

Final Exam MSBA

This is a dataset on crowdfunding. The market we consider is a crowd-funded platform based in San Francisco, California, which was established in November of 2021. The site enables new and established journalists to crowd-source capital in support of their work by pitching their ideas to the community, just as an entrepreneur might pitch to venture capitalists. The site functions as follows: individuals first register to become members of the community. Members can then establish a profile page where they can provide details about themselves along with a photograph and a link to a personal website. Any member in the community can choose to “pitch” a story to other members of the community. A pitch is a proposal to publish a story on the website, and includes a description of the story topic, as well as a budget. Other members of the community can then choose to contribute funds toward the pitch in any increment they wish. The contribution phase continues in this manner until the story is completed and published on the website. Once a community member contributes toward a pitch, this becomes public record. When reviewing a pitch, a list of all prior contribution events is provided to the observer as well as a summary indication of the aggregate contributions received to date. Thus, page visitors can easily assess cumulative contributions, and recent contributions to the campaign in question. The explanations for the different variables in the dataset are as follows:

1. story_id – unique pitch identifier
 2. abs_week - absolute week (observation period)
 3. rel_week - relative week (weeks since story was published)
 4. story_readtime - total time visitors spent at the story URL this week, in seconds
 5. story_views - total number of visitors to the story URL this week
 6. story_funding_duration - the number of days the associated pitch spent being funded
 7. story_length - the length of the story, in characters
 8. focal_page - a binary indicator of whether this story is on the first page or not (rank = 1-12)
 9. pitch_contributors - total number of contributors that funded the story
 10. tags - number of story topic tags associated with this story
 11. pitch_views - total number of visitors to the pitch URL
 12. arindex - Automated Readability index value for the story text
 13. pitchsum_budget - the funding target of the associated story, in cents
 14. video - binary indicator of whether the story includes video
 15. insights - Google Search Trends for pitch keywords (0-100, relative to peak)
- You would like to understand what factors will determine the attractiveness of a pitch. The pitch attractiveness is measured by whether the story has convinced other members to contribute. How would you approach this problem? Pick a dependent variable (or a few dependent variables) that will give you some insights on contribution behaviors. Discuss why you choose this variable (or this set of variables) as your dependent variable(s). Then, given each dependent variable, select an appropriate model and relevant independent variables. Justify your choices with a few sentences. Finally, show the output of your model in STATA or R and discuss your interpretations of the results.
 - Now, you want to build and estimate a regression model where the dependent variable is a measure of how much time people spent on reading a story or how many people actually read a story (story_read time or story_views, respectively). What variables would you choose as independent variables? Show the output of the final model in STATA or R. Provide in a few sentences on your main takeaways from the analyses.
 - Note that the dependent variable above is a count of number of seconds people took to read a story or a count of number of views for a given story. Would estimating an OLS regression, even with panel data models, give you precise estimates? If not, what kind of model would you want to run to get estimates that are more precise and less biased? Show the output of the final model in STATA or R. Provide in a few sentences on your main takeaways from the analyses.