R Notebook

Code

Este conjunto de datos ha sido extraido de Kaggle:
https://www.kaggle.com/namanmanchanda/entrepreneurial-competency-in-university-students
(https://www.kaggle.com/namanmanchanda/entrepreneurial-competency-in-university-students)

Hide

```
df = read.csv("data.csv")

df
```

cargamos las librerias

Hide

```
library(ggplot2)
library(summarytools)
```

```
Registered S3 method overwritten by 'pryr':
  method      from
  print.bytes Rcpp
```

Hide

```
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
```

Hide

```
library(plotrix)
library(tidyverse)
```

```
Registered S3 methods overwritten by 'dbplyr':
  method         from
  print.tbl_lazy
  print.tbl_sql
-- Attaching packages --------------------------------- tidyverse 1.3.1 --
v tibble  3.1.1     v purrr   0.3.4
v tidyr   1.1.3     v stringr 1.4.0
v readr   1.4.0     v forcats 0.5.1
-- Conflicts ------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
x tibble::view()  masks summarytools::view()
```

Hide

```
library(tokenizers)
```

Disponemos de un conjunto de datos sobre estudiantes universitarios, que han sido analizados, en base a ciertas caracteristicas o aptitudes, con el objetivo de predecir la probabilidad de que acaben siendo emprendedores o no.

El estudio ha establecido una variable objetivo denominada “whether the student is likely to become an entrepreneur or not” (probabilidad de que el estudiante se convierta en empresario o no) formada por dos elementos, 0 - No llega a serlo / 1 - Si llega a serlo.

El problema que intentaremos resolver es determinar, que caracterisiticas, segun este estudio, son determinantes para que un estudiante universitario se convierta en empresario.

Variables que contiene el dataset:

1 - EducationSector: Tipo de carrera que está cursando el estudiante

2 - IndividualProject: ¿Estaria dispuesto a llevar a cabo un proyecto individual? SI/NO

3 - Age: Edad del encuestado

4 - Gender: Genero del encuestado

5 - City: ¿Vive en una ciudad? SI/NO

6 - Influenced: ¿Le ha influido alguien a la hora de elegir la carrera que está cursando? SI/NO

7 - Perseverance: Nivel de perseverancia o constancia que tiene, del 1 al 5

8 - DesireToTakeInitiative: Nivel de deseo de tomar la iniciativa, del 1 al 5

9 - Competitiveness: Nivel de competitividad, del 1 al 5

10 - SelfReliance: Nivel de autosuficiencia, del 1 al 5

11 - StrongNeedToAchieve: Nivel de necesidad de lograr una meta, del 1 al 5

12 - SelfConfidence: Nivel de autoconfianza, del 1 al 5

13 - GoodPhysicalHealth: Nivel de salud fisica, del 1 al 5

14 - MentalDisorder: ¿Tiene algun trastorno mental? SI/NO

15 - KeyTraits: Cuales son los rasgos claves del encuestado

16 - ReasonsForLack: Cual seria el motivo de la falta de interes en ser emprendedor.

17 - y: Se trata de la variable dependiente que recoge la prediccion de si el estudiante se convertirá en empresario 1, o no se convertirá 0.

Clasifiacion de las variables en funcion de su escala de medida:

- Todas estas son variables independientes.

1 - Variables categoricas nominales: EducationSector, IndividualProject, Gender, City, Influenced, MentalDisorder, KeyTraits, ReasonsForLack

2 - Variables categoricas ordinales: Perseverance, DesireToTakeInitiative, Competitiveness, SelfReliance, StrongNeedToAchieve, GoodPhysicalHealth, SelfConfidence

3 - Variables cuantitativas discreta: Age

- Variable dependiente.

4 - Variable categorica binomial.

Preguntas:

¿Que caracterisitica o caracteristicas son determinantes a la hora de emprender?

¿El tipo de carrera que se esté cursando influye a la hora de ser emprendor?

Analisis descriptivo

Hide

```
summary(df)
```

```
 EducationSector    IndividualProject       Age           Gender
 Length:219         Length:219         Min.   :17.00   Length:219
 Class :character   Class :character   1st Qu.:19.00   Class :character
 Mode  :character   Mode  :character   Median :20.00   Mode  :character
                                       Mean   :19.75
                                       3rd Qu.:20.00
                                       Max.   :26.00
     City            Influenced         Perseverance
 Length:219         Length:219         Min.   :1.000
 Class :character   Class :character   1st Qu.:3.000
 Mode  :character   Mode  :character   Median :3.000
                                       Mean   :3.352
                                       3rd Qu.:4.000
                                       Max.   :5.000
 DesireToTakeInitiative Competitiveness  SelfReliance   StrongNeedToAchieve
 Min.   :1.000          Min.   :1.000   Min.   :1.000   Min.   :1.000
 1st Qu.:3.000          1st Qu.:3.000   1st Qu.:3.000   1st Qu.:3.000
 Median :4.000          Median :4.000   Median :4.000   Median :4.000
 Mean   :3.621          Mean   :3.589   Mean   :3.721   Mean   :3.909
 3rd Qu.:5.000          3rd Qu.:4.500   3rd Qu.:5.000   3rd Qu.:5.000
 Max.   :5.000          Max.   :5.000   Max.   :5.000   Max.   :5.000
 SelfConfidence  GoodPhysicalHealth MentalDisorder      KeyTraits
 Min.   :1.000   Min.   :1.000      Length:219         Length:219
 1st Qu.:3.000   1st Qu.:3.000      Class :character   Class :character
 Median :4.000   Median :4.000      Mode  :character   Mode  :character
 Mean   :3.575   Mean   :3.562
 3rd Qu.:4.000   3rd Qu.:4.000
 Max.   :5.000   Max.   :5.000
 ReasonsForLack           y
 Length:219         Min.   :0.0000
 Class :character   1st Qu.:0.0000
 Mode  :character   Median :0.0000
                    Mean   :0.4155
                    3rd Qu.:1.0000
                    Max.   :1.0000
```

Hide

```
str(df)
```

```
'data.frame':   219 obs. of  17 variables:
 $ EducationSector       : chr  "Engineering Sciences" "Engineering Sciences" "Engineering Sc
iences" "Engineering Sciences" ...
 $ IndividualProject     : chr  "No" "Yes" "No" "Yes" ...
 $ Age                   : int  19 22 18 20 19 19 19 20 20 17 ...
 $ Gender                : chr  "Male" "Male" "Male" "Male" ...
 $ City                  : chr  "Yes" "No" "Yes" "Yes" ...
 $ Influenced            : chr  "No" "Yes" "No" "Yes" ...
 $ Perseverance          : int  2 3 3 3 2 3 3 4 2 2 ...
 $ DesireToTakeInitiative: int  2 3 4 3 3 3 2 2 3 3 ...
 $ Competitiveness       : int  3 3 3 3 3 3 3 4 3 2 ...
 $ SelfReliance          : int  3 4 3 3 3 3 3 4 1 1 ...
 $ StrongNeedToAchieve   : int  2 4 3 4 4 3 4 5 2 4 ...
 $ SelfConfidence        : int  2 3 4 3 3 3 1 3 2 3 ...
 $ GoodPhysicalHealth    : int  3 4 4 3 2 3 1 4 2 3 ...
 $ MentalDisorder        : chr  "Yes" "Yes" "No" "No" ...
 $ KeyTraits             : chr  "Passion" "Vision" "Passion" "Resilience" ...
 $ ReasonsForLack        : chr  "" "Just not interested! (Want to work in the corporate secto
r, or for the government or pursue research or something else)" "Not willing to start a ventu
re in India and waiting for future relocation" "Not able to take a Financial Risk" ...
 $ y                     : int  1 0 0 0 1 1 1 0 0 1 ...
```

Disponemos de 219 filas y 17 variables

Hide

```
apply(is.na(df), 2, sum)
```

```
      EducationSector      IndividualProject                    Age
                    0                      0                      0
               Gender                   City             Influenced
                    0                      0                      0
         Perseverance DesireToTakeInitiative        Competitiveness
                    0                      0                      0
         SelfReliance    StrongNeedToAchieve         SelfConfidence
                    0                      0                      0
    GoodPhysicalHealth         MentalDisorder              KeyTraits
                    0                      0                      0
        ReasonsForLack                      y
                    0                      0
```

Hide

```
df$ReasonsForLack = vector(df$ReasonsForLack)
```

```
Error in vector(df$ReasonsForLack) : argumento 'mode' inválido
```

Comenzamos el analisis de las variables:

y

La primera variable que vamos a analizar es la variable dependiente

Hide

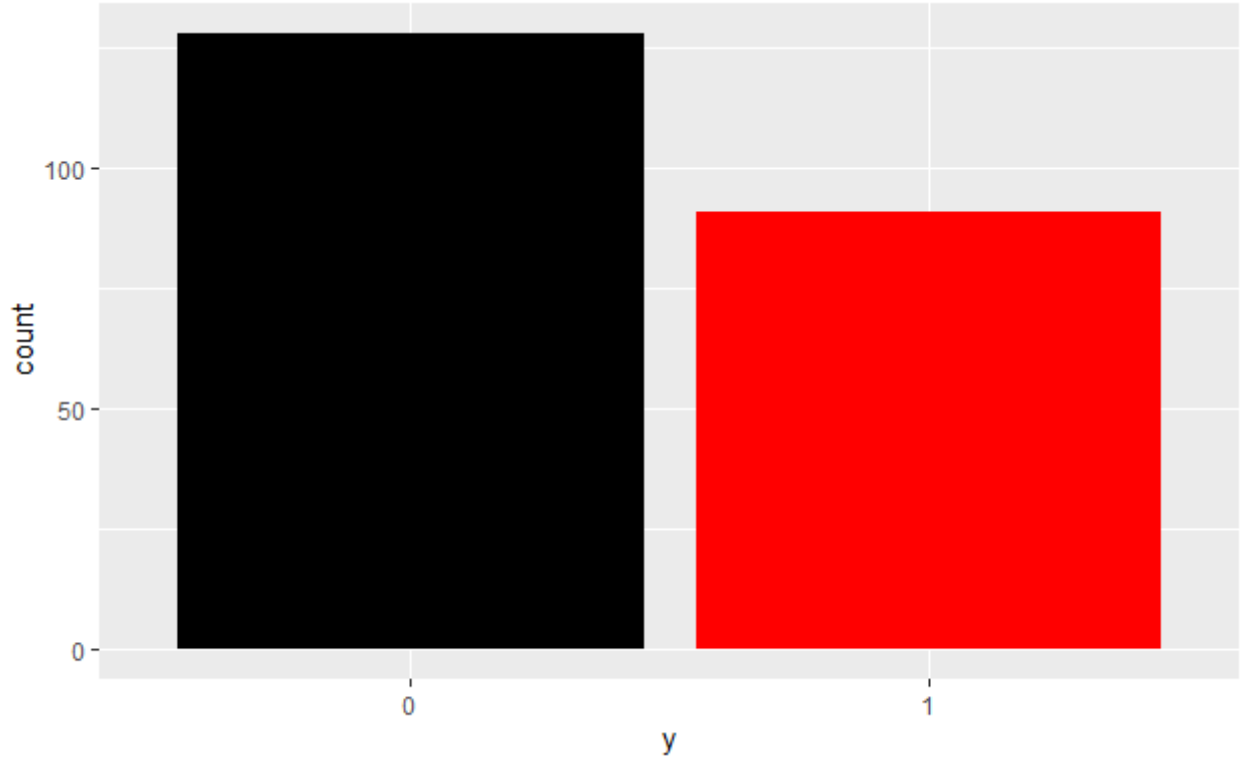
```
freq(df$y, style= "rmarkdown")
```

```
### Frequencies
#### df$y
**Type:** Factor

|       | Freq | % Valid | % Valid Cum. | % Total | % Total Cum. |
|-----------:|-----:|--------:|-------------:|--------:|-------------:|
|      **0** |  128 |   58.45 |        58.45 |   58.45 |        58.45 |
|      **1** |   91 |   41.55 |       100.00 |   41.55 |       100.00 |
| **\<NA\>** |    0 |         |              |    0.00 |       100.00 |
|  **Total** |  219 |  100.00 |       100.00 |  100.00 |       100.00 |
```

Hide

```
ggplot(df, aes(y)) + geom_bar(fill = c("black", "red"))
```



Vamos a comenzar a analizar las variables dependientes, tanto de forma individual como relacionadas con la variable dependiente

EducationSector

Hide

```
freq(df$EducationSector, style = "rmarkdown")
```

```
### Frequencies
#### df$EducationSector
**Type:** Factor

|                                                            | Freq | % Valid | % Valid Cum. |
% Total | % Total Cum. |
|----------------------------------------------------------------:|-----:|--------:|-------------:|
--------:|-------------:|
|                                  **Art, Music or Design** |   21 |    9.59 |         9.59 |
9.59 |         9.59 |
| **Economic Sciences, Business Studies, Commerce and Law** |   32 |   14.61 |        24.20 |
14.61 |        24.20 |
|                                  **Engineering Sciences** |  123 |   56.16 |        80.37 |
56.16 |        80.37 |
|                          **Humanities and Social Sciences** |    5 |    2.28 |        82.65 |
2.28 |        82.65 |
|                           **Language and Cultural Studies** |    1 |    0.46 |        83.11 |
0.46 |        83.11 |
|                         **Mathematics or Natural Sciences** |    4 |    1.83 |        84.93 |
1.83 |        84.93 |
|                              **Medicine, Health Sciences** |   10 |    4.57 |        89.50 |
4.57 |        89.50 |
|                                              **Others** |   20 |    9.13 |        98.63 |
9.13 |        98.63 |
|                            **Teaching Degree (e.g., B.Ed)** |    3 |    1.37 |       100.00 |
1.37 |       100.00 |
|                                              **\<NA\>** |    0 |         |              |
0.00 |       100.00 |
|                                               **Total** |  219 |  100.00 |       100.00 |
100.00 |       100.00 |
```

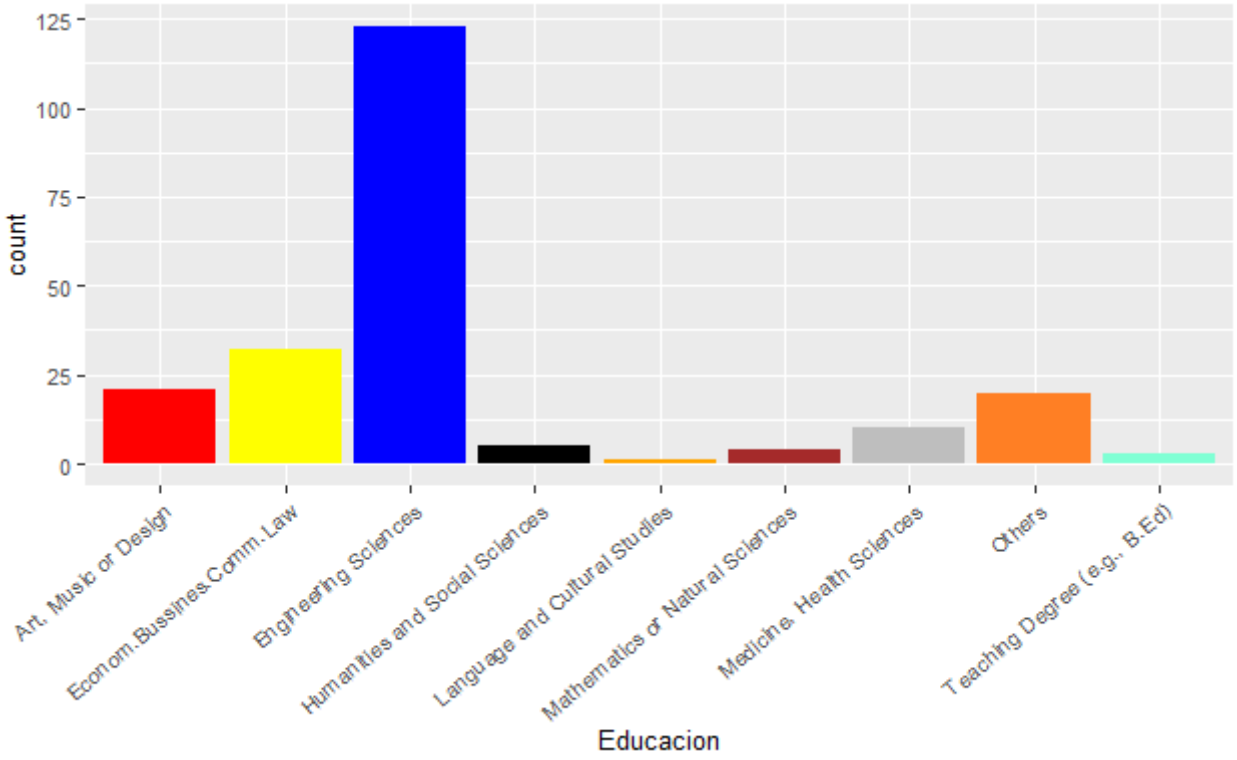
Hide

```
# El valor "Economic Sciences, Business Studies, Commerce and Law", "Econom.Bussines.Comm.Law
"de la variable EducationSector, es muy grande y causa algun problema, por lo que vamos a cre
ar una nueva variable pero modificando ese  valor.

df$EducationSector = as.character(df$EducationSector)

df$Educacion = replace(df$EducationSector, df$EducationSector == "Economic Sciences, Business
Studies, Commerce and Law", "Econom.Bussines.Comm.Law")

df$EducationSector = factor(df$EducationSector)
```



¿Como se asocian las carreras universitarias y la probabilidad de ser empresario?

Hide

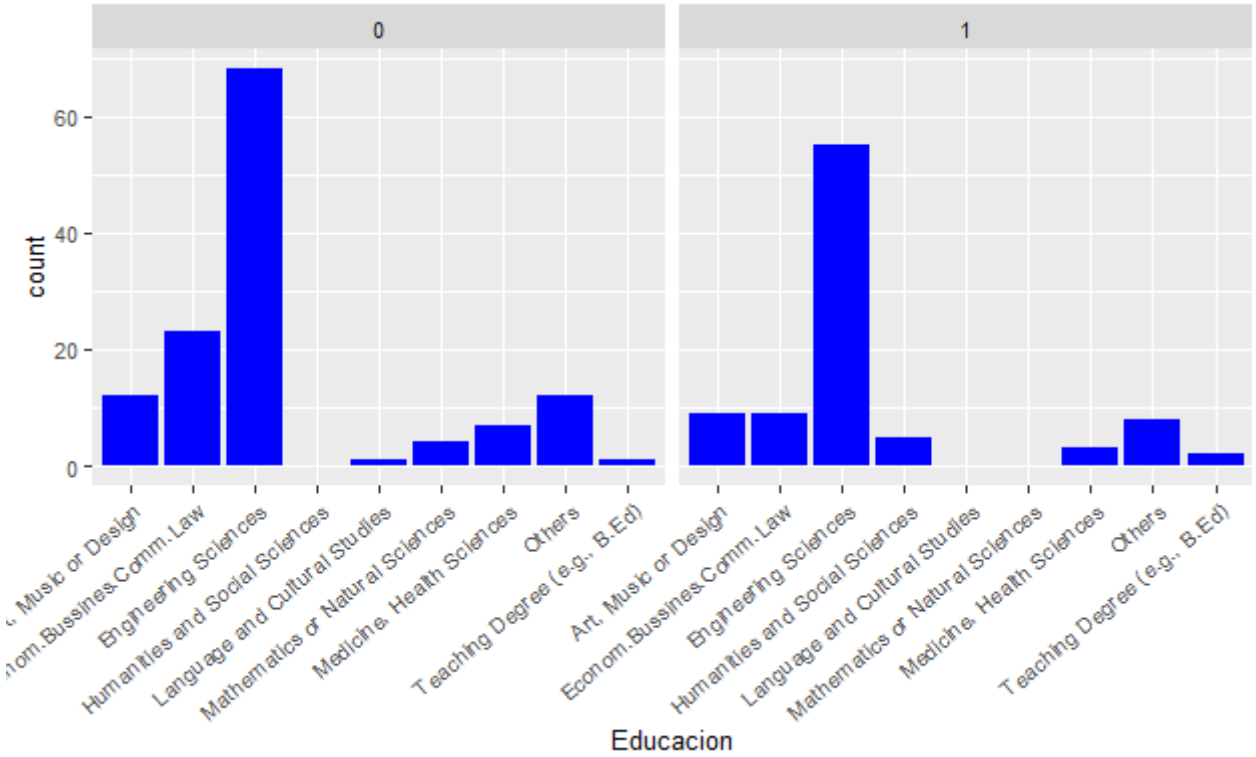
```
ctable(df$Educacion, df$y)
```

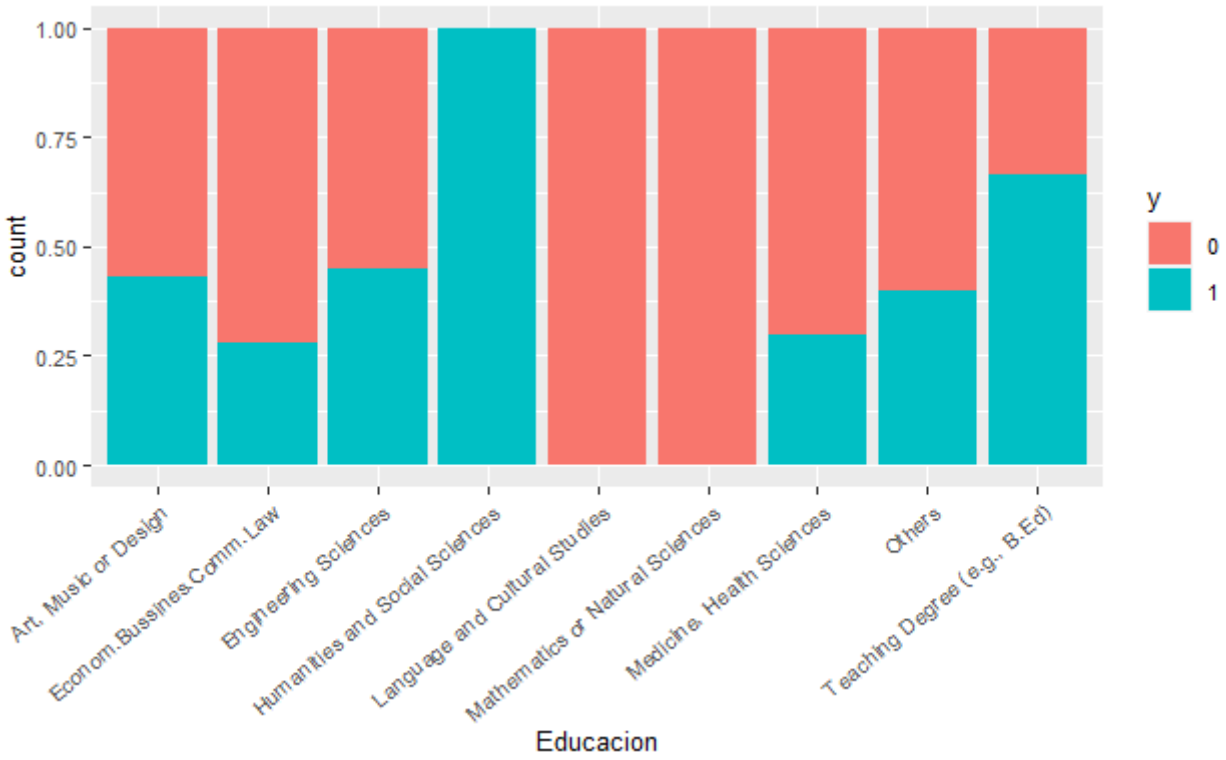
```
Cross-Tabulation, Row Proportions
Educacion * y
Data Frame: df

-------------------------------- --- -------------- ------------- --------------
                                  y              0             1          Total
                        Educacion
            Art, Music or Design         12 ( 57.1%)    9 ( 42.9%)    21 (100.0%)
        Econom.Bussines.Comm.Law         23 ( 71.9%)    9 ( 28.1%)    32 (100.0%)
            Engineering Sciences         68 ( 55.3%)   55 ( 44.7%)   123 (100.0%)
   Humanities and Social Sciences          0 (  0.0%)    5 (100.0%)     5 (100.0%)
    Language and Cultural Studies          1 (100.0%)    0 (  0.0%)     1 (100.0%)
  Mathematics or Natural Sciences          4 (100.0%)    0 (  0.0%)     4 (100.0%)
       Medicine, Health Sciences           7 ( 70.0%)    3 ( 30.0%)    10 (100.0%)
                          Others          12 ( 60.0%)    8 ( 40.0%)    20 (100.0%)
    Teaching Degree (e.g., B.Ed)          1 ( 33.3%)    2 ( 66.7%)     3 (100.0%)
                           Total        128 ( 58.4%)   91 ( 41.6%)   219 (100.0%)
-------------------------------- --- -------------- ------------- --------------
```

Hide

```
df %>%
group_by(y) %>%
ggplot(aes( x = Educacion)) +
geom_bar(fill = "blue") +
facet_wrap(~y) + theme(text = element_text(size=10),axis.text.x = element_text(angle=40, hjus
t=1))
```





IndividualProject

Vamos a ver primero cuantos han contestado Si/No a esta pregunta

Hide

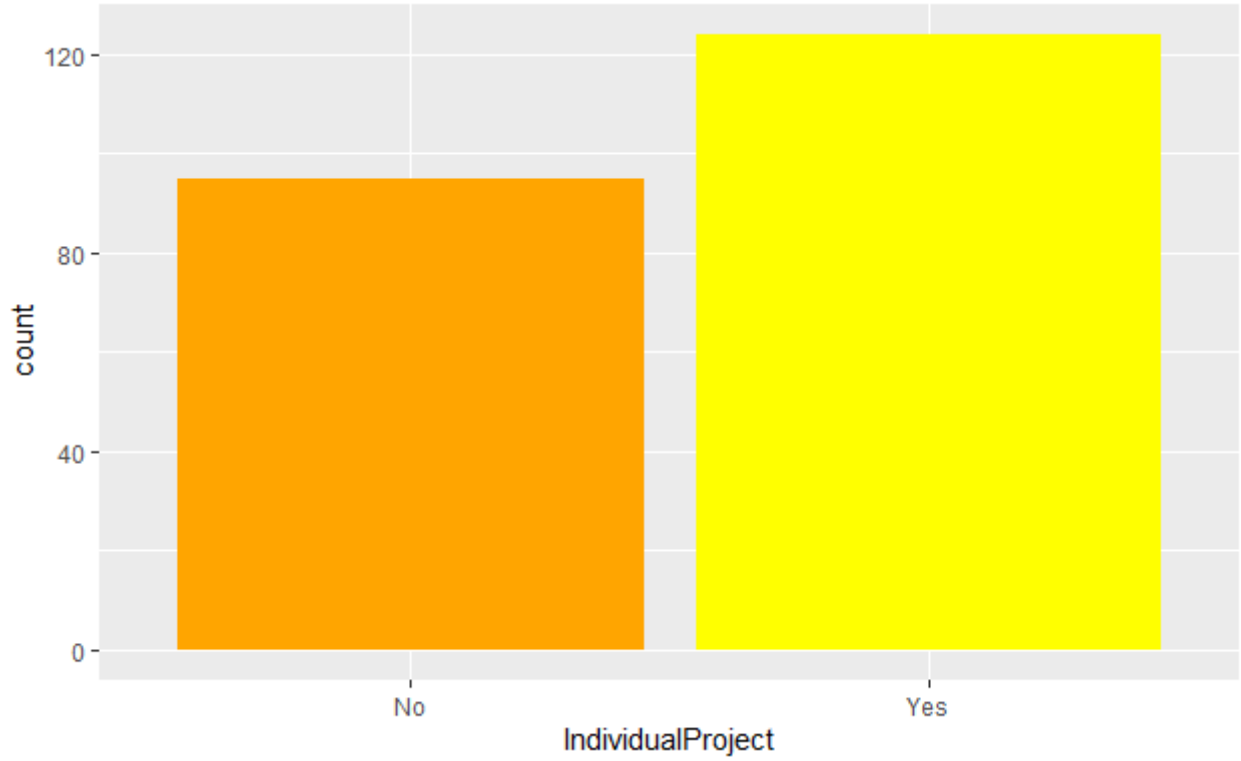
```
freq(df$IndividualProject, style = "rmarkdown")
```

```
### Frequencies
#### df$IndividualProject
**Type:** Factor

|       | Freq | % Valid | % Valid Cum. | % Total | % Total Cum. |
|-----------:|-----:|--------:|-------------:|--------:|-------------:|
|     **No** |   95 |   43.38 |        43.38 |   43.38 |        43.38 |
|    **Yes** |  124 |   56.62 |       100.00 |   56.62 |       100.00 |
| **\<NA\>** |    0 |         |              |    0.00 |       100.00 |
|  **Total** |  219 |  100.00 |       100.00 |  100.00 |       100.00 |
```

Hide

```
ggplot(df, aes(IndividualProject)) + geom_bar(fill = c("orange", "yellow"))
```



Mas de la mitad han respondido que si iniciarian un proyecto individual.

Vamos a ver como se relaciona esta pregunta con la variable dependiente

Hide

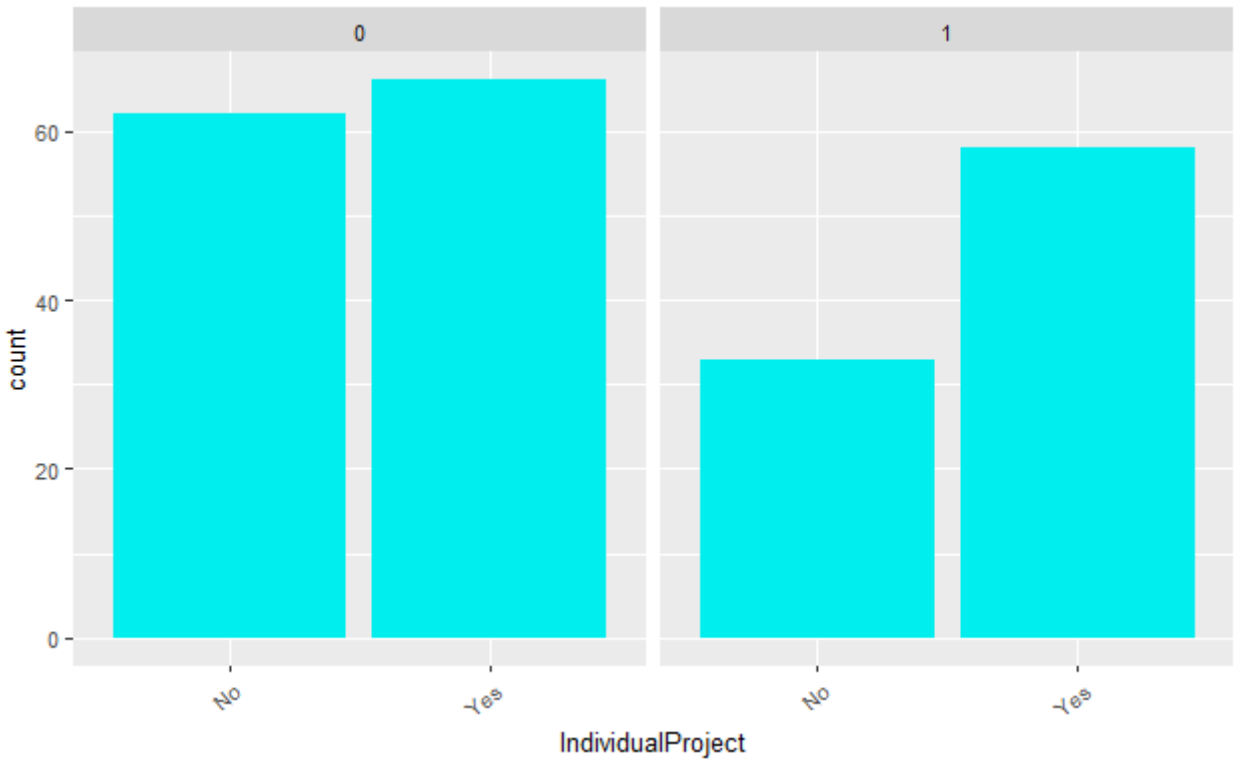
```
ctable(df$IndividualProject, df$y)
```

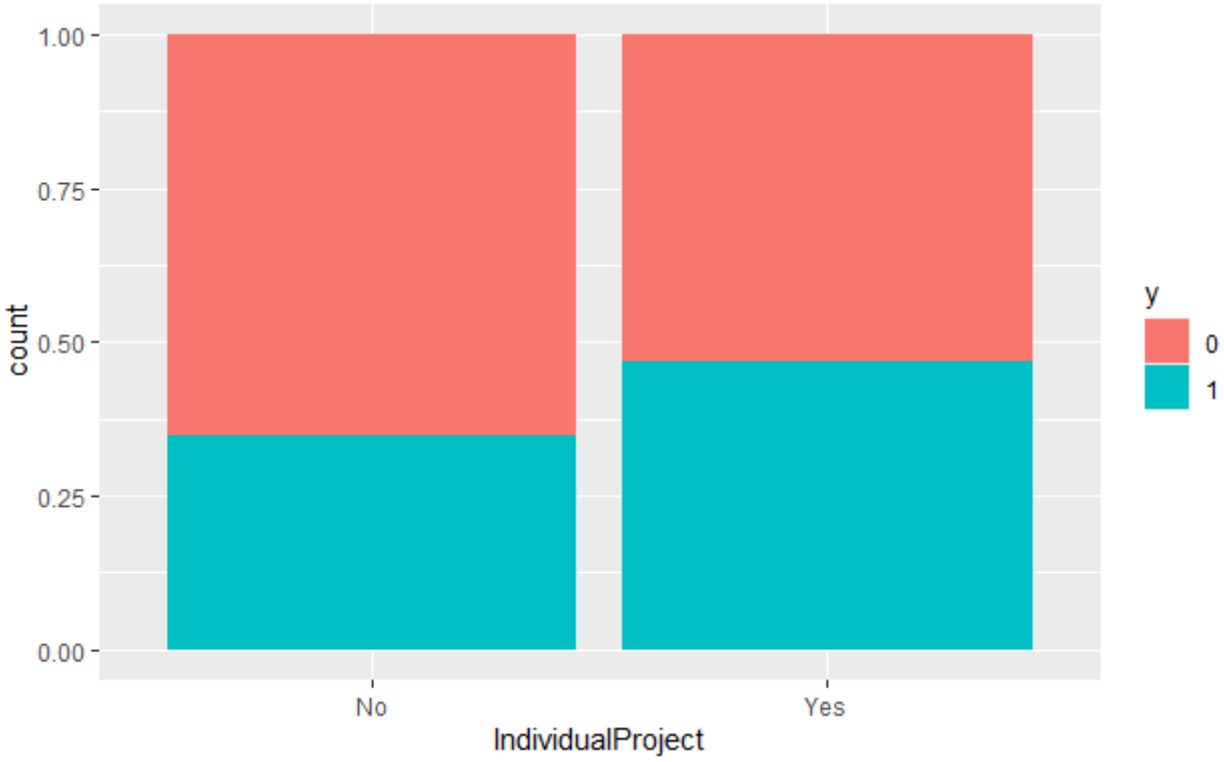
```
Cross-Tabulation, Row Proportions
IndividualProject * y
Data Frame: df

------------------- --- ------------ ------------ --------------
                      y            0            1          Total
  IndividualProject
                 No       62 (65.3%)   33 (34.7%)    95 (100.0%)
                Yes       66 (53.2%)   58 (46.8%)   124 (100.0%)
              Total      128 (58.4%)   91 (41.6%)   219 (100.0%)
------------------- --- ------------ ------------ --------------
```

Hide

```
df %>%
group_by(y) %>%
ggplot(aes( x = IndividualProject)) +
geom_bar(fill = "cyan2") +
facet_wrap(~y) + theme(text = element_text(size=10),axis.text.x = element_text(angle=40, hjus
t=1))
```





Age

En primer lugar, vamos a ver como se distribuyen las edades de los que han participado en esta encuensta

Hide

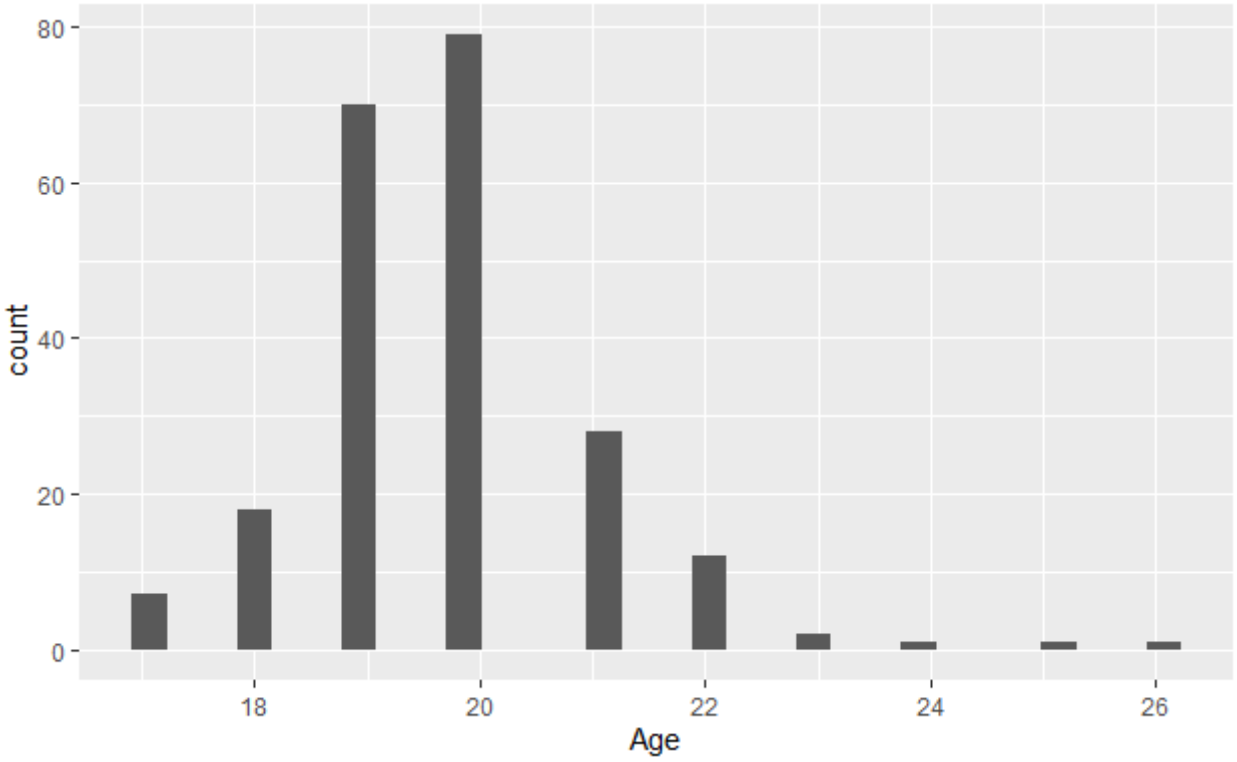
```
freq(df$Age, style = "rmarkdown")
```

```
### Frequencies
#### df$Age
**Type:** Integer

|       | Freq | % Valid | % Valid Cum. | % Total | % Total Cum. |
|-----------:|-----:|--------:|-------------:|--------:|-------------:|
|     **17** |    7 |    3.20 |         3.20 |    3.20 |         3.20 |
|     **18** |   18 |    8.22 |        11.42 |    8.22 |        11.42 |
|     **19** |   70 |   31.96 |        43.38 |   31.96 |        43.38 |
|     **20** |   79 |   36.07 |        79.45 |   36.07 |        79.45 |
|     **21** |   28 |   12.79 |        92.24 |   12.79 |        92.24 |
|     **22** |   12 |    5.48 |        97.72 |    5.48 |        97.72 |
|     **23** |    2 |    0.91 |        98.63 |    0.91 |        98.63 |
|     **24** |    1 |    0.46 |        99.09 |    0.46 |        99.09 |
|     **25** |    1 |    0.46 |        99.54 |    0.46 |        99.54 |
|     **26** |    1 |    0.46 |       100.00 |    0.46 |       100.00 |
| **\<NA\>** |    0 |         |              |    0.00 |       100.00 |
|  **Total** |  219 |  100.00 |       100.00 |  100.00 |       100.00 |
```

Hide

```
ggplot(df, aes(Age)) + geom_histogram()
```



¿Es la edad un factor determinante para determinar la probabilidad de un universitario de llegar a ser emprendor? Vamos a verlo

Hide

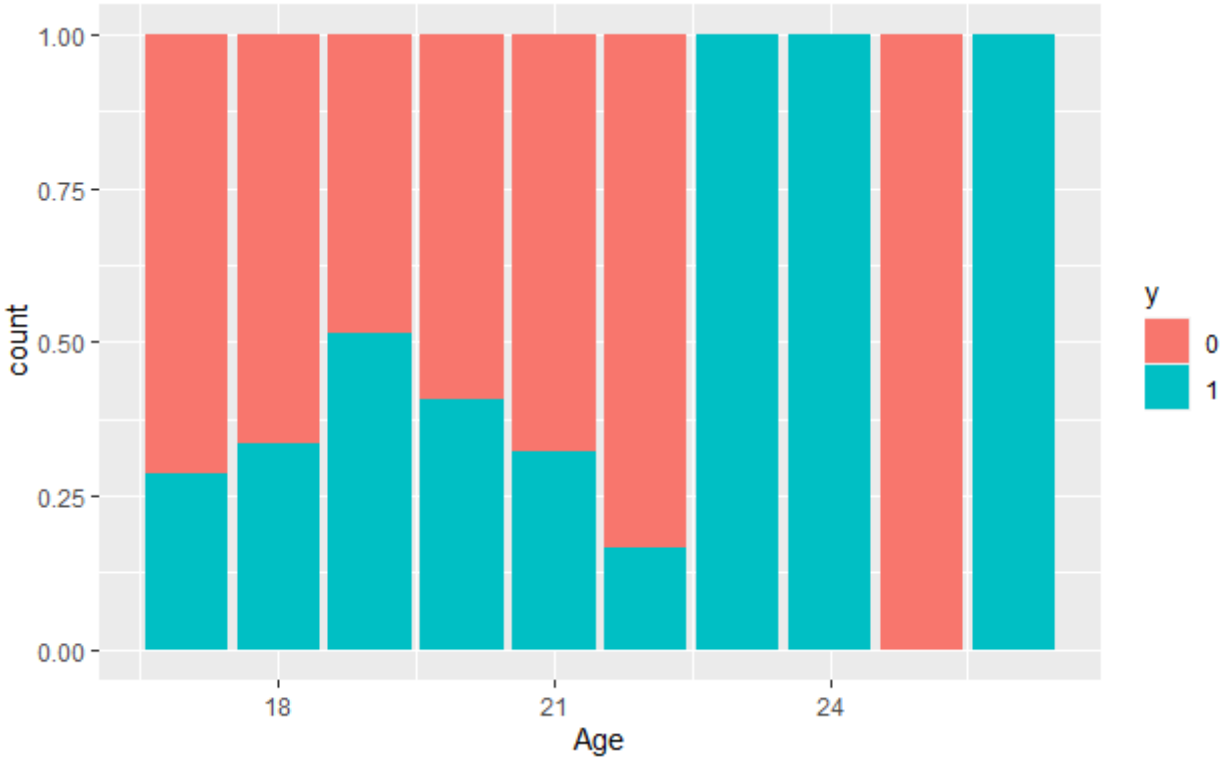
```
ctable(df$Age, df$y)
```

```
Cross-Tabulation, Row Proportions
Age * y
Data Frame: df

------- --- -------------- ------------- --------------
          y              0             1          Total
    Age
     17         5 ( 71.4%)    2 ( 28.6%)     7 (100.0%)
     18        12 ( 66.7%)    6 ( 33.3%)    18 (100.0%)
     19        34 ( 48.6%)   36 ( 51.4%)    70 (100.0%)
     20        47 ( 59.5%)   32 ( 40.5%)    79 (100.0%)
     21        19 ( 67.9%)    9 ( 32.1%)    28 (100.0%)
     22        10 ( 83.3%)    2 ( 16.7%)    12 (100.0%)
     23         0 (  0.0%)    2 (100.0%)     2 (100.0%)
     24         0 (  0.0%)    1 (100.0%)     1 (100.0%)
     25         1 (100.0%)    0 (  0.0%)     1 (100.0%)
     26         0 (  0.0%)    1 (100.0%)     1 (100.0%)
  Total       128 ( 58.4%)   91 ( 41.6%)   219 (100.0%)
------- --- -------------- ------------- --------------
```

Hide

```
ggplot(df, aes(x = Age, fill = y)) + geom_bar(position = "fill")
```



Hide

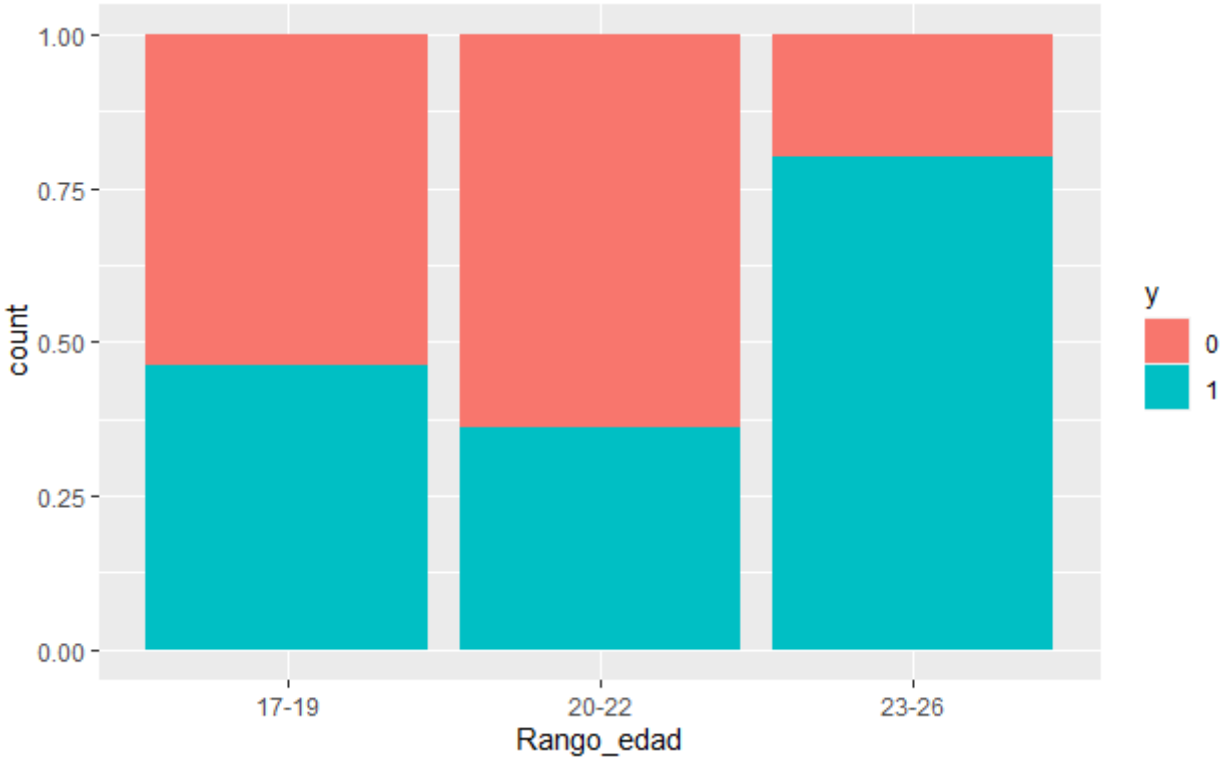
```
ctable(df$Rango_edad, df$y)
```

```
Cross-Tabulation, Row Proportions
Rango_edad * y
Data Frame: df

------------ --- ------------- ------------ --------------
               y             0            1          Total
  Rango_edad
       17-19        51 (53.7%)   44 (46.3%)    95 (100.0%)
       20-22        76 (63.9%)   43 (36.1%)   119 (100.0%)
       23-26         1 (20.0%)    4 (80.0%)     5 (100.0%)
       Total       128 (58.4%)   91 (41.6%)   219 (100.0%)
------------ --- ------------- ------------ --------------
```

Hide

```
ggplot(df, aes(x = Rango_edad, fill = y)) + geom_bar(position = "fill")
```



Como podemos ver, son los universitarios de mayor edad, los que mas probabilidades tienen de ser emprendedores, no obstante, solo se han encuestado a 5 personas en esta franja de edad, en contraposicion a los 95 y 119 de las otras dos franjas, aqui tenemos un problema de sesgo, seria necesario ampliar el numero de encuestados en la franaja de 23 a 26.

Gender

Vamos a estudiar ahora el genero, en primer lugar, vamos a ver cuantos han sido encuestados de cada sexo.

Hide

```
freq(df$Gender, style = "rmarkdown")
```

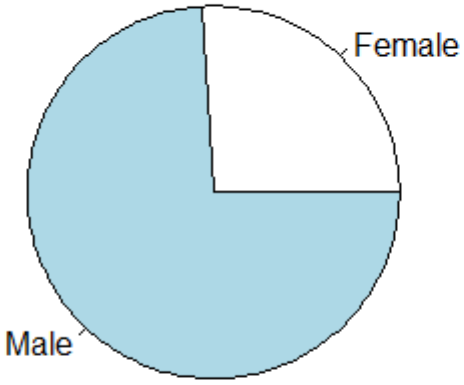
```
### Frequencies
#### df$Gender
**Type:** Factor

|       | Freq | % Valid | % Valid Cum. | % Total | % Total Cum. |
|-----------:|-----:|--------:|-------------:|--------:|-------------:|
| **Female** |   57 |   26.03 |        26.03 |   26.03 |        26.03 |
|   **Male** |  162 |   73.97 |       100.00 |   73.97 |       100.00 |
| **\<NA\>** |    0 |         |              |    0.00 |       100.00 |
|  **Total** |  219 |  100.00 |       100.00 |  100.00 |       100.00 |
```

Hide

```
tabla = table(df$Gender)

pie(tabla)
```



¿Tienen mayor probabilidad de ser emprendedor los hombres universitarios o las mujeres universitarias?

Hide

```
ctable(df$Gender, df$y)
```

```
Cross-Tabulation, Row Proportions
Gender * y
Data Frame: df

-------- --- ------------ ------------ --------------
           y            0            1          Total
  Gender
  Female        37 (64.9%)   20 (35.1%)    57 (100.0%)
    Male        91 (56.2%)   71 (43.8%)   162 (100.0%)
   Total       128 (58.4%)   91 (41.6%)   219 (100.0%)
-------- --- ------------ ------------ --------------
```

Hide

```
ggplot(df, aes(x = Gender, fill = y)) + geom_bar(position = "fill")
```

Como puede observarse en el grafico los hombres universitarios tienen mayor probabilidad de ser empresarios que las mujeres, no obstante, también podemos apreciar un pequeño problema de sesgo, en esta variable, aunque menor que en el caso anterior.

City

Vamos a ver cuantos de los universitarios encuestados viven en ciudad y cuantos no.

Hide

```
freq(df$City, style = "rmarkdown")
```

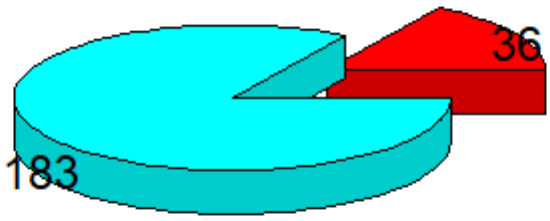
```
### Frequencies
#### df$City
**Type:** Factor

|       | Freq | % Valid | % Valid Cum. | % Total | % Total Cum. |
|-----------:|-----:|--------:|-------------:|--------:|-------------:|
|     **No** |   36 |   16.44 |        16.44 |   16.44 |        16.44 |
|    **Yes** |  183 |   83.56 |       100.00 |   83.56 |       100.00 |
| **\<NA\>** |    0 |         |              |    0.00 |       100.00 |
|  **Total** |  219 |  100.00 |       100.00 |  100.00 |       100.00 |
```

Hide

```
tabla2 = table(df$City)

pie3D(tabla2, labels = tabla2, explode = 0.25)
```



¿Es el lugar de residencia determinante a la hora de que un universitario llegue a ser emprendedor?

Hide

```
ctable(df$City, df$y)
```

```
Cross-Tabulation, Row Proportions
City * y
Data Frame: df

------- --- ------------ ------------ --------------
          y            0            1          Total
   City
     No        24 (66.7%)   12 (33.3%)    36 (100.0%)
    Yes       104 (56.8%)   79 (43.2%)   183 (100.0%)
  Total       128 (58.4%)   91 (41.6%)   219 (100.0%)
------- --- ------------ ------------ --------------
```

Hide

```
ggplot(df, aes(x = City, fill = y)) + geom_bar(position = "fill")
```

Parece que los universitarios que residen en ciudad tienen mas probabilidad de ser empresarios que los que viven en pueblos (43% - 33%)

Influenced

De todos los universitarios encuestados, ¿cuantos han elegido su carrera de forma voluntaria y cuantos han sido influenciados para que la elijan ?

Hide

```
freq(df$Influenced, style = "rmarkdown")
```

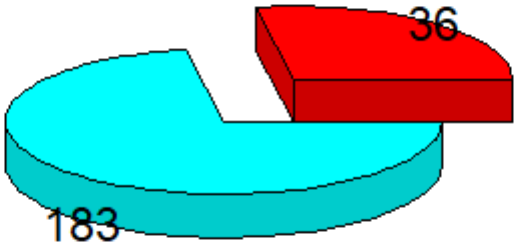
```
### Frequencies
#### df$Influenced
**Type:** Factor

|       | Freq | % Valid | % Valid Cum. | % Total | % Total Cum. |
|-----------:|-----:|--------:|-------------:|--------:|-------------:|
|     **No** |   61 |   27.85 |        27.85 |   27.85 |        27.85 |
|    **Yes** |  158 |   72.15 |       100.00 |   72.15 |       100.00 |
| **\<NA\>** |    0 |         |              |    0.00 |       100.00 |
|  **Total** |  219 |  100.00 |       100.00 |  100.00 |       100.00 |
```

Hide

```
tabla3 = table(df$Influenced)

pie3D(tabla3, labels = tabla2, explode = 0.25)
```



¿En cuanto afecta a la probabilida de ser empresario que la carrera elegida haya sido por influencia y no por propia decision ?

Hide

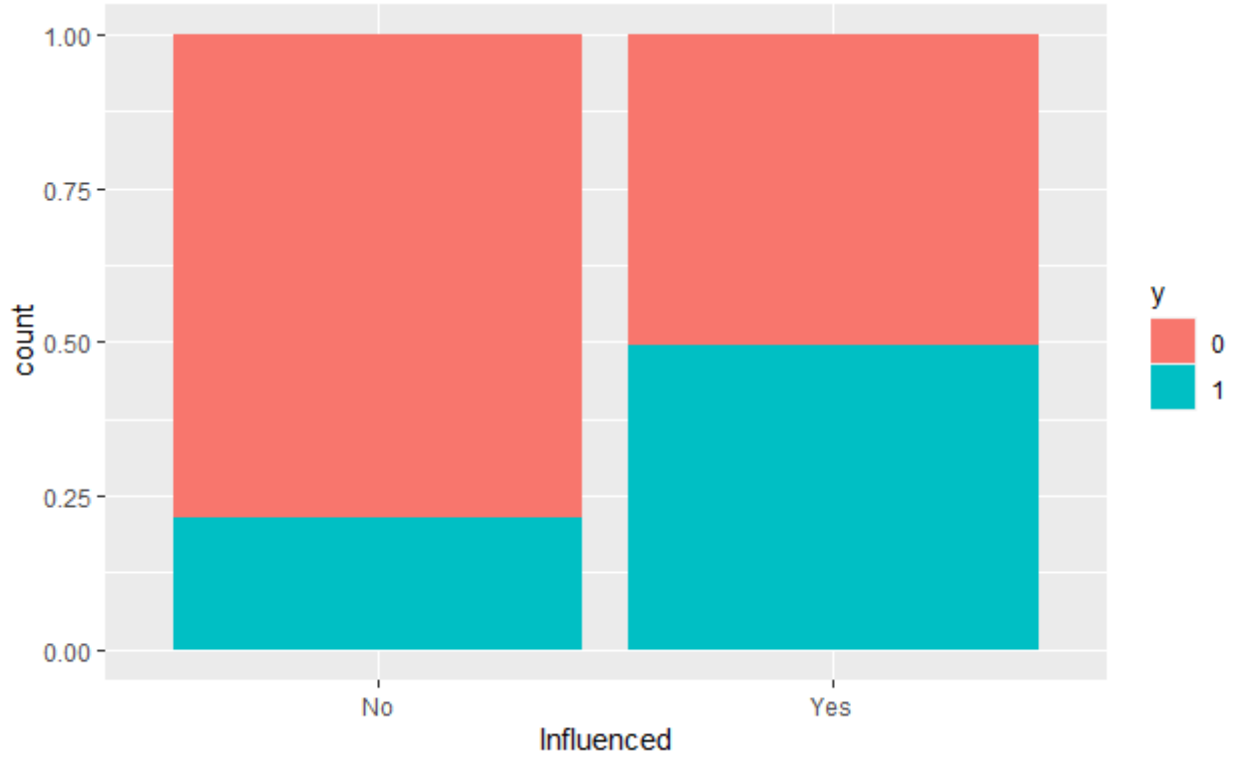
```
ctable(df$Influenced, df$y)
```

```
Cross-Tabulation, Row Proportions
Influenced * y
Data Frame: df

------------ --- ------------- ------------ --------------
               y             0            1          Total
  Influenced
          No        48 (78.7%)   13 (21.3%)    61 (100.0%)
         Yes        80 (50.6%)   78 (49.4%)   158 (100.0%)
       Total       128 (58.4%)   91 (41.6%)   219 (100.0%)
------------ --- ------------- ------------ --------------
```

Hide

```
ggplot(df, aes(x = Influenced, fill = y)) + geom_bar(position = "fill")
```



Por lo que se puede observar, los universitarios que son influenciados tienen mayor probabilidad (49.5%) de ser empresarios que los que no (21.3%).

Perseverance

¿Es la perseverancia un factor a tener en cuenta a la hora de ser empresario?

Hide

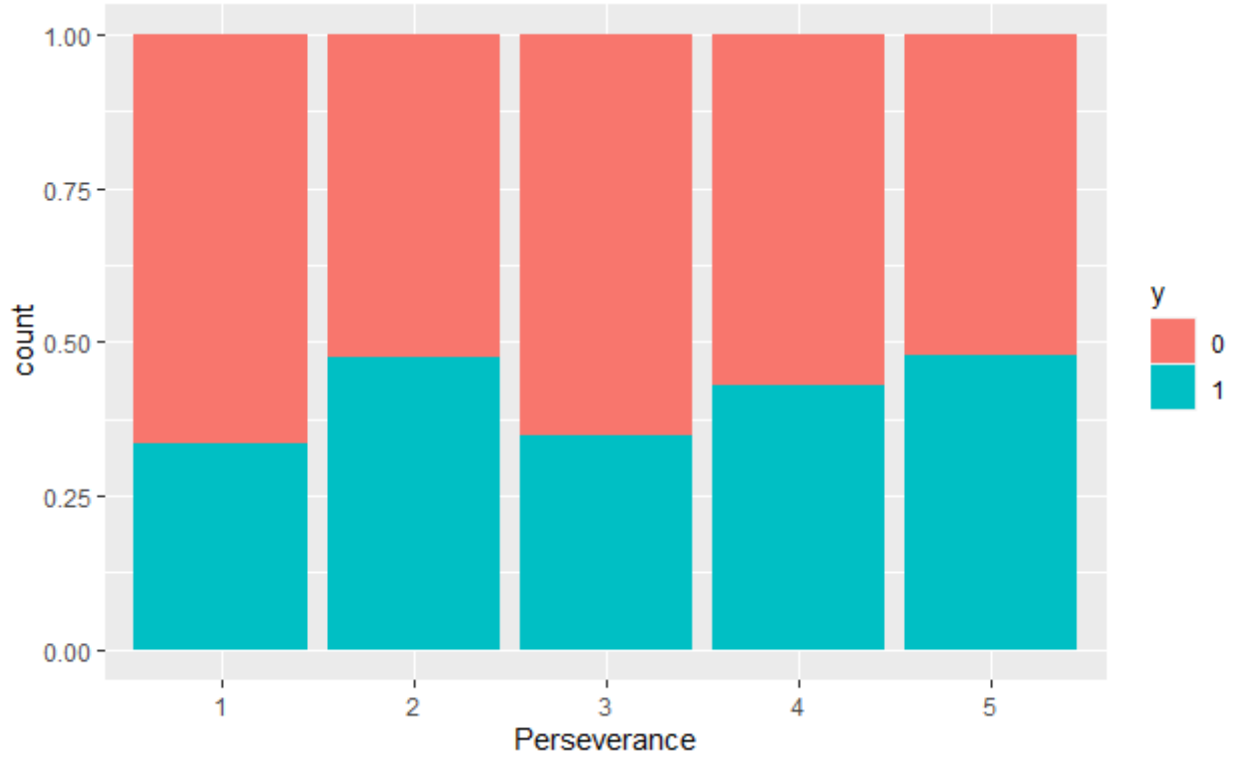
```
ctable(df$Perseverance, df$y)
```

```
Cross-Tabulation, Row Proportions
Perseverance * y
Data Frame: df

-------------- --- ------------ ------------ --------------
                y             0            1          Total
  Perseverance
             1         4 (66.7%)    2 (33.3%)     6 (100.0%)
             2        21 (52.5%)   19 (47.5%)    40 (100.0%)
             3        45 (65.2%)   24 (34.8%)    69 (100.0%)
             4        45 (57.0%)   34 (43.0%)    79 (100.0%)
             5        13 (52.0%)   12 (48.0%)    25 (100.0%)
         Total       128 (58.4%)   91 (41.6%)   219 (100.0%)
-------------- --- ------------ ------------ --------------
```

Hide

```
ggplot(df, aes(x = Perseverance, fill = y)) + geom_bar(position = "fill")
```



Hide

```
df = df %>%  mutate(Rango_perseverancia = case_when(between(Perseverance, 1,3) ~ "Bajo", Pers
everance >= 4 ~ "Alto"))
```

```
between() called on numeric vector with S3 class愀朠>=愀朢 not meaningful for factors
```

Hide

```
df$Rango_perseverancia[is.na(df$Rango_perseverancia)] <- "Alto"
```

Se observa una mayor tendencia de aquellos universitarios, que tienen como rasgo distintivo la perseverancia, a ser catalogados como empresarios (44% frente a 39%)

DesireToTakeInitiative

¿Es el deseo de tomar la iniciativa determimante para ser emprendedor?

Hide

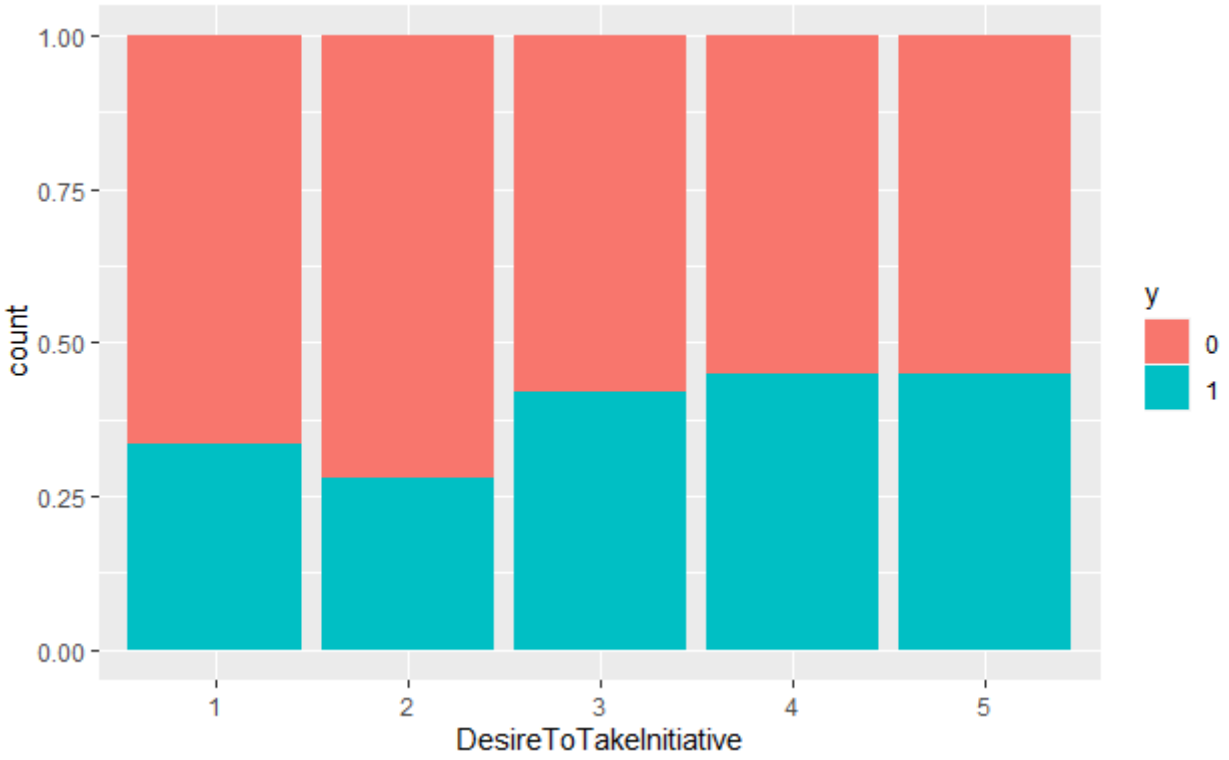
```
ctable(df$DesireToTakeInitiative, df$y)
```

```
Cross-Tabulation, Row Proportions
DesireToTakeInitiative * y
Data Frame: df

----------------------- --- ------------ ------------ --------------
                          y            0            1          Total
  DesireToTakeInitiative
                        1        8 (66.7%)    4 (33.3%)    12 (100.0%)
                        2       18 (72.0%)    7 (28.0%)    25 (100.0%)
                        3       32 (58.2%)   23 (41.8%)    55 (100.0%)
                        4       38 (55.1%)   31 (44.9%)    69 (100.0%)
                        5       32 (55.2%)   26 (44.8%)    58 (100.0%)
                    Total      128 (58.4%)   91 (41.6%)   219 (100.0%)
----------------------- --- ------------ ------------ --------------
```

Hide

```
ggplot(df, aes(x = DesireToTakeInitiative, fill = y)) + geom_bar(position = "fill")
```



Se observa una leve relacion entre el mayor deseo de tomar la iniciativa y la probabilidad de ser emprendedor.

Hide

```
df = df %>%  mutate(Rango_Iniciativa = case_when(between(DesireToTakeInitiative, 1,3) ~ "Baj
o", DesireToTakeInitiative >= 4 ~ "Alto"))
```

```
between() called on numeric vector with S3 class勄牯>=勄牷 not meaningful for factors
```

Hide

```
df$Rango_Iniciativa[is.na(df$Rango_Iniciativa)] <- "Alto"
```

Podemos ver como los universitarios con un nivel mas alto de deseo de tomar la iniciativa, tienen mas probabilidad de ser catalogados como empresarios (45% frente al 37%)

Competitiveness

¿Son los universitarios mas competitivos los que mas probabilidad tienen de ser empresarios?

Hide

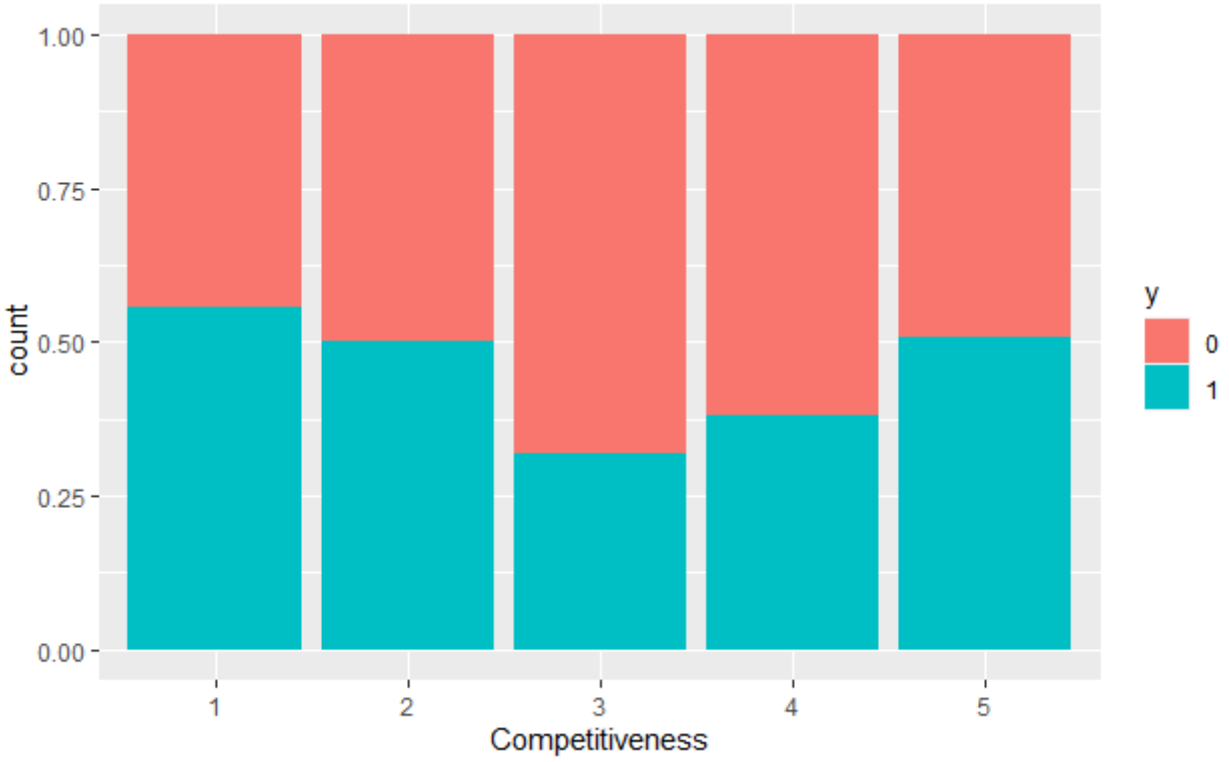
```
ctable(df$Competitiveness, df$y)
```

```
Cross-Tabulation, Row Proportions
Competitiveness * y
Data Frame: df

---------------- --- ------------ ------------ --------------
                   y            0            1          Total
  Competitiveness
                1         4 (44.4%)    5 (55.6%)     9 (100.0%)
                2        13 (50.0%)   13 (50.0%)    26 (100.0%)
                3        45 (68.2%)   21 (31.8%)    66 (100.0%)
                4        39 (61.9%)   24 (38.1%)    63 (100.0%)
                5        27 (49.1%)   28 (50.9%)    55 (100.0%)
            Total       128 (58.4%)   91 (41.6%)   219 (100.0%)
---------------- --- ------------ ------------ --------------
```

Hide

```
ggplot(df, aes(x = Competitiveness, fill = y)) + geom_bar(position = "fill")
```



No parece observarse una relacion entre la competitividad y la probabilida de ser empresario, no bostante, vamos a crear una variable binomial para verlo mejor

Hide

```
df = df %>%  mutate(Rango_Competitividad = case_when(between(Competitiveness, 1,3) ~ "Bajo",
 Competitiveness >= 4 ~ "Alto"))
```

```
between() called on numeric vector with S3 class惭=惭 not meaningful for factors
```

Hide

```
df$Rango_Competitividad[is.na(df$Rango_Competitividad)] <- "Alto"
```

En este grafico se puede observar como los universitarios con una nivel de competitividad mayor son mas propensos, aunque no de forma muy contundente, a ser catalogados como universitarios.

SelfReliance

¿Un mayor nivel de autosuficiencia determina una mayor probabilidad de ser empresario?

Hide

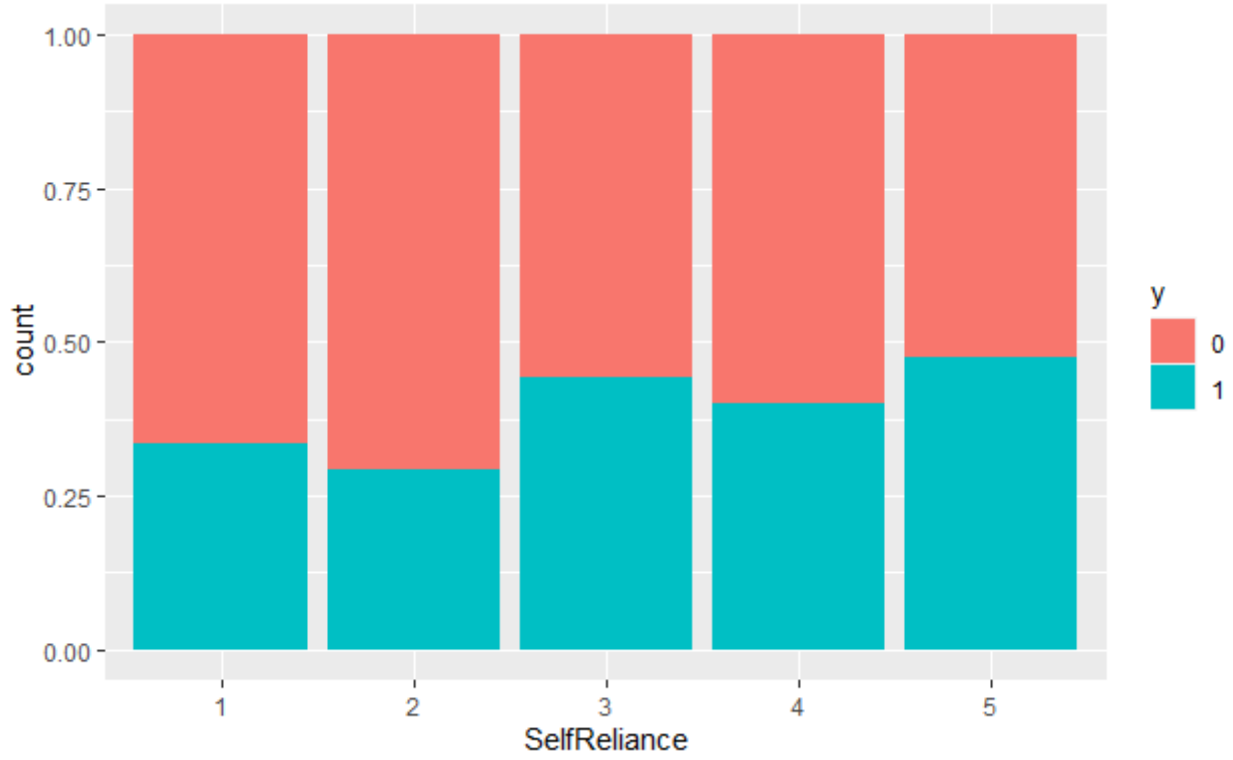
```
ctable(df$SelfReliance, df$y)
```

```
Cross-Tabulation, Row Proportions
SelfReliance * y
Data Frame: df

-------------- --- ------------ ------------ --------------
                y             0            1          Total
  SelfReliance
             1         4 (66.7%)    2 (33.3%)     6 (100.0%)
             2        17 (70.8%)    7 (29.2%)    24 (100.0%)
             3        29 (55.8%)   23 (44.2%)    52 (100.0%)
             4        48 (60.0%)   32 (40.0%)    80 (100.0%)
             5        30 (52.6%)   27 (47.4%)    57 (100.0%)
         Total       128 (58.4%)   91 (41.6%)   219 (100.0%)
-------------- --- ------------ ------------ --------------
```

Hide

```
ggplot(df, aes(x = SelfReliance, fill = y)) + geom_bar(position = "fill")
```



Por lo que se observa en la grafica, podemos decir que un mayor nivel de autosuficiencia, aumenta la probabilidad de ser empresario. Vamos a crear una variable binomial

Hide

```
df = df %>%  mutate(Rango_autosuficiencia = case_when(between(SelfReliance, 1,3) ~ "Bajo", Se
lfReliance >= 4 ~ "Alto"))
```

```
between() called on numeric vector with S3 class惭笔>=惭笔 not meaningful for factors
```

Hide

```
df$Rango_autosuficiencia[is.na(df$Rango_autosuficiencia)] <- "Alto"
```

En el grafico podemos ver también, que de forma leve, los univsersitarios con un nivel mayor de autosuficienca tienen mayor probabilidad de ser emprendedores

StrongNeedToAchieve

Hide

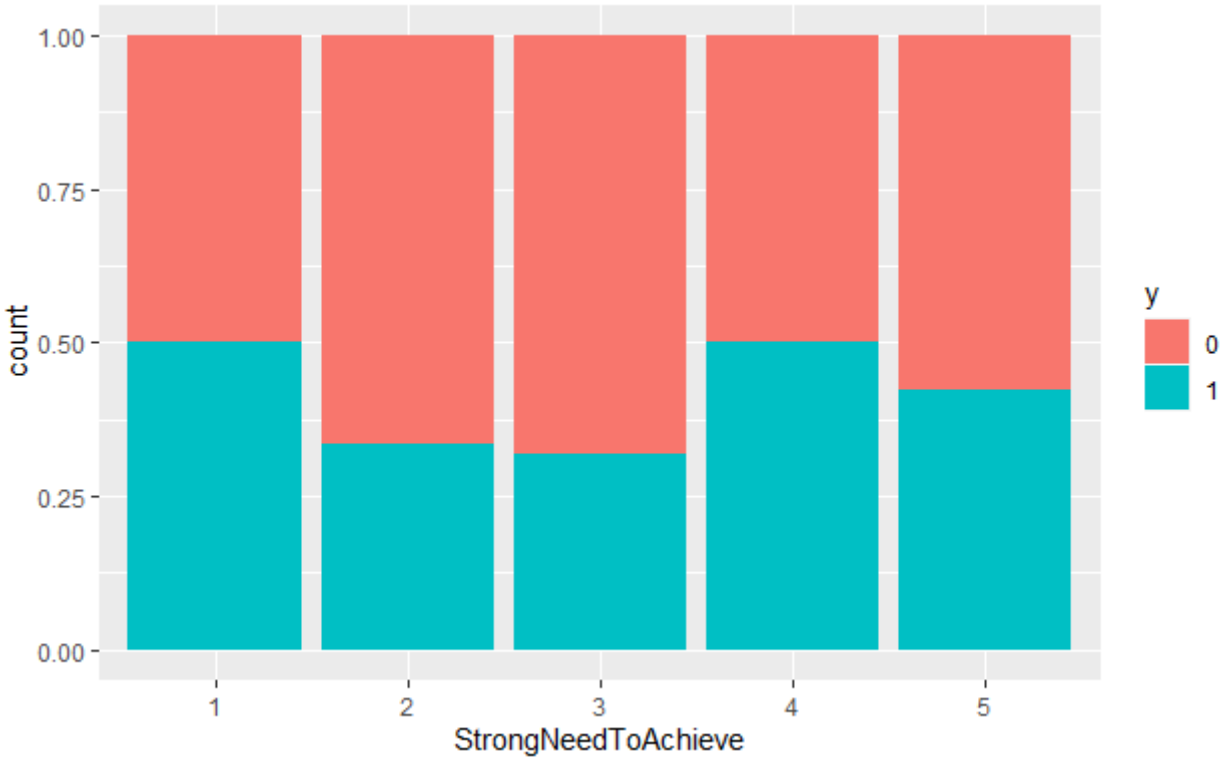
```
ctable(df$StrongNeedToAchieve, df$y)
```

```
Cross-Tabulation, Row Proportions
StrongNeedToAchieve * y
Data Frame: df

--------------------- --- ------------ ------------ --------------
                        y            0            1          Total
  StrongNeedToAchieve
                    1         1 (50.0%)    1 (50.0%)     2 (100.0%)
                    2        14 (66.7%)    7 (33.3%)    21 (100.0%)
                    3        34 (68.0%)   16 (32.0%)    50 (100.0%)
                    4        34 (50.0%)   34 (50.0%)    68 (100.0%)
                    5        45 (57.7%)   33 (42.3%)    78 (100.0%)
                Total       128 (58.4%)   91 (41.6%)   219 (100.0%)
--------------------- --- ------------ ------------ --------------
```

Hide

```
ggplot(df, aes(x = StrongNeedToAchieve, fill = y)) + geom_bar(position = "fill")
```



Viendo esta grafica no podemos decir que haya una relacion entre el mayor deseo de obtener una meta y la probablidad de ser empresario, vamos a crear una variable binomial

Hide

```
df = df%>% mutate(Rango_meta = case_when(between(StrongNeedToAchieve, 1,3) ~ "Bajo", StrongNe
edToAchieve >= 4 ~ "Alto"))
```

```
between() called on numeric vector with S3 class恸牠>=恸牢 not meaningful for factors
```

Hide

```
df$Rango_meta[is.na(df$Rango_meta)] <- "Alto"
```

En el grafico se observa como aquellos universitarios que tienen un nivel alto de necesidad, de lograr una meta, tienen mas probabilidad de ser empresarios.

SelfConfidence

¿Un mayor nivel de autoconfianza genera una mayor probabilidad de ser empreario?

Hide

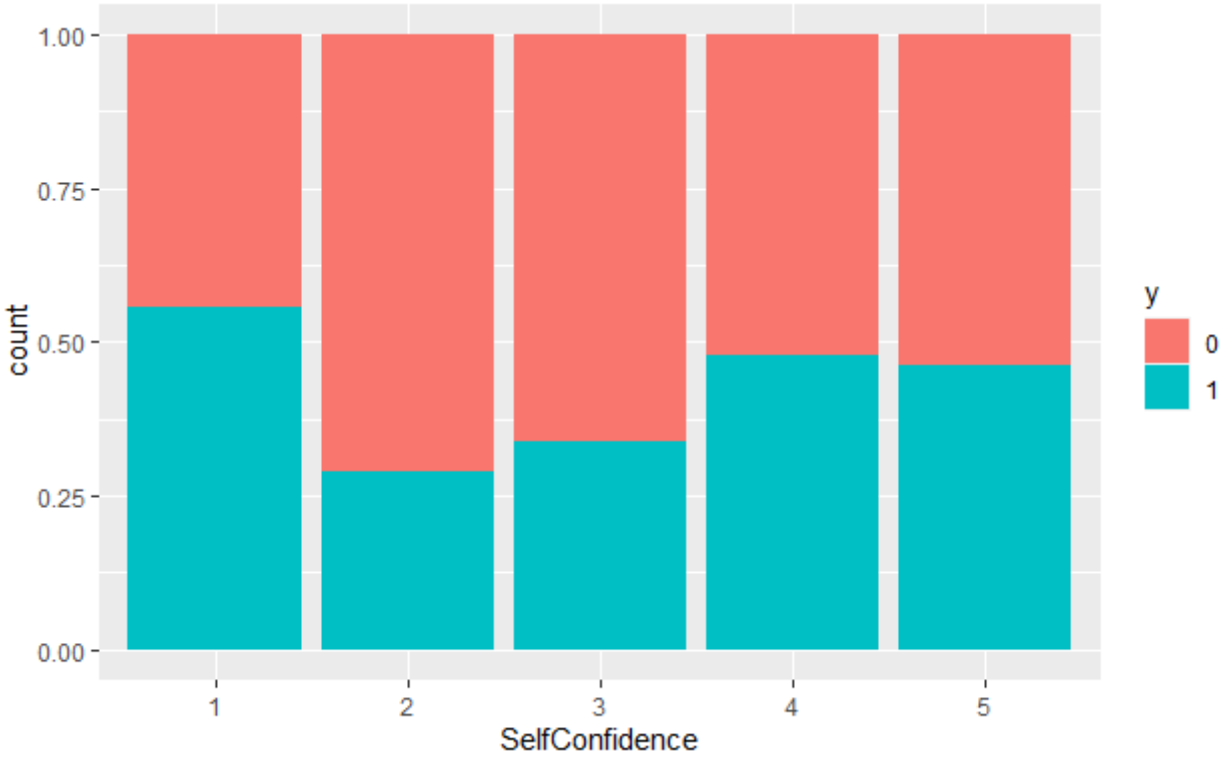
```
ctable(df$SelfConfidence, df$y)
```

```
Cross-Tabulation, Row Proportions
SelfConfidence * y
Data Frame: df

---------------- --- ------------ ------------ --------------
                   y            0            1          Total
  SelfConfidence
               1         4 (44.4%)    5 (55.6%)     9 (100.0%)
               2        22 (71.0%)    9 (29.0%)    31 (100.0%)
               3        37 (66.1%)   19 (33.9%)    56 (100.0%)
               4        37 (52.1%)   34 (47.9%)    71 (100.0%)
               5        28 (53.8%)   24 (46.2%)    52 (100.0%)
           Total       128 (58.4%)   91 (41.6%)   219 (100.0%)
---------------- --- ------------ ------------ --------------
```

Hide

```
ggplot(df, aes(x = SelfConfidence, fill = y)) + geom_bar(position = "fill")
```



Los resultados, como se pueden ver en la grafica, no son relevantes, vamos a hacer lo mismo que en la variable anterior, divir los niveles en dos y observar de nuevo los resultados.

Hide

```
df = df%>% mutate(Rango_autoconfianza = case_when(between(SelfConfidence, 1,3) ~ "Bajo", Self
Confidence >= 4 ~ "Alto"))
```

```
between() called on numeric vector with S3 class恤牠>=恤牷 not meaningful for factors
```

Hide

```
df$Rango_autoconfianza[is.na(df$Rango_autoconfianza)] <- "Alto"
```

Ahora si, podemos observar como los universitarios con un nivel de autoconfianza mayor, tienen mas probabilidades de ser emoresarios (47% frente a un 34,5%)

GoodPhysicalHealth

¿Esta relacionada la buena salud fisica con la probabilidad de ser empresario?

Antes de nada, vamos a crear la categoria doble (baja/alta)

Hide

```
df = df%>% mutate(Rango_salud_fisica = case_when(between(GoodPhysicalHealth, 1,3) ~ "Bajo", G
oodPhysicalHealth >= 4 ~ "Alto"))
```

```
between() called on numeric vector with S3 class恤牠>=恤牷 not meaningful for factors
```

Hide

```
df$Rango_salud_fisica[is.na(df$Rango_salud_fisica)] <- "Alto"
```

Como podemos ver en el grafico, tanto una buena salud fisica, como no tan buena, no parece determinante a la hora de llegar a ser empresario, por tanto, esta variable la vamos a descartar.

MentalDisorder

¿Tener algun trastorno mental determina la probabilidad de ser empresario?

Hide

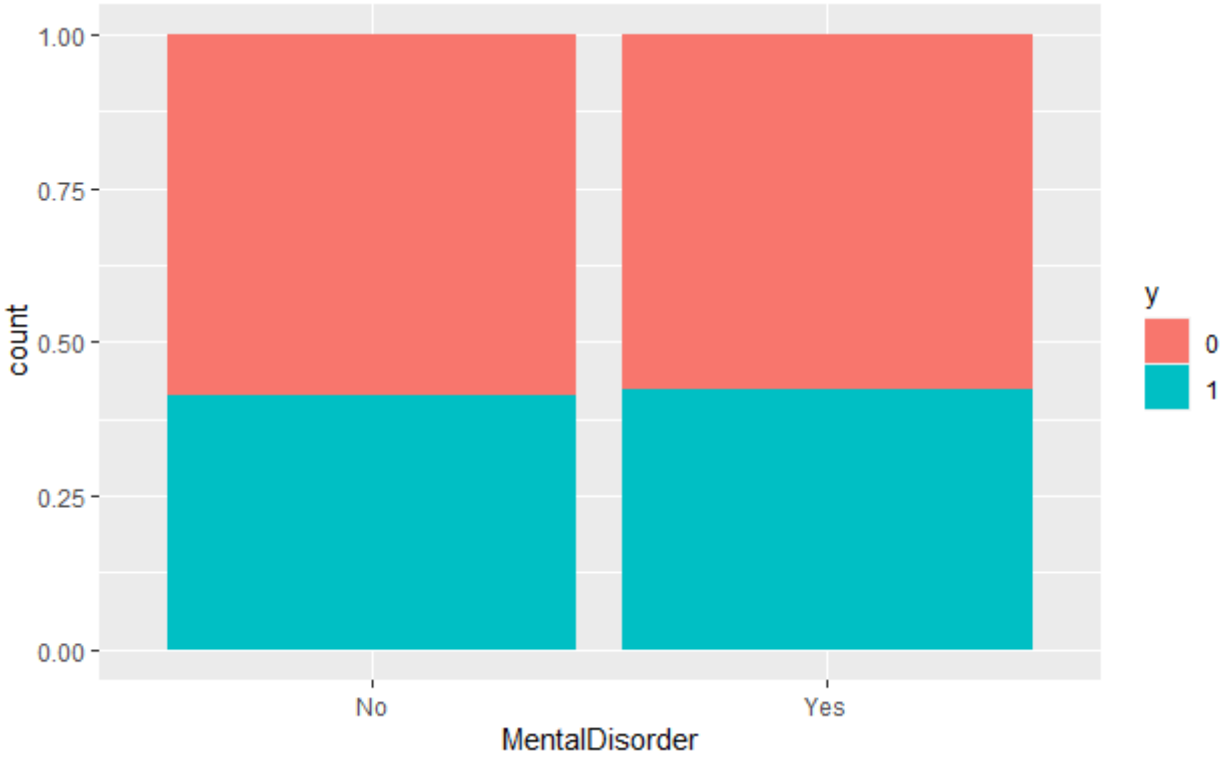
```
ctable(df$MentalDisorder, df$y)
```

```
Cross-Tabulation, Row Proportions
MentalDisorder * y
Data Frame: df

---------------- --- ------------ ------------ --------------
                   y            0            1          Total
  MentalDisorder
              No        91 (58.7%)   64 (41.3%)   155 (100.0%)
             Yes        37 (57.8%)   27 (42.2%)    64 (100.0%)
           Total       128 (58.4%)   91 (41.6%)   219 (100.0%)
---------------- --- ------------ ------------ --------------
```

Hide

```
ggplot(df, aes(x = MentalDisorder, fill = y)) + geom_bar(position = "fill")
```



Se observa unos valores muy parejos, tanto para aquellos universitarios que presentan algun transtorno mental, como para los que no. Esta variable la vamos a descartar al no ser muy relevante.

KeyTraits

¿De los rasgos analizados en esta pregunta, hay alguno que sea mas propenso para los empresarios?

Hide

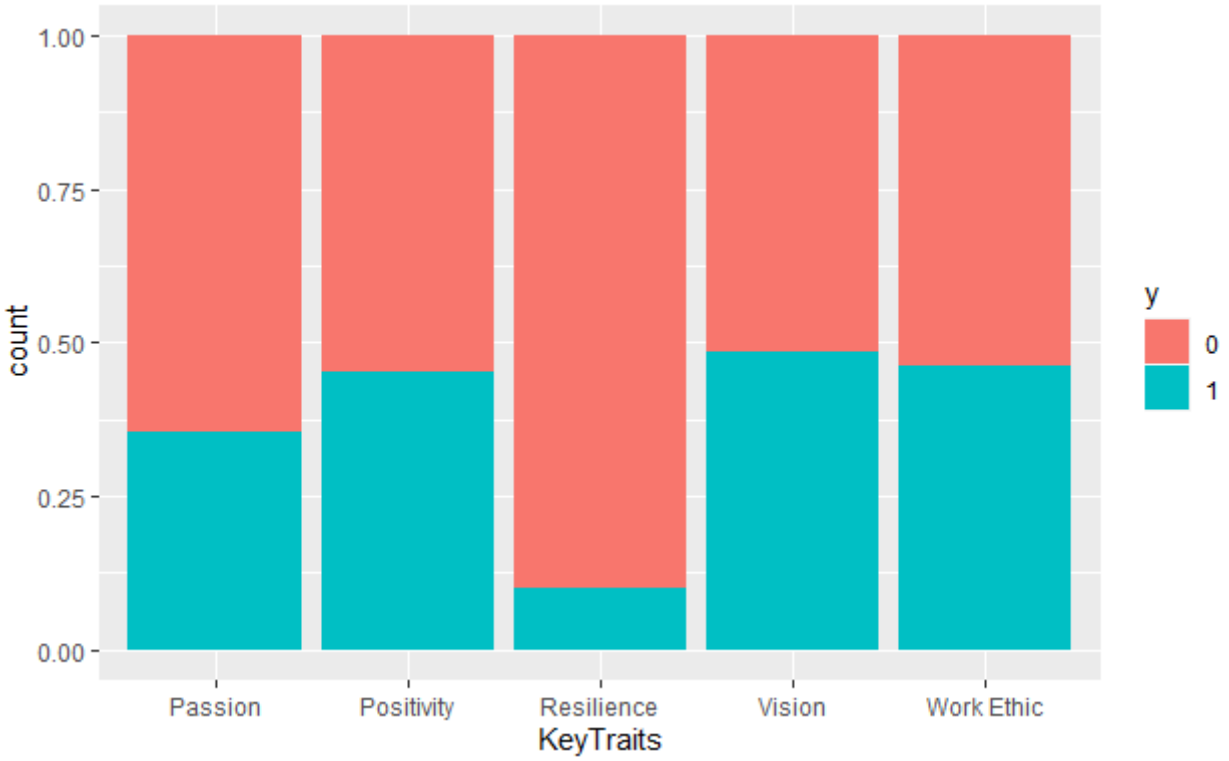
```
ctable(df$KeyTraits, df$y)
```

```
Cross-Tabulation, Row Proportions
KeyTraits * y
Data Frame: df

------------ --- ------------ ------------ --------------
           y             0            1          Total
  KeyTraits
    Passion       40 (64.5%)   22 (35.5%)    62 (100.0%)
 Positivity       40 (54.8%)   33 (45.2%)    73 (100.0%)
 Resilience        9 (90.0%)    1 (10.0%)    10 (100.0%)
     Vision       18 (51.4%)   17 (48.6%)    35 (100.0%)
 Work Ethic       21 (53.8%)   18 (46.2%)    39 (100.0%)
      Total      128 (58.4%)   91 (41.6%)   219 (100.0%)
------------ --- ------------ ------------ --------------
```

Hide

```
ggplot(df, aes(x = KeyTraits, fill = y)) + geom_bar(position = "fill")
```



Podemos ver como la positividad, vision y trabajo etico, son rasgos que destacan de los emprendedores, mientras que la resiliencia es un rasgo poco usual en los emprendedores.

ReasonsForLack

MOtivos en la falta de interes en ser emprendedor

¿Hay algun motivo determinante, a la hora de que sean catalogados los universitarios sin probabilidad de ser emprendedores?

Hide

```
unique(df$ReasonsForLack)
```

```
 [1]
 [2] Just not interested! (Want to work in the corporate sector, or for the government or pur
sue research or something else)
 [3] Not willing to start a venture in India and waiting for future relocation
 [4] Not able to take a Financial Risk
 [5] Just not interested! (Want to work in the corporate sector, or for the government or pur
sue research or something else), Academic Pressure, Lack of Knowledge, Not able to take a Fin
ancial Risk
 [6] Academic Pressure, Lack of Knowledge, Not able to take a Financial Risk
 [7] Just not interested! (Want to work in the corporate sector, or for the government or pur
sue research or something else), Academic Pressure, Parental Pressure
 [8] Academic Pressure, Lack of Knowledge, Mental Block
 [9] Lack of Knowledge
[10] Academic Pressure, Unwillingness to take risk, Lack of Knowledge
[11] Just not interested! (Want to work in the corporate sector, or for the government or pur
sue research or something else), Lack of Knowledge, Parental Pressure, Mental Block
[12] Unwillingness to take risk, Lack of Knowledge, Not able to take a Financial Risk
[13] Just not interested! (Want to work in the corporate sector, or for the government or pur
sue research or something else), Unwillingness to take risk, Lack of Knowledge
[14] Just not interested! (Want to work in the corporate sector, or for the government or pur
sue research or something else), Academic Pressure, Unwillingness to take risk, Lack of Knowl
edge, Parental Pressure, Not able to take a Financial Risk
[15] Academic Pressure, Lack of Knowledge
[16] Unwillingness to take risk
[17] Academic Pressure, Lack of Knowledge, Parental Pressure, Mental Block, Not able to take
a Financial Risk
[18] Lack of Knowledge, Not willing to start a venture in India and waiting for future reloca
tion
[19] Academic Pressure
[20] Just not interested! (Want to work in the corporate sector, or for the government or pur
sue research or something else), Academic Pressure, Unwillingness to take risk, Lack of Knowl
edge, Parental Pressure, Mental Block, Not able to take a Financial Risk, Not willing to star
t a venture in India and waiting for future relocation
[21] Just not interested! (Want to work in the corporate sector, or for the government or pur
sue research or something else), Lack of Knowledge
[22] Lack of Knowledge, Not able to take a Financial Risk
[23] Lack of Knowledge, Parental Pressure
[24] Just not interested! (Want to work in the corporate sector, or for the government or pur
sue research or something else), Not willing to start a venture in India and waiting for futu
re relocation
[25] Just not interested! (Want to work in the corporate sector, or for the government or pur
sue research or something else), Academic Pressure, Not able to take a Financial Risk
[26] Parental Pressure, Not able to take a Financial Risk, Not willing to start a venture in
India and waiting for future relocation

[27] Just not interested! (Want to work in the corporate sector, or for the government or pur
sue research or something else), Academic Pressure, Lack of Knowledge
[28] Academic Pressure, Not able to take a Financial Risk
[29] Mental Block
[30] Just not interested! (Want to work in the corporate sector, or for the government or pur
sue research or something else), Lack of Knowledge, Not able to take a Financial Risk, Not wi
lling to start a venture in India and waiting for future relocation
[31] Just not interested! (Want to work in the corporate sector, or for the government or pur
sue research or something else), Lack of Knowledge, Not able to take a Financial Risk
[32] Academic Pressure, Unwillingness to take risk, Lack of Knowledge, Parental Pressure, Men
tal Block, Not able to take a Financial Risk, Not willing to start a venture in India and wai
```

```
ting for future relocation
[33] Academic Pressure, Lack of Knowledge, Not able to take a Financial Risk, Not willing to
start a venture in India and waiting for future relocation
[34] Academic Pressure, Lack of Knowledge, Parental Pressure
[35] Just not interested! (Want to work in the corporate sector, or for the government or pur
sue research or something else), Unwillingness to take risk
[36] Academic Pressure, Unwillingness to take risk, Parental Pressure, Not able to take a Fin
ancial Risk
[37] Not able to take a Financial Risk, Not willing to start a venture in India and waiting f
or future relocation
[38] Lack of Knowledge, Parental Pressure, Mental Block, Not able to take a Financial Risk
[39] Academic Pressure, Unwillingness to take risk, Not able to take a Financial Risk
[40] Parental Pressure
[41] Just not interested! (Want to work in the corporate sector, or for the government or pur
sue research or something else), Not able to take a Financial Risk
[42] Just not interested! (Want to work in the corporate sector, or for the government or pur
sue research or something else), Academic Pressure, Lack of Knowledge, Parental Pressure, Not
able to take a Financial Risk
[43] Lack of Knowledge, Mental Block, Not able to take a Financial Risk
[44] Lack of Knowledge, Mental Block
[45] Just not interested! (Want to work in the corporate sector, or for the government or pur
sue research or something else), Academic Pressure
45 Levels:  Academic Pressure ... Unwillingness to take risk, Lack of Knowledge, Not able to
take a Financial Risk
```

Hide

```
table(df$ReasonsForLack)
```

```
91

Academic Pressure

11

Academic Pressure, Lack of Knowledge

1

Academic Pressure, Lack of Knowledge, Mental Block

1

Academic Pressure, Lack of Knowledge, Not able to take a Financial Risk

2

Academic Pressure, Lack of Knowledge, Not able to take a Financial Risk, Not willing to start
a venture in India and waiting for future relocation

1

Academic Pressure, Lack of Knowledge, Parental Pressure

1

Academic Pressure, Lack of Knowledge, Parental Pressure, Mental Block, Not able to take a Fin
ancial Risk

1

Academic Pressure, Not able to take a Financial Risk

2

Academic Pressure, Unwillingness to take risk, Lack of Knowledge

4

Academic Pressure, Unwillingness to take risk, Lack of Knowledge, Parental Pressure, Mental B
lock, Not able to take a Financial Risk, Not willing to start a venture in India and waiting
for future relocation

1

Academic Pressure, Unwillingness to take risk, Not able to take a Financial Risk

1

Academic Pressure, Unwillingness to take risk, Parental Pressure, Not able to take a Financia
l Risk
```

```
1

Just not interested! (Want to work in the corporate sector, or for the government or pursue r
esearch or something else)

41

Just not interested! (Want to work in the corporate sector, or for the government or pursue r
esearch or something else), Academic Pressure

1

Just not interested! (Want to work in the corporate sector, or for the government or pursue r
esearch or something else), Academic Pressure, Lack of Knowledge

1

Just not interested! (Want to work in the corporate sector, or for the government or pursue r
esearch or something else), Academic Pressure, Lack of Knowledge, Not able to take a Financia
l Risk

2

Just not interested! (Want to work in the corporate sector, or for the government or pursue r
esearch or something else), Academic Pressure, Lack of Knowledge, Parental Pressure, Not able
to take a Financial Risk

1

Just not interested! (Want to work in the corporate sector, or for the government or pursue r
esearch or something else), Academic Pressure, Not able to take a Financial Risk

1

Just not interested! (Want to work in the corporate sector, or for the government or pursue r
esearch or something else), Academic Pressure, Parental Pressure

1
Just not interested! (Want to work in the corporate sector, or for the government or pursue r
esearch or something else), Academic Pressure, Unwillingness to take risk, Lack of Knowledge,
Parental Pressure, Mental Block, Not able to take a Financial Risk, Not willing to start a ve
nture in India and waiting for future relocation

1
                                                                                    Just
not interested! (Want to work in the corporate sector, or for the government or pursue resear
ch or something else), Academic Pressure, Unwillingness to take risk, Lack of Knowledge, Pare
ntal Pressure, Not able to take a Financial Risk

1

Just not interested! (Want to work in the corporate sector, or for the government or pursue r
esearch or something else), Lack of Knowledge

2
```

```
Just not interested! (Want to work in the corporate sector, or for the government or pursue r
esearch or something else), Lack of Knowledge, Not able to take a Financial Risk

4

                                                                                Just not inte
rested! (Want to work in the corporate sector, or for the government or pursue research or so
mething else), Lack of Knowledge, Not able to take a Financial Risk, Not willing to start a v
enture in India and waiting for future relocation

1

Just not interested! (Want to work in the corporate sector, or for the government or pursue r
esearch or something else), Lack of Knowledge, Parental Pressure, Mental Block

1

Just not interested! (Want to work in the corporate sector, or for the government or pursue r
esearch or something else), Not able to take a Financial Risk

2

Just not interested! (Want to work in the corporate sector, or for the government or pursue r
esearch or something else), Not willing to start a venture in India and waiting for future re
location

1

Just not interested! (Want to work in the corporate sector, or for the government or pursue r
esearch or something else), Unwillingness to take risk

1

Just not interested! (Want to work in the corporate sector, or for the government or pursue r
esearch or something else), Unwillingness to take risk, Lack of Knowledge

1

Lack of Knowledge

7

Lack of Knowledge, Mental Block

1

Lack of Knowledge, Mental Block, Not able to take a Financial Risk

1

Lack of Knowledge, Not able to take a Financial Risk

3

Lack of Knowledge, Not willing to start a venture in India and waiting for future relocation
```

```
1

Lack of Knowledge, Parental Pressure

2

Lack of Knowledge, Parental Pressure, Mental Block, Not able to take a Financial Risk

1

Mental Block

1

Not able to take a Financial Risk

3

Not able to take a Financial Risk, Not willing to start a venture in India and waiting for fu
ture relocation

1

Not willing to start a venture in India and waiting for future relocation

10

Parental Pressure

1

Parental Pressure, Not able to take a Financial Risk, Not willing to start a venture in India
and waiting for future relocation

1

Unwillingness to take risk

4

Unwillingness to take risk, Lack of Knowledge, Not able to take a Financial Risk

1
```

No se trata de una variable que aporte mucha informacion, por lo que la vamos a eliminar.

Indicios encontrados

Podemos decir que hay sectores de estudios muy propensos a que los estudiantes acaben siendo emprendedores, como son las Humanidades y Ciencias Sociales y el grado de enseñanza(Teaching degree). En cuanto a la edad, claramente, son los encuestados de la franja de edad mayor, los que tienen una mayor probabilidad de ser emprendedores. En cuanto al genero, son los hombres los que mas probabilidad tienen de ser empresarios, al igual como los que viven en ciudades. El resto de variables que analizan distintos rasgos

estan bastante parejas entre los que son empresarios y los que no, de tal forma que retiraremos de la folmula las variables GoodPhysicalHealth y MentalDisorder, por no presentar practicamente diferencias. Eliminaremos también la variable Reasonforluck, por no ser significativa.

Creacion de un modelo

Hide

```
df_2 = select(df,EducationSector, IndividualProject, Rango_edad, Gender, City, Influenced, Ra
ngo_perseverancia,
Rango_Iniciativa, Rango_Competitividad, Rango_autosuficiencia,
Rango_meta, Rango_autoconfianza, KeyTraits, y)
```

```
Error: Can't subset columns that don't exist.
x Column `Rango_edad` doesn't exist.
Run `rlang::last_error()` to see where the error occurred.
```

Vamos a generar un modelo con estas variables

Hide

```
modelo = glm(y ~ EducationSector + IndividualProject + Rango_edad + Gender + City + Influence
d + Rango_perseverancia + Rango_Iniciativa + Rango_Competitividad + Rango_autosuficiencia + R
ango_meta + Rango_autoconfianza + KeyTraits,  data = df, family = binomial)

summary(modelo)
```

```
Call:
glm(formula = y ~ EducationSector + IndividualProject + Rango_edad +
    Gender + City + Influenced + Rango_perseverancia + Rango_Iniciativa +
    Rango_Competitividad + Rango_autosuficiencia + Rango_meta +
    Rango_autoconfianza + KeyTraits, family = binomial, data = df)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8569  -0.9026  -0.4390   1.0309   1.9762

Coefficients:
                                                                  Estimate
(Intercept)                                                       -2.07306
EducationSectorEconomic Sciences, Business Studies, Commerce and Law   -0.74964
EducationSectorEngineering Sciences                                0.30528
EducationSectorHumanities and Social Sciences                     18.32424
EducationSectorLanguage and Cultural Studies                     -18.05193
EducationSectorMathematics or Natural Sciences                   -17.87018
EducationSectorMedicine, Health Sciences                          -0.10795
EducationSectorOthers                                              0.26601
EducationSectorTeaching Degree (e.g., B.Ed)                        0.49863
IndividualProjectYes                                               0.82245
Rango_edad20-22                                                   -0.79690
Rango_edad23-26                                                    1.81658
GenderMale                                                         0.19414
CityYes                                                            0.36995
InfluencedYes                                                      1.25582
Rango_perseveranciaBajo                                            0.02155
Rango_IniciativaBajo                                              -0.15835
Rango_CompetitividadBajo                                           0.13913
Rango_autosuficienciaBajo                                          0.38233
Rango_metaBajo                                                    -0.43590
Rango_autoconfianzaBajo                                           -0.32454
KeyTraitsPositivity                                                0.46730
KeyTraitsResilience                                               -1.57109
KeyTraitsVision                                                    0.56596
KeyTraitsWork Ethic                                                0.26249
                                                                Std. Error
(Intercept)                                                        0.87354
EducationSectorEconomic Sciences, Business Studies, Commerce and Law    0.71335
EducationSectorEngineering Sciences                                0.55742
EducationSectorHumanities and Social Sciences                   1616.28650
EducationSectorLanguage and Cultural Studies                    3956.18045
EducationSectorMathematics or Natural Sciences                  1798.40276
EducationSectorMedicine, Health Sciences                           0.93018
EducationSectorOthers                                              0.70602
EducationSectorTeaching Degree (e.g., B.Ed)                        1.41363
IndividualProjectYes                                               0.35258
Rango_edad20-22                                                    0.35318
Rango_edad23-26                                                    1.31441
GenderMale                                                         0.39659
CityYes                                                            0.48307
InfluencedYes                                                      0.39833
Rango_perseveranciaBajo                                            0.45970
```

```
Rango_IniciativaBajo                                                    0.46289
Rango_CompetitividadBajo                                                0.49793
Rango_autosuficienciaBajo                                               0.53646
Rango_metaBajo                                                          0.55907
Rango_autoconfianzaBajo                                                 0.48093
KeyTraitsPositivity                                                     0.43249
KeyTraitsResilience                                                     1.17056
KeyTraitsVision                                                         0.52750
KeyTraitsWork Ethic                                                     0.50185
                                                                       z value
(Intercept)                                                             -2.373
EducationSectorEconomic Sciences, Business Studies, Commerce and Law    -1.051
EducationSectorEngineering Sciences                                      0.548
EducationSectorHumanities and Social Sciences                            0.011
EducationSectorLanguage and Cultural Studies                            -0.005
EducationSectorMathematics or Natural Sciences                          -0.010
EducationSectorMedicine, Health Sciences                                -0.116
EducationSectorOthers                                                    0.377
EducationSectorTeaching Degree (e.g., B.Ed)                              0.353
IndividualProjectYes                                                     2.333
Rango_edad20-22                                                         -2.256
Rango_edad23-26                                                          1.382
GenderMale                                                               0.490
CityYes                                                                  0.766
InfluencedYes                                                            3.153
Rango_perseveranciaBajo                                                  0.047
Rango_IniciativaBajo                                                    -0.342
Rango_CompetitividadBajo                                                 0.279
Rango_autosuficienciaBajo                                                0.713
Rango_metaBajo                                                          -0.780
Rango_autoconfianzaBajo                                                 -0.675
KeyTraitsPositivity                                                      1.080
KeyTraitsResilience                                                     -1.342
KeyTraitsVision                                                          1.073
KeyTraitsWork Ethic                                                      0.523
                                                                      Pr(>|z|)
(Intercept)                                                            0.01764
EducationSectorEconomic Sciences, Business Studies, Commerce and Law   0.29331
EducationSectorEngineering Sciences                                    0.58391
EducationSectorHumanities and Social Sciences                          0.99095
EducationSectorLanguage and Cultural Studies                           0.99636
EducationSectorMathematics or Natural Sciences                         0.99207
EducationSectorMedicine, Health Sciences                               0.90761
EducationSectorOthers                                                  0.70634
EducationSectorTeaching Degree (e.g., B.Ed)                            0.72429
IndividualProjectYes                                                   0.01966
Rango_edad20-22                                                        0.02405
Rango_edad23-26                                                        0.16696
GenderMale                                                             0.62448
CityYes                                                                0.44377
InfluencedYes                                                          0.00162
Rango_perseveranciaBajo                                                0.96262
Rango_IniciativaBajo                                                   0.73228
Rango_CompetitividadBajo                                               0.77992
Rango_autosuficienciaBajo                                              0.47604
Rango_metaBajo                                                         0.43557
```

```
Rango_autoconfianzaBajo                                                  0.49979
KeyTraitsPositivity                                                      0.27992
KeyTraitsResilience                                                      0.17954
KeyTraitsVision                                                          0.28331
KeyTraitsWork Ethic                                                      0.60095

(Intercept)                                                             *
EducationSectorEconomic Sciences, Business Studies, Commerce and Law
EducationSectorEngineering Sciences
EducationSectorHumanities and Social Sciences
EducationSectorLanguage and Cultural Studies
EducationSectorMathematics or Natural Sciences
EducationSectorMedicine, Health Sciences
EducationSectorOthers
EducationSectorTeaching Degree (e.g., B.Ed)
IndividualProjectYes                                                    *
Rango_edad20-22                                                         *
Rango_edad23-26
GenderMale
CityYes
InfluencedYes                                                           **
Rango_perseveranciaBajo
Rango_IniciativaBajo
Rango_CompetitividadBajo
Rango_autosuficienciaBajo
Rango_metaBajo
Rango_autoconfianzaBajo
KeyTraitsPositivity
KeyTraitsResilience
KeyTraitsVision
KeyTraitsWork Ethic
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 297.32  on 218  degrees of freedom
Residual deviance: 239.55  on 194  degrees of freedom
AIC: 289.55

Number of Fisher Scoring iterations: 16
```

Hide

```
modelo_final= step(modelo, direction="backward")
```

```
Start:  AIC=289.55
y ~ EducationSector + IndividualProject + Rango_edad + Gender + 
    City + Influenced + Rango_perseverancia + Rango_Iniciativa + 
    Rango_Competitividad + Rango_autosuficiencia + Rango_meta + 
    Rango_autoconfianza + KeyTraits

                        Df Deviance    AIC
- KeyTraits              4   244.82 286.82
- Rango_perseverancia    1   239.56 287.56
- Rango_Competitividad   1   239.63 287.63
- Rango_Iniciativa       1   239.67 287.67
- Gender                 1   239.79 287.79
- Rango_autoconfianza    1   240.01 288.01
- Rango_autosuficiencia  1   240.07 288.07
- City                   1   240.15 288.15
- Rango_meta             1   240.17 288.17
<none>                       239.55 289.55
- IndividualProject      1   245.21 293.20
- Rango_edad             2   247.75 293.75
- EducationSector        8   261.95 295.95
- Influenced             1   250.52 298.52

Step:  AIC=286.82
y ~ EducationSector + IndividualProject + Rango_edad + Gender + 
    City + Influenced + Rango_perseverancia + Rango_Iniciativa + 
    Rango_Competitividad + Rango_autosuficiencia + Rango_meta + 
    Rango_autoconfianza

                        Df Deviance    AIC
- Rango_perseverancia    1   244.82 284.82
- Rango_Competitividad   1   244.99 284.99
- Rango_Iniciativa       1   245.02 285.02
- Gender                 1   245.28 285.28
- Rango_autosuficiencia  1   245.43 285.43
- Rango_meta             1   245.63 285.63
- Rango_autoconfianza    1   245.72 285.72
- City                   1   246.14 286.14
<none>                       244.82 286.82
- IndividualProject      1   249.75 289.75
- Rango_edad             2   252.75 290.75
- EducationSector        8   268.41 294.41
- Influenced             1   255.50 295.50

Step:  AIC=284.82
y ~ EducationSector + IndividualProject + Rango_edad + Gender + 
    City + Influenced + Rango_Iniciativa + Rango_Competitividad + 
    Rango_autosuficiencia + Rango_meta + Rango_autoconfianza

                        Df Deviance    AIC
- Rango_Competitividad   1   245.01 283.01
- Rango_Iniciativa       1   245.02 283.02
- Gender                 1   245.28 283.28
- Rango_autosuficiencia  1   245.46 283.45
- Rango_meta             1   245.67 283.67
- Rango_autoconfianza    1   245.74 283.74
```

```
- City                   1   246.15 284.15
<none>                       244.82 284.82
- IndividualProject      1   249.82 287.82
- Rango_edad             2   252.86 288.86
- EducationSector        8   268.44 292.44
- Influenced             1   255.62 293.62

Step:  AIC=283.01
y ~ EducationSector + IndividualProject + Rango_edad + Gender +
    City + Influenced + Rango_Iniciativa + Rango_autosuficiencia +
    Rango_meta + Rango_autoconfianza

                        Df Deviance    AIC
- Rango_Iniciativa       1   245.12 281.12
- Gender                 1   245.52 281.52
- Rango_meta             1   245.76 281.76
- Rango_autoconfianza    1   245.80 281.80
- Rango_autosuficiencia  1   245.83 281.83
- City                   1   246.43 282.43
<none>                       245.01 283.01
- IndividualProject      1   249.88 285.88
- Rango_edad             2   253.09 287.08
- EducationSector        8   268.46 290.46
- Influenced             1   256.14 292.14

Step:  AIC=281.12
y ~ EducationSector + IndividualProject + Rango_edad + Gender +
    City + Influenced + Rango_autosuficiencia + Rango_meta +
    Rango_autoconfianza

                        Df Deviance    AIC
- Gender                 1   245.59 279.59
- Rango_autosuficiencia  1   245.87 279.87
- Rango_meta             1   246.02 280.02
- Rango_autoconfianza    1   246.33 280.33
- City                   1   246.50 280.50
<none>                       245.12 281.12
- IndividualProject      1   250.07 284.07
- Rango_edad             2   253.26 285.26
- EducationSector        8   268.46 288.46
- Influenced             1   256.19 290.19

Step:  AIC=279.59
y ~ EducationSector + IndividualProject + Rango_edad + City +
    Influenced + Rango_autosuficiencia + Rango_meta + Rango_autoconfianza

                        Df Deviance    AIC
- Rango_autosuficiencia  1   246.49 278.49
- Rango_meta             1   246.54 278.54
- Rango_autoconfianza    1   246.78 278.78
- City                   1   246.83 278.83
<none>                       245.59 279.59
- IndividualProject      1   250.62 282.62
- Rango_edad             2   253.77 283.77
- EducationSector        8   268.93 286.93
- Influenced             1   257.48 289.48
```

```
Step:  AIC=278.49
y ~ EducationSector + IndividualProject + Rango_edad + City + 
    Influenced + Rango_meta + Rango_autoconfianza

                     Df Deviance    AIC
- Rango_meta          1   246.80 276.80
- Rango_autoconfianza 1   247.19 277.19
- City                1   247.80 277.80
<none>                    246.49 278.49
- IndividualProject   1   251.50 281.50
- Rango_edad          2   255.97 283.96
- EducationSector     8   269.28 285.28
- Influenced          1   258.86 288.86

Step:  AIC=276.8
y ~ EducationSector + IndividualProject + Rango_edad + City + 
    Influenced + Rango_autoconfianza

                     Df Deviance    AIC
- City                1   248.14 276.14
<none>                    246.80 276.80
- Rango_autoconfianza 1   248.87 276.87
- IndividualProject   1   251.92 279.92
- Rango_edad          2   255.98 281.98
- EducationSector     8   269.66 283.66
- Influenced          1   260.19 288.19

Step:  AIC=276.14
y ~ EducationSector + IndividualProject + Rango_edad + Influenced + 
    Rango_autoconfianza

                     Df Deviance    AIC
<none>                    248.14 276.14
- Rango_autoconfianza 1   250.25 276.25
- IndividualProject   1   253.62 279.62
- Rango_edad          2   257.08 281.08
- EducationSector     8   271.03 283.03
- Influenced          1   261.13 287.13
```

Hide

```
modelo_final
```

```
Call:  glm(formula = y ~ EducationSector + IndividualProject + Rango_edad +
    Influenced + Rango_autoconfianza, family = binomial, data = df)

Coefficients:
                                                          (Intercept)
                                                              -1.32825
EducationSectorEconomic Sciences, Business Studies, Commerce and Law
                                                              -0.83384
                                    EducationSectorEngineering Sciences
                                                               0.35263
                           EducationSectorHumanities and Social Sciences
                                                              17.31617
                            EducationSectorLanguage and Cultural Studies
                                                             -16.52554
                         EducationSectorMathematics or Natural Sciences
                                                             -16.31142
                               EducationSectorMedicine, Health Sciences
                                                               0.02408
                                                  EducationSectorOthers
                                                               0.17480
                             EducationSectorTeaching Degree (e.g., B.Ed)
                                                               0.67678
                                               IndividualProjectYes
                                                               0.76321
                                                    Rango_edad20-22
                                                              -0.77041
                                                    Rango_edad23-26
                                                               1.77163
                                                      InfluencedYes
                                                               1.28772
                                             Rango_autoconfianzaBajo
                                                              -0.45323

Degrees of Freedom: 218 Total (i.e. Null);  205 Residual
Null Deviance:      297.3
Residual Deviance: 248.1    AIC: 276.1
```