

Stats 101 Spring 2025 S4 Final Project

Mr Beast Video Analysis - Wordcount xxxx

Anonymous

2025-04-28

Table of contents

Assignment instructions	2
Specific requirements	3
Considerations	3
Points of emphasis	4
Introduction	6
Literature review and hypotheses	6
Summary statistics	6
Regression analysis	6
Regression form	6
Regression diagnostics	6
Interpret coefficients	6
Proposed plan for MrBeast	6



Assignment instructions

MrBeast is one of the most popular YouTube stars of the last few years. His [YouTube videos](#) often revolve around organizing contests among various players who compete for cash prizes, and the videos each have millions of views. Imagine you have been newly hired by MrBeast's production team. He asks you to analyze the characteristics of his videos to understand what features of the videos translate into more views. In particular, he is most interested in what features of a video's title and description plus publish time lead to the most views, though you can consider other characteristics as well.

To do this, you need to model the variable `viewCount` using a multiple variable regression. You may need to quantify various features of the `title` and `description` variables, using functions like `nchar()`. You may need to extract various parts of `publishTime` using the `lubridate` library as discussed [here](#).

You will need to use all the material we learned in class to:

1. Think carefully about the sampling frame, what your research expectations and questions are, and how the method of data collection can influence your findings
2. Interpret summary data
3. Examine important bi-variate relationships

4. Construct a high-quality regression model of `viewCount` and at least one variable related to each of: the title, description, and publish time.
5. Interpret this model, including ALL relevant diagnostics and p values.
6. Propose a plan to MrBeast about how to increase his view counts.
7. Also note the limitations of the report and what other data you think MrBeast should consider before reaching a final conclusion

Specific requirements

- Save this document as a new document (Save As...) and rename it **Mr Beast report**.
- Rename the title of your report to **Mr Beast report**
- Delete the **Assignment instructions** section and its subsections
- Final report should be between 2000 and 3000 words
 - Maximum 8 graphs
 - Maximum 6 tables
- Suggested structure:
 - Introduction
 - Literature review and hypotheses
 - Summary statistics
 - Regression interpretation
 - Regression diagnostics
 - Interpret coefficients
 - Proposed plan for MrBeast
 - Conclusion

Considerations

Here's a **non-exhaustive** list of things you might want to do:

- **Views Determinants:** Analyze the relationship between key predictor variables and views. Explain which features of a video are the most influential in determining the number of views.
- **Text Variable Analysis:** You may consider what features of a title are most important at getting visitors to click on a video, such as the length of the title or other text features of the title.
- **How to Model Description:** Does it matter whether a video has a description? Consider whether to model description as a categorical variable (has/does not have a description) or whether to model it by the description length.

- Time Features: What time aspect of a posting matters - is it time of day? Day of the week? Etc.
- Variable Transformations: Consider applying transformations (e.g., log, square root) to some of the variables if they show non-linear relationship between the predictor and response variable. This could improve the accuracy of your regression model.
- Recommendations: Based on your analysis, provide recommendations for MrBeast as to some easy things he could do to increase his video view counts. What features or details are important in predicting views?

Points of emphasis

- Literature review should include some expectations you develop from finding news articles or studies related to your research question. You should specifically state some hypotheses that seem likely that 1) follow from your literature review and 2) answer the question posed in the Assignment Instructions section.
- Do not exclude a variable just because it initially does not meet the regression requirements. However, consider carefully whether some variables are actually highly related to another predictor variable – do not include collinear variables.
- You should focus your graphs and tables on that illustrating the most important information for drawing your conclusion. Choose your tables carefully such that they convey the key information needed to arrive at your conclusion. Do not make tables and graphs of irrelevant information or points that do not need discussing.
- Make sure to also interpret the coefficients. You need to interpret the impact of a one unit change in the coefficients on the response variable. You additionally need to examine whether changes in the predictor variables lead to a substantively large or small change in the predictor variable. One way to do this is creating a table of the predictor variables, examining whether a Q1 to Q3 value leads to a large or small change in the response variable. Another way is by the marginal effect plots we have previously used.
- Your report should be a polished, quality product that you would be proud to show your boss/employer. No unnecessary printed code, poorly labeled graphs, or strange looking formatting. Use everything you have learned in this class to make a quality final product! Remember, this document is not a formal essay. Some of the differences between a business report and a formal essay are summarized in the table below and by this [link](#):

	Essays	Business Reports
Purpose	Articulate a point of view in relation to a particular proposition	Often recommend action to solve a specific problem
Format & Structure	Have introduction, body and conclusion sections that normally do not use headings	Always have sections clearly divided by numbered headings (and often sub-headings)
	Use cohesive paragraphs to link ideas rather than list dot-points	Use shorter, more concise paragraphs and dot-points where applicable
Abstract	Typically don't normally need an abstract as readers read the text carefully from start to finish	Always have an abstract (or executive summary) as readers are typically 'time poor' and skim and scan through the text quickly
Graphics	Rarely use graphics as written evidence	Feature graphics (such as tables and graphs) for supporting main points
Writer	Are generally the result of individual work	Are often the result of group work
Reader	Are written for the lecturer/tutor or other academic audiences	Are addressed to a specific audience (i.e. client or manager) established by the topic

The below is a suggested structure but you can modify it if the order does not work for you

Introduction

Literature review and hypotheses

Summary statistics

Must have these three sections (Introduction, Literature review, and Summary statistics) done by the homework check due date (Sunday, May 4)

Regression analysis

Regression form

Regression diagnostics

Interpret coefficients

Proposed plan for MrBeast

Conclusion