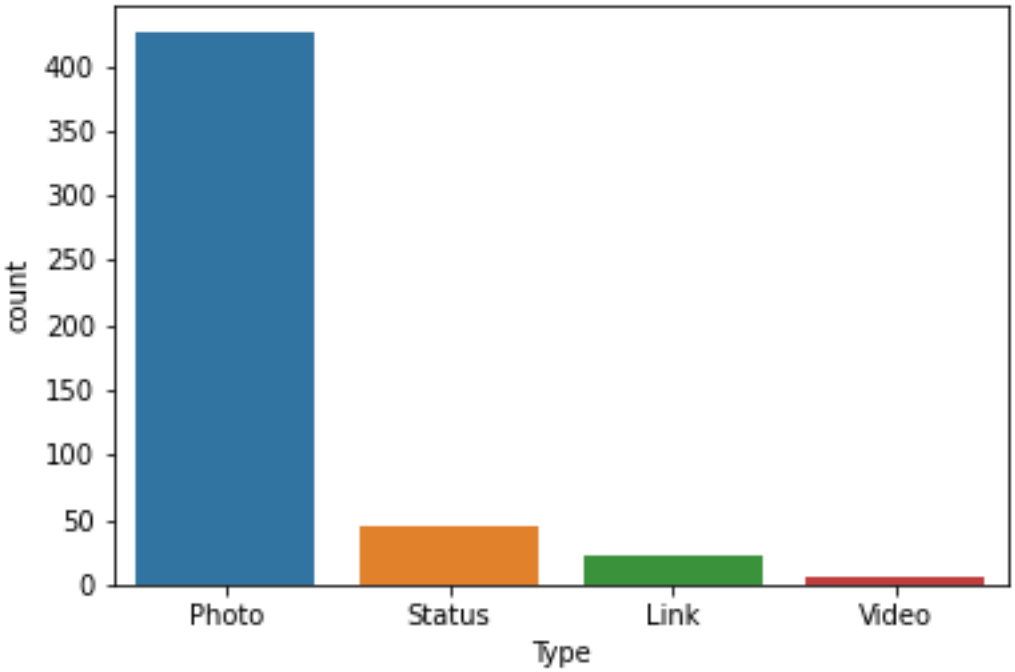
Count plots of other categorical variables



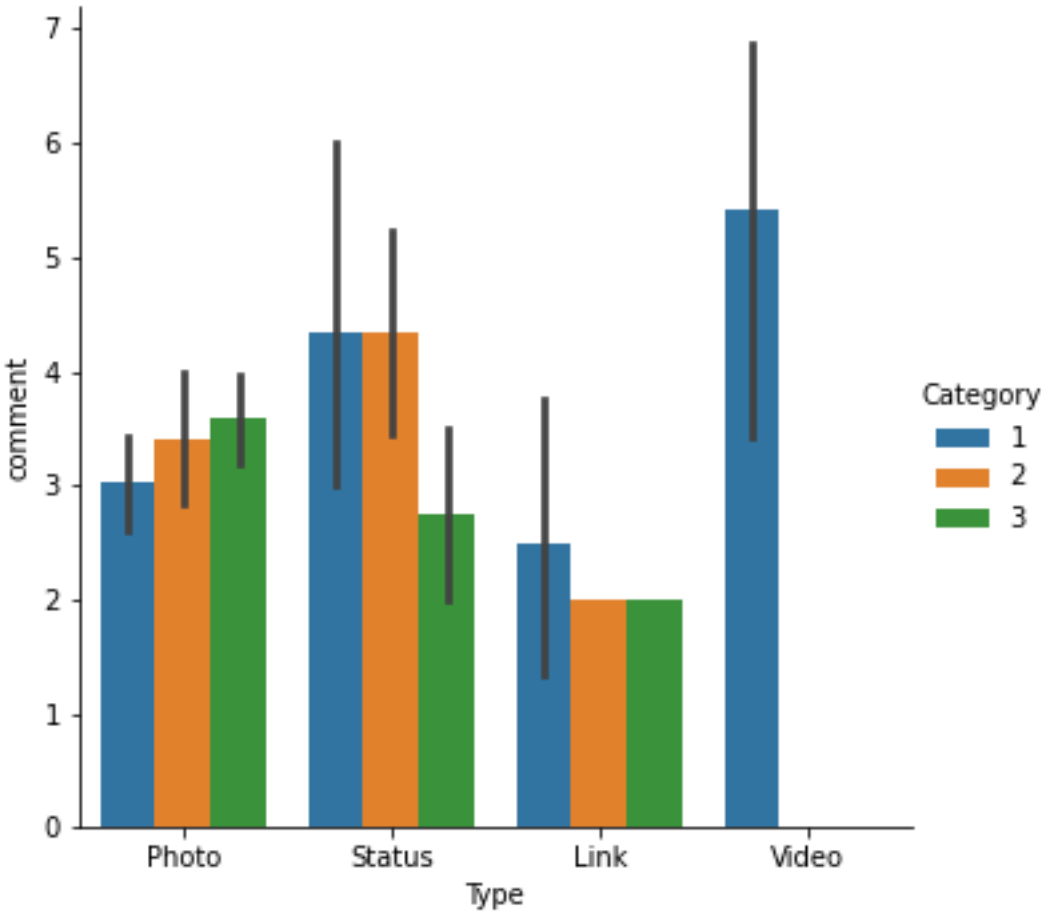


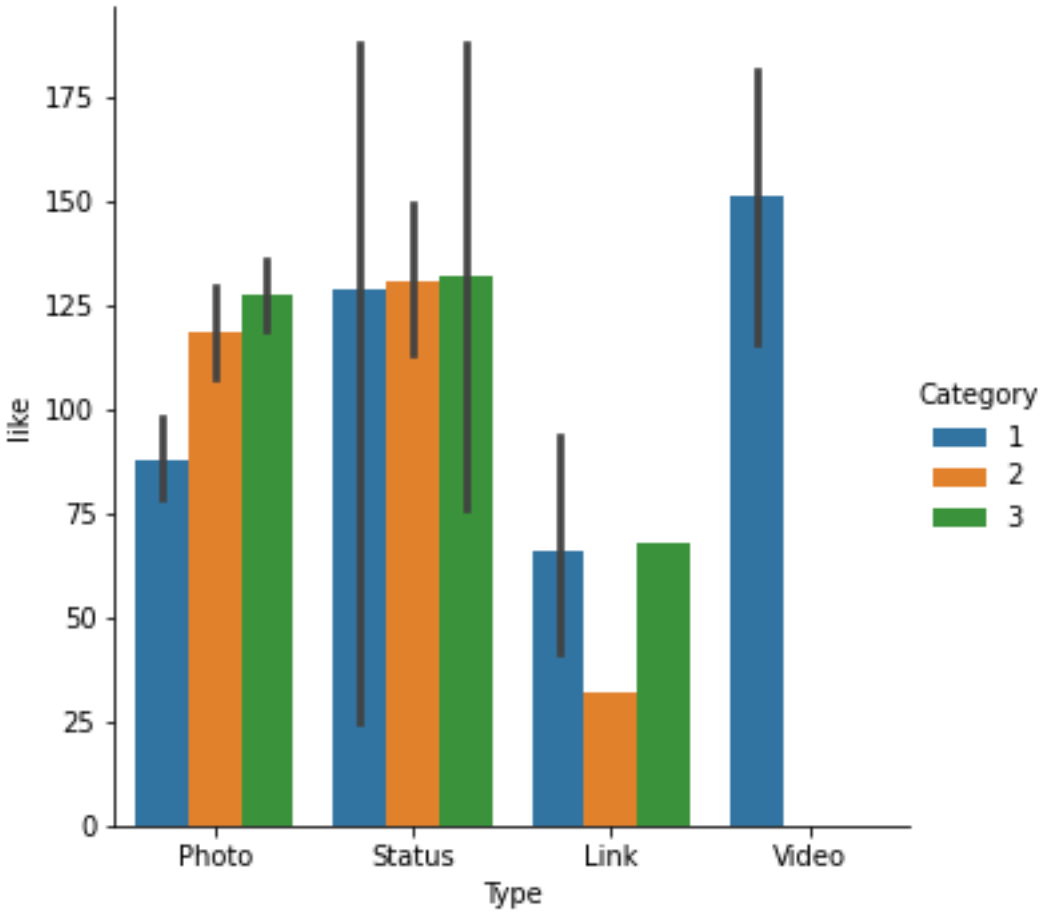

count
60
50
40
30
20
10
0
1
2
3
4
5
6
7
8
9
10
11
12
Post Month

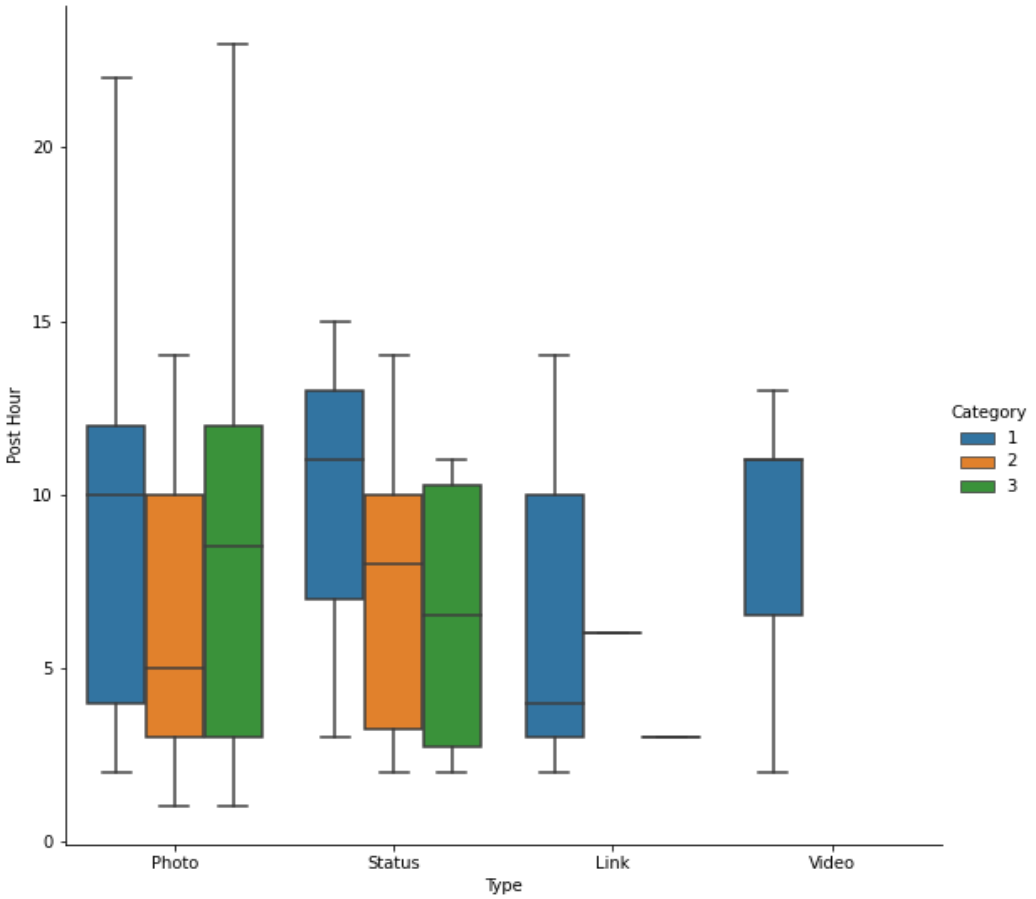

count
80
70
60
50
40
30
20
10
0
1
2
3
4
5
6
7
Post Weekday

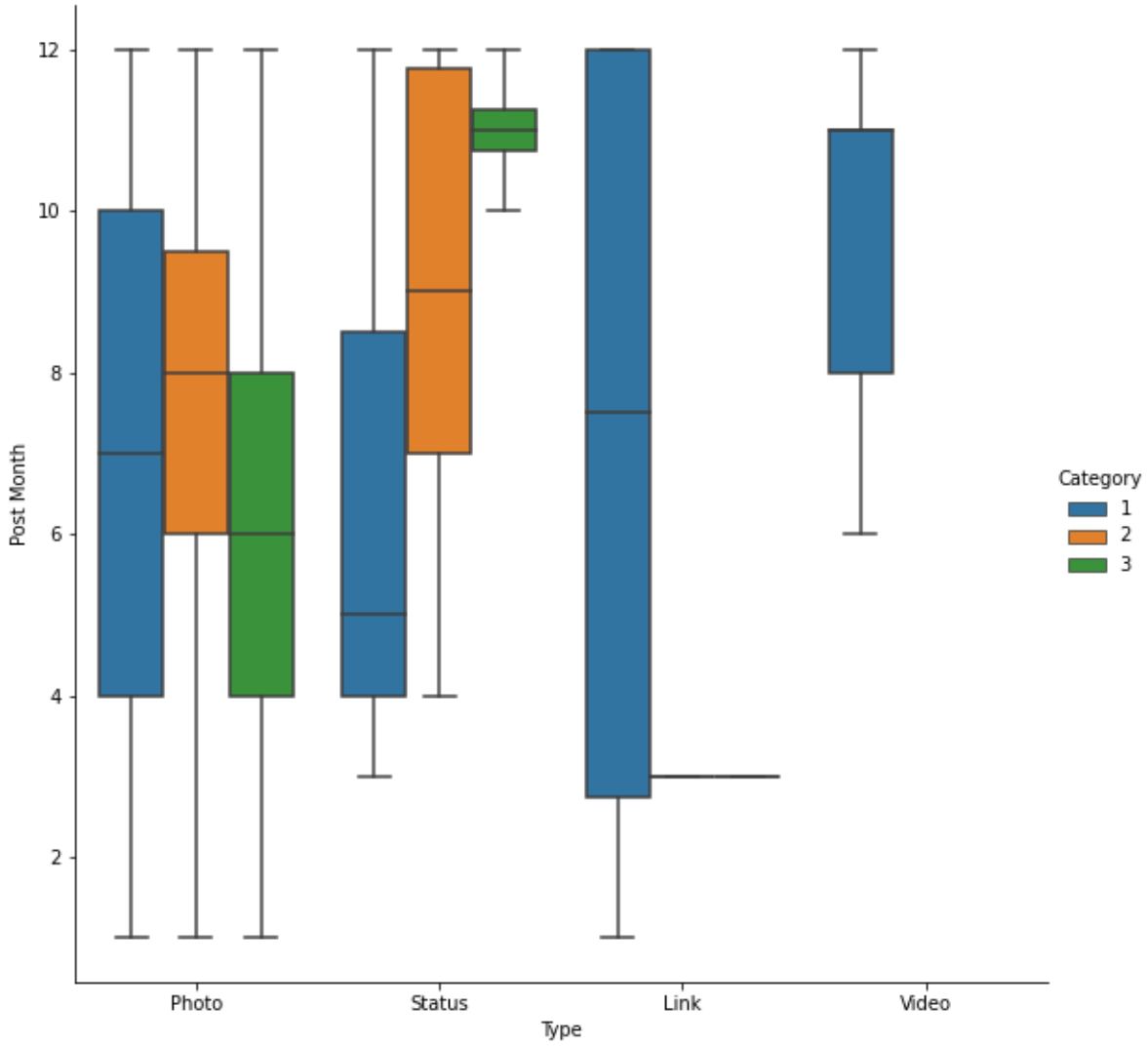


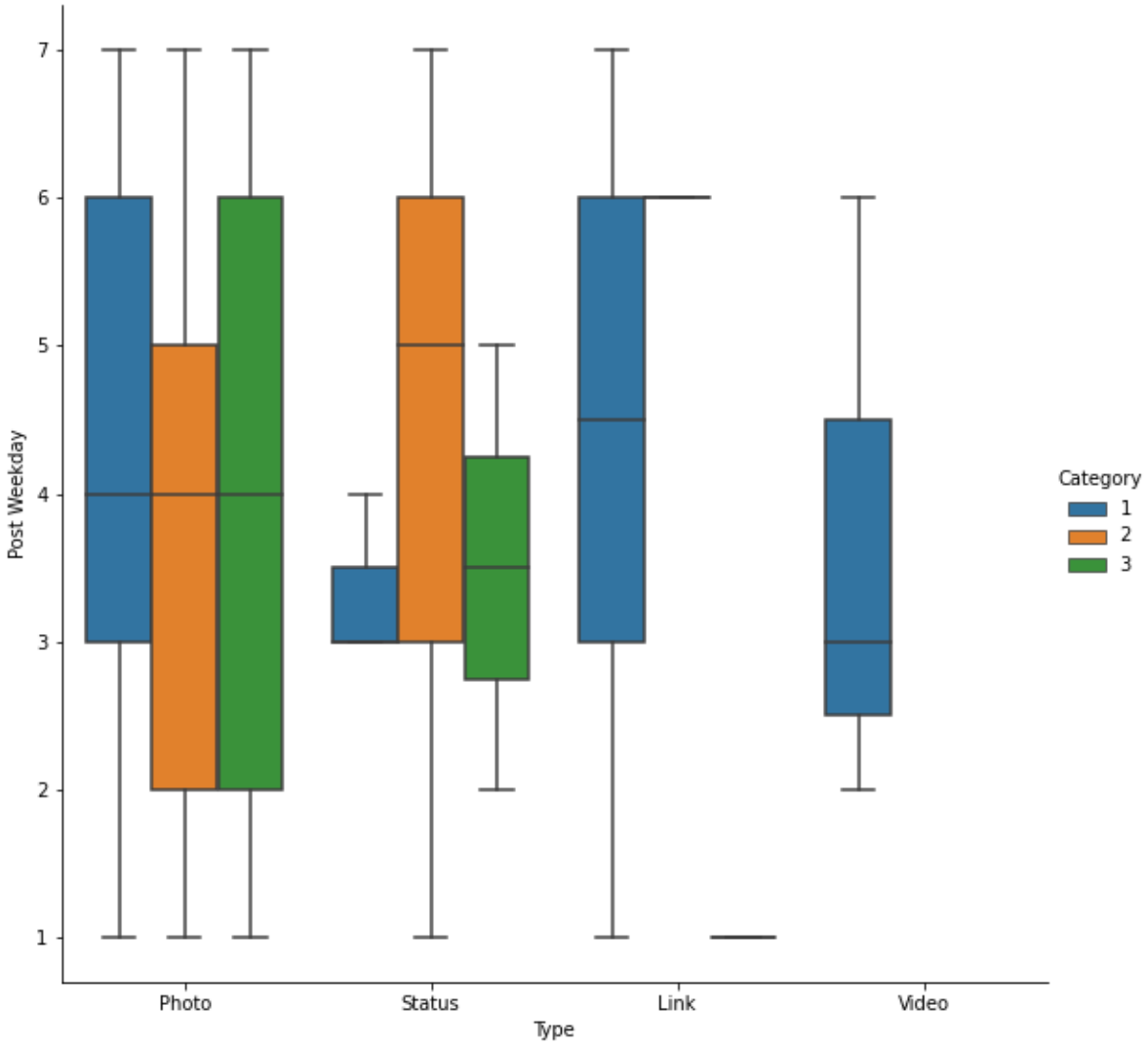
Bivariate analysis

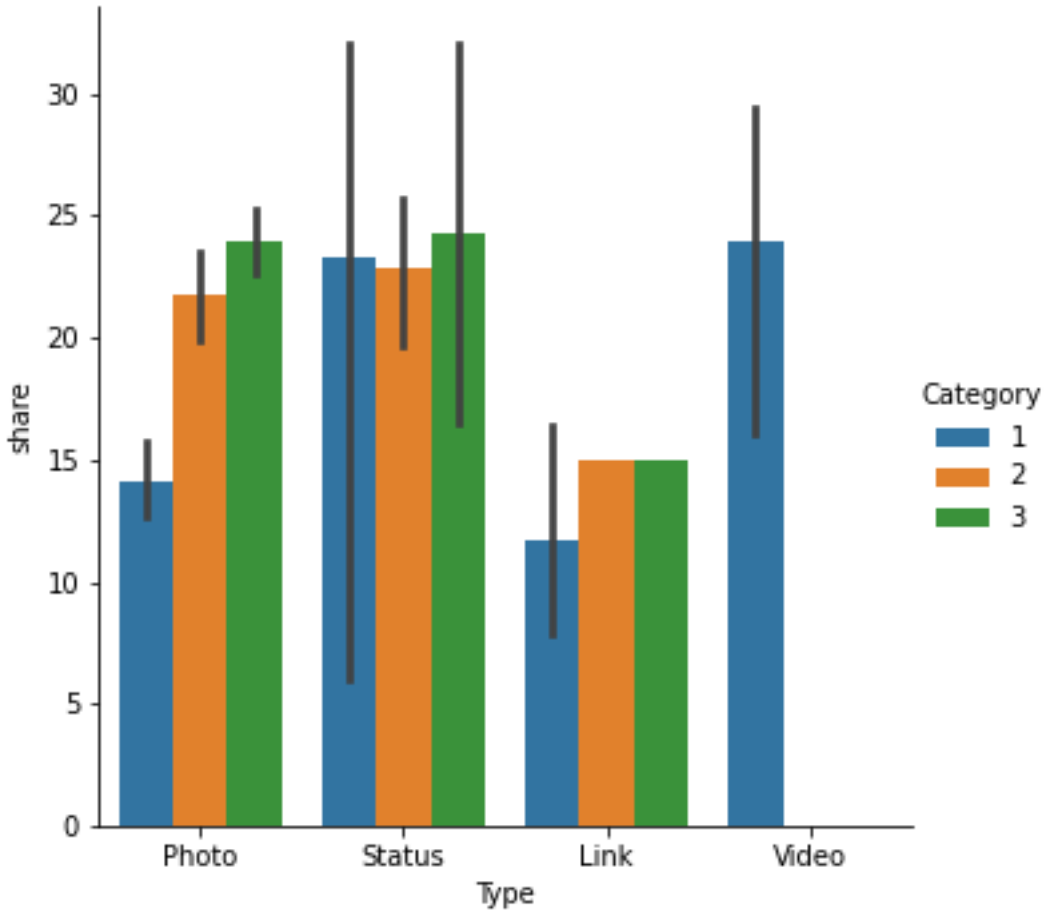



175
150
125
100
75
50
25
0
like
Photo
Status
Link
Video
Type
Category
1
2
3

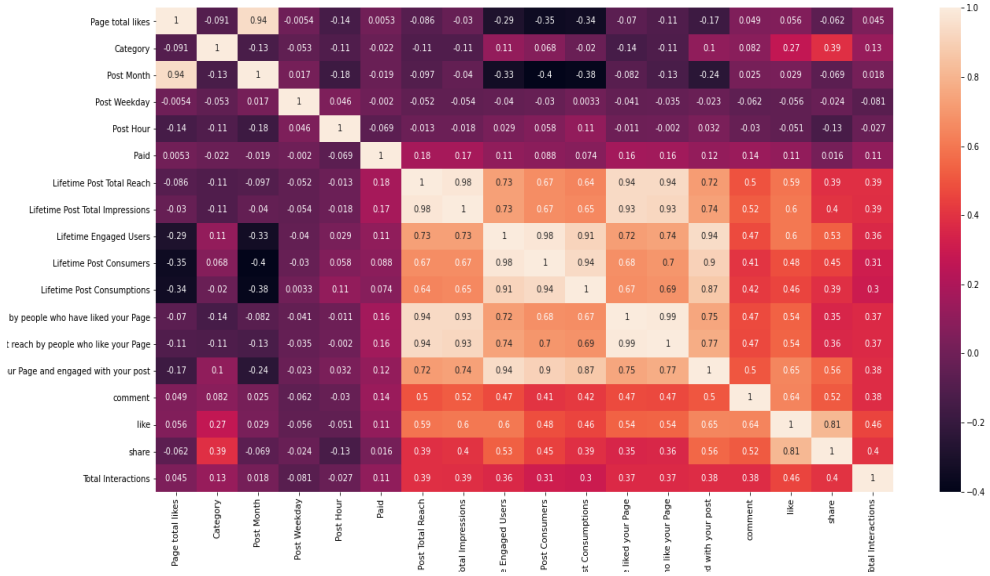

Post Hour
0
5
10
15
20
Photo
Status
Link
Video
Type
Category
1
2
3


12
10
8
6
4
2
Post Month
Photo
Status
Link
Video
Type
Category
1
2
3


7
6
5
4
3
2
1
Post Weekday
Photo
Status
Link
Video
Type
Category
1
2
3



Multivariate analysis using heatmap



Key Findings and Insights, which synthesizes the results of Exploratory Data Analysis in an insightful and actionable manner

The key finding from data visualization were the following: -

The greatest number of likes for a good post lies between 130000-140000.

The greatest number of comments lies 0-2.

On average 400-600 people engage with a post.

For a good post impression should be greater than 5000.

On average a post reaches 500 people.

The average like lies between 50-100.

Average share should at least reach between 10-20.

The greatest number of post belong to category 1.

The apt time to upload a most is 3 am in the morning.

A good month to post would be October.

The post should be uploaded on the weekends.

The most of the post are photos.

Formulating at least 3 hypothesis about this data

Hypothesis 1: - The average likes of people who like a post on Saturday is equal to people who like post on Sunday.

Hypothesis 2: - The photo type post from category 1 is equal to link type from category 1 between week 3-6.

Hypothesis 3: - The number of share for photo type post is equal to number of share for status type post both belonging to category 3.

Conducting a formal significance test for one of the hypotheses and discuss the results

I conducted a t-test method was conducted on first hypothesis and the conclusion was as following: -

t statistic was 0.6833989048166412 p value was 0.4953371233774665.

Here we see that the p value is greater than .05 at chosen level of significance at 5%

The p value of 0.4953371233774665 is not significant. Hence accept null hypothesis

Suggestions for next steps in analysing this data

The next steps for analysing the data would be to test other hypothesis related to the data.

Drive more hypothesis.

Apply the result of hypothesis in real life and verify the impacts.

A paragraph that summarizes the quality of this data set and a request for additional data if needed

Title: Facebook performance metrics

Sources: Created by: Sérgio Moro, Paulo Rita and Bernardo Vala (ISCTE-IUL) @ 2016

Past Usage: The full dataset was described and analyzed in S. Moro, P. Rita and B. Vala. Predicting social media performance metrics and evaluation of the impact on brand building a data mining approach. Journal of Business Research, Elsevier, In press, Available online since 28 February 2016.

Relevant Information: The data is related to posts' published during the year of 2014 on the Facebook's page of a renowned cosmetics brand. This dataset contains 500 of the 790 rows and part of the features analyzed by Moro et al. (2016). The remaining were omitted due to confidentiality issues.