Kickstarter Dataset Description

The table below provides a description of each variable in the Kickstarter dataset, with notes on data leakage risk and usability for project success prediction at launch time.

Column Name	Description	Usable at Launch	Notes
id	Unique identifier for projects	Yes	Metadata only, not predictive
name	Project name	Yes	Contains textual information about the project
goal	Goal amount requested by the project	Yes	In currency of origin country
pledged	Amount pledged at time of data scrape	No	**DATA LEAKAGE**: Only known after campaign ends
state	Status of the project (successful, failed, etc.)	No	**TARGET VARIABLE**
disable_communication	If communication with project owners was disabled	Yes	Project configuration setting
country	Origin country of project	Yes	Geographic indicator
currency	Currency of origin country	Yes	Financial indicator
deadline	End date of project funding period	Yes	Campaign timeframe information
state_changed_at	Date and time the project state was modified	No	**DATA LEAKAGE**: Happens after campaign ends
created_at	Date and time project was created	Yes	Project timeline information
launched_at	Date and time project was launched	Yes	Project timeline information
staff_pick	If the project was a staff pick	No	**DATA LEAKAGE**: Generally happens after launch based on early performance
backers_count	Number of backers	No	**DATA LEAKAGE**: Only known after campaign starts
static_usd_rate	Conversion rate of project country currency to USD	Yes	Financial context information
usd_pledged	Amount pledged in USD	No	**DATA LEAKAGE**: Only known after campaign starts
category	Category of project	Yes	Project classification
spotlight	If the project was featured on kickstarter spotlight page	No	**DATA LEAKAGE**: Only featured after success

Column Name	Description	Usable at Launch	Notes
name_len	Length of project name in word count	Yes	Text metadata
name_len_clean	Length of project name without non-key words	Yes	Text quality indicator
blurb_len	Length of project blurb in word count	Yes	Text metadata
blurb_len_clean	Length of project blurb without non-key words	Yes	Text quality indicator
deadline_weekday	Weekday of deadline date	Yes	Timing information
state_changed_at_weekday	Weekday of state change	No	**DATA LEAKAGE**: Only known after campaign ends
created_at_Weekday	Weekday of creation date	Yes	Timing information
launched_at_weekday	Weekday of launch date	Yes	Timing information
deadline_month	Month of the project deadline	Yes	Timing information
deadline_day	Day of the project deadline	Yes	Timing information
deadline_yr	Year of the project deadline	Yes	Timing information
deadline_hr	Hour of project deadline	Yes	Timing information
state_changed_at_month	Month of latest state change	No	**DATA LEAKAGE**: Only known after campaign ends
state_changed_at_day	Day of latest state change	No	**DATA LEAKAGE**: Only known after campaign ends
state_changed_at_yr	Year of latest state change	No	**DATA LEAKAGE**: Only known after campaign ends
state_changed_at_hr	Hour of latest state change	No	**DATA LEAKAGE**: Only known after campaign ends
created_at_month	Month of creation date	Yes	Timing information
created_at_day	Day of creation date	Yes	Timing information
created_at_yr	Year of creation date	Yes	Timing information
created_at_hr	Hour of creation date	Yes	Timing information
launched_at_month	Month of launch date	Yes	Timing information
launched_at_day	Day of launch date	Yes	Timing information
launched_at_yr	Year of launch date	Yes	Timing information

Column Name	Description	Usable at Launch	Notes
launched_at_hr	Hour of launch date	Yes	Timing information
main_category	Broader category of project	Yes	Project classification
staff_pick1	If project was selected for 'Projects We Love' page	No	**DATA LEAKAGE**: Selection happens after launch
show_feature_image	If project description includes product images	Yes	Project presentation quality
video	If project description includes video	Yes	Project presentation quality