

SBM Report Group 4

Bram Kapteijns, Diego van den Hoeven, Roland Vits, Thomas Boot, Borná Djavdan

Report SBM Group 4

Introduction

Kickstarter is a crowdfunding platform where creators can share their ideas for a product or service to bring it to life. Creators can set a pledge goal, which they believe will ensure that the project will be a success; other users can pledge money. Creators have several ways to attract donators, for instance they could promise perks to those who donate, they could promise an improved product depending on the amount reached, and they could be more transparent about progress. Creating a successful project on this site might be considered a matter of luck to many; however, with the availability of a data set of this platform, it is possible to study a factor that could influence the success of a project.

Creators in Kickstarter can be seen as entrepreneurs of a business with the goal of providing a product or a service. Thus, an organizational framework can be used to see what factors are at play in determining a project's success. The behavioral theory of the firm is one such framework, and it allows for the study of how learning affects the firm (Cyert and March 1963). It is commonly believed that more experience leads to a better chance of success for a business [Levitt and March (1988)](Parker 2013), however there are those who argue against this notion (Blank 2021), resulting in a contradicting literature. It is also interesting to study how experience affects the way that creators change their ideas, more specifically if their ideas become more or less innovative as they become more experienced. Entrepreneurial experience has been suggested to increase reported innovation (Vaillant and Lafuente 2019).

For this purpose, the following research questions were formulated:

RQ1: How does experience impact the success rate?

RQ2: How do business ideas change after experience?

Theory and Hypotheses

The Behavioural Theory of the Firm theorizes how organisations operate based on coalitions of various participants. Each participant has their own goals and aspirations, and the firm's behaviour as a whole is a result of the equilibrium reached by the choices of the individuals, bargaining process between those individuals, and the selections they make (Cyert and March 1963) (Levitt and March 1988).

However, the way an organization learns, though intertwined with these processes, can be framed differently. This framework, proposed by Levitt & March, focuses on three processes: behavior of organizations is based on routines, the actions of an organization are history-dependent, and organizations are goal oriented (Levitt and March 1988). Levitt & March also suggest that organizations can learn from direct experience, and indirectly from the experience of others. Therefore it is important to take these two possibilities into account when studying organizational learning, as this research intends to do. This means that data on direct experience and indirect experience is necessary for the goals of this study.

Research Question 1

The first research question relates to the idea of Organizational Learning developed by Levitt and March, given that it studies how KickStarter creators learn from their previous endeavours and the experience of others. Given this focus on KickStarter creators, the research question is centered in the case of an entrepreneurial organization. Therefore, it builds on the theory presented by Parker, where it was found that serial entrepreneurs have an performance improvement due to gained experience. The study also concluded that the benefits of this experience diminished over time (Parker 2013). This research question also builds upon a study by Blank which found that, in the case of student startups, previous experience in entrepreneurship did not result in significant benefits for the survival of the startup when compared to advice given by a mentor. The explanation given for this was that previous experience would cause a tendency to imitate processes and routines that were tested in the past, which may harm the survival of the organization (Blank 2021).

Given these contradictory results regarding past experience, we decided to formulate the hypothesis in line with the results of Parker, as the situation is more general and thus more applicable to the case of KickStarter. Thus the hypothesis is **people who have experience in starting companies will have more success than people who do not have this experience.**

Research Question 2

The theoretical foundation of this research is grounded in the dynamic nature of entrepreneurial cognition and the impact of experiential learning on the evolution of business ideas. En-

trepreneurial activities involve a continuous process of learning, adaptation, and refinement. According to cognitive theories of entrepreneurship, the experience gained through the implementation of an initial business idea contributes significantly to an entrepreneur's cognitive development (Marcati, Guido, and Peluso 2008). The theoretical framework posits that as entrepreneurs engage in practical ventures, they accumulate knowledge, skills, and insights, thereby influencing the generation and transformation of subsequent business ideas (Marcati, Guido, and Peluso 2008). The iterative nature of entrepreneurship suggests that learning from experience fosters a more sophisticated understanding of market dynamics, consumer needs, and effective business strategies. This accumulated knowledge is likely to lead to the generation of more innovative ideas over time. The theory is underpinned by the concept of the entrepreneurial learning cycle, emphasizing that the process of learning from experience enhances an entrepreneur's ability to conceive and implement innovative business ideas.

Building upon the theoretical framework, the hypothesis posits that **subsequent business ideas developed by entrepreneurs with prior experience are more innovative than their initial business idea**. This hypothesis is rooted in the belief that the knowledge gained from the practical implementation of the first idea contributes to an enhanced entrepreneurial mindset, characterized by a deeper understanding of market dynamics and a heightened ability to identify and capitalize on opportunities.

Assumptions

To test both research questions and to validate both hypothesis, multiple assumptions are made. Firstly it is assumed that an entrepreneur does not have any prior experience before joining the database. When entrepreneurs have prior experience which is not covered in the data, the results will be skewed since the entrepreneurs with prior experience and new entrepreneurs can not be distinguished in the data.

Secondly, we assume that the data in the database is sufficient and that there are no other, unmeasured, variables that influence the outcome of this research.

Thirdly, to measure the successrate, there is no explicit variable that measures successrate. So, to measure successrate, it is assumed that funding is a good proxy for successrate. How successful a start-up is, is measured by the total raised funding divided by the funding goal. This will be a good approximation since projects where the funding goal is not reached will not get the money from Kickstarter and therefore can not start with their start-up.

Fourthly, it is assumed that the data is unbiased because the data is collected by a third party. Fifthly, it is assumed that innovation-related words can also be taken as meaning innovation. By expanding the bucket of words referring to innovation we assure that the majority of the words related to innovation are covered.

Lastly, it is not assumed that employees influence business results, as we are taking a look at Kickstarter projects. And start-up culture is not the same as corporate culture. It is not assumed that the number of employees is a good proxy for experience, because the number

of employees does not say anything about the experience of the employees (Martin and Smith 2010).

Methods and Data

Describe the data set that you used, the way you translated the raw data into variables, and the methods you used to analyze the data (e.g., cluster analysis, regression analysis, panel data analysis, time series). Remember that most methods have a number of assumptions that you have to check prior to doing the analyses. Certain assumption violations can be fixed by transforming the focal variables, e.g., the dependent variable, in case of other violations you may need to choose a different method

```
getwd()
```

```
[1] "C:/Users/bvits/OneDrive/Desktop/2023/Masters/JADs/strat_bm/Project/SBM_assignment/report"
```

```
setwd("C:/Users/bvits/OneDrive/Desktop/2023/Masters/JADs/strat_bm/Project/SBM_assignment/report")
getwd()
```

```
[1] "C:/Users/bvits/OneDrive/Desktop/2023/Masters/JADs/strat_bm/Project/SBM_assignment/report"
```

Data Importing and cleaning, research question 1

In order to test the first hypothesis it is necessary to first clean the data. The first step is to select the data that has a deadline is before week 32 in 2019 in unix time and make sure that the variables that will be used in the regression are of the correct class. Next, a percentage is derived from the pledge goal of the project and how much was actually pledged, this will be used as a measure of success. For this reason the projects with a pledge goal of 0 were removed, this avoids division by zero. It was also chosen that values that exceeded 100% would be set to this value, as these might be detected as outliers and could cause issues. After this, we check the distribution of the data and remove cases in which we don't have enough data for the number of projects backed and the number of projects created, which are proxies for learning in this study. We consider that having less than 50 data points is when the data point is removed, this is because 50 data points are the minimum for a normal distribution to be present (find citation). Finally, the null values are removed. The below shows the distribution of the learning proxies and the proxy for success.

```
#inspect the data
#load .rds file named "info" into rstudio
```

```

info32<-readRDS('info.rds')

#create new data with only a few columns of info32 dataset
dataHyp1Sub <- info32[,c('Goal_USD','Pledge_USD','Launched_at',
                         'Deadline', 'Creator_nb_projects',
                         'Creator_nb_backed')]
#dataHyp1Sub

#make a subdataset where deadline is before week 32 in 2019 in unix time
dataHyp1 <- dataHyp1Sub[dataHyp1Sub$Deadline < 1565474400,]
#dataHyp1

#You drop only 1850 records out of 75000

#columns to integer
dataHyp1$Pledge_USD <- as.integer(dataHyp1$Pledge_USD)
dataHyp1$Goal_USD <- as.integer(dataHyp1$Goal_USD)

#drop the rows where Goal_USD is 0 -> then you might get inf percentage
dataHyp1 <- dataHyp1[dataHyp1$Goal_USD != 0,]
#dataHyp1

#Create a new column with percentage of goal reached
dataHyp1$Goal_percentage <- dataHyp1$Pledge_USD/dataHyp1$Goal_USD
dataHyp1$Goal_percentage <- round(dataHyp1$Goal_percentage, digits = 3)
dataHyp1$Goal_percentage[dataHyp1$Goal_percentage>100] = 100
#dataHyp1

#Get from Creator_nb_backed only the number with a regex
dataHyp1$Creator_nb_backed <- gsub("[^0-9]", "", dataHyp1$Creator_nb_backed)
#Seeing if this works better as a string rather than an integer
#dataHyp1

#drop redundant columns
dataHyp1 <- dataHyp1[,c('Creator_nb_projects', 'Creator_nb_backed',
                        'Goal_percentage')]

#Drop the rows where nan values occur
dataHyp1 <- dataHyp1[complete.cases(dataHyp1),]

```

```

#to find an appropriate cutoff point:

# we will only use values with at least 50 records,
# because for normal distribution you need a minimum of 50 values
# So then you have an substantial amount of values

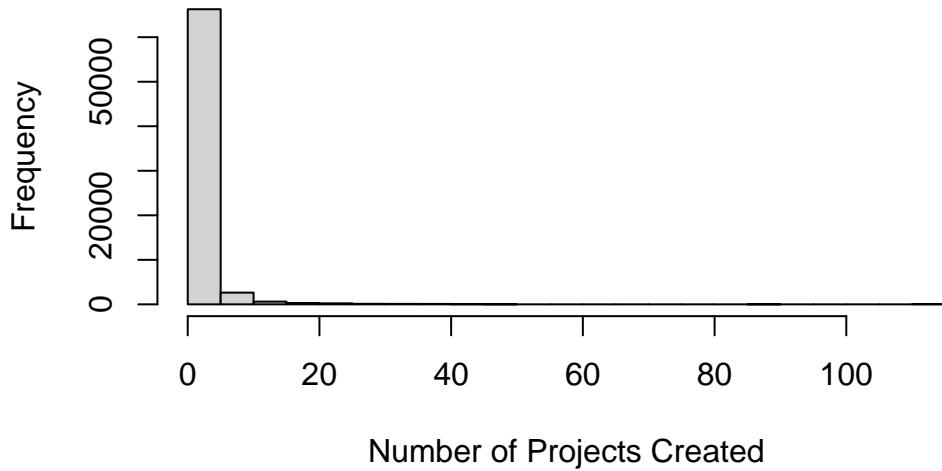
#iterate over the values of the following table:





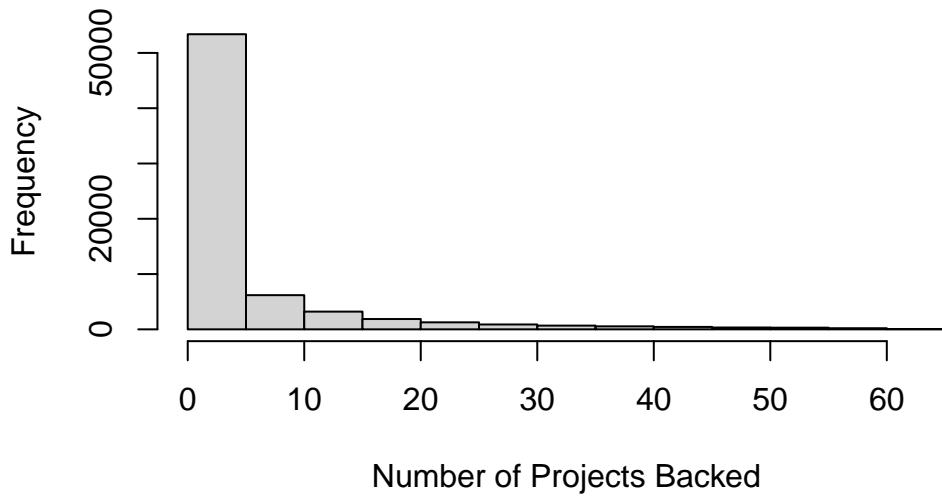
```

Histogram of Number of Projects Created (learning)



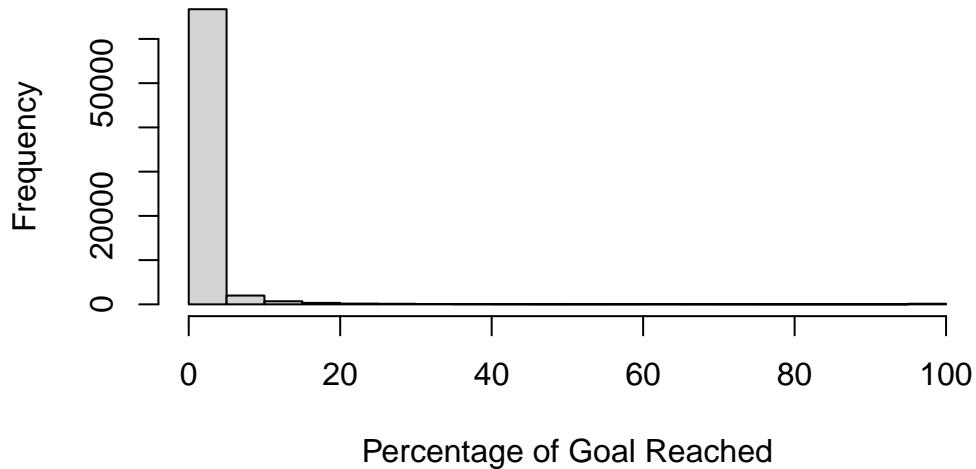
```
hist(as.numeric(dataHyp1_2$Creator_nb_backed),  
     main = 'Histogram of Number of Projects Backed (learning)',  
     xlab = 'Number of Projects Backed')
```

Histogram of Number of Projects Backed (learning)



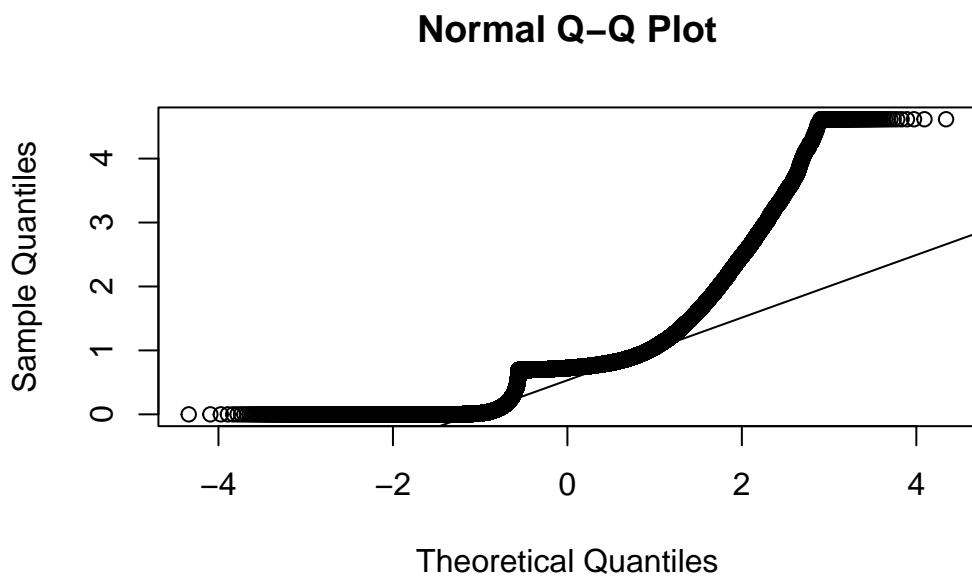
```
hist(dataHyp1_2$Goal_percentage,  
      main = 'Histogram of the Percentage of the Goal Reached (success)',  
      xlab = 'Percentage of Goal Reached')
```

Histogram of the Percentage of the Goal Reached (success)

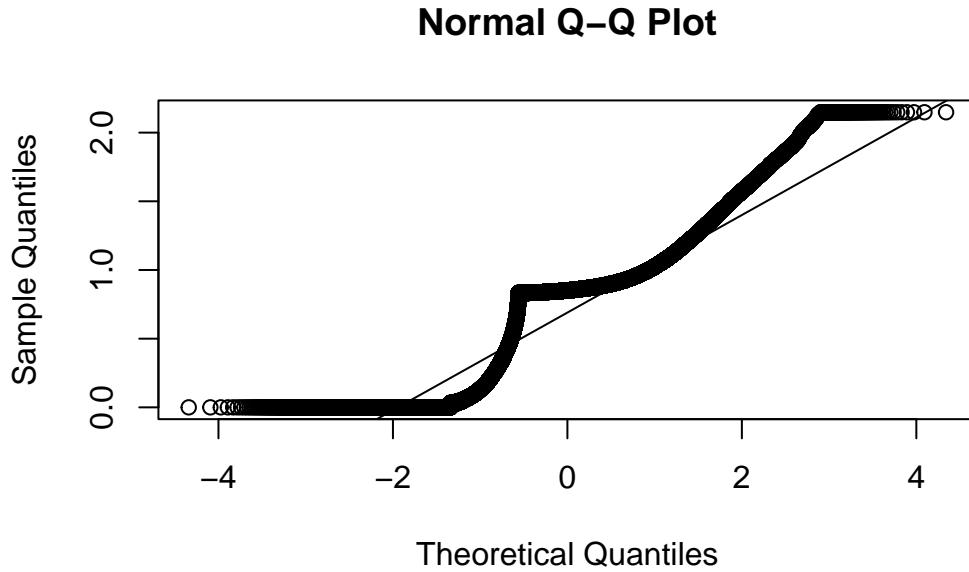


The above figure shows that the success proxy, the percentage of the Kickstarter goal that was reached, is not normal, so a couple of transformations are tried, as for linear regression a normal outcome variable is required. First $\text{Log}(x+1)$ is tried, given that 0% is still viable, this shift is necessary to avoid undefined values. Next $\text{sqrt}(\text{Log}(x+1))$ is applied, but even though it is better, with the QQ-plots below we notice that it is still not normally distributed. The first QQ-plot corresponds to $\text{Log}(x+1)$ and the second is for $\text{sqrt}(\text{Log}(x+1))$. Therefore a Linear Regression model cannot be used.

```
x<-dataHyp1_2$Goal_percentage  
#looks exponential, so we try log transform  
y<-log(dataHyp1_2$Goal_percentage+1)  
qqnorm(y)  
qqline(y)
```



```
#still skewed, so we try to include sqrt transform  
y<-sqrt(log(dataHyp1_2$Goal_percentage+1))  
qqnorm(y)  
qqline(y)
```



Regression and Anova test, research question 1

Given that the data is not normally distributed, and the tested transformations are not satisfactory a linear regression is not appropriate. By looking at the previous histogram, we can asses that the data seems to be from an exponential distribution with an inflated 0 value. Therefore, Poisson regression Gamma regression, and zero inflated regression are tried. For poisson regression, we round the goal percentage to the nearest integer, given that the model works for discrete outcome variables. It is also important to note that the zero inflated value model did not converge, regardless of the preprocessing steps taken for this regression. The first step was hot-encoding the proxies for learning because the model doesn't take factors as inputs, then the percentages for the success were rounded to integers because the model only outputs integers. The possible reasons as to why this error still occurred will be presented in the discussion section.

```
##non integer

##try Gamma distribution

dataHyp1_2$nonzero_percentage <- dataHyp1_2$Goal_percentage + 1e-6
modelGamma <- glm(nonzero_percentage ~ Creator_nb_projects +
                    Creator_nb_backed,
```

```

    data = dataHyp1_2,family =
      Gamma(link = 'identity'))

dataHyp1_2$Goal_percentage_int <- as.integer(round(dataHyp1_2$Goal_percentage,0))
##poisson distribution int

modelPoissonInt <- glm(Goal_percentage_int ~ Creator_nb_projects + Creator_nb_backed,
                        data = dataHyp1_2,
                        family = poisson(link = 'log'))

# trying zero inflation model
##can't use factors, need to one hot encode

hot_encode <- model.matrix(~ Creator_nb_projects + Creator_nb_backed - 1,
                           data = dataHyp1_2)

hot_encode<-as.data.frame(hot_encode)

##can't use non integer values for the dependent variable

inflDS <- cbind(dataHyp1_2$Goal_percentage,
                  hot_encode)

#don't run this if the second zeroinfl will be runned (it will cause issues with removing
inflDS$percent_int<- as.integer(round(inflDS$dataHyp1_2$Goal_percentage` ,0))

null_values_per_column <- sapply(inflDS,
                                    function(x) any(is.na(x)))

# Print the result

##doesn't converge
#modelZeroINFL_1 <- pscl:::zeroinfl(percent_int ~ . - `dataHyp1_2$Goal_percentage` | . - `d

##try to remove skeweness by ignoring certain values (between 0 an 1, to avoid having too

#inflDS<- inflDS[!(inflDS$dataHyp1_2$Goal_percentage` > 0 & inflDS$dataHyp1_2$Goal_perce
#inflDS$percent_int<- as.integer(round(inflDS$dataHyp1_2$Goal_percentage` ,0))

```

```

##doesn't converge either
#modelZeroINFL_2 <- pscl::zeroinfl(percent_int ~ . - `dataHyp1_2$Goal_percentage` | . - `d
#summary(modelZeroINFL)

```

Data Importing and cleaning, research question 2

First let us describe the data necessary for the second research question. We use the same dataset as the for the first research question: the kickstarter dataset. Besides a score for innovation, which we will discuss later, we used the following variables: - The number of projects that preceded this one (for the same entrepreneur). We made this both a number and a factor, to account for a decreasing effect as the number of observations decreased; - The number of dollars that the project aimed to achieve; - The duration of the project; - The success of the preceding project, which we defined as the number of dollars that project received divided by the number of dollars it aimed to achieve; - The category of the project (which is similar to the industry of the project) as control variables. Because of the pareto distribution that is generally considered to be present in income, we took the log values of the number of dollars that projects aimed to achieve and achieved (Arnold 2015).

```

# Install and load required packages
# install.packages(c("tidytext", "dplyr", "ggplot2"))
library(tidytext)

```

Warning: package 'tidytext' was built under R version 4.3.2

```
library(dplyr)
```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

```
filter, lag
```

The following objects are masked from 'package:base':

```
intersect, setdiff, setequal, union
```

```

library(ggplot2)

#inspect the data
info_filename <- file.choose()
info <- readRDS(info_filename)

#create a dataset with only creator_slug that appear more than once in the dataset
dataHyp2Double <- info[info$creator_slug %in% names(which(table(info$creator_slug) > 1)),]

dataHyp2sub<- dataHyp2Double[,c('creator_slug','Goal_USD',
                                'Pledge_USD','Launched_at',
                                'Deadline','Project_description',
                                'Creator_nb_projects','Creator_nb_backed',
                                'Category')]

dataHyp2filtered <- na.omit(dataHyp2sub)

#make a subdataset where deadline is before week 32 in 2019 in unix time
dataHyp2 <- dataHyp2filtered[dataHyp2filtered$Deadline < 1565474400,]

#make a new column with boolean if goal_USD >= Pledge_USD
dataHyp2$Pledge_USD <- log(as.numeric(dataHyp2$Pledge_USD) + 1)
dataHyp2$Goal_USD <- log(as.numeric(dataHyp2$Goal_USD) + 1)
dataHyp2$success_rate <- dataHyp2$Pledge_USD / dataHyp2$Goal_USD

dataHyp2 <- subset(dataHyp2, is.finite(dataHyp2$Pledge_USD)
                    & is.finite(dataHyp2$Goal_USD)
                    & is.finite(dataHyp2$success_rate))

# make a new column with order which orders the projects per creator_slug based on launched_at
# So the first project of a creator should get 1, the second 2 ...
dataHyp2 <- dataHyp2 %>%
  arrange(Launched_at) %>%
  group_by(creator_slug) %>%
  mutate(Order = row_number())

dataHyp2$duration <- as.numeric(dataHyp2$Deadline) - as.numeric(dataHyp2$Launched_at)

dataHyp2$orderFactor <- as.factor(dataHyp2$Order)

```

```

dataHyp2 <- dataHyp2 %>%
  arrange(creator_slug, Order) %>% # Sort the dataframe by 'id' and 'order'
  group_by(creator_slug) %>%
  mutate(previousSuccess = lag(success_rate, default = 1))

```

Defining innovation, research question 2

Since our dependent variable is innovation, we need to have a measure for innovation. We got our measure for innovation from the description of the project. Using a CAT scanner for innovation, which is a corpus of scientifically verified innovation-related words, we gave an innovation metric to the description, which became the innovation metric for the entire project (McKenny et al. 2018). We used the existence of these words as the measure, and more of these words means a higher measure.

```

word_dict <- c( "ad lib", "adroit", "adroitness", "bright idea", "clever", "cleverness", "c
innovation_ratios <- vector(mode='numeric', length=nrow(dataHyp2))
for (i in 1:nrow(dataHyp2)) {
  innovation_counter <- 0
  for (word in word_dict) {
    innovation_counter <- innovation_counter + stringr::str_count(dataHyp2$Project_descrip
  }
  innovation_ratios[i] <- innovation_counter/stringr::str_count(dataHyp2$Project_descripti
}

dataHyp2$innovation <- innovation_ratios * 1000

```

Defining linear models, research question 2

The variables mentioned in the section discussing the importing and cleaning of the data discussed the variables that we took into consideration. We investigated the impact on innovation for each of those variables separately (mod0 to mod5), and then added the variables together iteratively. We did not investigate all combinations, because this would take too much effort. We made a couple of models, based on our intuition of relevant parameters. We then plot the models.

```

mod0 <- lm(innovation ~ Order, data=dataHyp2)
mod1 <- lm(innovation ~ orderFactor, data=dataHyp2)
mod2 <- lm(innovation ~ Goal_USD, data=dataHyp2)

```

```

mod3 <- lm(innovation ~ duration, data=dataHyp2)
mod4 <- lm(innovation ~ previousSuccess, data=dataHyp2)
mod5 <- lm(innovation ~ Category, data=dataHyp2)

modA <- lm(innovation ~ Order + Goal_USD +
            duration, data=dataHyp2)
modB <- lm(innovation ~ orderFactor + Goal_USD +
            duration, data=dataHyp2)
modC <- lm(innovation ~ Order + Goal_USD +
            duration + previousSuccess,
            data=dataHyp2)
modD <- lm(innovation ~ orderFactor + Goal_USD +
            duration + previousSuccess,
            data=dataHyp2)
modE <- lm(innovation ~ Order + Goal_USD +
            duration + previousSuccess +
            previousSuccess * Order,
            data=dataHyp2)
modF <- lm(innovation ~ orderFactor + Goal_USD +
            duration + previousSuccess * orderFactor
            , data=dataHyp2)
modG <- lm(innovation ~ Order + Goal_USD + duration +
            previousSuccess + previousSuccess * Order
            + Category, data=dataHyp2)
modH <- lm(innovation ~ orderFactor + Goal_USD + duration +
            previousSuccess * orderFactor + Category,
            data=dataHyp2)

texreg::screenreg(list(mod0, mod1, mod2, mod3, mod4, mod5, modA,
                      modB, modC, modD, modE, modF, modG, modH))

```

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	3.33 *** (0.05)	3.23 *** (0.06)	0.21 (0.15)	2.16 *** (0.09)	3 (0.08)
Order	-0.11 *** (0.01)				
orderFactor2		-0.05 (0.08)			
orderFactor3			-0.43 ***		

	(0.12)
orderFactor4	-0.54 ** (0.17)
orderFactor5	-0.54 * (0.23)
orderFactor6	-0.55 (0.29)
orderFactor7	-0.40 (0.37)
orderFactor8	-1.04 * (0.45)
orderFactor9	-0.24 (0.50)
orderFactor10	-1.18 (0.61)
orderFactor11	-1.49 * (0.69)
orderFactor12	-1.13 (0.73)
orderFactor13	-1.27 (0.84)
orderFactor14	-1.92 * (0.94)
orderFactor15	-2.00 * (0.99)
orderFactor16	-2.47 * (1.02)
orderFactor17	-2.28 * (1.12)
orderFactor18	-2.36 * (1.17)
orderFactor19	-2.87 * (1.27)
orderFactor20	-1.60 (1.27)
orderFactor21	-3.10 * (1.49)
orderFactor22	-3.18 (1.71)
orderFactor23	-3.19 (1.71)
orderFactor24	-3.15 (1.88)

orderFactor25	-3.18 (1.88)
orderFactor26	-2.97 (2.10)
orderFactor27	-2.99 (2.10)
orderFactor28	-2.77 (2.10)
orderFactor29	-2.90 (2.10)
orderFactor30	-2.48 (2.42)
orderFactor31	-2.70 (2.42)
orderFactor32	-2.82 (2.42)
orderFactor33	-2.53 (2.97)
orderFactor34	-2.66 (2.97)
orderFactor35	-2.59 (2.97)
orderFactor36	-2.61 (2.97)
orderFactor37	-3.02 (2.97)
orderFactor38	-3.23 (4.20)
orderFactor39	-3.23 (4.20)
orderFactor40	-3.23 (4.20)
orderFactor41	-3.23 (4.20)
orderFactor42	-3.23 (4.20)
orderFactor43	-3.23 (4.20)
orderFactor44	-3.23 (4.20)
orderFactor45	-3.23 (4.20)
orderFactor46	-3.23

		(4.20)
orderFactor47	-3.23	
	(4.20)	
orderFactor48	-3.23	
	(4.20)	
orderFactor49	-3.23	
	(4.20)	
orderFactor50	-3.23	
	(4.20)	
orderFactor51	-3.23	
	(4.20)	
orderFactor52	-3.23	
	(4.20)	
Goal_USD	0.38 ***	
	(0.02)	
duration	0.00 ***	
	(0.00)	
previousSuccess	-0	
	(0)	
CategoryAcademic		
CategoryAccessories		
CategoryAction		
CategoryAnimals		
CategoryAnimation		
CategoryAnthologies		
CategoryApparel		
CategoryApps		
CategoryArchitecture		
CategoryArt		
CategoryArt Books		
CategoryAudio		

CategoryCalendars

CategoryCamera Equipment

CategoryCandles

CategoryCeramics

CategoryClassical Music

CategoryComedy

CategoryComic Books

CategoryComics

CategoryConceptual Art

CategoryCookbooks

CategoryCountry & Folk

CategoryCrafts

CategoryCrochet

CategoryDance

CategoryDesign

CategoryDigital Art

CategoryDIY

CategoryDIY Electronics

CategoryDocumentary

CategoryDrama

CategoryDrinks

CategoryElectronic Music

CategoryEmbroidery
CategoryEvents
CategoryExperimental
CategoryFabrication Tools
CategoryFaith
CategoryFamily
CategoryFantasy
CategoryFarms
CategoryFashion
CategoryFestivals
CategoryFiction
CategoryFilm & Video
CategoryFine Art
CategoryFood
CategoryFood Trucks
CategoryFootwear
CategoryGadgets
CategoryGames
CategoryGaming Hardware
CategoryGlass
CategoryGraphic Design

CategoryGraphic Novels

CategoryHardware

CategoryHip-Hop

CategoryHorror

CategoryIllustration

CategoryImmersive

CategoryIndie Rock

CategoryInstallations

CategoryInteractive Design

CategoryJazz

CategoryJewelry

CategoryJournalism

CategoryKids

CategoryKnitting

CategoryLetterpress

CategoryLiterary Journals

CategoryLiterary Spaces

CategoryLive Games

CategoryMakerspaces

CategoryMetal

CategoryMixed Media

CategoryMobile Games

CategoryMusic
CategoryMusic Videos
CategoryMusical
CategoryNarrative Film
CategoryNature
CategoryNonfiction
CategoryPainting
CategoryPeople
CategoryPerformance Art
CategoryPerformances
CategoryPeriodicals
CategoryPhoto
CategoryPhotobooks
CategoryPhotography
CategoryPlaces
CategoryPlaying Cards
CategoryPlays
CategoryPoetry
CategoryPop
CategoryPottery
CategoryPrint

CategoryPrinting

CategoryProduct Design

CategoryPublic Art

CategoryPublishing

CategoryPunk

CategoryPuzzles

CategoryQuilts

CategoryRadio & Podcasts

CategoryReady-to-wear

CategoryResidencies

CategoryRestaurants

CategoryRobots

CategoryRock

CategoryScience Fiction

CategorySculpture

CategoryShorts

CategorySmall Batch

CategorySoftware

CategorySound

CategorySpace Exploration

CategorySpaces

CategoryStationery

CategoryTabletop Games

CategoryTechnology

CategoryTelevision

CategoryTextiles

CategoryTheater

CategoryThrillers

CategoryTypography

CategoryVegan

CategoryVideo

CategoryVideo Games

CategoryWearables

CategoryWeaving

CategoryWeb

CategoryWebcomics

CategoryWebseries

CategoryWoodworking

CategoryWorkshops

CategoryWorld Music

CategoryYoung Adult

CategoryZines

Order:previousSuccess

orderFactor2:previousSuccess
orderFactor3:previousSuccess
orderFactor4:previousSuccess
orderFactor5:previousSuccess
orderFactor6:previousSuccess
orderFactor7:previousSuccess
orderFactor8:previousSuccess
orderFactor9:previousSuccess
orderFactor10:previousSuccess
orderFactor11:previousSuccess
orderFactor12:previousSuccess
orderFactor13:previousSuccess
orderFactor14:previousSuccess
orderFactor15:previousSuccess
orderFactor16:previousSuccess
orderFactor17:previousSuccess
orderFactor18:previousSuccess
orderFactor19:previousSuccess
orderFactor20:previousSuccess
orderFactor21:previousSuccess
orderFactor22:previousSuccess
orderFactor23:previousSuccess

```

orderFactor24:previousSuccess

orderFactor25:previousSuccess

orderFactor26:previousSuccess

orderFactor27:previousSuccess

orderFactor28:previousSuccess

orderFactor29:previousSuccess

orderFactor30:previousSuccess

orderFactor31:previousSuccess

orderFactor32:previousSuccess

orderFactor33:previousSuccess

orderFactor34:previousSuccess

orderFactor35:previousSuccess

orderFactor36:previousSuccess

```

R^2	0.01	0.01	0.03	0.01	0
Adj. R^2	0.01	0.00	0.03	0.01	-0
Num. obs.	14043	14043	14043	14043	14043

*** p < 0.001; ** p < 0.01; * p < 0.05

```
texreg::screenreg(list(mod0, mod2, mod3, mod4, modA, modC, modE))
```

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	3.33 ***	0.21	2.16 ***	3.07 ***	0.31

	(0.05)	(0.15)	(0.09)	(0.04)	(0.17)
Order	-0.11 *** (0.01)				-0.07 *** (0.01)
Goal_USD		0.38 *** (0.02)			0.34 *** (0.02)
duration			0.00 *** (0.00)		0.00 *** (0.00)
previousSuccess				-0.01 (0.01)	
Order:previousSuccess					
R^2	0.01	0.03	0.01	0.00	0.03
Adj. R^2	0.01	0.03	0.01	-0.00	0.03
Num. obs.	14043	14043	14043	14043	14043
=====					
*** p < 0.001; ** p < 0.01; * p < 0.05					

Plotting innovation graphs

Besides fitting a linear model, we also made graphs showing the impact of having multiple projects on the mean innovation and change in innovation compared to having one less project.

```

dataHyp2 <- dataHyp2 %>%
  arrange(creator_slug, Order) %>% # Sort the dataframe by 'id' and 'order'
  group_by(creator_slug) %>%
  mutate(change = innovation -
    lag(innovation, default = first(innovation)))

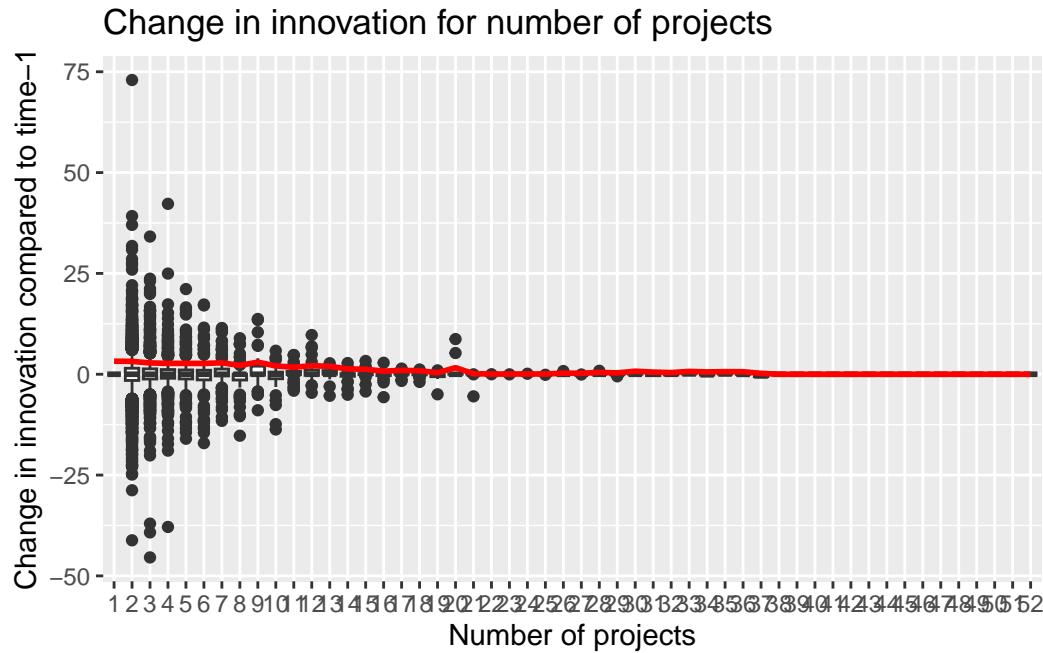
dataHyp2 %>% ggplot(aes(x=factor(Order), y=change)) +
  geom_boxplot() +
  geom_line(stat="summary", fun="mean",
            aes(group=1, y=innovation),
            color="red", size=1) +
  labs(title="Change in innovation for number of projects",
       x="Number of projects",
       y="Change in innovation compared to time-1")

```

Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.

Warning: Removed 12 rows containing non-finite values (`stat_boxplot()`).

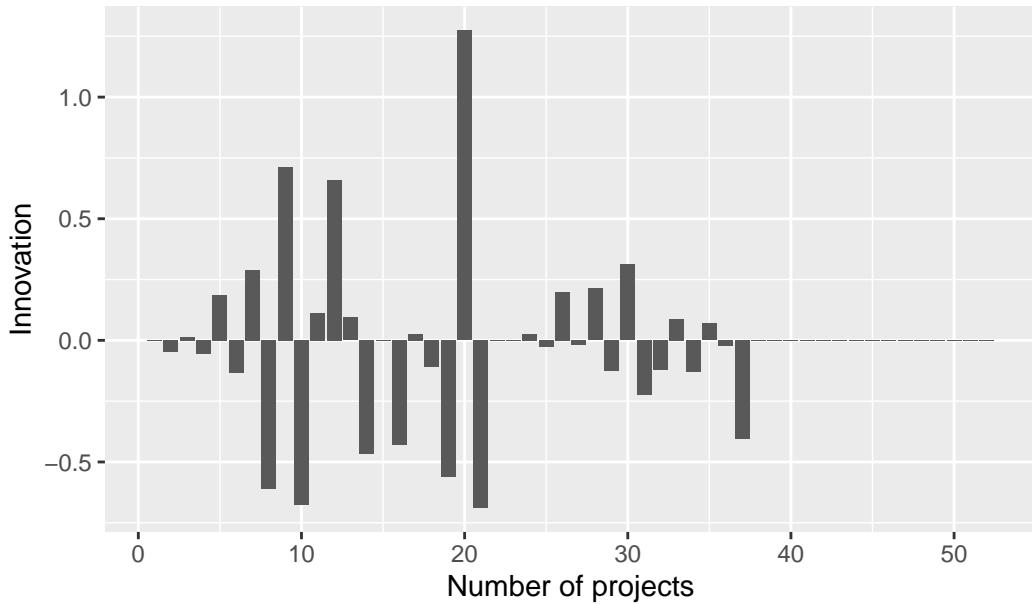
Warning: Removed 7 rows containing non-finite values (`stat_summary()`).



```
ggplot(dataHyp2, aes(x = Order, y = change)) +  
  geom_bar(stat = "summary", fun = "mean",  
           position = "dodge") +  
  labs(title = "Mean change in innovation for number of projects",  
        x = "Number of projects", y = "Innovation")
```

Warning: Removed 12 rows containing non-finite values (`stat_summary()`).

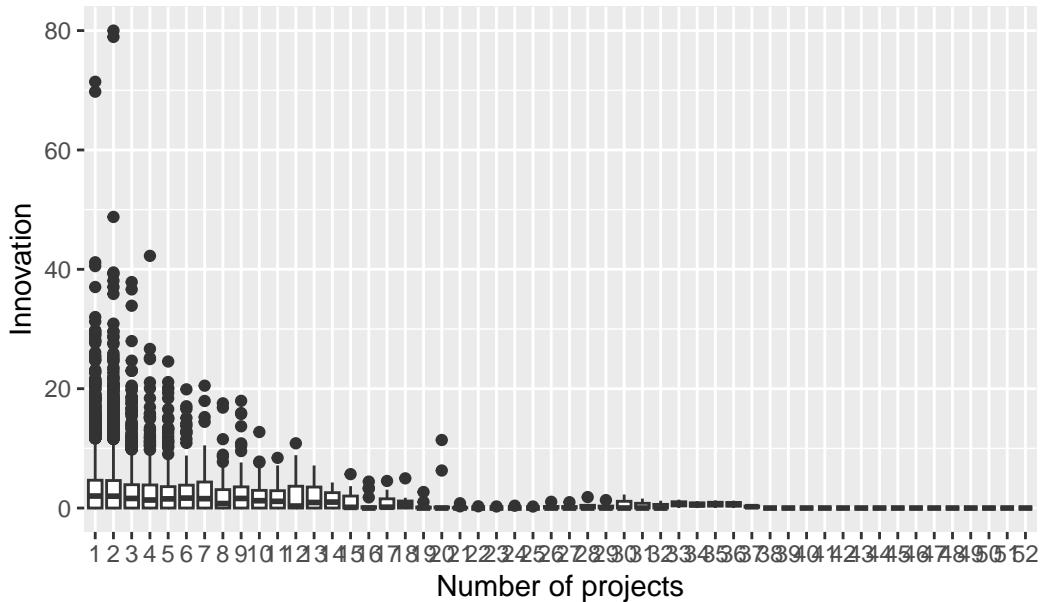
Mean change in innovation for number of projects



```
dataHyp2 %>% ggplot(aes(x=factor(Order), y=innovation)) +  
  geom_boxplot() +  
  labs(title="Innovation for number of projects",  
       x="Number of projects",  
       y="Innovation")
```

Warning: Removed 7 rows containing non-finite values (`stat_boxplot()`).

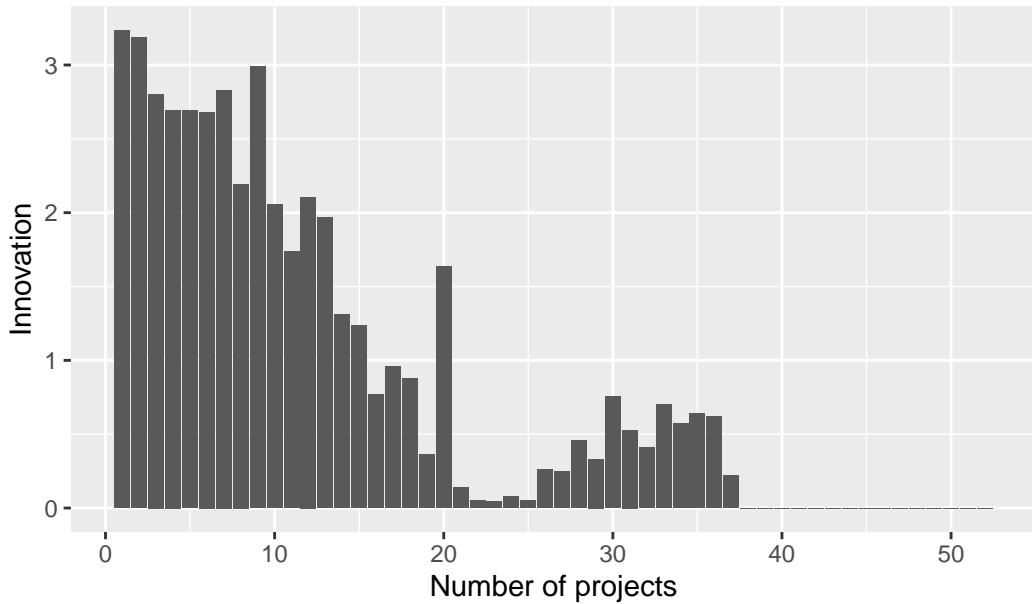
Innovation for number of projects



```
ggplot(dataHyp2, aes(x = Order, y = innovation)) +  
  geom_bar(stat = "summary", fun = "mean", position = "dodge") +  
  labs(title = "Mean innovation for number of projects",  
       x = "Number of projects", y = "Innovation")
```

Warning: Removed 7 rows containing non-finite values (`stat_summary()`).

Mean innovation for number of projects



Results

Hypothesis 1 results

Linear models

Given that the zero inflated model doesn't produce results, we proceed by studying the Poisson and Gamma models. First we look at the results for the Poisson Model.

```
##Poisson summary  
summary(modelPoissonInt)
```

```
Call:  
glm(formula = Goal_percentage_int ~ Creator_nb_projects + Creator_nb_backed,  
family = poisson(link = "log"), data = dataHyp1_2)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.484358	0.032766	-14.782	< 2e-16 ***
Creator_nb_projects10	0.970510	0.029803	32.564	< 2e-16 ***
Creator_nb_projects11	1.323840	0.029213	45.316	< 2e-16 ***

Creator_nb_projects11	0.190971	0.130855	1.459	0.144453	
Creator_nb_projects12	1.333329	0.032049	41.603	< 2e-16	***
Creator_nb_projects13	1.519120	0.031916	47.598	< 2e-16	***
Creator_nb_projects14	1.496603	0.037025	40.421	< 2e-16	***
Creator_nb_projects15	1.913954	0.039595	48.338	< 2e-16	***
Creator_nb_projects16	0.959446	0.043219	22.199	< 2e-16	***
Creator_nb_projects17	1.199196	0.059702	20.086	< 2e-16	***
Creator_nb_projects18	1.269823	0.051356	24.726	< 2e-16	***
Creator_nb_projects19	1.410542	0.057838	24.388	< 2e-16	***
Creator_nb_projects2	0.385425	0.007966	48.382	< 2e-16	***
Creator_nb_projects20	1.574735	0.065015	24.221	< 2e-16	***
Creator_nb_projects21	0.868416	0.080849	10.741	< 2e-16	***
Creator_nb_projects22	1.077218	0.070061	15.375	< 2e-16	***
Creator_nb_projects23	1.615828	0.060163	26.857	< 2e-16	***
Creator_nb_projects24	0.457387	0.093820	4.875	1.09e-06	***
Creator_nb_projects25	1.500792	0.057358	26.165	< 2e-16	***
Creator_nb_projects26	0.849905	0.099060	8.580	< 2e-16	***
Creator_nb_projects27	-0.049399	0.104824	-0.471	0.637454	
Creator_nb_projects28	2.800223	0.040252	69.567	< 2e-16	***
Creator_nb_projects29	1.692689	0.226079	7.487	7.04e-14	***
Creator_nb_projects3	0.606283	0.010527	57.591	< 2e-16	***
Creator_nb_projects30	0.621388	0.163417	3.802	0.000143	***
Creator_nb_projects31	0.956605	0.088319	10.831	< 2e-16	***
Creator_nb_projects32	0.703395	0.098781	7.121	1.07e-12	***
Creator_nb_projects33	0.085094	0.579067	0.147	0.883172	
Creator_nb_projects34	1.232163	0.336815	3.658	0.000254	***
Creator_nb_projects35	0.694530	0.111339	6.238	4.43e-10	***
Creator_nb_projects36	1.241970	0.060135	20.653	< 2e-16	***
Creator_nb_projects38	0.752018	0.123365	6.096	1.09e-09	***
Creator_nb_projects39	0.940241	0.095597	9.835	< 2e-16	***
Creator_nb_projects4	0.839143	0.012701	66.070	< 2e-16	***
Creator_nb_projects40	-0.289444	1.000126	-0.289	0.772269	
Creator_nb_projects42	1.925956	0.045265	42.548	< 2e-16	***
Creator_nb_projects49	0.171043	0.237197	0.721	0.470846	
Creator_nb_projects5	0.985661	0.014227	69.279	< 2e-16	***
Creator_nb_projects50	2.453810	0.174687	14.047	< 2e-16	***
Creator_nb_projects52	1.339198	0.415013	3.227	0.001251	**
Creator_nb_projects6	1.003268	0.016512	60.760	< 2e-16	***
Creator_nb_projects7	0.929724	0.020509	45.333	< 2e-16	***
Creator_nb_projects8	0.852312	0.023922	35.629	< 2e-16	***
Creator_nb_projects88	-0.577492	1.000900	-0.577	0.563958	
Creator_nb_projects89	-0.894210	0.577480	-1.548	0.121509	
Creator_nb_projects9	1.060078	0.025602	41.405	< 2e-16	***

Creator_nb_backed0	0.577828	0.033062	17.477	< 2e-16	***
Creator_nb_backed1	0.690691	0.033752	20.464	< 2e-16	***
Creator_nb_backed10	0.864926	0.040047	21.598	< 2e-16	***
Creator_nb_backed11	1.196799	0.038669	30.950	< 2e-16	***
Creator_nb_backed12	1.071574	0.039532	27.107	< 2e-16	***
Creator_nb_backed13	0.901101	0.041838	21.538	< 2e-16	***
Creator_nb_backed14	1.197472	0.040308	29.708	< 2e-16	***
Creator_nb_backed15	1.219767	0.039293	31.043	< 2e-16	***
Creator_nb_backed16	1.067656	0.043061	24.794	< 2e-16	***
Creator_nb_backed17	0.891568	0.044943	19.838	< 2e-16	***
Creator_nb_backed18	1.046087	0.045823	22.829	< 2e-16	***
Creator_nb_backed19	1.056437	0.044746	23.610	< 2e-16	***
Creator_nb_backed2	0.802310	0.034292	23.396	< 2e-16	***
Creator_nb_backed20	1.100849	0.045527	24.180	< 2e-16	***
Creator_nb_backed21	1.168717	0.044519	26.252	< 2e-16	***
Creator_nb_backed22	1.068879	0.046546	22.964	< 2e-16	***
Creator_nb_backed23	1.084586	0.050314	21.556	< 2e-16	***
Creator_nb_backed24	0.804730	0.055056	14.617	< 2e-16	***
Creator_nb_backed25	0.700544	0.053101	13.193	< 2e-16	***
Creator_nb_backed26	0.918567	0.052336	17.551	< 2e-16	***
Creator_nb_backed27	1.422655	0.046541	30.568	< 2e-16	***
Creator_nb_backed28	1.333736	0.052408	25.449	< 2e-16	***
Creator_nb_backed29	1.180291	0.052314	22.562	< 2e-16	***
Creator_nb_backed3	0.867743	0.034715	24.996	< 2e-16	***
Creator_nb_backed30	1.061850	0.053351	19.903	< 2e-16	***
Creator_nb_backed31	1.205576	0.057315	21.034	< 2e-16	***
Creator_nb_backed32	0.871443	0.062869	13.861	< 2e-16	***
Creator_nb_backed33	0.935607	0.063032	14.843	< 2e-16	***
Creator_nb_backed34	1.205795	0.058489	20.616	< 2e-16	***
Creator_nb_backed35	1.253387	0.053419	23.463	< 2e-16	***
Creator_nb_backed36	1.361238	0.052204	26.076	< 2e-16	***
Creator_nb_backed37	1.430297	0.054583	26.204	< 2e-16	***
Creator_nb_backed38	1.245577	0.059517	20.928	< 2e-16	***
Creator_nb_backed39	0.739734	0.065334	11.322	< 2e-16	***
Creator_nb_backed4	0.833909	0.035616	23.414	< 2e-16	***
Creator_nb_backed40	1.449420	0.058112	24.942	< 2e-16	***
Creator_nb_backed41	1.814699	0.051697	35.102	< 2e-16	***
Creator_nb_backed42	0.999048	0.067248	14.856	< 2e-16	***
Creator_nb_backed43	2.113780	0.044994	46.979	< 2e-16	***
Creator_nb_backed44	1.362498	0.062887	21.666	< 2e-16	***
Creator_nb_backed45	1.556374	0.052652	29.560	< 2e-16	***
Creator_nb_backed46	1.767394	0.057524	30.724	< 2e-16	***
Creator_nb_backed47	1.119002	0.064515	17.345	< 2e-16	***

```

Creator_nb_backed48    0.936920  0.081311  11.523 < 2e-16 ***
Creator_nb_backed49    1.284011  0.070419  18.234 < 2e-16 ***
Creator_nb_backed5     0.773803  0.036150  21.405 < 2e-16 ***
Creator_nb_backed50    2.508425  0.048006  52.253 < 2e-16 ***
Creator_nb_backed51    0.736593  0.076250  9.660 < 2e-16 ***
Creator_nb_backed52    1.322452  0.067986  19.452 < 2e-16 ***
Creator_nb_backed54    1.645189  0.064784  25.395 < 2e-16 ***
Creator_nb_backed55    1.756205  0.055481  31.654 < 2e-16 ***
Creator_nb_backed56    0.946325  0.073508  12.874 < 2e-16 ***
Creator_nb_backed58    0.858853  0.078141  10.991 < 2e-16 ***
Creator_nb_backed59    2.050878  0.047924  42.795 < 2e-16 ***
Creator_nb_backed6      0.936734  0.036403  25.732 < 2e-16 ***
Creator_nb_backed61    0.940873  0.088337  10.651 < 2e-16 ***
Creator_nb_backed7     1.025420  0.036877  27.806 < 2e-16 ***
Creator_nb_backed8     1.149193  0.037219  30.876 < 2e-16 ***
Creator_nb_backed9     0.853349  0.039087  21.832 < 2e-16 ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 324908  on 70489  degrees of freedom
Residual deviance: 275525  on 70385  degrees of freedom
AIC: 395399

```

Number of Fisher Scoring iterations: 6

From the results we see that most of the coefficients for both the number of backed projects and the number of projects created are significant. This suggests that the effects of the proxies we selected for learning are significant. In regards to the sign of the coefficients, we notice that none of the negative values are significant, with the exception of the intercept. Since the significant coefficients are positive, though small, we conclude that learning has a positive effect on success, which is in line with (find citation). In this case we don't notice an increasing pattern in of the coefficients, wherein more experience would have significantly higher values, as sometimes the coefficients decrease again.

```

#Anova Poisson
anova(modelPoissonInt)

```

Analysis of Deviance Table

Model: poisson, link: log

```
Response: Goal_percentage_int
```

```
Terms added sequentially (first to last)
```

	Df	Deviance	Resid. Df	Resid. Dev
NULL			70489	324908
Creator_nb_projects	45	38918	70444	285991
Creator_nb_backed	59	10465	70385	275525

In the analysis of deviance table we notice that the inclusion of the number of projects backed reduces the deviance by a large margin, this justifies including these variables in the model, as it results in a better model.

```
#Summary for Gamma model  
summary(modelGamma)
```

Call:

```
glm(formula = nonzero_percentage ~ Creator_nb_projects + Creator_nb_backed,  
family = Gamma(link = "identity"), data = dataHyp1_2)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.47702	0.05323	8.961	< 2e-16 ***
Creator_nb_projects10	2.45192	0.83133	2.949	0.003185 **
Creator_nb_projects11	2.61519	1.09477	2.389	0.016906 *
Creator_nb_projects111	-0.99142	1.49522	-0.663	0.507292
Creator_nb_projects12	4.39404	1.63824	2.682	0.007316 **
Creator_nb_projects13	6.28918	2.52673	2.489	0.012810 *
Creator_nb_projects14	1.95381	1.19632	1.633	0.102433
Creator_nb_projects15	9.52753	4.76430	2.000	0.045528 *
Creator_nb_projects16	2.86984	1.46914	1.953	0.050774 .
Creator_nb_projects17	3.64356	2.72903	1.335	0.181843
Creator_nb_projects18	4.16088	2.66558	1.561	0.118537
Creator_nb_projects19	4.07335	2.57359	1.583	0.113482
Creator_nb_projects2	0.58984	0.06466	9.123	< 2e-16 ***
Creator_nb_projects20	5.97452	4.30087	1.389	0.164794
Creator_nb_projects21	1.75103	1.33439	1.312	0.189448
Creator_nb_projects22	1.27609	1.50905	0.846	0.397768
Creator_nb_projects23	5.03077	3.07581	1.636	0.101930
Creator_nb_projects24	1.66014	2.58275	0.643	0.520368

Creator_nb_projects25	4.00556	2.33387	1.716	0.086117	.
Creator_nb_projects26	3.31913	4.26133	0.779	0.436044	
Creator_nb_projects27	-0.05482	0.78193	-0.070	0.944112	
Creator_nb_projects28	29.71981	21.44784	1.386	0.165849	
Creator_nb_projects29	7.29686	22.09647	0.330	0.741229	
Creator_nb_projects3	1.14062	0.13666	8.346	< 2e-16	***
Creator_nb_projects30	1.78175	3.30486	0.539	0.589798	
Creator_nb_projects31	2.35291	2.15624	1.091	0.275183	
Creator_nb_projects32	1.55039	2.55401	0.607	0.543824	
Creator_nb_projects33	0.10745	4.31377	0.025	0.980127	
Creator_nb_projects34	5.15591	30.30396	0.170	0.864901	
Creator_nb_projects35	1.32984	1.80534	0.737	0.461359	
Creator_nb_projects36	3.25917	2.18547	1.491	0.135889	
Creator_nb_projects38	1.82152	4.69743	0.388	0.698187	
Creator_nb_projects39	2.55425	3.05858	0.835	0.403658	
Creator_nb_projects4	1.72973	0.24054	7.191	6.50e-13	***
Creator_nb_projects40	-0.26305	4.06774	-0.065	0.948439	
Creator_nb_projects42	12.26538	7.27411	1.686	0.091768	.
Creator_nb_projects49	0.07423	1.84086	0.040	0.967836	
Creator_nb_projects5	2.53598	0.38055	6.664	2.69e-11	***
Creator_nb_projects50	14.91381	38.01987	0.392	0.694865	
Creator_nb_projects52	3.93888	18.63956	0.211	0.832640	
Creator_nb_projects6	2.59277	0.46399	5.588	2.31e-08	***
Creator_nb_projects7	2.26201	0.51423	4.399	1.09e-05	***
Creator_nb_projects8	1.81782	0.54857	3.314	0.000921	***
Creator_nb_projects88	-0.98533	4.01060	-0.246	0.805930	
Creator_nb_projects89	-0.99283	0.90235	-1.100	0.271220	
Creator_nb_projects9	2.40109	0.74868	3.207	0.001341	**
Creator_nb_backed0	0.62741	0.05763	10.887	< 2e-16	***
Creator_nb_backed1	0.81937	0.07212	11.361	< 2e-16	***
Creator_nb_backed10	1.22452	0.23372	5.239	1.62e-07	***
Creator_nb_backed11	2.05495	0.35333	5.816	6.05e-09	***
Creator_nb_backed12	1.50790	0.30495	4.945	7.64e-07	***
Creator_nb_backed13	1.21275	0.28879	4.199	2.68e-05	***
Creator_nb_backed14	1.70983	0.37714	4.534	5.81e-06	***
Creator_nb_backed15	2.08037	0.45345	4.588	4.49e-06	***
Creator_nb_backed16	1.57752	0.40191	3.925	8.68e-05	***
Creator_nb_backed17	1.54373	0.41226	3.745	0.000181	***
Creator_nb_backed18	1.72639	0.45732	3.775	0.000160	***
Creator_nb_backed19	1.73725	0.51359	3.383	0.000718	***
Creator_nb_backed2	0.97195	0.08941	10.871	< 2e-16	***
Creator_nb_backed20	1.85212	0.53923	3.435	0.000593	***
Creator_nb_backed21	1.59011	0.52268	3.042	0.002349	**

Creator_nb_backed22	1.75947	0.56570	3.110	0.001870	**
Creator_nb_backed23	1.60450	0.57229	2.804	0.005054	**
Creator_nb_backed24	1.27303	0.52055	2.446	0.014465	*
Creator_nb_backed25	1.32311	0.52310	2.529	0.011429	*
Creator_nb_backed26	1.43292	0.56400	2.541	0.011067	*
Creator_nb_backed27	2.98434	0.98057	3.043	0.002340	**
Creator_nb_backed28	2.28945	0.95994	2.385	0.017083	*
Creator_nb_backed29	2.28390	0.92331	2.474	0.013378	*
Creator_nb_backed3	1.13101	0.11092	10.197	< 2e-16	***
Creator_nb_backed30	1.71431	0.79076	2.168	0.030169	*
Creator_nb_backed31	2.08959	0.94583	2.209	0.027160	*
Creator_nb_backed32	1.27232	0.62945	2.021	0.043251	*
Creator_nb_backed33	1.35462	0.76168	1.778	0.075332	.
Creator_nb_backed34	1.28405	0.67674	1.897	0.057777	.
Creator_nb_backed35	2.40923	1.04774	2.299	0.021482	*
Creator_nb_backed36	2.56560	1.15339	2.224	0.026124	*
Creator_nb_backed37	2.11322	1.08288	1.951	0.051005	.
Creator_nb_backed38	2.20728	1.14924	1.921	0.054781	.
Creator_nb_backed39	1.11370	0.69429	1.604	0.108697	
Creator_nb_backed4	1.09967	0.12537	8.771	< 2e-16	***
Creator_nb_backed40	3.64407	1.84075	1.980	0.047745	*
Creator_nb_backed41	3.90018	2.07388	1.881	0.060027	.
Creator_nb_backed42	1.47726	0.96867	1.525	0.127255	
Creator_nb_backed43	8.52575	3.31873	2.569	0.010202	*
Creator_nb_backed44	2.45128	1.47126	1.666	0.095696	.
Creator_nb_backed45	4.05141	1.89621	2.137	0.032635	*
Creator_nb_backed46	5.03571	2.66157	1.892	0.058494	.
Creator_nb_backed47	1.85418	1.09379	1.695	0.090044	.
Creator_nb_backed48	1.29110	1.08684	1.188	0.234861	
Creator_nb_backed49	2.43701	1.64639	1.480	0.138822	
Creator_nb_backed5	1.03303	0.13327	7.751	9.20e-15	***
Creator_nb_backed50	10.77501	5.31291	2.028	0.042556	*
Creator_nb_backed51	2.10415	1.31211	1.604	0.108798	
Creator_nb_backed52	2.68821	1.68243	1.598	0.110088	
Creator_nb_backed54	3.58360	2.20501	1.625	0.104124	
Creator_nb_backed55	7.68948	4.41989	1.740	0.081909	.
Creator_nb_backed56	1.42138	1.14023	1.247	0.212558	
Creator_nb_backed58	0.92806	1.07294	0.865	0.387055	
Creator_nb_backed59	8.71661	4.22414	2.064	0.039067	*
Creator_nb_backed6	1.35877	0.17202	7.899	2.86e-15	***
Creator_nb_backed61	1.21247	1.07981	1.123	0.261504	
Creator_nb_backed7	1.32316	0.19127	6.918	4.63e-12	***
Creator_nb_backed8	1.31279	0.21024	6.244	4.28e-10	***

```

Creator_nb_backed9      1.15493    0.21489   5.374 7.71e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 10.63113)

Null deviance: 313015  on 70489  degrees of freedom
Residual deviance: 292975  on 70385  degrees of freedom
AIC: 117267

Number of Fisher Scoring iterations: 10

```

For this model, we notice that the coefficients are not mostly significant in the case of the number of projects created; contrastingly, the number of projects backed does have mostly significant values. It is also important to note that this Gamma distribution model has a lower AIC than the Poisson distribution model. We also notice that if the values are placed in increasing order of the amounts projects back and of the number of projects created there is an increase in the coefficients. This suggests that the more experience one has, the better they tend to perform. We also notice that the values that don't follow this tendency don't have significant p-values.

```

#Anova Gamma
anova(modelGamma)

```

Analysis of Deviance Table

```

Model: Gamma, link: identity

Response: nonzero_percentage

Terms added sequentially (first to last)

```

	Df	Deviance	Resid. Df	Resid. Dev
NULL			70489	313015
Creator_nb_projects	45	15623.3	70444	297391
Creator_nb_backed	59	4416.6	70385	292975

In this second analysis of deviance table, a similar effect is noticed to the previous, including the number of projects backed improves the model by reducing the deviance, which suggests that it's a good variable for the model. Therefore we can say that the success is not just explained by the experience one gains by doing more projects, but also observing the projects of others.

Justification of Results

The hypothesis that was studied stated that the more experience one has, whether direct or indirect, the more success one will have. However, the benefit that one sees from this has diminishing returns. A possible explanation for this could be similar to what Parker stated, that is if re-entry to entrepreneurship takes too long, the benefits one receives diminishes (Parker 2013). This could be due to forgetting past experiences, or the experiences becoming outdated in the ever changing environment in markets. This means that having too many past experiences implies that not all of the experiences will be used, as the gap between the earliest and the latest could be too wide, explaining why the coefficients in the regression seem to increase minimally after having some experience. Another explanation for this could simply be that there is not enough data on those who have a high amount of experience.

Hypothesis 2 results

Linear models

Let us first discuss the results from the linear models. Using a significance level of 0.05, which is justified since we are not dealing with a critical situation, we can see that in all linear models, the number of previous projects (the ‘Order’ term) has a significant effect on the innovation score of projects. We can also see that this effect is negative. This means that the more projects a creator has done, the less innovative a new project is expected to be. As said in the methods section, we also added the number of previous projects as a factor (the ‘orderFactor’ term). Without any control variables, we can see that as projects get more previous projects, the innovation score gets lower. However, after adding the control variables, the effect of ‘orderFactor’ was no longer determined to be significant.

Comparing the models, it stands out that the ‘category’ term greatly increases the R-squared metric. We also see that a lot of categories significantly differ from the baseline category. We hypothesize that this means that some categories are more likely to be innovative than others.

Plots

Looking at the plots, we are unable to see a clear trend. From our visual interpretations of the plots of the change in innovativeness. In the first plot, the red line seems to show that the number of projects slightly increases the innovation: the projects are - on average - slightly more innovative than the previous one, for the same entrepreneur. However, the effect seems to be very small, and there are lots of projects that are a lot less innovative than the previous ones for the same entrepreneur.

From the plots showing the mean innovation score for every project, we can see that as entrepreneurs have more previous projects, the projects become less innovative, on average. This seems to be in contrast with the first plots, showing the change in innovation. This could be

explained by the fact that innovative products are more difficult to produce and thus take longer. This would mean that people that make more innovative products are not able to do as many projects. This also implies that projects with a lot of previous projects only occur when the entrepreneur does not do innovative projects (because they simply take too long to complete).

Justification of the results

We hypothesized that people with more previous projects, would be more innovative than people with fewer previous projects. However, we found the opposite effect. This could be due to the fact that people with lots of previous projects may be entrenched in their communities. This could make them behave as ‘imitative entrepreneurs’ (Cliff, Jennings, and Greenwood 2006).

As we said in the previous section, the effect can also be explained by the fact that projects with more previous projects can only be done by entrepreneurs that are less innovative. So this would mean that an increase in the number of projects is not the cause of an increase in innovativeness, but that an increase in number of projects is only possible for entrepreneurs that are not as innovative: instead of describing the effect of the number of projects, we described a feature of the data.

Discussion

In this report we have found that, in the context of Kickstarter, entrepreneurs who become more experienced tend to become more successful in the endeavor of obtaining funding from users of the platform, though the benefits gained from the experience are not permanent. As mentioned previously, this is in accordance to the results of Parker’s article (Parker 2013) and seem to contradict the fact that experience does not affect success, as shown by Blank (Blank 2021). This study however does not definitively disprove or prove the previous results though, as it is centered around the context of Kickstarter, and thus suggests that more research must be done in the area of organizational learning to determine if more experience improves the odds of success of an entrepreneur’s project.

It was also shown that Kickstarter entrepreneurs tend towards being less innovative as they become more experienced. This might be caused by them becoming “imitative entrepreneurs” as proposed by (Cliff, Jennings, and Greenwood 2006). Therefore, this report solidifies Cliff’s findings and expands the literature into the digital entrepreneurship which Kickstarter is a part of. It is then important that entrepreneurs are able to keep this in mind, as in the search for success one might try to gain experience on what successful projects may look like which could foster less innovative ideas. So a balance between experience and innovativeness must be struck, as without innovativeness it seems intuitive to say that the market will become saturated, thus decreasing chances of success.

This contradictory nature of becoming more experienced to achieve success, while becoming less innovative, which saturates the market, creates interesting room for discussion about the balance between experience and innovativeness: how does innovation impact the probabilities for success. A study by Hult (Hult, Hurley, and Knight 2004) found that innovativeness improves performance of businesses and is thus an aspect businesses must retain. It is therefore of interest to future research to study the effect of both experience and innovativeness on success, for example: does one supersede the other?

Limitations

The study of the first research question in this project was limited by the fact that there aren't many people who back a large amount of projects and are creators of projects at the same time. This means that the data gets skewed, and not many creators make it past backing 10 other projects, meaning that it is difficult to measure indirect experience. Similarly, most creators don't make many projects, the great majority making just 1 project. This again produces biases in the data as we can't say that there are many experienced users. An attempt to mitigate this was attempted through the use of generalized linear models such as the poisson linear model and the gamma linear model as these account for right skewed distributions.

In regards to the second research question it is important to consider that an innovative project may take more time to complete as building the product may take longer, regardless of what the campaign duration is. This means that entrepreneurs that are more likely to do innovative projects will have done fewer projects in total, and on the contrary creators with a lot of previous projects are almost all not innovative. This also produces biases in the data set, and given that the Kickstarter data set doesn't contain data on how long it takes for the product to be launched after the pledge goal is reached, more information is needed for an improvement to be made.

The fact that very few creators make a large amount of projects is also a limitation for the second research question given that it looks at the number of previous successes. Therefore, if there isn't a representative enough sample of creators that made a large amount of projects the data set contains a bias. This bias affects the results of the model and must be considered when viewing the results provided by the models that were created.

Conclusion

In conclusion, this report set out to answer questions about whether or not experience impacts entrepreneurs in regards to success and to innovativeness. The results show that experience is a positive predictor of success, but a negative predictor of innovativeness in entrepreneurs in the kickstarter dataset. The results regarding success suggest that further research needs to be made in order to conclusively decide whether this is the case of entrepreneurs as a whole and not just in Kickstarter. The results regarding innovativeness suggest that research can be

made in order to see what the trade off between experience and innovativeness is, and how to maximize the success of a business by taking this into account.

References

1. Arnold, Barry C. 2015. "Pareto Distribution." In Wiley StatsRef: Statistics Reference Online, 1–10. John Wiley & Sons, Ltd. <https://doi.org/https://doi.org/10.1002/9781118445112.stat01100.pub2>.
2. Blank, Tali Hadasa. 2021. "When Incubator Resources Are Crucial: Survival Chances of Student Startups Operating in an Academic Incubator." *The Journal of Technology Transfer* 46 (6): 1845–68. <https://doi.org/10.1007/s10961-020-09831-4>.
3. Cliff, Jennifer E., P. Devereaux Jennings, and Royston Greenwood. 2006. "New to the Game and Questioning the Rules: The Experiences and Beliefs of Founders Who Start Imitative Versus Innovative Firms." *Journal of Business Venturing* 21 (5): 633–63. <https://doi.org/https://doi.org/10.1016/j.jbusvent.2005.02.010>.
4. Cyert, Richard Michael, and James G. March. 1963. *A Behavioral Theory of the Firm*. Prentice-Hall.
5. Hult, G.Tomas M., Robert F. Hurley, and Gary A. Knight. 2004. "Innovativeness: Its Antecedents and Impact on Business Performance." *Industrial Marketing Management* 33 (5): 429–38. <https://doi.org/https://doi.org/10.1016/j.indmarman.2003.08.015>.
6. Levitt, Barbara, and James G. March. 1988. "Organizational Learning." *Annual Review of Sociology* 14: 319–40. <http://www.jstor.org/stable/2083321>.
7. Marcati, Alberto, Gianluigi Guido, and Alessandro M Peluso. 2008. "The Role of SME Entrepreneurs' Innovativeness and Personality in the Adoption of Innovations." *Research Policy* 37 (9): 1579–90.
8. Martin, Frank, and Ronnie Smith. 2010. "What Is It That Entrepreneurs Learn from Experience?" *Industry and Higher Education* 24 (6): 505–12. <https://doi.org/10.5367/ihe.2010.0016>.
9. McKenny, Aaron F., Herman Aguinis, Jeremy C. Short, and Aaron H. Anglin. 2018. "What Doesn't Get Measured Does Exist: Improving the Accuracy of Computer-Aided Text Analysis." *Journal of Management* 44 (September): 2909–33. <https://doi.org/10.1177/0149206316657594>.
10. Parker, Simon C. 2013. "Do Serial Entrepreneurs Run Successively Better-Performing Businesses?" *Journal of Business Venturing* 28 (5): 652–66. <https://doi.org/https://doi.org/10.1016/j.jbusvent.2012.08.001>.

11. Vaillant, Yancy, and Esteban Lafuente. 2019. “Entrepreneurial Experience and the Innovativeness of Serial Entrepreneurs.” *Management Decision* 57 (November): 2869–89. <https://doi.org/10.1108/MD-06-2017-0592>.
- Arnold, Barry C. 2015. “Pareto Distribution.” In *Wiley StatsRef: Statistics Reference Online*, 1–10. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118445112.stat01100.pub2>.
- Blank, Tali Hadasa. 2021. “When Incubator Resources Are Crucial: Survival Chances of Student Startups Operating in an Academic Incubator.” *The Journal of Technology Transfer* 46 (6): 1845–68. <https://doi.org/10.1007/s10961-020-09831-4>.
- Cliff, Jennifer E., P. Devereaux Jennings, and Royston Greenwood. 2006. “New to the Game and Questioning the Rules: The Experiences and Beliefs of Founders Who Start Imitative Versus Innovative Firms.” *Journal of Business Venturing* 21 (5): 633–63. <https://doi.org/10.1016/j.jbusvent.2005.02.010>.
- Cyert, Richard Michael, and James G. March. 1963. *A Behavioral Theory of the Firm*. Prentice-Hall.
- Hult, G.Tomas M., Robert F. Hurley, and Gary A. Knight. 2004. “Innovativeness: Its Antecedents and Impact on Business Performance.” *Industrial Marketing Management* 33 (5): 429–38. <https://doi.org/10.1016/j.indmarman.2003.08.015>.
- Levitt, Barbara, and James G. March. 1988. “Organizational Learning.” *Annual Review of Sociology* 14: 319–40. <http://www.jstor.org/stable/2083321>.
- Marcati, Alberto, Gianluigi Guido, and Alessandro M Peluso. 2008. “The Role of SME Entrepreneurs’ Innovativeness and Personality in the Adoption of Innovations.” *Research Policy* 37 (9): 1579–90.
- Martin, Frank, and Ronnie Smith. 2010. “What Is It That Entrepreneurs Learn from Experience?” *Industry and Higher Education* 24 (6): 505–12. <https://doi.org/10.5367/ihe.2010.0016>.
- McKenny, Aaron F., Herman Aguinis, Jeremy C. Short, and Aaron H. Anglin. 2018. “What Doesn’t Get Measured Does Exist: Improving the Accuracy of Computer-Aided Text Analysis.” *Journal of Management* 44 (September): 2909–33. <https://doi.org/10.1177/0149206316657594>.
- Parker, Simon C. 2013. “Do Serial Entrepreneurs Run Successively Better-Performing Businesses?” *Journal of Business Venturing* 28 (5): 652–66. <https://doi.org/10.1016/j.jbusvent.2012.08.001>.
- Vaillant, Yancy, and Esteban Lafuente. 2019. “Entrepreneurial Experience and the Innovativeness of Serial Entrepreneurs.” *Management Decision* 57 (November): 2869–89. <https://doi.org/10.1108/MD-06-2017-0592>.

Appendix

The github can be accessed through the following link: https://github.com/Bero898/SBM_assignment.git