Data visualization and analysis using R

Jie Zhang

10/27/2019

Data Visualization and EDA

```
# Call libraries
library(tidyverse)
```

```
## -- Attaching packages ----------------------------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.2.1     v purrr   0.3.2
## v tibble  2.1.3     v dplyr   0.8.3
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## -- Conflicts -------------------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggthemes)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

```
#library(rworldmap)
library(gplots)
```

```
##
## Attaching package: 'gplots'
```

```
## The following object is masked from 'package:stats':
##
##     lowess
```

```
library(knitr)
```

```
ks_data <- read.csv("Kickstarter_data.csv", fileEncoding="latin1")
str(ks_data)

# remove the rows with mulls
sapply(ks_data, function(x) sum(is.na(x)))
#ks_data<-ks_data[-1,]
head(ks_data)

#=========================================================
#break down the number of projects by
#their status (e.g. successful, failed, cancelled, etc.).
#=========================================================
```

```r
state.freq <- ks_data %>%
  group_by(state) %>%
  summarize(count=n()) %>%
  arrange(desc(count))

state.freq.ratio <- ks_data %>%
  group_by(state) %>%
  summarize(ratio=(n()/nrow(ks_data))) %>%
  arrange(desc(ratio))

#state.freq.ratio

state.freq$state <- factor(state.freq$state, levels=state.freq$state)

ggplot(state.freq, aes(state, count, fill=count)) + geom_bar(stat="identity") +
  ggtitle("Projects by State") + xlab("Project State") + ylab("Frequency") +
  geom_text(aes(label=count), vjust=-0.5) + theme_economist() +
  theme(plot.title=element_text(hjust=0.5), axis.title=element_text(size=12, face="bold"),
        axis.text.x=element_text(size=12), legend.position="null") +
  scale_fill_gradient(low="skyblue1", high="royalblue4")


#=========================================================
# Which projects tend to succeed?
#=========================================================

#=========================================================
# state versus median project goal
#=========================================================

state.goal <- ks_data %>%
  filter(state %in% c("successful", "failed"))
state.goal<-data.frame(state.goal$goal, state.goal$state)

colnames(state.goal)<-c('goal','state')

state.goal.1<-filter(state.goal, state=="failed")
median(state.goal.1$goal) # 9600
state.goal.2<-filter(state.goal, state=="successful")
median(state.goal.2$goal) # 4500


c1<-c(9600, 4500)
c2<-c("failed","successful")
c3<-data.frame(cbind(c1,c2))
colnames(c3)<-c('median_goal','state')
c3


ggplot(data=c3, aes(x=state, y=median_goal,fill=state)) +
  geom_bar(stat="identity",width = 0.5)+ggtitle("State vs. Median project goal") +
  xlab("state") + ylab("median goal")+
  theme_minimal()
```

```r
#=======================================================
# state versus mean project duration
#=======================================================
# introduce the processed data by python
ks_data1 <- read.csv("Kickstarter_data_update2.csv", fileEncoding="latin1")
ks_data1 <- ks_data1 %>%
  filter(is_successful %in% c(0, 1))
ks_data1<-data.frame(ks_data1$duration, ks_data1$is_successful)
colnames(ks_data1)<-c('duration','state')
ks_data1<-na.omit(ks_data1)
state.duration.1<-filter(ks_data1, state==0)
mean(state.duration.1$duration) #34.44344
state.duration.2<-filter(ks_data1, state==1)
mean(state.duration.2$duration) #30.71463

c1<-c(34.44, 30.71)
c2<-c("failed","successful")
c3<-data.frame(cbind(c1,c2))
colnames(c3)<-c('mean_duration','state')
c3

ggplot(data=c3, aes(x=state, y=mean_duration,fill=state)) +
  geom_bar(stat="identity",width = 0.5)+ggtitle("State vs. Mean project duration") +
  xlab("state") + ylab("mean days")+
  theme_minimal()

#=====================================================
# project success vs. rate main category
#=====================================================

state.pct <- ks_data %>%
  filter(state %in% c("successful", "failed")) %>%
  group_by(main_category, state) %>%
  summarize(count=n()) %>%
  mutate(pct=count/sum(count)) %>%
  arrange(desc(state), pct)

state.pct$main_category <- factor(state.pct$main_category,
                                  levels=state.pct$main_category[1:(nrow(state.pct)/2)])

ggplot(state.pct, aes(main_category, pct, fill=state)) + geom_bar(stat="identity") +
  ggtitle("Success vs. Failure Rate by Project Main Category") +
  xlab("Project Main Category") + ylab("Percentage") + scale_y_continuous(labels=scales::percent) +
  scale_fill_discrete(name="Project Status", breaks=c("successful", "failed"),
                      labels=c("Success", "Failure")) +
  geom_text(aes(label=paste0(round(pct*100,1),"%")), position=position_stack(vjust=0.5),
            colour="white", size=5) + theme_economist() +
  theme(plot.title=element_text(hjust=0.5), axis.title=element_text(size=12, face="bold"),
        axis.text.x=element_text(size=12), legend.position="bottom",
        legend.title=element_text(size=12, face="bold")) + coord_flip()
```

```r
#==================================
# project success rate by country
#==================================
ks_data <- read.csv("Kickstarter_data.csv", fileEncoding="latin1")
ks_data<-na.omit(ks_data)
state.ctry <- ks_data %>%
  filter(country!='N,0"') %>%
  filter(state %in% c("successful", "failed")) %>%
  group_by(country, state) %>%
  summarize(count=n()) %>%
  mutate(pct=count/sum(count)) %>%
  arrange(desc(state), pct)

state.ctry$country <- factor(state.ctry$country,
                             levels=state.ctry$country[1:(nrow(state.ctry)/2)])

ggplot(state.ctry, aes(country, pct, fill=state)) + geom_bar(stat="identity") +
  ggtitle("Success vs. Failure Rate by Country") +
  xlab("Country") + ylab("Percentage") + scale_y_continuous(labels=scales::percent) +
  scale_fill_discrete(name="Project Status", breaks=c("successful", "failed"),
                      labels=c("Success", "Failure")) +
  geom_text(aes(label=paste0(round(pct*100,1),"%")), position=position_stack(vjust=0.5),
            colour="white", size=5) + theme_economist() +
  theme(plot.title=element_text(hjust=0.5), axis.title=element_text(size=12, face="bold"),
        axis.text.x=element_text(size=12), legend.position="bottom",
        legend.title=element_text(size=12, face="bold")) + coord_flip()

#============================================================
# Log  Project goals versus project success rate, dot plot
#============================================================
# Because goals are highly skewed, so we do log transform
ks_data1 <-
  read.csv("Kickstarter_data_update1.csv",
           fileEncoding="latin1")
ks_data1<-ks_data1[,-1]
ks_data1<-na.omit(ks_data1)
ks_data1<-mutate(ks_data1, log_goal=log(goal))
#names(ks_data1)

goal.pct <- ks_data1 %>%
  filter(is_successful %in% c(1, 0)) %>%
  group_by(log_goal, is_successful) %>%
  summarize(count=n()) %>%
  mutate(pct=count/sum(count))

ggplot(goal.pct[goal.pct$is_successful==1,], aes(log_goal, pct)) +
  geom_point(colour="royalblue4", size=2.5) + ggtitle("Success Rate vs. Log Project Goal") +
  xlab("Log Project Goal (Dollors)") + ylab("Success Rate (%)") +
  scale_x_continuous(breaks=c(0,2,4,6,4,8,10,12,14,16,18,20)) +
  theme_economist() +
  theme(plot.title=element_text(hjust=0.5), axis.title=element_text(size=12, face="bold"))
```

```r
#=======================================================
# Project duration versus project success rate, dot  plot
#=======================================================
ks_data1 <-
  read.csv("Kickstarter_data_update1.csv",
           fileEncoding="latin1")
ks_data1<-ks_data1[,-1]

#The maximum project duration according to Kickstarter's rules is 60 days
length.pct <- ks_data1 %>%
  filter(is_successful %in% c(1, 0), duration <= 61) %>%
  group_by(duration, is_successful) %>%
  summarize(count=n()) %>%
  mutate(pct=count/sum(count))

ggplot(length.pct[length.pct$is_successful==1,], aes(duration, pct)) +
  geom_point(colour="royalblue4", size=2.5) + ggtitle("Success Rate vs. Project Duration") +
  xlab("Project Duration (Days)") + ylab("Success Rate (%)") +
  scale_x_continuous(breaks=c(0,10,20,30,40,50,60)) + geom_vline(xintercept=30, colour="red") +
  theme_economist() +
  theme(plot.title=element_text(hjust=0.5), axis.title=element_text(size=12, face="bold"))


#=========================================
# state versus mean project name length
#=========================================

ks_data1 <- read.csv("Kickstarter_data_update1.csv", fileEncoding="latin1")
ks_data1 <- ks_data1 %>%
  filter(is_successful %in% c(0, 1))

ks_data1<-data.frame(ks_data1$name_length, ks_data1$is_successful)

colnames(ks_data1)<-c('name_length','state')

ks_data1<-na.omit(ks_data1)

state.name.1<-filter(ks_data1, state==0)
mean_0<-mean(state.name.1$name_length) # 5.39
state.name.2<-filter(ks_data1, state==1)
mean_1<-mean(state.name.2$name_length) # 6.10


c1<-c(5.39, 6.10)
c2<-c("failed","successful")
c3<-data.frame(cbind(c1,c2))
colnames(c3)<-c('mean_name_length','state')
c3


ggplot(data=c3, aes(x=state, y=mean_name_length,fill=state)) +
  geom_bar(stat="identity",width = 0.5)+ggtitle("State vs. Mean project name length") +
  xlab("state") + ylab("mean name length")+
```

```
  theme_minimal()



#=========================================
# success rate versus project name length
#=========================================
ks_data1 <-
  read.csv("Kickstarter_data_update1.csv",
           fileEncoding="latin1")
ks_data1<-ks_data1[,-1]
ks_data1<-na.omit(ks_data1)
ks_data1<-mutate(ks_data1, log_goal=log(goal))
#names(ks_data1)

name_length.pct <- ks_data1 %>%
  filter(is_successful %in% c(1, 0)) %>%
  group_by(name_length, is_successful) %>%
  summarize(count=n()) %>%
  mutate(pct=count/sum(count))

name_length.pct
#summary(ks_data1$name_length)

ggplot(name_length.pct[name_length.pct$is_successful==1,], aes(name_length, pct)) +
  geom_point(colour="royalblue4", size=2.5) + ggtitle("Success Rate vs. Project name length") +
  xlab("Project name length") + ylab("Success Rate (%)") +
  scale_x_continuous(breaks=c(0,5,10,15,20,25,30)) +
  theme_economist() +
  theme(plot.title=element_text(hjust=0.5), axis.title=element_text(size=12, face="bold"))









#============================
# find out the mean name length for both failed and successful projects
#============================
ks_data1 <- read.csv("Kickstarter_data_update1.csv", fileEncoding="latin1")
ks_data1 <- ks_data1 %>%
  filter(is_successful %in% c(0, 1))

ks_data1<-data.frame(ks_data1$name_length, ks_data1$is_successful)

colnames(ks_data1)<-c('name_length','state')

ks_data1<-na.omit(ks_data1)

state.name.1<-filter(ks_data1, state==0)
mean_0<-mean(state.name.1$name_length)
```

```r
state.name.2<-filter(ks_data1, state==1)
mean_1<-mean(state.name.2$name_length)


c1<-c(5.39, 6.10)
c2<-c("failed","successful")
c3<-data.frame(cbind(c1,c2))
colnames(c3)<-c('mean_name_length','state')
c3


ggplot(data=c3, aes(x=state, y=mean_name_length,fill=state)) +
  geom_bar(stat="identity",width = 0.5)+ggtitle("State vs. Mean project names") +
  xlab("state") + ylab("mean name length")+
  theme_minimal()



#============================
# logisitc regression using R
#============================

ks_data1 <-
  read.csv("Kickstarter_data_update1.csv",
           fileEncoding="latin1")

ks_data1<-ks_data1[,-1]
#head(ks_data1)
ks_data1<-na.omit(ks_data1)

#log_goal, duration, category, country

glm.fit <- glm(is_successful ~ log(goal)+duration+
                 factor(main_category)+factor(country),
               data = ks_data1, family = binomial)

summary(glm.fit)


# adding length of the name

glm.fit1 <- glm(is_successful ~ log(goal)+duration+ name_length +
                  factor(main_category)+factor(country),
                data = ks_data1, family = binomial)

summary(glm.fit1)


# GB vs. US

ks_data1 <- ks_data1 %>%
  filter(country %in% c("GB", "US"))
```

```r
glm.fit2 <- glm(is_successful ~ factor(country),
                data = ks_data1, family = binomial)

summary(glm.fit2)
```

```r
#=============================================================
# Project duration versus project success rate, plot
#=============================================================
ks_data1 <-
  read.csv("Kickstarter_data_update1.csv",
           fileEncoding="latin1")

ks_data1<-ks_data1[,-1]


#The maximum project duration according to Kickstarter's rules is 60 days
length.pct <- ks_data1 %>%
  filter(is_successful %in% c(1, 0), duration <= 61) %>%
  group_by(duration, is_successful) %>%
  summarize(count=n()) %>%
  mutate(pct=count/sum(count))



ggplot(length.pct[length.pct$is_successful==1,], aes(duration, pct)) +
  geom_point(colour="royalblue4", size=2.5) + ggtitle("Success Rate vs. Project Duration") +
  xlab("Project Duration (Days)") + ylab("Success Rate (%)") +
  scale_x_continuous(breaks=c(0,10,20,30,40,50,60)) + geom_vline(xintercept=30, colour="red") +
  theme_economist() +
  theme(plot.title=element_text(hjust=0.5), axis.title=element_text(size=12, face="bold"))
```

Logistic Regression (base model)

```r
#=============================================================
# logisitc regression using R (basic)
#=============================================================
ks_data1 <-
  read.csv("Kickstarter_data_update1.csv",
           fileEncoding="latin1")
ks_data1<-ks_data1[,-1]
ks_data1<-na.omit(ks_data1)
```

```r
#log_goal, category, country, duration
glm.fit <- glm(is_successful ~ log(goal)+duration+
                 factor(category)+factor(country),
               data = ks_data1, family = binomial)
summary(glm.fit)
```

```
##
## Call:
## glm(formula = is_successful ~ log(goal) + duration + factor(category) +
##     factor(country), family = binomial, data = ks_data1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5896  -0.8879  -0.5730   1.0619   2.7427
##
## Coefficients:
##                                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)                       2.3948180  0.1903926  12.578  < 2e-16
## log(goal)                        -0.2847691  0.0046561 -61.161  < 2e-16
## duration                         -0.0185146  0.0006496 -28.500  < 2e-16
## factor(category)Academic         -1.3393815  0.1828589  -7.325 2.39e-13
## factor(category)Accessories      -0.6256784  0.1427227  -4.384 1.17e-05
## factor(category)Action           -1.4759240  0.2126337  -6.941 3.89e-12
## factor(category)Animals          -1.3191913  0.3209219  -4.111 3.95e-05
## factor(category)Animation        -0.7371951  0.1658194  -4.446 8.76e-06
## factor(category)Anthologies       0.7817404  0.1696544   4.608 4.07e-06
## factor(category)Apparel          -1.3074484  0.1401579  -9.328  < 2e-16
## factor(category)Apps             -2.4523491  0.1550192 -15.820  < 2e-16
## factor(category)Architecture     -0.6747861  0.1925087  -3.505 0.000456
## factor(category)Art              -0.5898390  0.1405865  -4.196 2.72e-05
## factor(category)Art Books         0.2710400  0.1513393   1.791 0.073303
## factor(category)Audio            -1.1818682  0.2191136  -5.394 6.90e-08
## factor(category)Bacon            -2.1558526  0.6455523  -3.340 0.000839
## factor(category)Blues            -0.1698878  0.2392748  -0.710 0.477697
## factor(category)Calendars        -0.5349734  0.2203087  -2.428 0.015170
## factor(category)Camera Equipment  0.2841343  0.1938010   1.466 0.142617
## factor(category)Candles          -2.2138035  0.2588866  -8.551  < 2e-16
## factor(category)Ceramics         -0.3058899  0.2179626  -1.403 0.160496
## factor(category)Children's Books -0.3879482  0.1412588  -2.746 0.006026
## factor(category)Childrenswear    -1.3678053  0.2212209  -6.183 6.29e-10
## factor(category)Chiptune          1.1164838  0.8331356   1.340 0.180213
## factor(category)Civic Design     -0.3571267  0.2531333  -1.411 0.158296
## factor(category)Classical Music   0.4247824  0.1641616   2.588 0.009665
## factor(category)Comedy           -0.2709890  0.1483723  -1.826 0.067788
## factor(category)Comic Books       0.3255482  0.1440769   2.260 0.023850
## factor(category)Comics            0.2053561  0.1575130   1.304 0.192322
## factor(category)Community Gardens -1.3757960  0.3001259  -4.584 4.56e-06
## factor(category)Conceptual Art   -1.2467096  0.2108808  -5.912 3.38e-09
## factor(category)Cookbooks        -0.8229164  0.2091239  -3.935 8.32e-05
## factor(category)Country & Folk    0.6799712  0.1536699   4.425 9.65e-06
## factor(category)Couture          -2.0142295  0.3489298  -5.773 7.81e-09
## factor(category)Crafts           -1.3800198  0.1463934  -9.427  < 2e-16
## factor(category)Crochet          -1.8617369  0.3273548  -5.687 1.29e-08
## factor(category)Dance             0.2093481  0.1971332   1.062 0.288253
```

```
## factor(category)Design             -0.4676836  0.1445762  -3.235 0.001217
## factor(category)Digital Art        -1.3510434  0.1780466  -7.588 3.25e-14
## factor(category)DIY                -1.7546050  0.1775364  -9.883  < 2e-16
## factor(category)DIY Electronics    -0.1354363  0.1690666  -0.801 0.423084
## factor(category)Documentary        -0.4408456  0.1398178  -3.153 0.001616
## factor(category)Drama              -0.3443440  0.1494881  -2.303 0.021251
## factor(category)Drinks             -0.7744405  0.1515075  -5.112 3.20e-07
## factor(category)Electronic Music   -1.0071695  0.1638631  -6.146 7.93e-10
## factor(category)Embroidery         -1.6888691  0.3752172  -4.501 6.76e-06
## factor(category)Events             -1.2318173  0.2000585  -6.157 7.40e-10
## factor(category)Experimental       -0.3953647  0.1746455  -2.264 0.023586
## factor(category)Fabrication Tools  -0.5580880  0.2548616  -2.190 0.028541
## factor(category)Faith              -0.1253486  0.1651299  -0.759 0.447798
## factor(category)Family             -1.3140025  0.2679117  -4.905 9.36e-07
## factor(category)Fantasy            -0.4128770  0.2180815  -1.893 0.058328
## factor(category)Farmer's Markets   -1.4289296  0.2506560  -5.701 1.19e-08
## factor(category)Farms              -1.2176239  0.1860280  -6.545 5.93e-11
## factor(category)Fashion            -1.3560651  0.1519190  -8.926  < 2e-16
## factor(category)Festivals          -0.0535124  0.1729201  -0.309 0.756969
## factor(category)Fiction            -1.2406050  0.1440890  -8.610  < 2e-16
## factor(category)Film & Video       -0.7220381  0.1449573  -4.981 6.32e-07
## factor(category)Fine Art           -0.5969880  0.1783570  -3.347 0.000816
## factor(category)Flight             -1.3462238  0.2742197  -4.909 9.14e-07
## factor(category)Food               -1.2471894  0.1456547  -8.563  < 2e-16
## factor(category)Food Trucks        -1.7246074  0.1802643  -9.567  < 2e-16
## factor(category)Footwear           -0.2007305  0.1638857  -1.225 0.220643
## factor(category)Gadgets            -0.3641392  0.1450104  -2.511 0.012035
## factor(category)Games              -1.1073622  0.1529217  -7.241 4.44e-13
## factor(category)Gaming Hardware    -0.5781243  0.2117654  -2.730 0.006333
## factor(category)Glass              -1.1831486  0.3860780  -3.065 0.002180
## factor(category)Graphic Design     -0.7306277  0.1600623  -4.565 5.00e-06
## factor(category)Graphic Novels      0.3242440  0.1484601   2.184 0.028959
## factor(category)Hardware           -0.2583032  0.1499594  -1.722 0.084981
## factor(category)Hip-Hop            -2.1069071  0.1665814 -12.648  < 2e-16
## factor(category)Horror             -0.5479621  0.1617468  -3.388 0.000705
## factor(category)Illustration       -0.1423156  0.1451622  -0.980 0.326894
## factor(category)Immersive           0.1147303  0.2270125   0.505 0.613284
## factor(category)Indie Rock          0.3884982  0.1581895   2.456 0.014053
## factor(category)Installations      -0.0976676  0.2058020  -0.475 0.635093
## factor(category)Interactive Design -1.1188584  0.2366895  -4.727 2.28e-06
## factor(category)Jazz                0.5566959  0.1756678   3.169 0.001530
## factor(category)Jewelry            -1.1997113  0.1642235  -7.305 2.77e-13
## factor(category)Journalism         -1.3155951  0.2043057  -6.439 1.20e-10
## factor(category)Kids               -0.2733844  0.2415602  -1.132 0.257742
## factor(category)Knitting           -0.3494090  0.3100930  -1.127 0.259832
## factor(category)Latin              -1.2751528  0.3581837  -3.560 0.000371
## factor(category)Letterpress         0.0424797  0.5146692   0.083 0.934219
## factor(category)Literary Journals  -0.2949905  0.2334528  -1.264 0.206374
## factor(category)Literary Spaces     0.5119201  0.4640872   1.103 0.269997
## factor(category)Live Games         -1.2529301  0.1819458  -6.886 5.73e-12
## factor(category)Makerspaces        -0.8572066  0.2747157  -3.120 0.001806
## factor(category)Metal              -0.3242439  0.1939655  -1.672 0.094592
## factor(category)Mixed Media        -1.0128878  0.1600722  -6.328 2.49e-10
## factor(category)Mobile Games       -2.2403939  0.1809904 -12.379  < 2e-16
```

```
## factor(category)Movie Theaters     -0.7466487  0.3138378  -2.379 0.017355
## factor(category)Music              -0.2104028  0.1382882  -1.521 0.128139
## factor(category)Music Videos       -0.8026476  0.1874333  -4.282 1.85e-05
## factor(category)Musical             0.0795241  0.1723970   0.461 0.644594
## factor(category)Narrative Film      0.1874197  0.1656135   1.132 0.257773
## factor(category)Nature             -1.6969304  0.2385857  -7.112 1.14e-12
## factor(category)Nonfiction         -0.9041982  0.1432334  -6.313 2.74e-10
## factor(category)Painting           -1.0332872  0.1550093  -6.666 2.63e-11
## factor(category)People             -1.0380429  0.1812383  -5.728 1.02e-08
## factor(category)Performance Art    -0.5787867  0.1812756  -3.193 0.001409
## factor(category)Performances        0.7649960  0.1718362   4.452 8.51e-06
## factor(category)Periodicals        -0.0557788  0.1952366  -0.286 0.775109
## factor(category)Pet Fashion        -1.4918184  0.3425505  -4.355 1.33e-05
## factor(category)Photo              -1.3911174  0.3274946  -4.248 2.16e-05
## factor(category)Photobooks         -0.0218304  0.1540245  -0.142 0.887290
## factor(category)Photography        -0.8356812  0.1655272  -5.049 4.45e-07
## factor(category)Places             -1.5315810  0.2206935  -6.940 3.92e-12
## factor(category)Playing Cards      -0.3221285  0.1446478  -2.227 0.025948
## factor(category)Plays               0.3312944  0.1617551   2.048 0.040548
## factor(category)Poetry             -0.8260797  0.1804625  -4.578 4.70e-06
## factor(category)Pop                -0.1816445  0.1586462  -1.145 0.252223
## factor(category)Pottery            -0.3242525  0.3543772  -0.915 0.360196
## factor(category)Print              -1.1447772  0.1953051  -5.861 4.59e-09
## factor(category)Printing           -1.8733327  0.2869595  -6.528 6.66e-11
## factor(category)Product Design     -0.1508073  0.1351166  -1.116 0.264367
## factor(category)Public Art         -0.3286629  0.1626269  -2.021 0.043284
## factor(category)Publishing         -0.7206364  0.1432109  -5.032 4.85e-07
## factor(category)Punk               -0.1934299  0.2254922  -0.858 0.390996
## factor(category)Puzzles            -0.5380002  0.2612636  -2.059 0.039473
## factor(category)Quilts             -0.8071551  0.4839917  -1.668 0.095374
## factor(category)R&B                -1.2512834  0.2296972  -5.448 5.11e-08
## factor(category)Radio & Podcasts   -0.8879703  0.1914370  -4.638 3.51e-06
## factor(category)Ready-to-wear      -1.4993232  0.1931889  -7.761 8.43e-15
## factor(category)Residencies         1.9740472  0.5606372   3.521 0.000430
## factor(category)Restaurants        -1.0370931  0.1547268  -6.703 2.05e-11
## factor(category)Robots             -0.0299564  0.1932382  -0.155 0.876803
## factor(category)Rock               -0.1597662  0.1506035  -1.061 0.288763
## factor(category)Romance            -1.0452036  0.2888367  -3.619 0.000296
## factor(category)Science Fiction    -0.3600783  0.1801579  -1.999 0.045643
## factor(category)Sculpture          -0.8627086  0.1773721  -4.864 1.15e-06
## factor(category)Shorts              0.1156720  0.1447821   0.799 0.424326
## factor(category)Small Batch        -0.5610159  0.1570150  -3.573 0.000353
## factor(category)Software           -1.9867783  0.1747041 -11.372  < 2e-16
## factor(category)Sound               0.2339085  0.1735988   1.347 0.177849
## factor(category)Space Exploration  -0.4159844  0.2400642  -1.733 0.083130
## factor(category)Spaces             -0.3472037  0.1724627  -2.013 0.044093
## factor(category)Stationery         -0.8562235  0.2405206  -3.560 0.000371
## factor(category)Tabletop Games      0.4163019  0.1360186   3.061 0.002209
## factor(category)Taxidermy          -1.3448234  1.1895007  -1.131 0.258233
## factor(category)Technology         -1.0440696  0.1434491  -7.278 3.38e-13
## factor(category)Television         -1.4761680  0.1978274  -7.462 8.53e-14
## factor(category)Textiles           -1.0886460  0.2345181  -4.642 3.45e-06
## factor(category)Theater             0.2979527  0.1654276   1.801 0.071686
## factor(category)Thrillers          -0.8312416  0.1812849  -4.585 4.53e-06
```

```
## factor(category)Translations       -1.0976275  0.3352189  -3.274 0.001059
## factor(category)Typography         -0.2731454  0.3502853  -0.780 0.435521
## factor(category)Vegan              -0.6175664  0.1854555  -3.330 0.000868
## factor(category)Video              -1.9603970  0.2682346  -7.309 2.70e-13
## factor(category)Video Art          -1.2348373  0.2971756  -4.155 3.25e-05
## factor(category)Video Games        -1.2508421  0.1401418  -8.926  < 2e-16
## factor(category)Wearables          -0.1497302  0.1582925  -0.946 0.344195
## factor(category)Weaving            -0.7416695  0.3841823  -1.931 0.053543
## factor(category)Web                -2.2853851  0.1580841 -14.457  < 2e-16
## factor(category)Webcomics           0.0810016  0.1755333   0.461 0.644468
## factor(category)Webseries          -1.1033257  0.1543984  -7.146 8.94e-13
## factor(category)Woodworking        -1.1639822  0.1675665  -6.946 3.75e-12
## factor(category)Workshops          -0.6271734  0.3213564  -1.952 0.050980
## factor(category)World Music        -0.4714729  0.1668395  -2.826 0.004715
## factor(category)Young Adult        -1.3192794  0.1857233  -7.103 1.22e-12
## factor(category)Zines              -0.3639194  0.1965990  -1.851 0.064159
## factor(country)AU                   0.5621499  0.1347895   4.171 3.04e-05
## factor(country)BE                   0.4391440  0.1704093   2.577 0.009966
## factor(country)CA                   0.6096327  0.1319588   4.620 3.84e-06
## factor(country)CH                   0.6353525  0.1612489   3.940 8.14e-05
## factor(country)DE                   0.2803025  0.1362448   2.057 0.039653
## factor(country)DK                   1.1865946  0.1582268   7.499 6.41e-14
## factor(country)ES                   0.0784514  0.1419595   0.553 0.580516
## factor(country)FR                   0.7539713  0.1378912   5.468 4.55e-08
## factor(country)GB                   0.6633081  0.1299456   5.105 3.32e-07
## factor(country)HK                   1.2016416  0.1577105   7.619 2.55e-14
## factor(country)IE                   0.2998407  0.1687426   1.777 0.075583
## factor(country)IT                  -0.2436140  0.1417728  -1.718 0.085734
## factor(country)JP                   1.5053402  0.4812658   3.128 0.001761
## factor(country)LU                   0.6189554  0.3622270   1.709 0.087497
## factor(country)MX                   0.7801107  0.1427407   5.465 4.62e-08
## factor(country)NL                   0.4525636  0.1463331   3.093 0.001983
## factor(country)NO                   0.8814936  0.1803771   4.887 1.02e-06
## factor(country)NZ                   0.7563766  0.1578067   4.793 1.64e-06
## factor(country)SE                   1.2044223  0.1474651   8.168 3.15e-16
## factor(country)SG                   0.5806449  0.1613781   3.598 0.000321
## factor(country)US                   0.6175187  0.1285968   4.802 1.57e-06
##
## (Intercept)                       ***
## log(goal)                         ***
## duration                          ***
## factor(category)Academic          ***
## factor(category)Accessories       ***
## factor(category)Action            ***
## factor(category)Animals           ***
## factor(category)Animation         ***
## factor(category)Anthologies       ***
## factor(category)Apparel           ***
## factor(category)Apps              ***
## factor(category)Architecture      ***
## factor(category)Art               ***
## factor(category)Art Books         .
## factor(category)Audio             ***
## factor(category)Bacon             ***
```

```
## factor(category)Blues
## factor(category)Calendars          *
## factor(category)Camera Equipment
## factor(category)Candles            ***
## factor(category)Ceramics
## factor(category)Children's Books   **
## factor(category)Childrenswear      ***
## factor(category)Chiptune
## factor(category)Civic Design
## factor(category)Classical Music    **
## factor(category)Comedy             .
## factor(category)Comic Books        *
## factor(category)Comics
## factor(category)Community Gardens  ***
## factor(category)Conceptual Art     ***
## factor(category)Cookbooks          ***
## factor(category)Country & Folk     ***
## factor(category)Couture            ***
## factor(category)Crafts             ***
## factor(category)Crochet            ***
## factor(category)Dance
## factor(category)Design             **
## factor(category)Digital Art        ***
## factor(category)DIY                ***
## factor(category)DIY Electronics
## factor(category)Documentary        **
## factor(category)Drama              *
## factor(category)Drinks             ***
## factor(category)Electronic Music   ***
## factor(category)Embroidery         ***
## factor(category)Events             ***
## factor(category)Experimental       *
## factor(category)Fabrication Tools  *
## factor(category)Faith
## factor(category)Family             ***
## factor(category)Fantasy            .
## factor(category)Farmer's Markets   ***
## factor(category)Farms              ***
## factor(category)Fashion            ***
## factor(category)Festivals
## factor(category)Fiction            ***
## factor(category)Film & Video       ***
## factor(category)Fine Art           ***
## factor(category)Flight             ***
## factor(category)Food               ***
## factor(category)Food Trucks        ***
## factor(category)Footwear
## factor(category)Gadgets            *
## factor(category)Games              ***
## factor(category)Gaming Hardware    **
## factor(category)Glass              **
## factor(category)Graphic Design     ***
## factor(category)Graphic Novels     *
## factor(category)Hardware           .
```

```
## factor(category)Hip-Hop           ***
## factor(category)Horror            ***
## factor(category)Illustration
## factor(category)Immersive
## factor(category)Indie Rock        *
## factor(category)Installations
## factor(category)Interactive Design ***
## factor(category)Jazz              **
## factor(category)Jewelry           ***
## factor(category)Journalism        ***
## factor(category)Kids
## factor(category)Knitting
## factor(category)Latin             ***
## factor(category)Letterpress
## factor(category)Literary Journals
## factor(category)Literary Spaces
## factor(category)Live Games        ***
## factor(category)Makerspaces       **
## factor(category)Metal             .
## factor(category)Mixed Media       ***
## factor(category)Mobile Games      ***
## factor(category)Movie Theaters    *
## factor(category)Music
## factor(category)Music Videos      ***
## factor(category)Musical
## factor(category)Narrative Film
## factor(category)Nature            ***
## factor(category)Nonfiction        ***
## factor(category)Painting          ***
## factor(category)People            ***
## factor(category)Performance Art   **
## factor(category)Performances      ***
## factor(category)Periodicals
## factor(category)Pet Fashion       ***
## factor(category)Photo             ***
## factor(category)Photobooks
## factor(category)Photography       ***
## factor(category)Places            ***
## factor(category)Playing Cards     *
## factor(category)Plays             *
## factor(category)Poetry            ***
## factor(category)Pop
## factor(category)Pottery
## factor(category)Print             ***
## factor(category)Printing          ***
## factor(category)Product Design
## factor(category)Public Art        *
## factor(category)Publishing        ***
## factor(category)Punk
## factor(category)Puzzles           *
## factor(category)Quilts            .
## factor(category)R&B               ***
## factor(category)Radio & Podcasts  ***
## factor(category)Ready-to-wear     ***
```

```
## factor(category)Residencies        ***
## factor(category)Restaurants        ***
## factor(category)Robots
## factor(category)Rock
## factor(category)Romance            ***
## factor(category)Science Fiction    *
## factor(category)Sculpture          ***
## factor(category)Shorts
## factor(category)Small Batch        ***
## factor(category)Software           ***
## factor(category)Sound
## factor(category)Space Exploration  .
## factor(category)Spaces             *
## factor(category)Stationery         ***
## factor(category)Tabletop Games     **
## factor(category)Taxidermy
## factor(category)Technology         ***
## factor(category)Television         ***
## factor(category)Textiles           ***
## factor(category)Theater            .
## factor(category)Thrillers          ***
## factor(category)Translations       **
## factor(category)Typography
## factor(category)Vegan              ***
## factor(category)Video              ***
## factor(category)Video Art          ***
## factor(category)Video Games        ***
## factor(category)Wearables
## factor(category)Weaving            .
## factor(category)Web                ***
## factor(category)Webcomics
## factor(category)Webseries          ***
## factor(category)Woodworking        ***
## factor(category)Workshops          .
## factor(category)World Music        **
## factor(category)Young Adult        ***
## factor(category)Zines              .
## factor(country)AU                  ***
## factor(country)BE                  **
## factor(country)CA                  ***
## factor(country)CH                  ***
## factor(country)DE                  *
## factor(country)DK                  ***
## factor(country)ES
## factor(country)FR                  ***
## factor(country)GB                  ***
## factor(country)HK                  ***
## factor(country)IE                  .
## factor(country)IT                  .
## factor(country)JP                  **
## factor(country)LU                  .
## factor(country)MX                  ***
## factor(country)NL                  **
## factor(country)NO                  ***
```

```
## factor(country)NZ                  ***
## factor(country)SE                  ***
## factor(country)SG                  ***
## factor(country)US                  ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 141906  on 109991  degrees of freedom
## Residual deviance: 123398  on 109810  degrees of freedom
## AIC: 123762
##
## Number of Fisher Scoring iterations: 5
```

```
# adding the project name length
glm.fit1 <- glm(is_successful ~ log(goal)+duration + name_length +
                  factor(category)+factor(country),
                data = ks_data1, family = binomial)

summary(glm.fit1)
```

```
##
## Call:
## glm(formula = is_successful ~ log(goal) + duration + name_length +
##     factor(category) + factor(country), family = binomial, data = ks_data1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.7551  -0.8788  -0.5521   1.0471   2.8262
##
## Coefficients:
##                                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)                       1.7709071  0.1921100   9.218  < 2e-16
## log(goal)                        -0.2973451  0.0047262 -62.914  < 2e-16
## duration                         -0.0180649  0.0006558 -27.547  < 2e-16
## name_length                       0.1045352  0.0026399  39.599  < 2e-16
## factor(category)Academic         -1.2537171  0.1841433  -6.808 9.87e-12
## factor(category)Accessories      -0.5525469  0.1437556  -3.844 0.000121
## factor(category)Action           -1.2807690  0.2142089  -5.979 2.24e-09
## factor(category)Animals          -1.2211202  0.3222852  -3.789 0.000151
## factor(category)Animation        -0.5642911  0.1671808  -3.375 0.000737
## factor(category)Anthologies       0.8834434  0.1710039   5.166 2.39e-07
## factor(category)Apparel          -1.1595967  0.1412129  -8.212  < 2e-16
## factor(category)Apps             -2.3141899  0.1560460 -14.830  < 2e-16
## factor(category)Architecture     -0.5373933  0.1939787  -2.770 0.005599
## factor(category)Art              -0.4456851  0.1417024  -3.145 0.001660
## factor(category)Art Books         0.3406005  0.1524912   2.234 0.025511
## factor(category)Audio            -1.0441278  0.2208909  -4.727 2.28e-06
## factor(category)Bacon            -2.1166988  0.6528204  -3.242 0.001185
## factor(category)Blues            -0.1285637  0.2413231  -0.533 0.594210
## factor(category)Calendars        -0.4052630  0.2221771  -1.824 0.068144
## factor(category)Camera Equipment  0.3152426  0.1951198   1.616 0.106173
## factor(category)Candles          -2.0826164  0.2605842  -7.992 1.33e-15
## factor(category)Ceramics         -0.1247635  0.2194522  -0.569 0.569680
```

```
## factor(category)Children's Books   -0.3434462  0.1423539  -2.413 0.015838
## factor(category)Childrenswear      -1.2760307  0.2224950  -5.735 9.75e-09
## factor(category)Chiptune            1.2394287  0.8400048   1.476 0.140078
## factor(category)Civic Design       -0.2037795  0.2534874  -0.804 0.421453
## factor(category)Classical Music     0.5091498  0.1653980   3.078 0.002082
## factor(category)Comedy             -0.0911126  0.1495831  -0.609 0.542451
## factor(category)Comic Books         0.4773065  0.1452472   3.286 0.001016
## factor(category)Comics              0.3649167  0.1589019   2.296 0.021648
## factor(category)Community Gardens  -1.2349247  0.3017512  -4.093 4.27e-05
## factor(category)Conceptual Art     -1.1294282  0.2123439  -5.319 1.04e-07
## factor(category)Cookbooks          -0.7137023  0.2105164  -3.390 0.000698
## factor(category)Country & Folk      0.7619374  0.1548229   4.921 8.59e-07
## factor(category)Couture            -1.8670094  0.3500921  -5.333 9.67e-08
## factor(category)Crafts             -1.2315382  0.1474974  -8.350  < 2e-16
## factor(category)Crochet            -1.6851886  0.3291446  -5.120 3.06e-07
## factor(category)Dance               0.3557614  0.1985709   1.792 0.073196
## factor(category)Design             -0.3811418  0.1456792  -2.616 0.008889
## factor(category)Digital Art        -1.1847259  0.1793083  -6.607 3.92e-11
## factor(category)DIY                -1.6218262  0.1787586  -9.073  < 2e-16
## factor(category)DIY Electronics    -0.0939228  0.1705163  -0.551 0.581761
## factor(category)Documentary        -0.2898763  0.1409185  -2.057 0.039681
## factor(category)Drama              -0.0920056  0.1506723  -0.611 0.541442
## factor(category)Drinks             -0.6402878  0.1527082  -4.193 2.75e-05
## factor(category)Electronic Music   -0.9144380  0.1652806  -5.533 3.15e-08
## factor(category)Embroidery         -1.5665608  0.3786161  -4.138 3.51e-05
## factor(category)Events             -1.0872655  0.2014572  -5.397 6.78e-08
## factor(category)Experimental       -0.2144055  0.1759140  -1.219 0.222917
## factor(category)Fabrication Tools  -0.4689278  0.2572625  -1.823 0.068340
## factor(category)Faith               0.0006689  0.1664853   0.004 0.996794
## factor(category)Family             -1.2296373  0.2703628  -4.548 5.41e-06
## factor(category)Fantasy            -0.2281990  0.2198209  -1.038 0.299217
## factor(category)Farmer's Markets   -1.2634999  0.2523837  -5.006 5.55e-07
## factor(category)Farms              -1.1013486  0.1873898  -5.877 4.17e-09
## factor(category)Fashion            -1.2146654  0.1530112  -7.938 2.05e-15
## factor(category)Festivals           0.0453552  0.1741818   0.260 0.794563
## factor(category)Fiction            -1.1316841  0.1452089  -7.793 6.52e-15
## factor(category)Film & Video       -0.5050757  0.1461137  -3.457 0.000547
## factor(category)Fine Art           -0.4931331  0.1806070  -2.730 0.006325
## factor(category)Flight             -1.2265712  0.2757806  -4.448 8.68e-06
## factor(category)Food               -1.1040953  0.1467635  -7.523 5.36e-14
## factor(category)Food Trucks        -1.5738170  0.1814636  -8.673  < 2e-16
## factor(category)Footwear           -0.1167439  0.1651216  -0.707 0.479555
## factor(category)Gadgets            -0.3433398  0.1461046  -2.350 0.018775
## factor(category)Games              -0.9344241  0.1541652  -6.061 1.35e-09
## factor(category)Gaming Hardware    -0.5035041  0.2129380  -2.365 0.018052
## factor(category)Glass              -1.0403195  0.3891102  -2.674 0.007505
## factor(category)Graphic Design     -0.6130264  0.1613686  -3.799 0.000145
## factor(category)Graphic Novels      0.4301362  0.1497167   2.873 0.004066
## factor(category)Hardware           -0.2101184  0.1510813  -1.391 0.164297
## factor(category)Hip-Hop            -1.9405991  0.1677860 -11.566  < 2e-16
## factor(category)Horror             -0.3350441  0.1631408  -2.054 0.040004
## factor(category)Illustration       -0.0227750  0.1462605  -0.156 0.876257
## factor(category)Immersive           0.2675355  0.2295689   1.165 0.243864
## factor(category)Indie Rock          0.4778776  0.1593829   2.998 0.002715
```

```
## factor(category)Installations       0.0059910  0.2075619   0.029 0.976973
## factor(category)Interactive Design -0.9777220  0.2373513  -4.119 3.80e-05
## factor(category)Jazz                0.6226056  0.1769977   3.518 0.000435
## factor(category)Jewelry            -1.1137466  0.1654604  -6.731 1.68e-11
## factor(category)Journalism         -1.1756029  0.2057786  -5.713 1.11e-08
## factor(category)Kids               -0.2146375  0.2429082  -0.884 0.376904
## factor(category)Knitting           -0.2283719  0.3130625  -0.729 0.465710
## factor(category)Latin              -1.1471199  0.3613866  -3.174 0.001502
## factor(category)Letterpress         0.2199915  0.5215010   0.422 0.673140
## factor(category)Literary Journals  -0.1485556  0.2356094  -0.631 0.528357
## factor(category)Literary Spaces     0.6394796  0.4662175   1.372 0.170178
## factor(category)Live Games         -1.1104001  0.1836062  -6.048 1.47e-09
## factor(category)Makerspaces        -0.7710969  0.2779475  -2.774 0.005533
## factor(category)Metal              -0.2020934  0.1954170  -1.034 0.301059
## factor(category)Mixed Media        -0.8748104  0.1612013  -5.427 5.74e-08
## factor(category)Mobile Games       -2.0619787  0.1821547 -11.320  < 2e-16
## factor(category)Movie Theaters     -0.5637322  0.3159164  -1.784 0.074353
## factor(category)Music              -0.0836221  0.1393910  -0.600 0.548566
## factor(category)Music Videos       -0.6865478  0.1890106  -3.632 0.000281
## factor(category)Musical             0.1711467  0.1739325   0.984 0.325124
## factor(category)Narrative Film      0.4006442  0.1669478   2.400 0.016403
## factor(category)Nature             -1.5978277  0.2401633  -6.653 2.87e-11
## factor(category)Nonfiction         -0.8728353  0.1443234  -6.048 1.47e-09
## factor(category)Painting           -0.9251260  0.1561379  -5.925 3.12e-09
## factor(category)People             -0.8824227  0.1826015  -4.833 1.35e-06
## factor(category)Performance Art    -0.4186339  0.1825247  -2.294 0.021815
## factor(category)Performances        0.9379902  0.1734411   5.408 6.37e-08
## factor(category)Periodicals         0.1472552  0.1967748   0.748 0.454253
## factor(category)Pet Fashion        -1.4868901  0.3437693  -4.325 1.52e-05
## factor(category)Photo              -1.3058477  0.3306274  -3.950 7.83e-05
## factor(category)Photobooks          0.0831081  0.1553027   0.535 0.592556
## factor(category)Photography        -0.6667442  0.1666214  -4.002 6.29e-05
## factor(category)Places             -1.4017024  0.2221113  -6.311 2.78e-10
## factor(category)Playing Cards      -0.1754041  0.1458351  -1.203 0.229071
## factor(category)Plays               0.4600498  0.1631371   2.820 0.004802
## factor(category)Poetry             -0.7012909  0.1819805  -3.854 0.000116
## factor(category)Pop                -0.0841424  0.1600400  -0.526 0.599056
## factor(category)Pottery            -0.1861565  0.3573012  -0.521 0.602362
## factor(category)Print              -1.0479304  0.1968861  -5.323 1.02e-07
## factor(category)Printing           -1.7331665  0.2890379  -5.996 2.02e-09
## factor(category)Product Design     -0.1175259  0.1361470  -0.863 0.388012
## factor(category)Public Art         -0.1988853  0.1639687  -1.213 0.225150
## factor(category)Publishing         -0.6184510  0.1443464  -4.284 1.83e-05
## factor(category)Punk               -0.1213497  0.2271020  -0.534 0.593106
## factor(category)Puzzles            -0.4086970  0.2636529  -1.550 0.121110
## factor(category)Quilts             -0.6179254  0.4878226  -1.267 0.205262
## factor(category)R&B                -1.0906969  0.2317719  -4.706 2.53e-06
## factor(category)Radio & Podcasts   -0.7845167  0.1926981  -4.071 4.68e-05
## factor(category)Ready-to-wear      -1.3848198  0.1943325  -7.126 1.03e-12
## factor(category)Residencies         2.1279262  0.5662672   3.758 0.000171
## factor(category)Restaurants        -0.8784539  0.1560025  -5.631 1.79e-08
## factor(category)Robots              0.0481014  0.1947877   0.247 0.804953
## factor(category)Rock               -0.0943846  0.1518005  -0.622 0.534095
## factor(category)Romance            -0.8501881  0.2904515  -2.927 0.003421
```

```
## factor(category)Science Fiction    -0.1501869  0.1814613  -0.828 0.407867
## factor(category)Sculpture          -0.7470666  0.1786869  -4.181 2.90e-05
## factor(category)Shorts              0.3667233  0.1460442   2.511 0.012037
## factor(category)Small Batch        -0.4327035  0.1581642  -2.736 0.006223
## factor(category)Software           -1.8549559  0.1758329 -10.550  < 2e-16
## factor(category)Sound               0.2545200  0.1747559   1.456 0.145273
## factor(category)Space Exploration  -0.3220747  0.2418725  -1.332 0.182995
## factor(category)Spaces             -0.2189146  0.1738991  -1.259 0.208081
## factor(category)Stationery         -0.7662935  0.2433672  -3.149 0.001640
## factor(category)Tabletop Games      0.5555241  0.1371307   4.051 5.10e-05
## factor(category)Taxidermy          -1.4342186  1.1894024  -1.206 0.227883
## factor(category)Technology         -0.9527939  0.1445514  -6.591 4.36e-11
## factor(category)Television         -1.2834658  0.1986957  -6.459 1.05e-10
## factor(category)Textiles           -0.9683376  0.2367583  -4.090 4.31e-05
## factor(category)Theater             0.4640021  0.1667943   2.782 0.005404
## factor(category)Thrillers          -0.5941126  0.1826370  -3.253 0.001142
## factor(category)Translations       -1.0814607  0.3369420  -3.210 0.001329
## factor(category)Typography         -0.1603818  0.3523478  -0.455 0.648980
## factor(category)Vegan              -0.5498828  0.1869076  -2.942 0.003261
## factor(category)Video              -1.8278588  0.2696438  -6.779 1.21e-11
## factor(category)Video Art          -1.0781532  0.2992938  -3.602 0.000315
## factor(category)Video Games        -1.0366831  0.1412921  -7.337 2.18e-13
## factor(category)Wearables          -0.1107853  0.1594882  -0.695 0.487287
## factor(category)Weaving            -0.5719854  0.3837140  -1.491 0.136052
## factor(category)Web                -2.1301636  0.1592056 -13.380  < 2e-16
## factor(category)Webcomics           0.2185282  0.1770520   1.234 0.217106
## factor(category)Webseries          -0.9203537  0.1555909  -5.915 3.31e-09
## factor(category)Woodworking        -1.0288937  0.1691418  -6.083 1.18e-09
## factor(category)Workshops          -0.5474125  0.3219994  -1.700 0.089123
## factor(category)World Music        -0.3595436  0.1681988  -2.138 0.032548
## factor(category)Young Adult        -1.2396892  0.1873229  -6.618 3.64e-11
## factor(category)Zines              -0.1617285  0.1982687  -0.816 0.414670
## factor(country)AU                   0.5417298  0.1354830   3.999 6.37e-05
## factor(country)BE                   0.4430655  0.1713953   2.585 0.009736
## factor(country)CA                   0.6050166  0.1325935   4.563 5.04e-06
## factor(country)CH                   0.6373177  0.1621081   3.931 8.44e-05
## factor(country)DE                   0.2498320  0.1369192   1.825 0.068051
## factor(country)DK                   1.2533839  0.1588217   7.892 2.98e-15
## factor(country)ES                   0.0988227  0.1426883   0.693 0.488575
## factor(country)FR                   0.7411986  0.1385952   5.348 8.90e-08
## factor(country)GB                   0.6464910  0.1305622   4.952 7.36e-07
## factor(country)HK                   1.1434033  0.1585488   7.212 5.53e-13
## factor(country)IE                   0.2930240  0.1698401   1.725 0.084475
## factor(country)IT                  -0.2618377  0.1425514  -1.837 0.066240
## factor(country)JP                   1.5018825  0.4839139   3.104 0.001912
## factor(country)LU                   0.6360254  0.3636692   1.749 0.080306
## factor(country)MX                   0.8333674  0.1435452   5.806 6.41e-09
## factor(country)NL                   0.4365568  0.1471181   2.967 0.003003
## factor(country)NO                   0.9202753  0.1814637   5.071 3.95e-07
## factor(country)NZ                   0.7520218  0.1585379   4.743 2.10e-06
## factor(country)SE                   1.2393336  0.1482687   8.359  < 2e-16
## factor(country)SG                   0.5352971  0.1620177   3.304 0.000953
## factor(country)US                   0.6137693  0.1292011   4.750 2.03e-06
##
```

```
## (Intercept)                          ***
## log(goal)                            ***
## duration                             ***
## name_length                          ***
## factor(category)Academic             ***
## factor(category)Accessories          ***
## factor(category)Action               ***
## factor(category)Animals              ***
## factor(category)Animation            ***
## factor(category)Anthologies          ***
## factor(category)Apparel              ***
## factor(category)Apps                 ***
## factor(category)Architecture         **
## factor(category)Art                  **
## factor(category)Art Books            *
## factor(category)Audio                ***
## factor(category)Bacon                **
## factor(category)Blues
## factor(category)Calendars            .
## factor(category)Camera Equipment
## factor(category)Candles              ***
## factor(category)Ceramics
## factor(category)Children's Books     *
## factor(category)Childrenswear        ***
## factor(category)Chiptune
## factor(category)Civic Design
## factor(category)Classical Music      **
## factor(category)Comedy
## factor(category)Comic Books          **
## factor(category)Comics               *
## factor(category)Community Gardens    ***
## factor(category)Conceptual Art       ***
## factor(category)Cookbooks            ***
## factor(category)Country & Folk       ***
## factor(category)Couture              ***
## factor(category)Crafts               ***
## factor(category)Crochet              ***
## factor(category)Dance                .
## factor(category)Design               **
## factor(category)Digital Art          ***
## factor(category)DIY                  ***
## factor(category)DIY Electronics
## factor(category)Documentary          *
## factor(category)Drama
## factor(category)Drinks               ***
## factor(category)Electronic Music     ***
## factor(category)Embroidery           ***
## factor(category)Events               ***
## factor(category)Experimental
## factor(category)Fabrication Tools    .
## factor(category)Faith
## factor(category)Family               ***
## factor(category)Fantasy
## factor(category)Farmer's Markets     ***
```

```
## factor(category)Farms              ***
## factor(category)Fashion            ***
## factor(category)Festivals
## factor(category)Fiction            ***
## factor(category)Film & Video       ***
## factor(category)Fine Art           **
## factor(category)Flight             ***
## factor(category)Food               ***
## factor(category)Food Trucks        ***
## factor(category)Footwear
## factor(category)Gadgets            *
## factor(category)Games              ***
## factor(category)Gaming Hardware    *
## factor(category)Glass              **
## factor(category)Graphic Design     ***
## factor(category)Graphic Novels     **
## factor(category)Hardware
## factor(category)Hip-Hop            ***
## factor(category)Horror             *
## factor(category)Illustration
## factor(category)Immersive
## factor(category)Indie Rock         **
## factor(category)Installations
## factor(category)Interactive Design ***
## factor(category)Jazz               ***
## factor(category)Jewelry            ***
## factor(category)Journalism         ***
## factor(category)Kids
## factor(category)Knitting
## factor(category)Latin              **
## factor(category)Letterpress
## factor(category)Literary Journals
## factor(category)Literary Spaces
## factor(category)Live Games         ***
## factor(category)Makerspaces        **
## factor(category)Metal
## factor(category)Mixed Media        ***
## factor(category)Mobile Games       ***
## factor(category)Movie Theaters     .
## factor(category)Music
## factor(category)Music Videos       ***
## factor(category)Musical
## factor(category)Narrative Film     *
## factor(category)Nature             ***
## factor(category)Nonfiction         ***
## factor(category)Painting           ***
## factor(category)People             ***
## factor(category)Performance Art    *
## factor(category)Performances       ***
## factor(category)Periodicals
## factor(category)Pet Fashion        ***
## factor(category)Photo              ***
## factor(category)Photobooks
## factor(category)Photography        ***
```

```
## factor(category)Places             ***
## factor(category)Playing Cards
## factor(category)Plays              **
## factor(category)Poetry             ***
## factor(category)Pop
## factor(category)Pottery
## factor(category)Print              ***
## factor(category)Printing           ***
## factor(category)Product Design
## factor(category)Public Art
## factor(category)Publishing         ***
## factor(category)Punk
## factor(category)Puzzles
## factor(category)Quilts
## factor(category)R&B                ***
## factor(category)Radio & Podcasts   ***
## factor(category)Ready-to-wear      ***
## factor(category)Residencies        ***
## factor(category)Restaurants        ***
## factor(category)Robots
## factor(category)Rock
## factor(category)Romance            **
## factor(category)Science Fiction
## factor(category)Sculpture          ***
## factor(category)Shorts             *
## factor(category)Small Batch        **
## factor(category)Software           ***
## factor(category)Sound
## factor(category)Space Exploration
## factor(category)Spaces
## factor(category)Stationery         **
## factor(category)Tabletop Games     ***
## factor(category)Taxidermy
## factor(category)Technology         ***
## factor(category)Television         ***
## factor(category)Textiles           ***
## factor(category)Theater            **
## factor(category)Thrillers          **
## factor(category)Translations       **
## factor(category)Typography
## factor(category)Vegan              **
## factor(category)Video              ***
## factor(category)Video Art          ***
## factor(category)Video Games        ***
## factor(category)Wearables
## factor(category)Weaving
## factor(category)Web                ***
## factor(category)Webcomics
## factor(category)Webseries          ***
## factor(category)Woodworking        ***
## factor(category)Workshops          .
## factor(category)World Music        *
## factor(category)Young Adult        ***
## factor(category)Zines
```

```
## factor(country)AU                  ***
## factor(country)BE                  **
## factor(country)CA                  ***
## factor(country)CH                  ***
## factor(country)DE                  .
## factor(country)DK                  ***
## factor(country)ES
## factor(country)FR                  ***
## factor(country)GB                  ***
## factor(country)HK                  ***
## factor(country)IE                  .
## factor(country)IT                  .
## factor(country)JP                  **
## factor(country)LU                  .
## factor(country)MX                  ***
## factor(country)NL                  **
## factor(country)NO                  ***
## factor(country)NZ                  ***
## factor(country)SE                  ***
## factor(country)SG                  ***
## factor(country)US                  ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 141906  on 109991  degrees of freedom
## Residual deviance: 121805  on 109809  degrees of freedom
## AIC: 122171
##
## Number of Fisher Scoring iterations: 5
```

```r
# comparision of GB and USA
# GB vs. US
ks_data1 <- ks_data1 %>%
  filter(country %in% c("GB", "US"))
glm.fit2 <- glm(is_successful ~ factor(country),
                data = ks_data1, family = binomial)
summary(glm.fit2)
```

```
##
## Call:
## glm(formula = is_successful ~ factor(country), family = binomial,
##     data = ks_data1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0415  -0.9339  -0.9339   1.4422   1.4422
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -0.32833    0.01851  -17.74   <2e-16 ***
## factor(country)US -0.27558    0.02009  -13.72   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 109896  on 83890  degrees of freedom
## Residual deviance: 109710  on 83889  degrees of freedom
## AIC: 109714
##
## Number of Fisher Scoring iterations: 4
```

```
#=========================================
# additional analysis, logisitc regression
#=========================================
library(caTools)
ks_data1 <-
  read.csv("Kickstarter_data_update1.csv",
           fileEncoding="latin1")

ks_data1<-ks_data1[,-1]
#head(ks_data1)
ks_data1<-na.omit(ks_data1)

##  set seed to ensure you always have same random numbers generated
set.seed(123)
# splits the data in the ratio mentioned in SplitRatio.
#After splitting marks these rows as logical
#TRUE and the the remaining are marked as logical FALSE
sample = sample.split(ks_data1,SplitRatio = 0.7)

# creates a training dataset named train1 with rows which are marked as TRUE
train1 =subset(ks_data1,sample ==TRUE)
test1=subset(ks_data1, sample==FALSE)


glm.fit1 <- glm(is_successful ~ log(goal) + duration + name_length +
                  factor(main_category)+factor(country),
                data = train1, family = binomial)

summary(glm.fit1)
```

```
##
## Call:
## glm(formula = is_successful ~ log(goal) + duration + name_length +
##     factor(main_category) + factor(country), family = binomial,
##     data = train1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5873  -0.9009  -0.6316   1.1369   2.6466
##
## Coefficients:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    1.0869212  0.1634967   6.648 2.97e-11
## log(goal)                     -0.2789330  0.0054558 -51.126  < 2e-16
## duration                      -0.0197720  0.0007764 -25.467  < 2e-16
## name_length                    0.1169187  0.0031120  37.570  < 2e-16
```

```
## factor(main_category)Comics        0.9553065  0.0518312  18.431  < 2e-16
## factor(main_category)Crafts       -0.7816615  0.0592519 -13.192  < 2e-16
## factor(main_category)Dance         0.9187445  0.0961591   9.554  < 2e-16
## factor(main_category)Design        0.2487817  0.0395937   6.283 3.31e-10
## factor(main_category)Fashion      -0.5042683  0.0440261 -11.454  < 2e-16
## factor(main_category)Film & Video  0.1823982  0.0390471   4.671 2.99e-06
## factor(main_category)Food         -0.3906395  0.0475676  -8.212  < 2e-16
## factor(main_category)Games         0.3046515  0.0379279   8.032 9.56e-16
## factor(main_category)Journalism   -0.8570546  0.0895615  -9.569  < 2e-16
## factor(main_category)Music         0.3469169  0.0394727   8.789  < 2e-16
## factor(main_category)Photography  -0.0772807  0.0636300  -1.215 0.224544
## factor(main_category)Publishing   -0.1163934  0.0396959  -2.932 0.003366
## factor(main_category)Technology   -0.4456577  0.0420954 -10.587  < 2e-16
## factor(main_category)Theater       0.8389299  0.0649497  12.917  < 2e-16
## factor(country)AU                  0.5191979  0.1617746   3.209 0.001330
## factor(country)BE                  0.3088238  0.2048717   1.507 0.131708
## factor(country)CA                  0.5668206  0.1582294   3.582 0.000341
## factor(country)CH                  0.5886025  0.1919003   3.067 0.002161
## factor(country)DE                  0.2467104  0.1633410   1.510 0.130941
## factor(country)DK                  1.2585635  0.1885858   6.674 2.49e-11
## factor(country)ES                  0.1500902  0.1700257   0.883 0.377371
## factor(country)FR                  0.7231837  0.1650960   4.380 1.18e-05
## factor(country)GB                  0.7091516  0.1558233   4.551 5.34e-06
## factor(country)HK                  1.2297321  0.1891760   6.500 8.01e-11
## factor(country)IE                  0.4065904  0.1998607   2.034 0.041914
## factor(country)IT                 -0.2226111  0.1703987  -1.306 0.191412
## factor(country)JP                  0.5318445  0.7830316   0.679 0.497003
## factor(country)LU                  1.0395483  0.4172612   2.491 0.012725
## factor(country)MX                  0.7491488  0.1719965   4.356 1.33e-05
## factor(country)NL                  0.5153295  0.1757038   2.933 0.003358
## factor(country)NO                  0.8933739  0.2158376   4.139 3.49e-05
## factor(country)NZ                  0.7513255  0.1894059   3.967 7.29e-05
## factor(country)SE                  1.1308870  0.1772139   6.381 1.75e-10
## factor(country)SG                  0.5164991  0.1950992   2.647 0.008112
## factor(country)US                  0.6532881  0.1542221   4.236 2.28e-05
##
## (Intercept)                       ***
## log(goal)                         ***
## duration                          ***
## name_length                       ***
## factor(main_category)Comics       ***
## factor(main_category)Crafts       ***
## factor(main_category)Dance        ***
## factor(main_category)Design       ***
## factor(main_category)Fashion      ***
## factor(main_category)Film & Video ***
## factor(main_category)Food         ***
## factor(main_category)Games        ***
## factor(main_category)Journalism   ***
## factor(main_category)Music        ***
## factor(main_category)Photography
## factor(main_category)Publishing   **
## factor(main_category)Technology   ***
## factor(main_category)Theater      ***
```

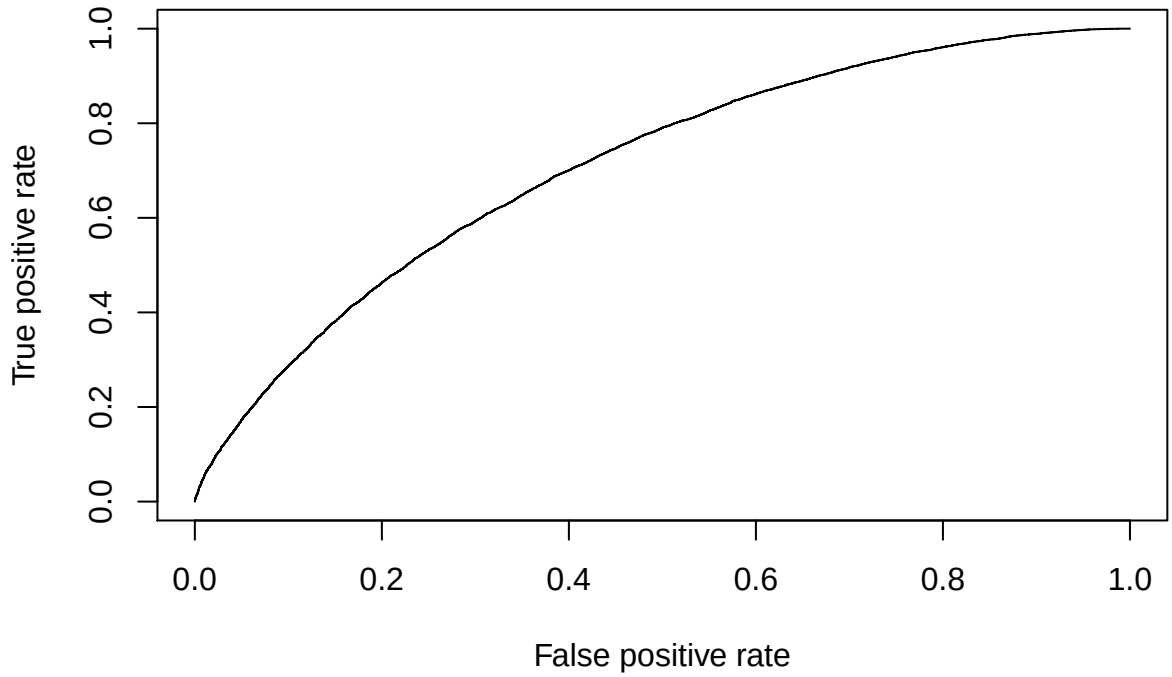
```
## factor(country)AU                 **
## factor(country)BE
## factor(country)CA                 ***
## factor(country)CH                 **
## factor(country)DE
## factor(country)DK                 ***
## factor(country)ES
## factor(country)FR                 ***
## factor(country)GB                 ***
## factor(country)HK                 ***
## factor(country)IE                 *
## factor(country)IT
## factor(country)JP
## factor(country)LU                 *
## factor(country)MX                 ***
## factor(country)NL                 **
## factor(country)NO                 ***
## factor(country)NZ                 ***
## factor(country)SE                 ***
## factor(country)SG                 **
## factor(country)US                 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 94636  on 73326  degrees of freedom
## Residual deviance: 85149  on 73288  degrees of freedom
## AIC: 85227
##
## Number of Fisher Scoring iterations: 4
```

```
#anova(glm.fit1, test="Chisq")


fitted.results <- predict(glm.fit1, newdata=test1,type='response')
fitted.results <- ifelse(fitted.results > 0.5,1,0)
misClasificError <- mean(fitted.results != test1$is_successful)
print(paste('Accuracy is',1-misClasificError)) # accuracy is about 68.8%
```

```
## [1] "Accuracy is 0.688013091504159"
```

```
# plot the ROC curve and find out the area under ROC curve
#install.packages("ROCR")
library(ROCR)
p <- predict(glm.fit1, newdata=test1, type="response")
pr <- prediction(p, test1$is_successful)
prf <- performance(pr, measure = "tpr", x.measure = "fpr")
plot(prf)
```




```
auc <- performance(pr, measure = "auc")
auc <- auc@y.values[[1]]
auc # about 70.9%, means the model still can be improved, because it's not very close to 1
```

```
## [1] 0.7099281
```

Logistic Regression with additional terms

```
#============================================
# logisitc regression adding polynomal terms and interactions
#============================================
ks_data1 <-
  read.csv("Kickstarter_data_update1.csv",
           fileEncoding="latin1")

ks_data1<-ks_data1[,-1]
#head(ks_data1)
ks_data1<-na.omit(ks_data1)


# full  model with name length
glm.fit2 <- glm(is_successful ~ log(goal)+ I((log(goal))^2) +duration + I(duration^2) + name_length +
                  factor(main_category)+factor(country),
                data = ks_data1, family = binomial)

summary(glm.fit2)
```

```
##
## Call:
## glm(formula = is_successful ~ log(goal) + I((log(goal))^2) +
##     duration + I(duration^2) + name_length + factor(main_category) +
```

```
##     factor(country), family = binomial, data = ks_data1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1192  -0.9113  -0.6189   1.1299   2.7192
##
## Coefficients:
##                                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)                       -5.210e-01  1.649e-01  -3.160 0.001580
## log(goal)                         -3.253e-02  2.393e-02  -1.359 0.174098
## I((log(goal))^2)                  -1.601e-02  1.449e-03 -11.046  < 2e-16
## duration                           2.334e-02  2.660e-03   8.774  < 2e-16
## I(duration^2)                     -5.813e-04  3.525e-05 -16.489  < 2e-16
## name_length                        1.138e-01  2.543e-03  44.743  < 2e-16
## factor(main_category)Comics        9.014e-01  4.228e-02  21.318  < 2e-16
## factor(main_category)Crafts       -7.710e-01  4.807e-02 -16.039  < 2e-16
## factor(main_category)Dance         9.767e-01  7.832e-02  12.470  < 2e-16
## factor(main_category)Design        2.570e-01  3.231e-02   7.954 1.81e-15
## factor(main_category)Fashion      -5.024e-01  3.589e-02 -13.999  < 2e-16
## factor(main_category)Film & Video  1.765e-01  3.180e-02   5.552 2.83e-08
## factor(main_category)Food         -4.327e-01  3.907e-02 -11.076  < 2e-16
## factor(main_category)Games         2.963e-01  3.095e-02   9.573  < 2e-16
## factor(main_category)Journalism   -8.665e-01  7.351e-02 -11.788  < 2e-16
## factor(main_category)Music         3.256e-01  3.231e-02  10.077  < 2e-16
## factor(main_category)Photography  -8.127e-02  5.108e-02  -1.591 0.111600
## factor(main_category)Publishing   -1.442e-01  3.244e-02  -4.444 8.84e-06
## factor(main_category)Technology   -4.561e-01  3.452e-02 -13.213  < 2e-16
## factor(main_category)Theater       8.546e-01  5.315e-02  16.078  < 2e-16
## factor(country)AU                  5.559e-01  1.331e-01   4.176 2.96e-05
## factor(country)BE                  4.393e-01  1.673e-01   2.626 0.008649
## factor(country)CA                  6.252e-01  1.303e-01   4.798 1.60e-06
## factor(country)CH                  6.382e-01  1.593e-01   4.005 6.20e-05
## factor(country)DE                  2.653e-01  1.345e-01   1.973 0.048551
## factor(country)DK                  1.369e+00  1.561e-01   8.768  < 2e-16
## factor(country)ES                  1.583e-01  1.399e-01   1.132 0.257620
## factor(country)FR                  7.239e-01  1.360e-01   5.322 1.03e-07
## factor(country)GB                  7.252e-01  1.283e-01   5.651 1.59e-08
## factor(country)HK                  1.457e+00  1.572e-01   9.268  < 2e-16
## factor(country)IE                  3.742e-01  1.665e-01   2.248 0.024601
## factor(country)IT                 -1.774e-01  1.400e-01  -1.268 0.204922
## factor(country)JP                  1.945e+00  4.732e-01   4.110 3.96e-05
## factor(country)LU                  7.877e-01  3.558e-01   2.214 0.026842
## factor(country)MX                  9.375e-01  1.413e-01   6.634 3.28e-11
## factor(country)NL                  5.035e-01  1.442e-01   3.493 0.000478
## factor(country)NO                  9.545e-01  1.789e-01   5.336 9.49e-08
## factor(country)NZ                  7.783e-01  1.553e-01   5.011 5.41e-07
## factor(country)SE                  1.341e+00  1.463e-01   9.167  < 2e-16
## factor(country)SG                  6.425e-01  1.596e-01   4.026 5.68e-05
## factor(country)US                  6.695e-01  1.270e-01   5.270 1.36e-07
##
## (Intercept)                       **
## log(goal)
## I((log(goal))^2)                  ***
## duration                          ***
```

```
## I(duration^2)                     ***
## name_length                       ***
## factor(main_category)Comics       ***
## factor(main_category)Crafts       ***
## factor(main_category)Dance        ***
## factor(main_category)Design       ***
## factor(main_category)Fashion      ***
## factor(main_category)Film & Video ***
## factor(main_category)Food         ***
## factor(main_category)Games        ***
## factor(main_category)Journalism   ***
## factor(main_category)Music        ***
## factor(main_category)Photography
## factor(main_category)Publishing   ***
## factor(main_category)Technology   ***
## factor(main_category)Theater      ***
## factor(country)AU                 ***
## factor(country)BE                 **
## factor(country)CA                 ***
## factor(country)CH                 ***
## factor(country)DE                 *
## factor(country)DK                 ***
## factor(country)ES
## factor(country)FR                 ***
## factor(country)GB                 ***
## factor(country)HK                 ***
## factor(country)IE                 *
## factor(country)IT
## factor(country)JP                 ***
## factor(country)LU                 *
## factor(country)MX                 ***
## factor(country)NL                 ***
## factor(country)NO                 ***
## factor(country)NZ                 ***
## factor(country)SE                 ***
## factor(country)SG                 ***
## factor(country)US                 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 141906  on 109991  degrees of freedom
## Residual deviance: 127194  on 109951  degrees of freedom
## AIC: 127276
##
## Number of Fisher Scoring iterations: 4
```

```
# this is interesting, it shows that,
# the success rate becomes higher when duration is low
# but it goes down when duration is high enough enough
```

```r
# full  model with log(goal) and duration interaction
glm.fit2 <- glm(is_successful ~ log(goal)+ duration +  I(log(goal)*duration) + name_length +
                  factor(main_category)+factor(country),
                data = ks_data1, family = binomial)

summary(glm.fit2)
```

```
##
## Call:
## glm(formula = is_successful ~ log(goal) + duration + I(log(goal) *
##     duration) + name_length + factor(main_category) + factor(country),
##     family = binomial, data = ks_data1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4975  -0.9034  -0.6296   1.1379   2.6406
##
## Coefficients:
##                                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)                       0.8672419  0.1601279   5.416 6.10e-08
## log(goal)                        -0.2521898  0.0114705 -21.986  < 2e-16
## duration                         -0.0120682  0.0028375  -4.253 2.11e-05
## I(log(goal) * duration)          -0.0009047  0.0003321  -2.724 0.006451
## name_length                       0.1144097  0.0025388  45.064  < 2e-16
## factor(main_category)Comics       0.9384715  0.0422710  22.201  < 2e-16
## factor(main_category)Crafts      -0.7555829  0.0482178 -15.670  < 2e-16
## factor(main_category)Dance        0.9986516  0.0779872  12.805  < 2e-16
## factor(main_category)Design       0.2719421  0.0322962   8.420  < 2e-16
## factor(main_category)Fashion     -0.4809343  0.0359258 -13.387  < 2e-16
## factor(main_category)Film & Video  0.1741993  0.0318231   5.474 4.40e-08
## factor(main_category)Food        -0.4343053  0.0390751 -11.115  < 2e-16
## factor(main_category)Games        0.3056486  0.0310059   9.858  < 2e-16
## factor(main_category)Journalism  -0.8703658  0.0736061 -11.825  < 2e-16
## factor(main_category)Music        0.3430057  0.0323247  10.611  < 2e-16
## factor(main_category)Photography -0.0759533  0.0512797  -1.481 0.138564
## factor(main_category)Publishing  -0.1175210  0.0324507  -3.622 0.000293
## factor(main_category)Technology  -0.4679266  0.0344929 -13.566  < 2e-16
## factor(main_category)Theater      0.8619747  0.0529787  16.270  < 2e-16
## factor(country)AU                 0.5388992  0.1328063   4.058 4.95e-05
## factor(country)BE                 0.4101236  0.1671459   2.454 0.014140
## factor(country)CA                 0.6137791  0.1300116   4.721 2.35e-06
## factor(country)CH                 0.6123824  0.1587676   3.857 0.000115
## factor(country)DE                 0.2542857  0.1342083   1.895 0.058131
## factor(country)DK                 1.2563562  0.1553764   8.086 6.17e-16
## factor(country)ES                 0.1371840  0.1396512   0.982 0.325936
## factor(country)FR                 0.7117235  0.1357091   5.244 1.57e-07
## factor(country)GB                 0.7147531  0.1280417   5.582 2.38e-08
## factor(country)HK                 1.3199475  0.1560885   8.456  < 2e-16
## factor(country)IE                 0.3595504  0.1662539   2.163 0.030567
## factor(country)IT                -0.1967639  0.1396830  -1.409 0.158940
## factor(country)JP                 1.4467855  0.4641322   3.117 0.001826
## factor(country)LU                 0.7347972  0.3550668   2.069 0.038503
## factor(country)MX                 0.8190545  0.1406654   5.823 5.79e-09
## factor(country)NL                 0.4968597  0.1438432   3.454 0.000552
```

```
## factor(country)NO                  0.8342904  0.1774630   4.701 2.59e-06
## factor(country)NZ                  0.7555107  0.1549268   4.877 1.08e-06
## factor(country)SE                  1.2095210  0.1453255   8.323  < 2e-16
## factor(country)SG                  0.6203683  0.1593718   3.893 9.92e-05
## factor(country)US                  0.6542007  0.1267346   5.162 2.44e-07
##
## (Intercept)                       ***
## log(goal)                         ***
## duration                          ***
## I(log(goal) * duration)           **
## name_length                       ***
## factor(main_category)Comics       ***
## factor(main_category)Crafts       ***
## factor(main_category)Dance        ***
## factor(main_category)Design       ***
## factor(main_category)Fashion      ***
## factor(main_category)Film & Video ***
## factor(main_category)Food         ***
## factor(main_category)Games        ***
## factor(main_category)Journalism   ***
## factor(main_category)Music        ***
## factor(main_category)Photography
## factor(main_category)Publishing   ***
## factor(main_category)Technology   ***
## factor(main_category)Theater      ***
## factor(country)AU                 ***
## factor(country)BE                 *
## factor(country)CA                 ***
## factor(country)CH                 ***
## factor(country)DE                 .
## factor(country)DK                 ***
## factor(country)ES
## factor(country)FR                 ***
## factor(country)GB                 ***
## factor(country)HK                 ***
## factor(country)IE                 *
## factor(country)IT
## factor(country)JP                 **
## factor(country)LU                 *
## factor(country)MX                 ***
## factor(country)NL                 ***
## factor(country)NO                 ***
## factor(country)NZ                 ***
## factor(country)SE                 ***
## factor(country)SG                 ***
## factor(country)US                 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 141906  on 109991  degrees of freedom
## Residual deviance: 127615  on 109952  degrees of freedom
## AIC: 127695
```

```
##
## Number of Fisher Scoring iterations: 4
# it shows that, the interaction term is significant negative
```

Other models for comparision

Tree based model

```
#======================
# Decision Tree
#======================
library(caTools)
ks_data1 <-
  read.csv("Kickstarter_data_update1.csv",
           fileEncoding="latin1")

ks_data1<-ks_data1[,-1]
#head(ks_data1)
ks_data1<-na.omit(ks_data1)

##  set seed to ensure you always have same random numbers generated
set.seed(123)
# splits the data in the ratio mentioned in SplitRatio.
#After splitting marks these rows as logical
#TRUE and the the remaining are marked as logical FALSE
sample = sample.split(ks_data1,SplitRatio = 0.7)

# creates a training dataset named train1 with rows which are marked as TRUE
train1 =subset(ks_data1,sample ==TRUE)
test1=subset(ks_data1, sample==FALSE)
#install.packages("tree")
library(tree)
```

```
## Registered S3 method overwritten by 'tree':
##   method     from
##   print.tree cli
```

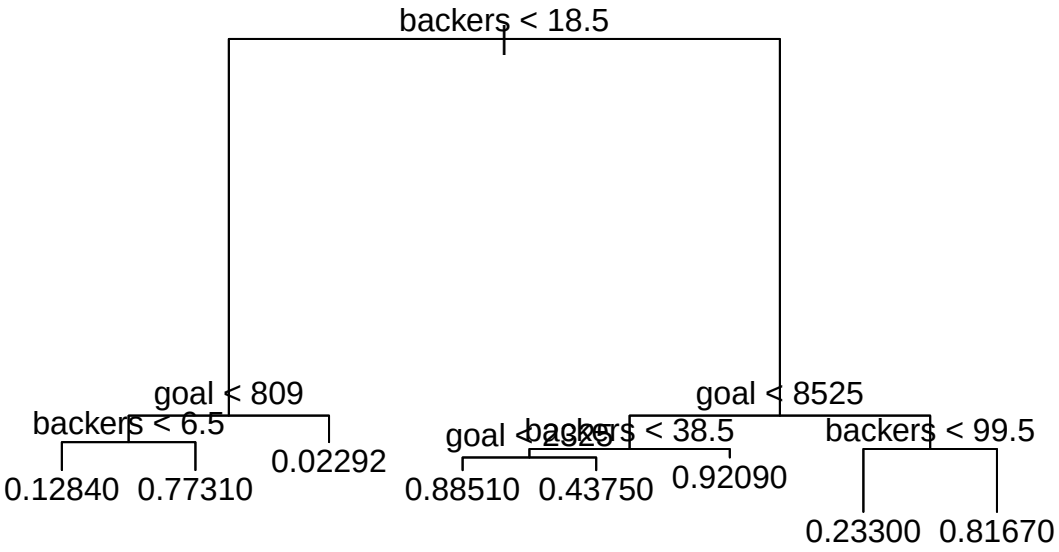
```
tree1 <- tree(is_successful ~ goal + duration + name_length +main_category + + country + backers , data

summary(tree1)
```

```
##
## Regression tree:
## tree(formula = is_successful ~ goal + duration + name_length +
##     main_category + +country + backers, data = train1)
## Variables actually used in tree construction:
## [1] "backers" "goal"
## Number of terminal nodes:  8
## Residual mean deviance:  0.07765 = 5693 / 73320
## Distribution of residuals:
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -0.92090 -0.02292 -0.02292  0.00000  0.07905  0.97710
```

```r
# look into decision tree rules:

plot(tree1)
text(tree1 ,pretty = 0)
```




```r
tree1
```

```
## node), split, n, deviance, yval
##       * denotes terminal node
##
##  1) root 73327 16610.0 0.34660
##    2) backers < 18.5 41810  2688.0 0.06905
##      4) goal < 809 6030  1358.0 0.34280
##        8) backers < 6.5 4025   450.6 0.12840 *
##        9) backers > 6.5 2005   351.7 0.77310 *
##      5) goal > 809 35780   801.2 0.02292 *
##    3) backers > 18.5 31517  6426.0 0.71470
##      6) goal < 8525 16592  2088.0 0.85240
##       12) backers < 38.5 5245  1093.0 0.70410
##         24) goal < 2325 3124   317.7 0.88510 *
##         25) goal > 2325 2121   522.0 0.43750 *
##       13) backers > 38.5 11347   826.1 0.92090 *
##      7) goal > 8525 14925  3674.0 0.56170
##       14) backers < 99.5 6520  1165.0 0.23300 *
##       15) backers > 99.5 8405  1258.0 0.81670 *
```

```r
# number of backers, goal are the msot significant factors



#Re-applying the tree rules to the training set itself, we can validate our model:

Predt <- predict(tree1, train1)

validf <- data.frame( kickstarter_id = train1$ID, orig_status = train1$is_successful, new_status = Predt
validf$new = ifelse(validf$new_status < 0.5, 0, 1)
```

```r
# contingency Tables:
table(validf$orig_status, validf$new)
```

```
##
##         0     1
##   0 44662  3252
##   1  3784 21629
```

```r
# find out the Area under the curve
#install.packages("pROC")
library(pROC)
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
## Attaching package: 'pROC'
```

```
## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```r
auc(validf$orig_status, validf$new)
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```
## Area under the curve: 0.8916
```

```
## Area under the curve: 0.89
```

```r
#Finally applying the tree rules to the test set, we get the following stats:

Pred1 <- predict(tree1, test1)


validf <- data.frame( kickstarter_id = test1$ID, orig_status = test1$is_successful, new_status = Pred1)
validf$new = ifelse(validf$new_status < 0.5, 0, 1)

# contingency Tables:
table(validf$orig_status, validf$new)
```

```
##
##         0     1
##   0 22345  1652
##   1  1967 10701
```

```r
auc(validf$orig_status, validf$new)
```

```
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```

```
## Area under the curve: 0.8879
```

```r
# the area under the curve is 0.88, still as good as the one in traning data set
```