

Feature name	Source	Value type	Scale	Data type	Unique count	Null	Notes
backers_count				int64	3326	NA	number of backers
blurb				object	187330	NA	short description of the project Natural language processing can be helpful to find out if the description of the project can effect the number of backers.
category				object	170	NA	
converted_pledged_amount				int64	33045	NA	most currencies are converted into USD in this column, but not all. Instead, the 'usd_pledged' column will be used as these all use the same currency (the dollar).
country				object	25	NA	country pledged from
country_displayable_name				object	25	NA	
created_at				int64	188890	NA	
creator				object	212501	NA	most projects are by different people, and so this cannot be usefully used to group or categorise projects, and is not useful in a machine learning context
currency				object	15	NA	all currency values will be used as/converted to dollars, so that they can be evaluated together. It is not necessary to keep the original record because of this, and because it will be highly correlated with country (which will be kept).
currency_symbol				object	7	NA	as above
currency_trailing_code				bool	2	NA	as above
current_currency				object	3	NA	as above
deadline				int64	177707	NA	deadline for crowdfunding
disable_communication				bool	1	NA	99.7% of project owners did not disable communication with their backers (unsurprisingly). Because nearly all projects have the same value for this variable, it will be dropped as it does not provide much information.
friends				object	1		contain very few non-null entries, and can be dropped:
fx_rate				float64	52	NA	this is used to create 'converted_pledged_amount' from 'pledged', but does not always convert to dollars so can be dropped in favour of 'static_usd_rate' which always converts to dollars
goal				float64	5699	NA	fundraising goal- In order to allow for fair comparisons between projects, goals will be converted into dollars (as amount pledged already is) & 'goal' will be dropped
id				int64	188984	NA	
is_backing				object	2		contain very few non-null entries, and can be dropped:
is_starrable				bool	2	NA	Only 2.9% of projects were starrable by users. Although this is only a very small proportion, whether or not a project was liked and saved by users is likely to be informative about whether or not a project was successful, so the variable will be kept for now and assessed again once irrelevant rows have been dropped, to check it is still useful.
is_starred				object	2		contain very few non-null entries, and can be dropped:
launched_at				int64	188803	NA	
location				object	16865		There are a large number (16,865) of unique locations. Because the country is already recorded separately in the country field, and there are such a large number of unique categories (making one-hot encoding not useful, particularly as there are likely to be a lot of smaller towns and cities with very few projects), the column will be dropped.
name				object	188376	NA	The length of project names will be calculated, in case this is useful for the model. The original name variable will then be dropped.
permissions				object	1		contain very few non-null entries, and can be dropped:
photo				object	188957	NA	image processing/computer vision will not be used in this project
pledged				float64	48669	NA	data in this column is stored in native currencies, so this will be dropped in favour of 'usd_pledged' which is all in the same currency (dollars)
profile				object	188989	NA	this column contains a combination of information from other columns (e.g. id, state, dates, url).
slug				object	188984	NA	this is simply the 'name' column with hyphens instead of spaces
source_url				object	167	NA	the sites that the rows were each scraped from is not useful for building a model, as each is unique to an id
spotlight				bool	2	NA	projects can only be spotlighted after they are already successful, so this will be entirely correlated with successful projects
staff_pick				bool	2	NA	
state				object	4	NA	current condition the project is in
state_changed_at				int64	178829	NA	this is the same as deadline for most projects. The only exceptions are for projects which were cancelled before their deadline, but they will not be included in this analysis.
static_usd_rate				float64	14277	NA	
urls				object	189025	NA	as with source_url.
usd_pledged				float64	87153	NA	amount of money pledged
usd_type				object	2		it is unclear what this column means, but it is unlikely to be necessary since all currency values will be converted to dollars, and other currency information has been dropped