

MEDIA TREND ANALYSIS

Shaoyu Zhang

The Barstow School, Ningbo Campus, China

DOI: 10.46609/IJSSER.2021.v06i06.010 URL: <https://doi.org/10.46609/IJSSER.2021.v06i06.010>

ABSTRACT

With the advent of information era, media has become more and more important in our daily life. In this project, with the statistics of a commercial TV show as an example, I statistically analyzed the potential driving forces behind the decline of the TV show viewership. Specifically, I first examined the linear relationship between potential variables and the TV show viewership, based upon which, I come up with some statistical factor to make TV show viewership prediction. After an examination and some analysis on the regression result, I conclude that the presence of a specific character and the presence of some competing shows don't have significant effects on the decline of the TV viewership. I also conclude that the budget of the show has positive but minor effect on the boost of the viewership. Those findings are also consistent with my background research.

Keywords: Linear regression, media data, TV show conversions.

1. Motivation

With the advent of information era, media has become more and more important in our daily life. Recent studies have found that adult spent 2 hours on average on social medias, among which, Facebook and Youtube have become the major part of the adult time allocation. In China, recent development of short video platform such as Tiktok and Bilibili have been very successful in capturing teenagers' attentions. From both company profit and efficient information prospective, understanding what are the important factors behind the soaring media participation has become more and more important. In this project, I use the viewership of a digital company as a quantification of the media participation. As the viewership fluctuate according to the other statistics, I analyze the relevant driving forces behind the media participation.

2. Relevant Digital Company Data

An India media company launched a TV show in 2017. The show has been very popular and

commercially successfully at the beginning but then the viewership experiences a decrease as it approaches the ends. The companies have released some statistics that are very useful for my studies. Here I briefly summarize the statistics below:

- Date: Date of the statistics
- Views of the show: The viewership of the show
- Visitors: The number of visitors to the platform, who doesn't necessarily watch a show
- Views of the platform: the number of times videos on the platform got played
- Ad impression: number of ads' impressions, which represent the budget of the show
- Cricket match India: whether or not the popular cricket match is played on that day
- Character A: whether or not a famous character is present in the show

3. Background Research and Data Visualization

In this section, I summarize the background research I conducted about the Media influence and propose a few conjectures with data visualization to support my preliminary conclusion.

3.1 Leisure Time Hypothesis

My first hypothesis is that the time of day a program is broadcast has important impact on the popularity of the media show. Because adult viewers usually watch programs after work, when they get home from work, and in the evening; children watch TV around the time of day, after dinner or before bed. Seniors are more likely to watch the news in the morning. So it is very important for a TV show to have a good time slot. If it airs at midnight, early in the morning, or during most people's working hours, TV views will suffer in some amount, no matter what the program's appealing subject matter is. The reason is only few people will watch a program during non-regular entertainment hours just because it is very appealing and hot. For example, China Central Television (here in after referred to as CCTV) usually broadcasts the morning news at 6:00 a.m. sharp to provide fresh information for early commuters; at 12:00 p.m., broadcasting the news for 30 minutes for those who have free time during lunch break; and finally at 7:00 p.m. broadcasting the news bulletin. These three programs have been the same schedule of CCTV for many years. Based upon this conjecture, I plotted the TV show number according to the different weekday in Figure 1.

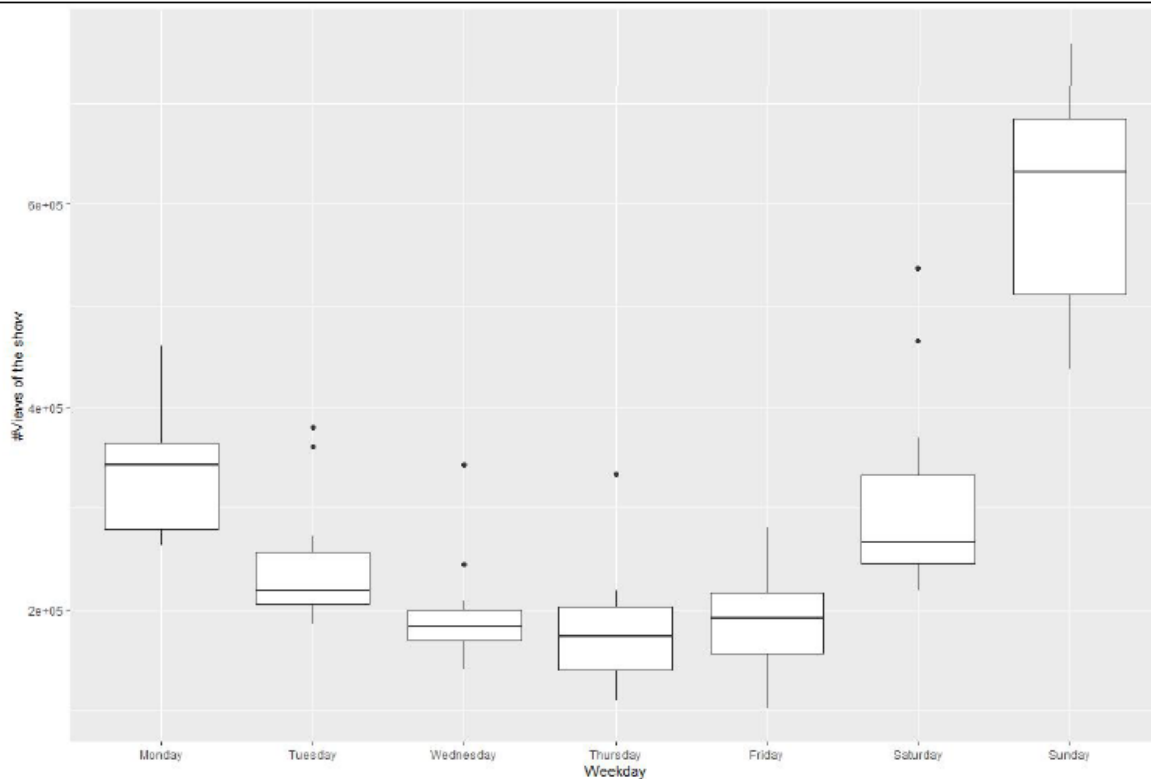


Figure 1: Weekday effect

As it is suggested by Figure 1, one important pattern I observe is the weekend effect. The number of people watching the show tends to be much higher over the weekend and much slower during the weekdays.

3.2 Star Effect Hypothesis

Another factor that could be influential to the views of the show - superstar. This term was first described in a study titled "Quitters Never Win: The (Adverse) Incentive Effects of Competing with Superstars" by Jennifer Brown. In the NBA, for example, games with superstars not only generate higher TV views, but also higher game attendance. This increases the revenue of both the superstar's team and the opposing team. Another example of the influence of celebrities on their fans is a TV show starring a popular celebrity, which gets significantly more views during the time period it is on that network. I thus also plot the effect of presence of some celebrity in Figure 2, which seems to be consistent with my background research.

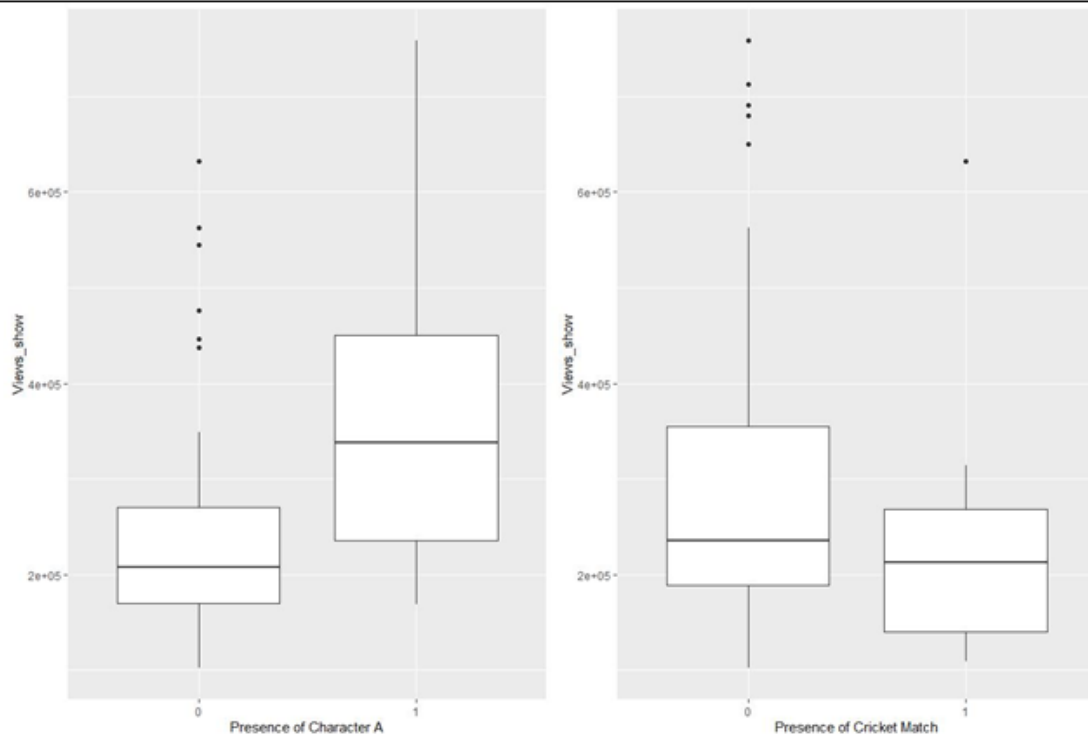


Figure 2: Presence of Character A/ Presence of Cricket Match

3.3 Audience within the Progra

The third speculation about the factors affecting the ratings is the subject matter and audience of the program. The formula for calculating ratings is

$$\text{Ratings} = \frac{\text{number of people watching the program in the region}}{\text{total number of people in the region}}$$

So for each TV station, they cannot simply broadcast popular or fixed programs, they need to meet the needs of different people for different topics on TV. For example, older people may prefer military dramas, younger people prefer news, American dramas, movies, and romance dramas, while children prefer cartoons. So a popular program subject matter and novel production methods in meeting a wider audience group at the same time drive the views of show. One of the representator of the number of people in the region could be the number of visitors to the companies platform and Figure 3 seems to confirm their linear relationship since the #Views of the show is positively correlated with respect to the number of vistoris and number of views in the platform.

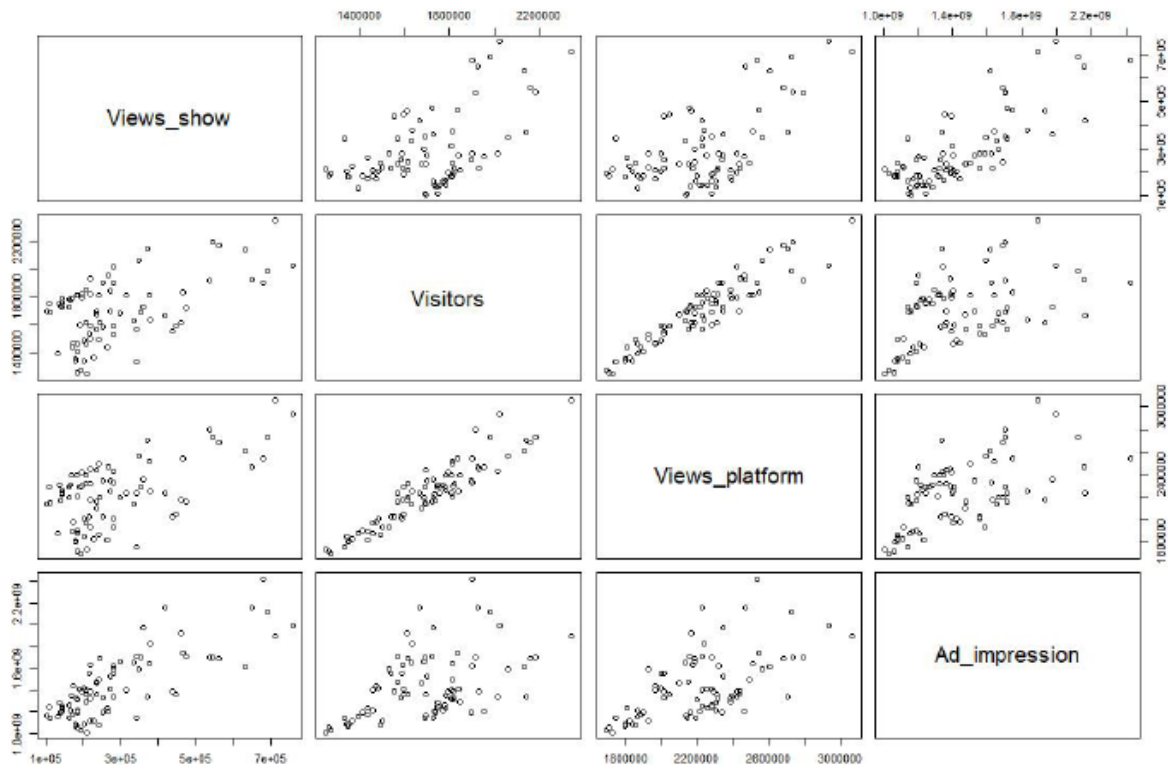


Figure 3: Exploration of Linear Relationship

4. Model

4.1 Model setup

As it is confirmed in section 3, the data demonstrate great linearity, which falls perfectly under the linear regression setup:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon$$

where:

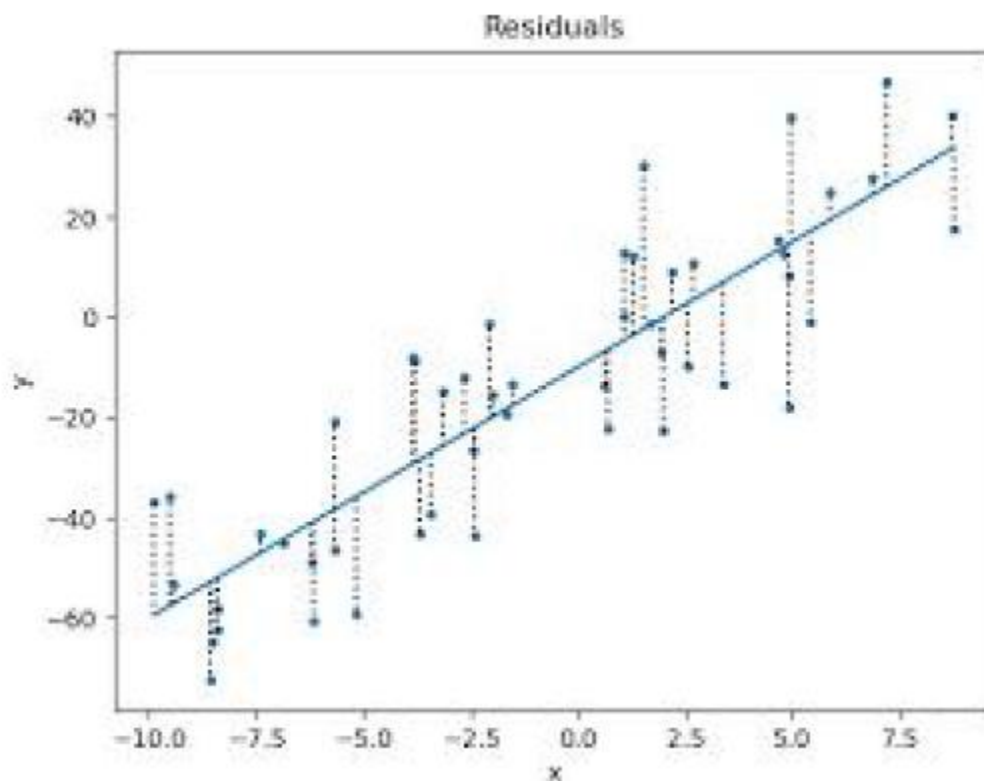
- Y is the viewership
- X1, X2, ...Xp are p different variables relevant to the viewership
- $\epsilon \sim N(0, \sigma^2)$ is the residual, which follow a normal distribution

The model is thus assuming that one unit change of certain variable X_p will change β_p unit of Y.

4.2 Model solution

The model is defined by the pairs of the coefficients and is found by minimizing the squared error distance between the observation Y and model fit \hat{Y} . Specifically, we hope to find $\{(\beta_0)^*, \beta_1^*, \dots, \beta_p^*\}$ that minimizes the distance from n data

$$L\{(\beta_0)^*, \beta_1^*, \dots, \beta_p^*\} = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \dots - \beta_p X_{pi})^2$$



Using calculus (when a function reaches the minimum at a point, its derivative equals 0 at this point), we could solve $\beta_0^*, \beta_1^*, \dots, \beta_p^*$ by solving the following equations:

$$\frac{dL(\dots)}{d\beta_0} = 0, \frac{dL(\dots)}{d\beta_1} = 0, \dots, \frac{dL(\dots)}{d\beta_p} = 0$$

For example, if we only have one variable for prediction, the solutions β_0^*, β_1^* are:

$$\beta_1^* = \frac{\sum_{i=1}^n X_i Y_i - n \bar{X} \bar{Y}}{\sum_{i=1}^n X_i^2 - n \bar{X}^2}$$

$$\beta_0^* = \bar{Y} - \beta_1^* \bar{X}$$

For a multivariate regression, I used R to solve for the solutions.

Model interpretation

Now I attach the output of the R regression for model interpretation, the corresponding variable with “*” indicates significant influence on the viewership prediction

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.570e+05  8.373e+04  -4.264 5.93e-05 ***
Visitors     -2.593e-01  1.400e-01  -1.852 0.06811 .
Views_platform  3.329e-01  1.189e-01   2.801 0.00652 **
Ad_impression  2.189e-04  5.310e-05   4.122 9.81e-05 ***
Cricket_match_india  3.427e+03  3.025e+04   0.113 0.91011
Character_A   -2.646e+04  3.097e+04  -0.854 0.39577
isweekend     1.508e+05  2.577e+04   5.852 1.28e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 75520 on 73 degrees of freedom
Multiple R-squared:  0.7804,    Adjusted R-squared:  0.7624
F-statistic: 43.25 on 6 and 73 DF,  p-value: < 2.2e-16

```

From the fitted model, we could conclude that the presence of Character A and the presence of a cricket match don't have significant effects on the viewership. The model confirms the weekend effect and has demonstrated minor but strong dependence on the Ad impression variable. This is in general consist with my background research that the commercial budget of the show typically has positive but limited effects on its popularity. Compared to the number of visitors, the number of views in the platform is more significant since the people who visit the platform don't necessarily get converted into the real user of the platform, which suggests potential improvement on capturing the visitor's attentions to the platform. Quantitatively, we have the presence of weekend date increase the show viewership 150,000 on average, tenths units change of the of the Ad impression will boost the viewership by 2 units, and that about 30% of the views occurred on the platform is converted into the viewership of this specific show.

5. Result Analysis---Normality Check

The exist of the linear regression model solution doesn't necessarily require any statistical assumption, but in order to validate the statistical conclusion in section 4, we need to put extra efforts on the normality examination. Specifically, in section 4, we assumed the residual follows a normal distribution $\epsilon \sim N(0, \sigma^2)$, without which the statistical conclusion (significance of the slope) doesn't hold. To conduct the examination, I here first estimate the σ^2 with the residual

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-p}$$

where p is the number of parameters included in the linear regression model.

Then with the estimated $\hat{\sigma}^2$, I simulate the normal distribution with $N(0, \hat{\sigma}^2)$ and overlay it with the density plot of the residual Y_i, \hat{Y}_i to compare how similar they are:

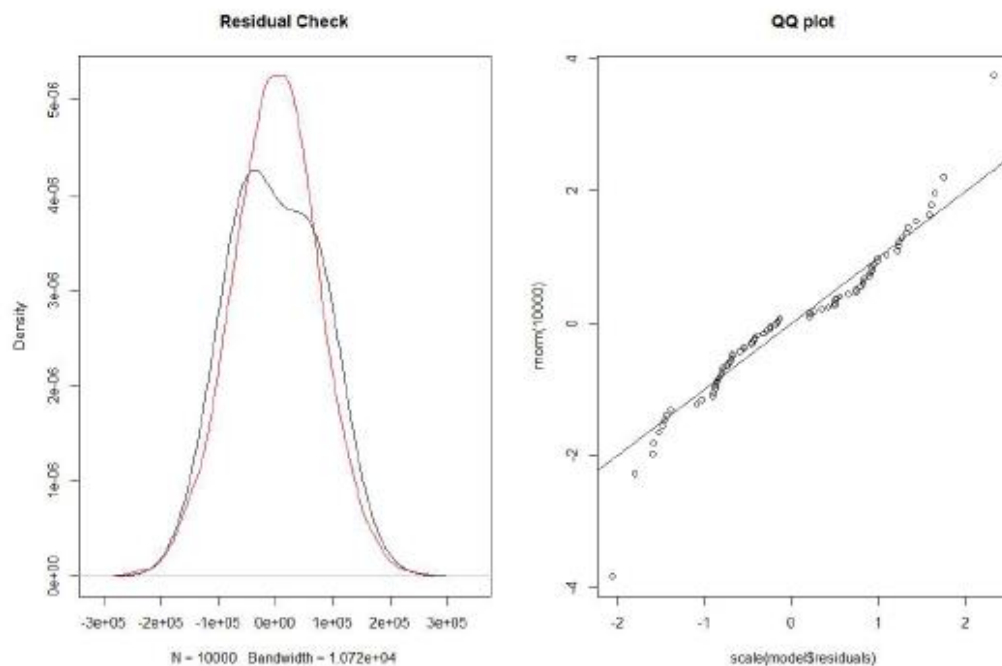


Figure 4: Presence of Character A/ Presence of Cricket Match

The result indicates great normality and thus conclude the linear regression with its correct assumption.

6. Conclusion

In this project, I employed linear regression to analyze the commercial statistics of a TV show. The result of the model confirms the weekend effect and quantitatively measured the conversion rate of the platform viewership to the viewership of the show. The model also eliminates some plausible influencer such as the presence of a famous character A and the presence of a popular local match. It also provided important guidance on the viewership improvement by quantifying and comparing the effect of number of visitors and effect of number of views in the platform. The examination of the normality assumption also suggests great normality and thus validate the model conclusion.

Reference

- Linear Regression Using R: An Introduction to Data Modeling, David J. Lilja, 2016
- R tutorial: https://www.tutorialspoint.com/r/r_mean_median_mode.htm
- Digital Marketing Data from Kaggle: <https://www.kaggle.com/ashydv/media-company-case-study-linear-regression/data>
- All of statistics: A concise course in statistical inference. Wasserman, L. (2010), New York: Springer.
- Social Media Utilization in Business-to-Business Relationships of Technology IndustryFirms. Computers in Human Behavior, 2014
- TV News: Whose Bias?: A Casebook Analysis of Strikes, Television and Media Studies.
- Routledge, 2020.