**Summary**

Prior to building a project the category an idea belongs to is of prime importance. The success mainly depends on people’s acceptance which makes us explore the pulse of project backers. Several categories are doing extremely well like dance and Theatre. They have success probability of 60% approximately. Contrastingly Technology has lowest probability of success. I would recommend a project in Theatre, dance or Music. As the market is diverse financial investment for various categories differs. It would be better to understand which category raises more money. Apparently, Technology sector requires on an average 100000$ per project. We can infer Technology as a high demanding and less likely to succeed. On the other hand, Technology raises more money than any other category which confirms backers’ allure. Overall, it’s a high-risk high gain segment.

Next question is when to launch project? We see a sudden surge in the job market after the Financial year end this is because company comes with new plan for next year soon after the year end. Similarly, investors plan on investing during this period. We proceeded with an assumption that a successful money raise for a project happens after year end (June). It turned out that most of the successful projects were launched at this period. So, project creators should focus on June as the deadline to bring it on to the platform. Another question following this is deadline for the fund raise,

More time more money, not in this case few tests were conducted which showed us good projects sell like hot cake in the market. Ideal time frame to raise fund would be 32 days from the launch.

The outer picture would be to asses the platform itself by checking is it worth to use kickstart platform to raise fund. This is by comparing whether difference in pledged and goal is significantly greater than 0. The test results came up to be positive indicating Kickstarter is a promising platform.

**Introduction**

Kickstarter is a platform to raise money for the creative, imaginative and ambitious projects. It covers a wide variety of projects like Music, Dance, Technology etc. A creator of the project will pitch in ideas with a clear goal and budget. Backers are people who are willing to invest in a project. When forecasted amount reaches the pledged amount a project is considered to be successfully funded in the platform.

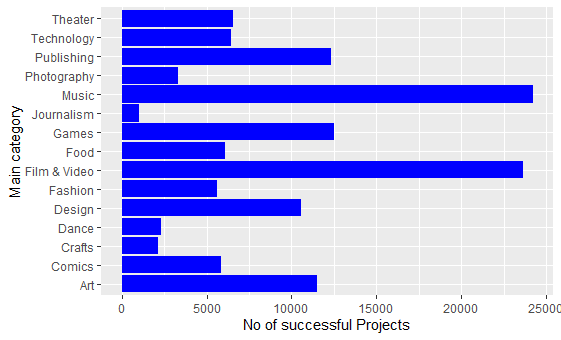
It’s a great platform to express an idea. Kickstarter takes time to curate projects so that a ton of bad ones, can be filtered out to make the backer experience more pleasant. Feature for making a pledge is pretty straightforward and has enough information to scroll through to assess the perks before making a pledge. Creating a project to Kickstarter is relatively easy. Not much technicality involved in creating a project which helps creators from different category to participate. It lights up unexplored demand in the market. Getting it online makes it wide reach and people to invest on future. Customized project proposal helps creators to express their complex ideas in a comprehensive manner.

Usually in the business there are two people involved producer and supplier but Kickstarter platform has three different prospective. A creator who believes in his project and seeks investment. And people exploring to invest in a venture and finally, platform owner to make stage friendly and lively. So, as a Data Analyst we must answer questions pertaining to each prospective. As a creator I would like to know popular and expensive categories, an ideal time to launch the project and wait time. A backer would like to invest on trending, unique and might get popular projects. Likewise, platform owner needs to know the overall progress of the ventures.

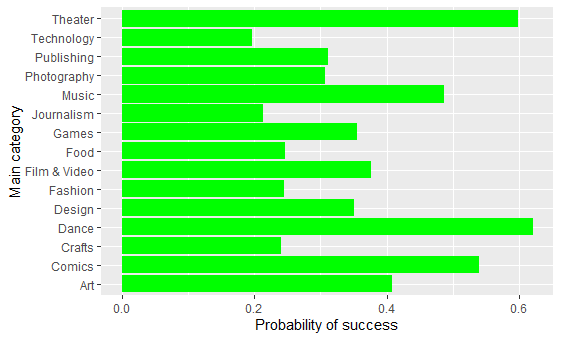
Insights

1. Which categories are more successful in reaching goal?

There are many categories to build a project on. But as an analyst we going to see how various categories are performing.



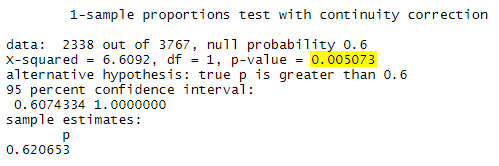
We can see that Music and Film-Video categories have highest successful projects it might be because of large number of projects in these categories. Let us plot a success probability for each category.



Graph clearly indicates probability of success in dance and theatre are around 60%. Surely these categories outperform others. We shall have a proportionality z -test for the confirmation.

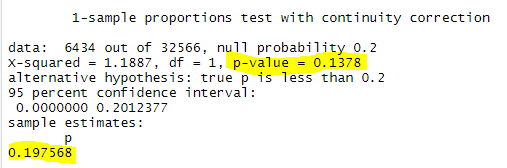
H0 : P <= 0.6

H1 : p>0.6



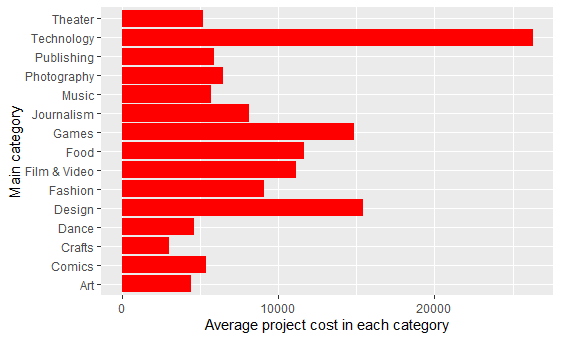
At significance level of 5% we reject the null hypothesis and conclude that probability of success of dance project is 60% and above.

Technology and Journalism have low probability of success typically below 20%.



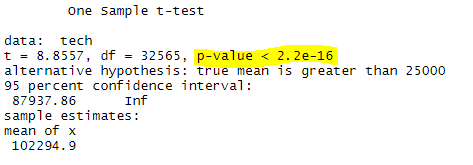
1. Which category raises more money?

Though technology has less probability of success it seems to have highest average investment for a project. The average fund raise for a tech project is above 100000$. We can prove the same by a hypothesis test.



H0 : Average fund raised for technology <= 25000

H1 : Average fund raised for technology > 25000



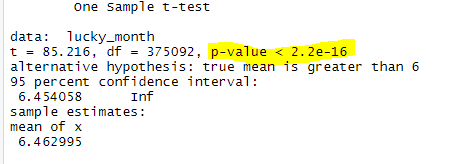
Thus, we have enough statistical evidence to reject null and say average money for successful project in technology is greater than 25000.

1. What time is suitable to launch project?

We are testing this hypothesis with an assumption that successful money raise happens for a project when it is launched June or July that is 6 or 7 months of the year.

H0 : month <=6

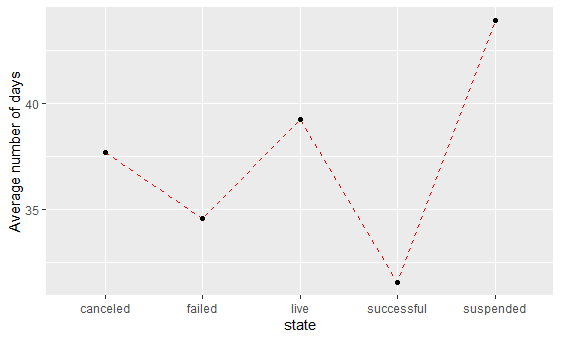
H1 : month >6



We reject null hypothesis and conclude projects launched after June are more likely to be successful.

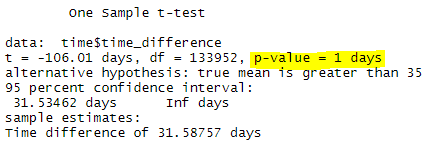
1. Typical deadline for a successful project.

An interesting, innovative and successful projects raise money earlier than other. Below graph shows successful projects take less time compared to failed and suspended. We conduct a hypothesis test to confirm whether successful projects take more than 35 days.



H0 : Average number of days <= 35

H1 : Average number of days > 35



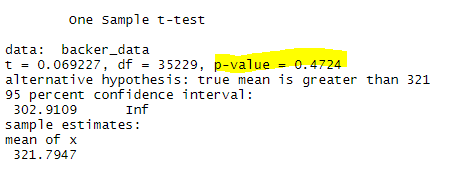
We can reject null hypothesis and conclude successful projects take less than 35 day to raise pledged money.

1. What categories are backers willing to invest?

Popular categories among backers are Games, Technology and Comics. These categories attract huge amount of money with average number of investors 321, 164 and 134. This clearly indicates Games as most popular one. A hypothesis test was conducted to confirm Games popularity.

H0 : Average no of backers for games >= 321

H1 : Average no of backers for games < 321



With a p-value of 0.4724 we cannot reject null hypothesis and conclude that Games is most popular category among backers.

1. Is the platform useful/popular ?

On the outer picture does platform really help project fund raising. This question can be answered joint hypothesis test. If on an average pledge amount is equal goal amount the platform is supposed to be successful.

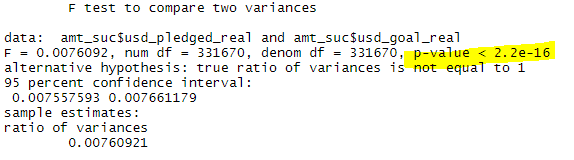
H0 : USD\_goal – USD\_pledge=0

H1 : USD\_goal – USD\_pledge!=0

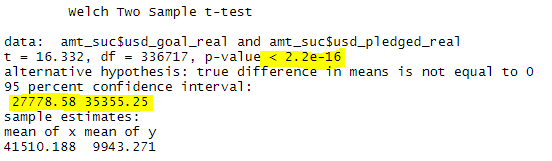
Test of Variance

H0 : equal Variance

H1 : unequal Variance



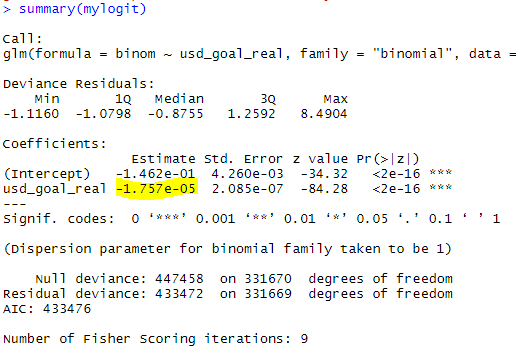
F-test reject null hypothesis to conclude unequal variance.



Two sample t-test with unequal variance rejects null hypothesis. Thus, we can infer there is significance difference between pledge and goal. In fact, confidence interval indicates a positive number. A huge difference indicates not many projects are successfully funded. Please note only successful and failed projects are considered.

1. Does success depend on amount?

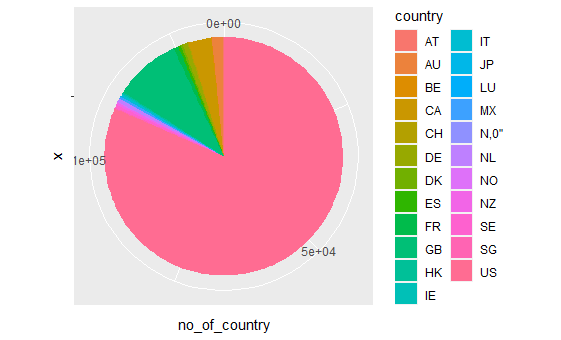
During the initial days people rely less on the online platform for investment. This sparks a question does low budget projects are funded more than costly ones. This analysis is carried out using a logistic regression which help us understand how project cost decides its success.



Co-efficient of usd\_goal\_real is -1.757e-05. Thus odds ratio = e^-1.757e-05 = 0.99999. Each unit increase in goal amount increases odds of success by 1. Which practically has no significance. Hence, we can say that goal amount does not impact much for the success of the project.

1. Most popular destination for projects success.

Undoubtedly US prevails in this section. There are a greater number of project in US compared to other country. Thus, more number of success over a margin. Below graph shows US market share in successful projects.



Appendix

All the analysis has been carried out in r and below is the relevant code.

**Data loading and exploration**

install.packages("ggplot2")

library("dplyr")

library('ggplot2')

kickstart <- read.csv("DATA.csv")

subset <- kickstart %>% select(main\_category,category,state)

distinct(kickstart,state)

ggplot(data=(agg1 %>% filter(state=='canceled')), aes( x=count,y=main\_category)) +

geom\_bar( stat="identity",fill='blue') +

xlab("No of canceled Projects") + ylab("Main category")

ggplot(data=(agg1 %>% filter(state=='failed')), aes( x=count,y=main\_category)) +

geom\_bar( stat="identity",fill='blue') +

xlab("No of failed Projects") + ylab("Main category")

ggplot(data=(agg1 %>% filter(state=='successful')), aes( x=count,y=main\_category)) +

geom\_bar( stat="identity",fill='blue') +

xlab("No of successful Projects") + ylab("Main category")

ggplot(data=(agg1 %>% filter(state=='live')), aes( x=count,y=main\_category)) +

geom\_bar( stat="identity",fill='blue') +

xlab("No of live Projects") + ylab("Main category")

###1#####

At first probability of each category is visualized and a hypothesis test was conducted to confirm.

prob\_suc <- kickstart %>%

filter(state=='successful') %>%

select(main\_category,state) %>%

group\_by(main\_category) %>%

summarise(count\_suc = n())

prob\_suc1 <- kickstart %>%

select(main\_category,state) %>%

group\_by(main\_category) %>%

summarise(count\_t = n())

dummy\_prob <- cbind(prob\_suc,prob\_suc1)

dummy\_prob$prob <- dummy\_prob$count\_suc/dummy\_prob$count\_t #probability of success in dance is 0.62

head(dummy\_prob2)

dummy\_prob2 <- as.data.frame(dummy\_prob$prob)

dummy\_prob2 <- cbind(dummy\_prob2,prob\_suc$main\_category)

ggplot(data=dummy\_prob2, aes( x=dummy\_prob$prob,y=prob\_suc$main\_category)) +

geom\_bar( stat="identity",fill='green') +

xlab("Probability of success") + ylab("Main category")

prop.test(2338,3767,p=0.6,alternative = "greater",correct = TRUE) #dance probability of success is more

prop.test(6434,32566,p=0.2,alternative = "less",correct = TRUE) #technology probability of success is less

#####2###########

Average cost for project in each category is calculated first and one sample t test was conducted on technology to confirm.

subset5 <- kickstart %>% select(main\_category,usd\_goal\_real)

agg6 <- subset5 %>%

select(main\_category,usd\_goal\_real) %>%

group\_by(main\_category) %>%

summarise(Average\_cost=mean(usd\_goal\_real))

ggplot(data=agg6, aes( y=main\_category,x=Average\_cost)) +

geom\_bar( stat="identity",fill='red') +

xlab("Average project cost in each category") + ylab("Main category")

t.test(tech,mu = 25000,alternative = "greater")

####3######

Number of projects in each month which were successful was tabulated and month with highest success rate was hypothesis tested

agg3 <- subset2 %>%

select(deadline,launched,state,time\_difference)%>%

group\_by(state)%>%

summarise(total=mean(time\_difference))

ggplot(data=agg3, aes(x=state, y=total , group=1)) +

geom\_line(linetype='dashed', color="red")+ xlab("state") + ylab("Average number of days")+

geom\_point()

lucky\_month <- month(subset2$launched)

mean(lucky\_month)

t.test(lucky\_month,mu=6,alternative = "greater")

########4#####

The average time taken for all state of projects was calculated and surprisingly successful projects take less time which was further confirmed by a t test.

time <- subset2 %>%

filter(state=='successful')

t.test(time$time\_difference,mu = 35,alternative = "greater")

########5##########

To check which category is popular among backers, average number of backers for each category was visualized and a single sample t test was conducted for confirmation.

backers <- kickstart %>%

select(main\_category,backers) %>%

group\_by(main\_category) %>%

summarise(average =mean(backers))

max(backers$average)#Games is the sector with more backers average for each project.

backer\_data <-kickstart %>%

filter(main\_category=='Games') %>%

select(backers)

mean(backer\_data$backers)

t.test(backer\_data,mu=321,alternative = "greater")

#########6############

To check overall platform efficiency each

sample\_table <- table(kickstart$state,kickstart$usd\_pledged\_real,kickstart$usd\_pledged\_real)

amt\_suc <- kickstart %>%

filter(state=='successful' | state=='failed') %>%

select(state,usd\_pledged\_real,usd\_goal\_real)

amt\_suc <-as.data.frame(amt\_suc)

group\_by(amt\_suc, state) %>%

summarise(

count = n(),

mean = mean(usd\_pledged\_real, na.rm = TRUE),

sd = sd(usd\_pledged\_real, na.rm = TRUE)

)

hist(amt\_suc$usd\_pledged\_real[amt\_suc$state == "successful"])

hist(amt\_suc$usd\_pledged\_real[amt\_suc$state == "failed"])

var.test(amt\_suc$usd\_pledged\_real,amt\_suc$usd\_goal\_real)

t.test(amt\_suc$usd\_goal\_real,amt\_suc$usd\_pledged\_real,,var.equal=FALSE,alternative="two.sided")

#######7##########

logit<- kickstart %>%

filter(state=='successful' | state=='failed') %>%

select(state,usd\_goal\_real)

logit$binom <- ifelse((logit$state=='successful'),1,0)

mylogit <- glm(binom ~ usd\_goal\_real , data = logit, family = "binomial")

summary(mylogit)

###8####

agg5\_updated <- filter(agg5,state=="successful")

ggplot(agg5\_updated, aes(x="", y=no\_of\_country ,fill=country)) +

geom\_bar(stat="identity", width=2)+

coord\_polar("y", start=0)