KickStarter successful projects

john myers

2/22/2022

The purpose of this project is to visualize the categories of projects that were successful and which projects seemed to be the most popular by category. There are many reasons why a project will not be successful, but here we will look at funding and backers as a way of measuring success.

Import the libraries

**library**("tidyverse")

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4

## v tibble 3.1.6 v dplyr 1.0.8

## v tidyr 1.2.0 v stringr 1.4.0

## v readr 2.1.2 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --

## x dplyr::filter() masks stats::filter()

## x dplyr::lag() masks stats::lag()

**library**("dplyr")

Import the dataset for 2018.These are all the projects on Kickstarter since it began in April 2009.

data <- read\_csv("ks-projects-201801.csv")

## Rows: 378661 Columns: 15

## -- Column specification --------------------------------------------------------

## Delimiter: ","

## chr (8): name, category, main\_category, currency, deadline, launched, state,...

## dbl (7): ID, goal, pledged, backers, usd pledged, usd\_pledged\_real, usd\_goal...

##

## i Use `spec()` to retrieve the full column specification for this data.

## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

head(data)

## # A tibble: 6 x 15

## ID name category main\_category currency deadline goal launched pledged

## <dbl> <chr> <chr> <chr> <chr> <chr> <dbl> <chr> <dbl>

## 1 1.00e9 The ~ Poetry Publishing GBP 10/9/20~ 1000 8/11/20~ 0

## 2 1.00e9 Gree~ Narrati~ Film & Video USD 11/1/20~ 30000 9/2/201~ 2421

## 3 1.00e9 Wher~ Narrati~ Film & Video USD 2/26/20~ 45000 1/12/20~ 220

## 4 1.00e9 Tosh~ Music Music USD 4/16/20~ 5000 3/17/20~ 1

## 5 1.00e9 Comm~ Film & ~ Film & Video USD 8/29/20~ 19500 7/4/201~ 1283

## 6 1.00e9 Mona~ Restaur~ Food USD 4/1/2016 50000 2/26/20~ 52375

## # ... with 6 more variables: state <chr>, backers <dbl>, country <chr>,

## # `usd pledged` <dbl>, usd\_pledged\_real <dbl>, usd\_goal\_real <dbl>

Check for missing values

data %>% is.na() %>% colMeans()

## ID name category main\_category

## 0.000000e+00 1.056354e-05 0.000000e+00 0.000000e+00

## currency deadline goal launched

## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00

## pledged state backers country

## 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00

## usd pledged usd\_pledged\_real usd\_goal\_real

## 1.002744e-02 0.000000e+00 0.000000e+00

Get name of all the columns

colnames(data)

## [1] "ID" "name" "category" "main\_category"

## [5] "currency" "deadline" "goal" "launched"

## [9] "pledged" "state" "backers" "country"

## [13] "usd pledged" "usd\_pledged\_real" "usd\_goal\_real"

Here we see that the amount of money pledged increased with the number of backers. All 15 categories were represented in different amounts of pledged money and number of backers. This graph takes a 10% sample of the data so that the data could fit on the graph. A design category company had the highest number of pledged money and backers, followed by two games category companies.

data %>%

sample\_frac(0.1) %>%

ggplot(aes(x= pledged , y= backers, color=main\_category )) + geom\_point() +

scale\_x\_continuous(trans='log10') +

scale\_y\_continuous(trans='log10')

## Warning: Transformation introduced infinite values in continuous x-axis

## Warning: Transformation introduced infinite values in continuous y-axis

Chart, scatter chart

Description automatically generated

These barcharts show that Games, Design, Technology and Comics had the greatest amount of backers and pledged funding with design in the lead of funding and games in the lead of backers.

data %>%

group\_by(main\_category) %>%

summarise(backers = mean(backers), pledged = mean(pledged)) %>%

gather(key, val, -main\_category) %>%

ggplot(aes(y=main\_category,x=val)) + geom\_col(position = 'dodge') +

facet\_wrap(~key, scales = 'free\_x') +

labs(title = "Backers and Pledged funding per category")

Chart, timeline, bar chart

Description automatically generated

finding the unique values for a project’s status.

unique(data$state)

## [1] "failed" "canceled" "successful" "live" "undefined"

## [6] "suspended"

###Looking for projects that were successful. This graph shows that most projects were not successful. Projeccts with more funding generally had a better chance of being successful depending on what their goal amount was.Music was the most successful category.

data %>%

ggplot(aes(state,main\_category, color = main\_category)) + geom\_col() +

labs(title= "measurement of Success of projects")

Chart, bar chart

Description automatically generated

Conclusion:

Using number of backers and funding, although predicted some success was not a great measure for success since different projects had different goal amounts and backers could invest a small amount or a large amount in a project so the amount of them was not a direct measure of success. Games, Design, Technology and Comics had the greatest amount of backers in their respective categories.