DATA Project

Welcome to play with real data!

Log Transformation in Regression

Linear Regression

In a linear regression, we assume that the relationship between the dependent variable and independent variable is linear, i.e., we specify the following relationship:

$$Y = a + bX$$

Log Transformations

But sometimes we also take the log-transformation of the linear regression. For example, consider the following relationship:

$$\log Y = a + b \log X$$

Here, we typically use the natural logarithms (base is $e \approx 2.718$) in log transformation.

The Log Function

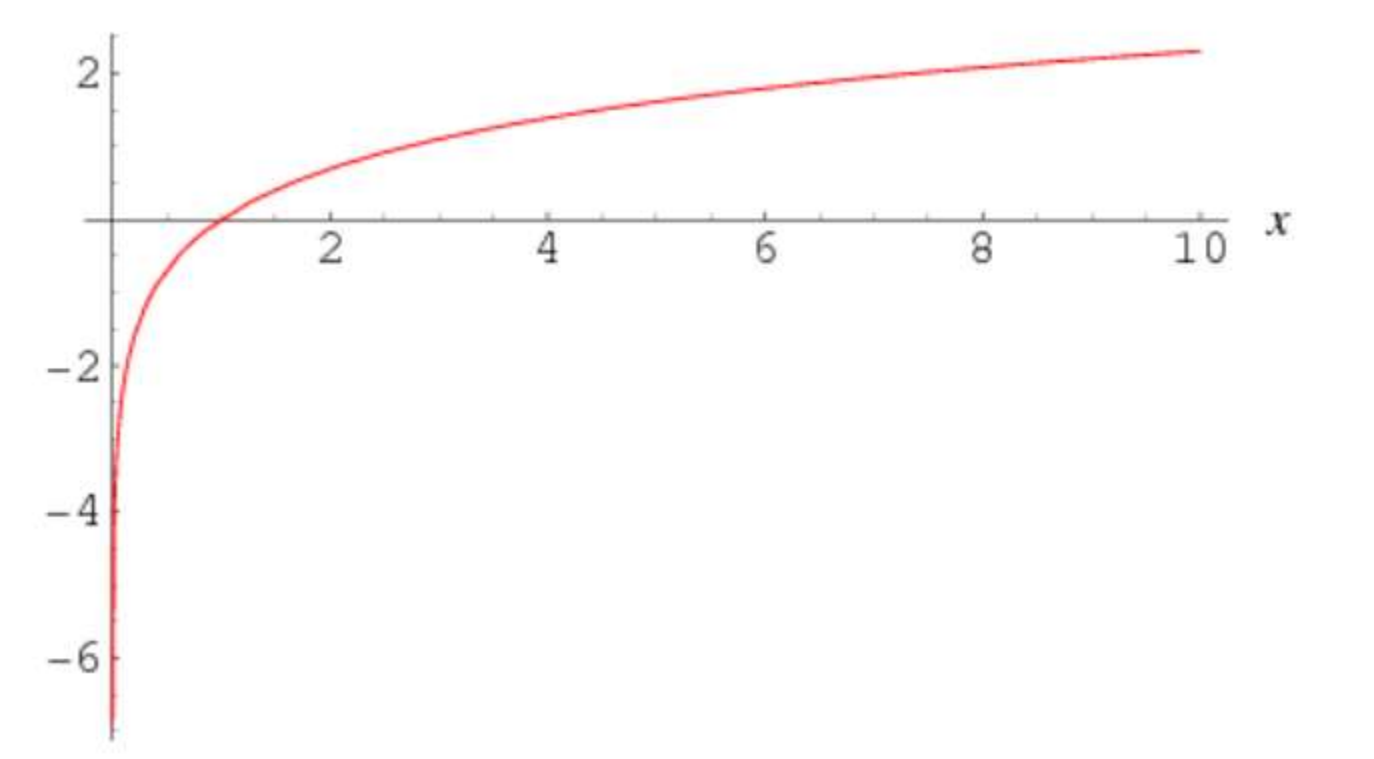
If $e^a = b$, then $\log(b) = a$, where $e \approx 2.718$

For example,

$$\log(10) \approx 2.3026$$

$$\log(100) \approx 4.6052$$

The Log Function

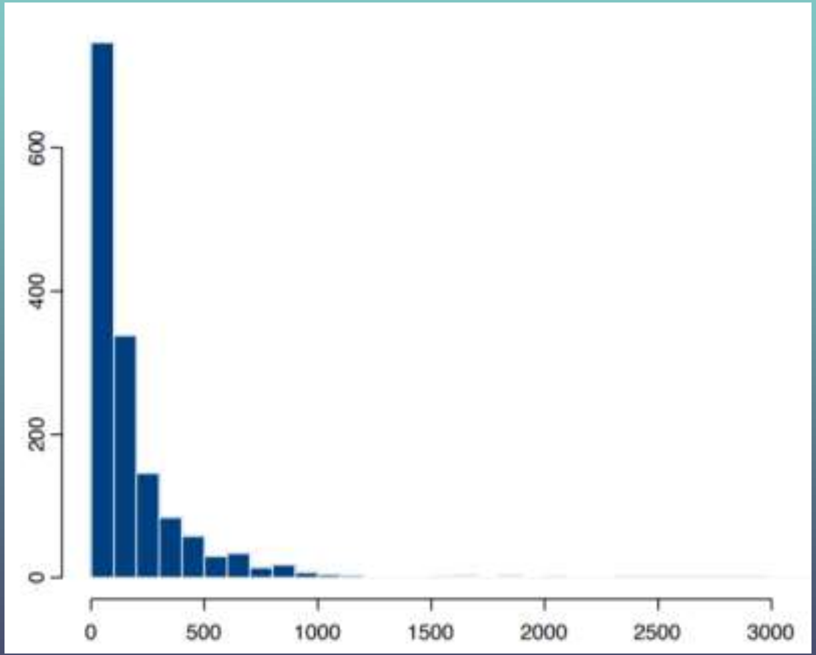


Question

Why do we take log transformation instead of directly using the linear regression?

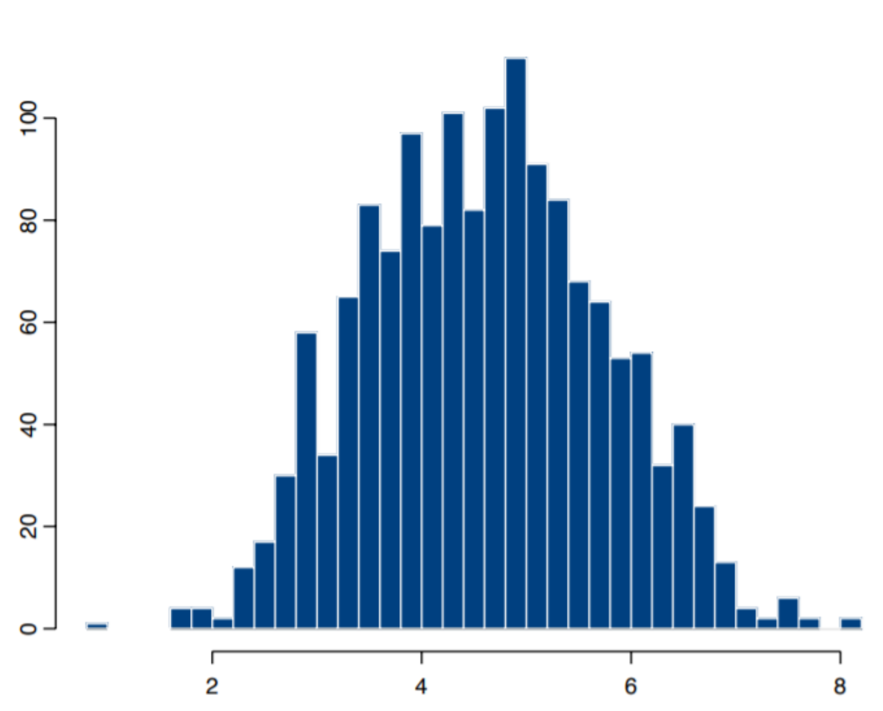
We typically take the log transformation when the original variable is right-skewed, i.e., when the right tail is much longer than the left tail.

A highly skewed distribution



We see a significant right skew in this data, meaning the mass of cases are bunched at lower values.

A highly skewed distribution



If you take the log transformation, the distribution looks like a normal distribution.

Fixed Effects

Fixed Effects

Suppose that we want to investigate the effect of month on consumers’ spending. Here, the independent variable is the month, e.g., January, February, March, …

How to run this regression?

Fixed Effects

One solution: We can assign Jan = 1, Feb = 2, March = 3, … and then we can simply run a linear regression as usual.

Any issues with the above regression?

Fixed Effects

One solution: We can assign Jan = 1, Feb = 2, March = 3, … and then we can simply run a linear regression as usual.

Suppose that you find that $Y = 100 + 15 \times Month$, how would you interpret this result?

When the month grows, Y also increases. But wait, what do you mean by “month grows”?

Fixed Effects

One solution: We can assign Jan = 1, Feb = 2, March = 3, … and then we can simply run a linear regression as usual.

In addition, the above transformation implicitly assumes that Feb = 2 Jan, March = Jan + Feb, which does not make any sense!

Need a better way to regress.

Fixed Effects

Here, we adopt fixed effects. Instead of using a single variable month, we create 11 variables:

$$\text{Feb} = \begin{cases} 1 & \text{if month is Feb} \\ 0 & \text{otherwise} \end{cases}$$

Then, we can run a regression on these 11 variables. Note that we do not need to include Jan in the regression because, when $\text{Feb} = \text{Mar} = \cdots = \text{Dec} = 0,$ the month must be Jan.

Data Project

Question

Suppose that you have a brilliant idea, and you believe that your idea can change the world.

But you need resources to implement the idea and turn it into reality. This may cost you hundreds of thousands of dollars.

But you do not have much money yourself. What should you do?

Question

If you have a rich dad, then ask him to fund you.

If you have a rich friend, then ask him/her to support you.

If you are famous in the industry, then you can seek help from venture capitalists or private equities.

But what should you do if you do not have any of the above?

Solution

Now, you have a new option: crowdfunding (in Chinese: “众筹”). Crowdfunding is the practice of funding a project or venture by raising small amounts of money from a large number of people, typically via the Internet.

Many hands make light work. ---English proverb

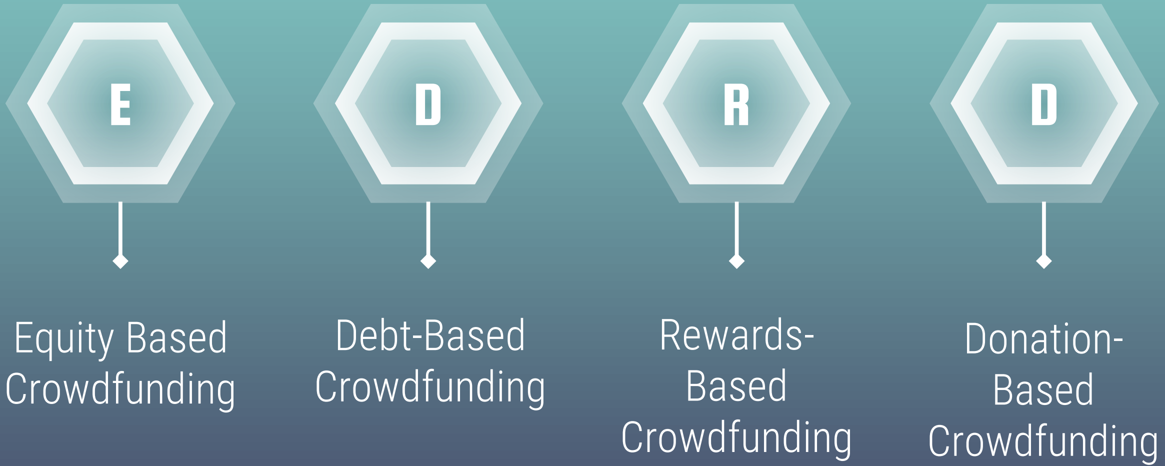
Crowdfunding

Are you familiar with these platforms?

INDIEGOGO
KICKSTARTER
PLEDGEMUSIC
WEFUNDER
FLASHFUNDERS
seed engine
Republic
SEED&SPARK

Understanding Crowdfunding

4 Types of crowdfunding



E

Equity Based Crowdfunding

D

Debt-Based Crowdfunding

R

Rewards-Based Crowdfunding

D

Donation-Based Crowdfunding

Equity Based Crowdfunding

The backer receives shares of a company, usually in its early stages, in exchange for the money pledged.

Example: WEFUNDER

Debt Based Crowdfunding

Debt-based crowdfunding is a crowdfunding model used to raise capital by taking loans from several investors (lenders) who expect to be repaid their loan with an added interest over the period that the loan was “used”. The entire process takes place through a crowdfunding platform.

Example: PROSPER

Donation Based Crowdfunding

Donation-based crowdfunding is when money is raised to support a good cause. As the name suggests, funding is raised through a crowd of people who decide to donate a certain amount of money to the cause, normally via online platforms specifically designed for the purpose.

Example:

Rewards Based Crowdfunding

Rewards-based, or seed, crowdfunding is a type of small-business financing in which entrepreneurs solicit financial donations from individuals in return for a product or service. There are about 19 times as many rewards campaigns as there are for its closely related counterpart, equity-based crowdfunding.

It is closely related to marketing and we focus on it in our class.

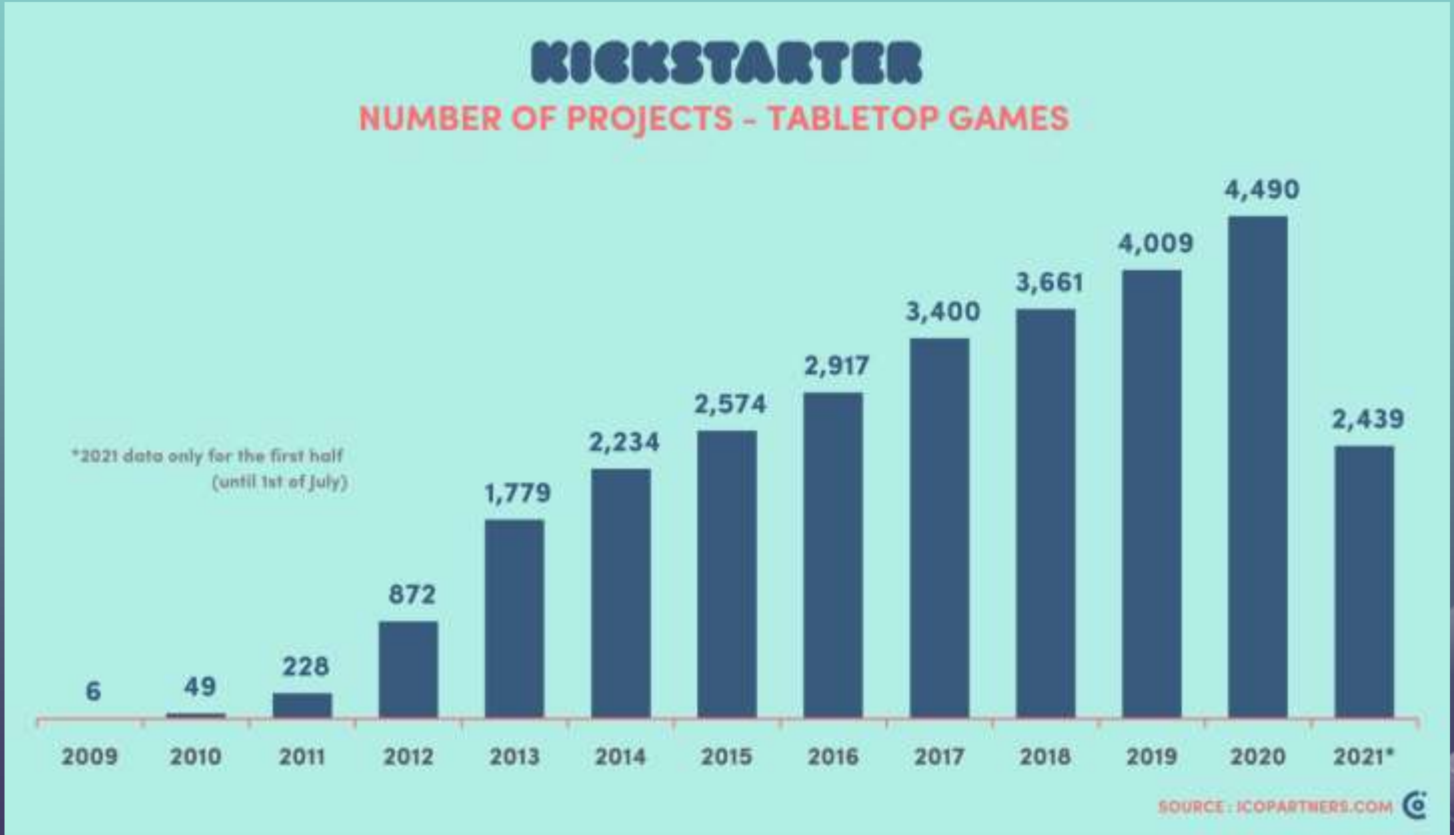
WHY KICKSTARTER?

Kickstart is THE largest leading crowdfunding platform.

As of October 2022, Kickstarter has received more than $6.9 billion in pledges from 21.5 million backers to fund 228,654 projects, such as films, music, stage shows, comics, journalism, video games, technology, publishing, and food-related projects.

It is a great example of the emerging crowdfunding market.

Kickstarter



SOURCE : ICOPARTNERS.COM

Product Categories in Kickstarter

Kickstarter supports almost all kinds of product categories including Art, Comics, Crafts, Dance, Design, Fashion, Film & Video, Food, Games, Journalism, Music, Photography, Publishing, Technology, and Theater.

Within each category, there are also several subcategories. For example, within the technology category, we have subcategories including gadgets, hardware, DIY electronics, flight, 3D printing, apps, camera equipment, etc.

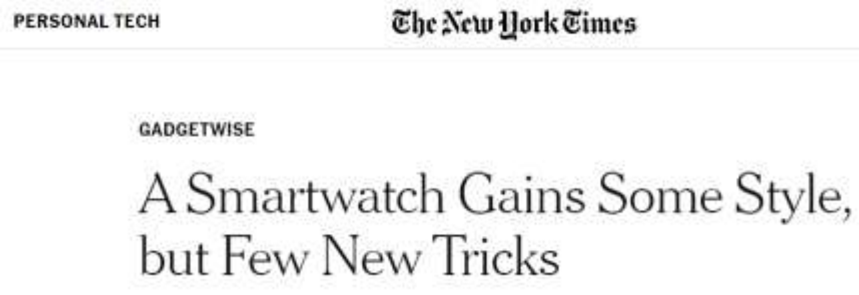
Pebble Watch

Pebble Watch was a smartwatch developed by the Pebble Technology Corporation. Funding was conducted through a Kickstarter campaign running from April 11, 2012 to May 18, 2012, which raised $10.3 million; it was the most funded project in Kickstarter history, at the time.

Let's visit Pebble Watch's initial crowdfunding webpage to know more about here. Click here to go.

Recall that it is in 2012.

Pebble Watch

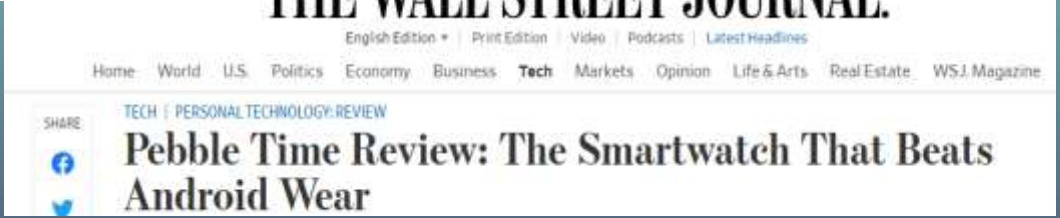


PERSONAL TECH

The New York Times

GADGETWISE

A Smartwatch Gains Some Style, but Few New Tricks



THE WALL STREET JOURNAL.

English Edition ▾ | Print Edition | Video | Podcasts | Latest Headlines

Home World U.S. Politics Economy Business **Tech** Markets Opinion Life & Arts Real Estate WSJ. Magazine

SHARE

TECH | PERSONAL TECHNOLOGY: REVIEW

Pebble Time Review: The Smartwatch That Beats Android Wear

Harvard Business Review

Pebble: Wearables Pioneer

Pebble Watch

In 2015, Pebble launched its second generation of smartwatches: the Pebble Time and Time Steel. The devices were similarly funded through Kickstarter, raising $20.3 million from over 75,000 backers and breaking records for the site. See the Kickstarter webpage here.

In 2016, Pebble shut down their subsequent Time 2 series watches and refunded Kickstarter backers, citing financial issues. It was purchased by Fitbit later.

Life on the Line

Cristian Barnett is a professional photographer living in Cambridge, England. Mr. Barnett was so fascinated with the Arctic Circle that in 2006 he started visiting the countries intersected by the circle. After seven years and a dozen trips to that area, he decided to create a book called *Life on the Line*, which would contain a selection of portraits he had taken over the years. See here for more details about the project.

“All-Or-Nothing”

Most crowdfunding platforms like Kickstarter strictly implements an “all-or-nothing” policy. That is, the creator (entrepreneur) must set up a target for the project. If the collected fund exceeds the target, the project is successful, and the creator uses the fund to run the project. Otherwise, the project fails, and all the money will be fully refunded to the investors (backers, consumers).

LEARNING ABOUT KICKSTARTER

THE DATA

I started to collect data from Kickstarter in its early age. At that time, crowdfunding was growing very fast, and many people wanted to figure out how crowdfunding really works.

I was one of the first a few people analyzing online crowdfunding using scientific methods.

Kickstarter has updated several times later and much of the data is no longer available to us. So the dataset is pretty unique.

BEFORE SEEING THE DATA…

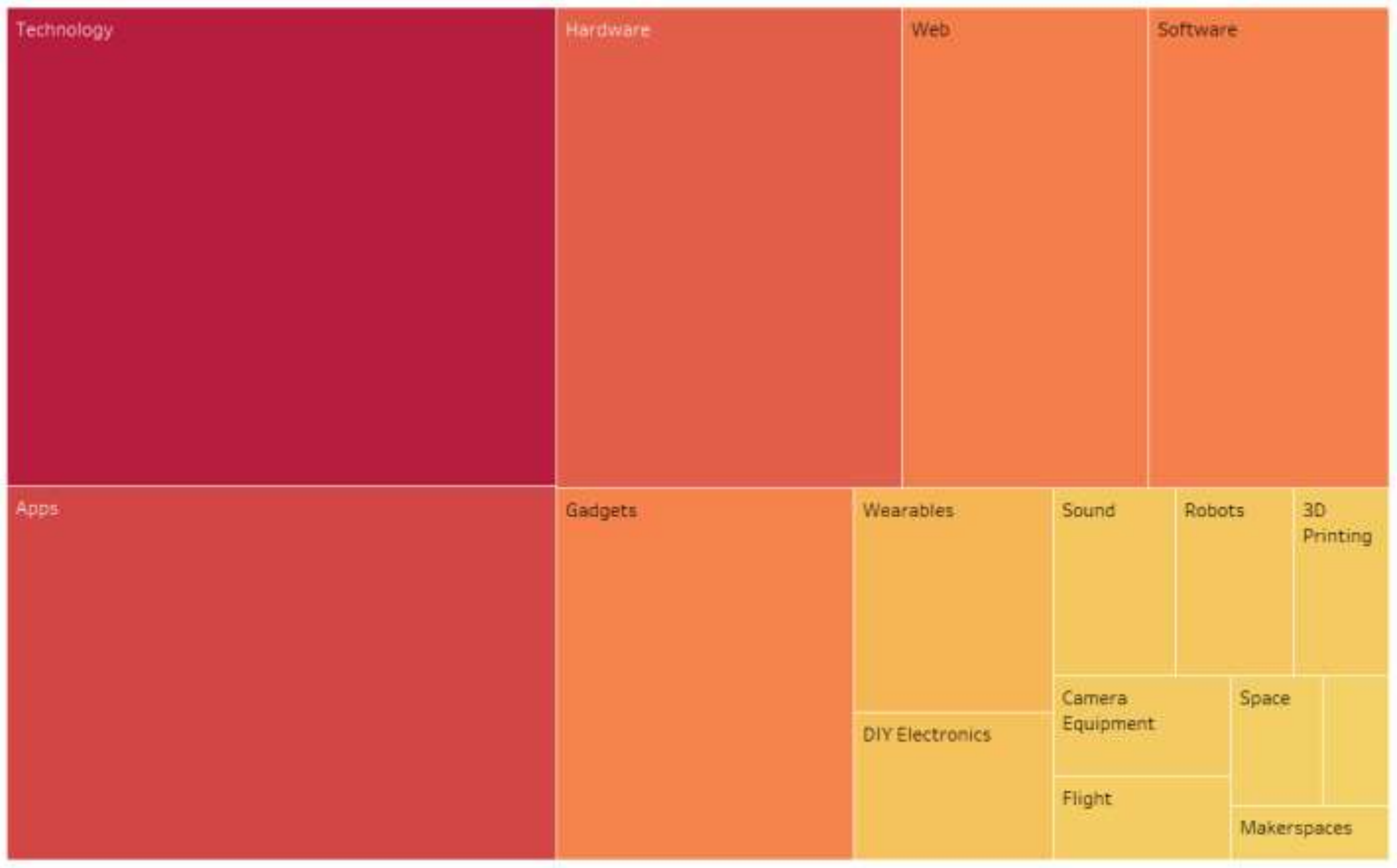
Please go to the Kickstarter website (https://www.kickstarter.com/), browse a few Kickstarter projects.

If you an entrepreneur trying to launch a successful crowdfunding campaign, what do you want to learn from Kickstarter?

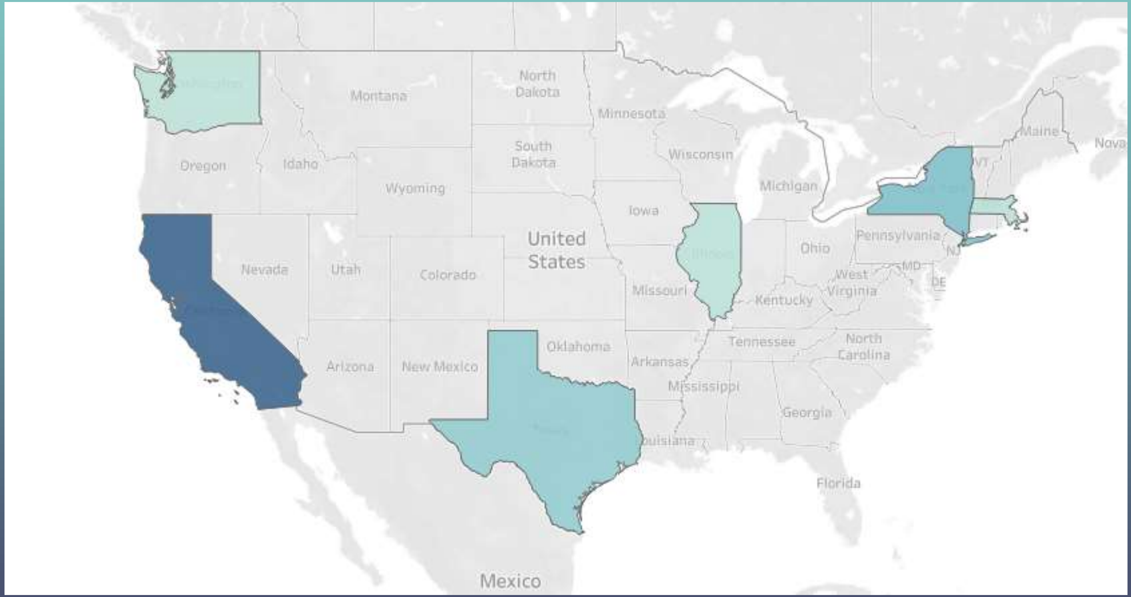
To answer these questions, which data do you need to collect?

The Data is Available at the Course Website

Subtype


Technology
Hardware
Web
Software
Apps
Gadgets
Wearables
Sound
Robots
3D Printing
DIY Electronics
Camera Equipment
Space
Flight
Makerspaces

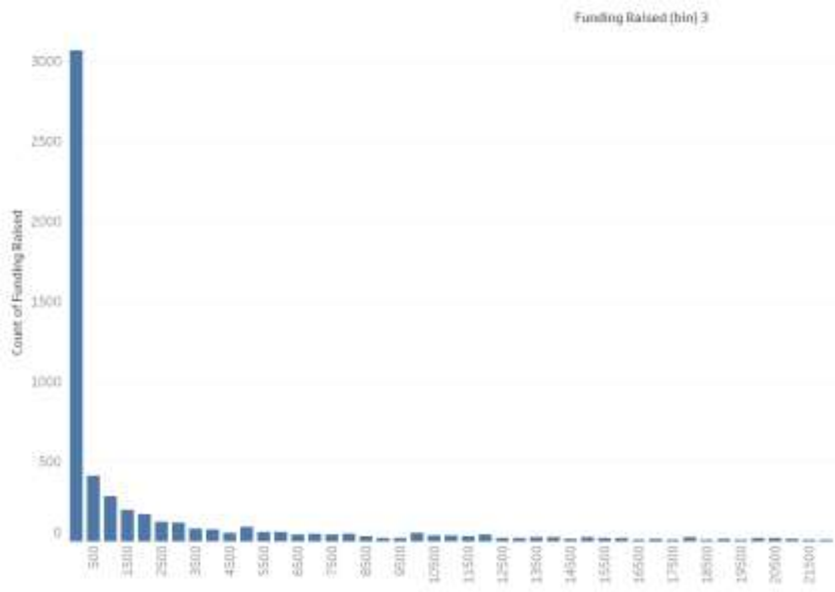
Location



The dataset covers projects from six US states: California (CA), New York (NY), Texas (TX), Massachusetts (MA), Washington (WA), and Illinois (IL).

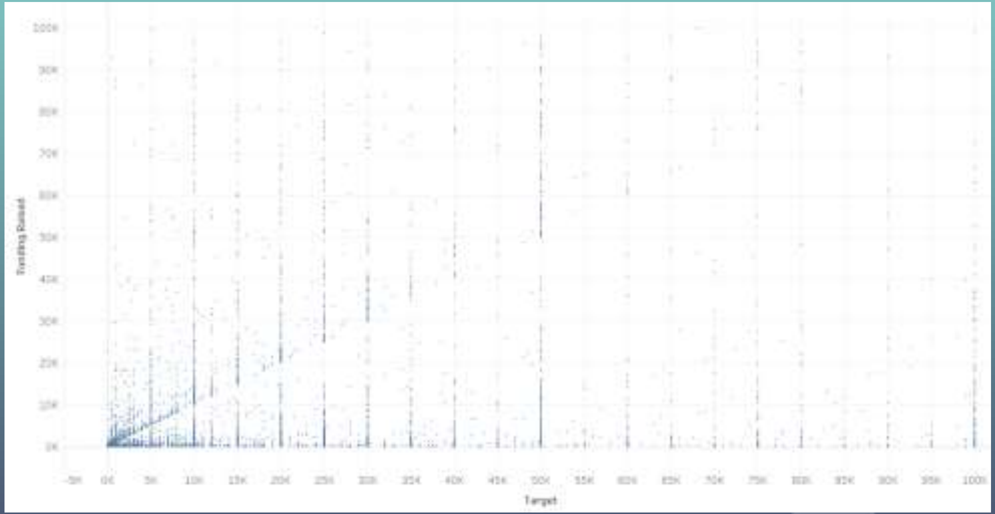
Total Funding Raised

The total funding raised by an individual project, measured by US$. You can see from the histogram that the total funding raised is really an L-shaped distribution: Most projects received almost $0 while some projects are very successful.



Target

At Kickstarter, each entrepreneur needs to specify a target for the project. The project is successful when the funds raised exceeds the target. Otherwise, the project fails and all the funds will be returned to the consumers.



Other Measures of Project Result

Outcome: Whether or not the project succeeded. It is a binary variable (1 = success, 0 = failure).

Backers: Number of people supporting the project. If you divide funding raised by the number of backers, you will get the average fund contributed by a backer.

Entrepreneurs' Personal History

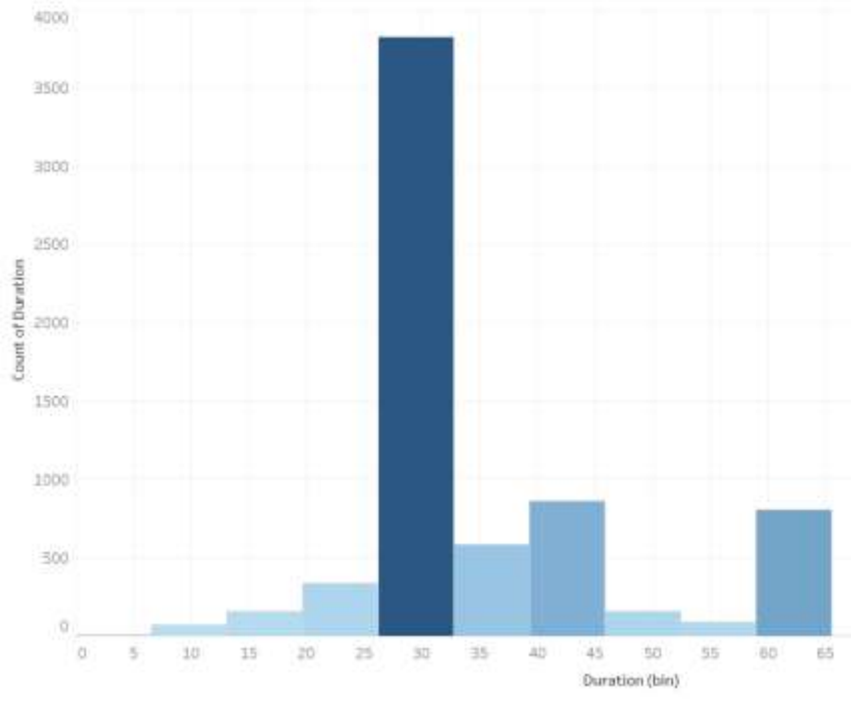
Created: Number of projects created by the same entrepreneur in the past. For example, 4 means the same entrepreneur had already created another 4 projects on Kickstarter.

Backed: Number of projects backed by the same entrepreneur in the past (i.e., the entrepreneur supporting others' projects on Kickstarter).

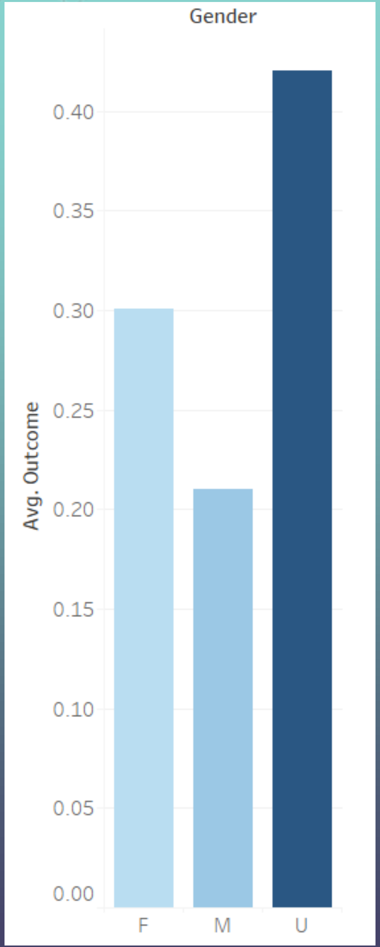
FbNumber: Number of Facebook friends the entrepreneur has.

Duration

The duration of a project’s fund raising period (in days). Most projects have a duration of around one month.



Gender



In the dataset, we have three genders: males, female, and unknown. The gender is obtained by analyzing the creators’ first name. Unknow refers to the case in which the name cannot be identified (e.g., a team name such as “marketing”).

Some Other Variables

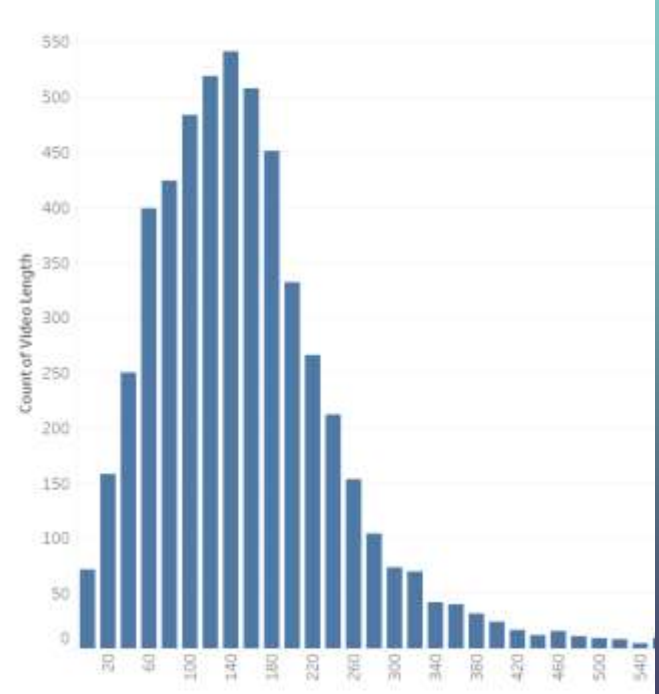
Video: Whether or not the project has a video. In the dataset, 76% of the projects have a video. Here, 1 means has video and 0 means no video.

Human: Whether or not the project's video features human-beings (usually the entrepreneurs themselves). 1 means has human-beings and 0 means no human-beings. This variable is set to 0 is the project does not have a video.

Computer: Whether or not the project's video features a computer. 1 means has computers and 0 means no computers. This variable is set to 0 is the project does not have a video.

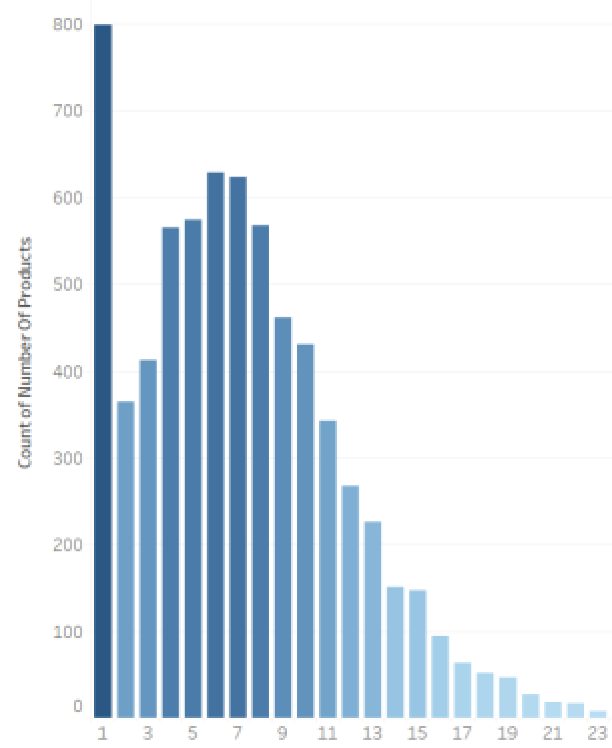
Some Other Variables

Video length: the duration of the video measured in seconds. For project without a video this variable is set to 0. The following is a histogram of video length (for projects with a video).



Some Other Variables

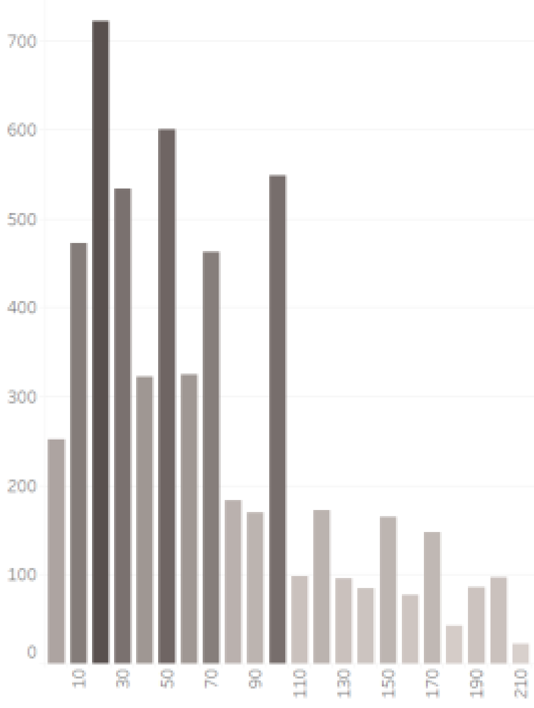
Number of Products: In a project the entrepreneurs often offer consumers a number of products to choose from. This variable measures how many products are offered in the project.



Some Other Variables

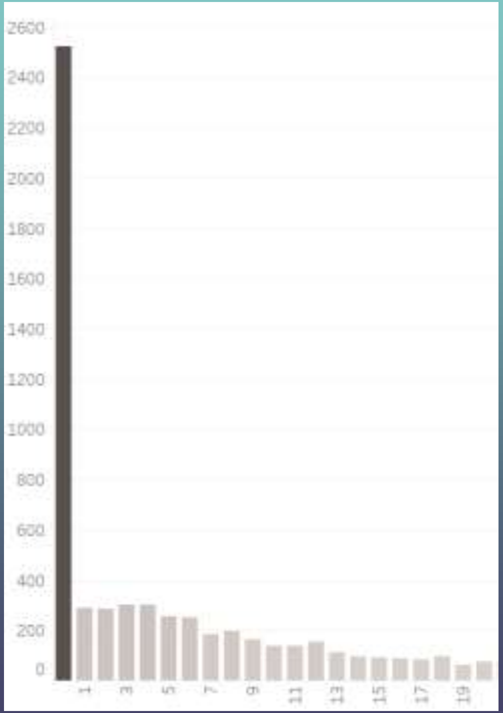
Price: A project may offer multiple products with different prices. Here, price means the median price among all product offerings.

Comments: Number of comments posted by consumers



Some Other Variables

Photo Number: Entrepreneurs often upload photos to their project description. This variable measures the number of photos uploaded to the project webpage.



Audio Measures

For projects with a video, we analyze their audio information:

Energy: Whether or not the audio pitch sounds energetic. A large number of an energetic audio pitch.

Content: Whether or not the audio pitch shows signs of content.

Upset: Whether or not the audio pitch shows signs of upset.

Angry: Whether or not the audio pitch shows signs of anger.

MaxAmpVol: The max sound volume. A greater number means louder sound.

What should we do?

Use the data to provide recommendations for the platform or the entrepreneurs. You can focus on anything that can be helpful for the platform or the entrepreneurs:

- How do males and females entrepreneurs behave differently on Kickstarter? (e.g., compared to females, males may be too aggressive in setting high targets.)
- Which type of video is most productive in terms of generating funds? (e.g., is having a lengthy video always beneficial?)
- What makes a successful crowdfunding project?

You need to submit:

10-15 pages slides documenting your findings from the data project. No presentation; no report.

In the slides: Please illustrate:

a. Your research question(s) --- At most two questions
b. The results from data analysis (including regression results)
c. Implications: What can others learn from the project?
d. That’s all. Don’t put anything extra in your slides.

Deadline: Nov 9 (Class B) and Nov 12 (Class A)