Universitat Oberta de Catalunya

Master Ciencia de Datos

Tratamiento de datos: *Kickstarter Projects*.

Irene López Ruiz

Javier Samir Rey

Índice

1 Descripción del dataset

Kickstarter (https://www.kickstarter.com) es una plataforma de mecenazgo para proyectos creativos (*Kickstarter e impuestos — Kickstarter*, s. f.). Los dueños del proyecto eligen una fecha límite y un mínimo objetivo de fondos a recaudar y, si el objetivo elegido no es recolectado en el plazo, no se perciben fondos gracias a un contrato de garantía (*Kickstarter - Wikipedia, la enciclopedia libre*, s. f.).

Como es lógico, no todos los proyectos de Kickstarter consiguen recaudar los fondos que necesitan, mientras que otros rebasan por grandes cantidades la suma mínima que requerían sus autores. Esto plantea la necesidad de esclarecer las características de los proyectos que consiguen recaudar el dinero necesario y las diferencias que éstos tienen con los proyectos que son finalmente cancelados. Con un estudio de este calibre se podrán realizar estrategias de lanzamiento de proyectos para que los autores de estos tengan una mayor probabilidad de éxito, beneficiando así tanto a los proyectos como a la propia plataforma.

Es por esto por lo que se ha decidido realizar un proyecto de análisis de datos sobre un conjunto de datos, provisto en la plataforma Kaggle (https://www.kaggle.com), sobre la financiación de diferentes proyectos de Kickstarter que permita discriminar los puntos clave de los proyectos que más éxito tienen, en cuanto a la cantidad de dinero que consiguen recaudar. El conjunto de datos tratado ha sido creado por el usuario de Kaggle Mickaël Mouillé (https://www.kaggle.com/kemical), y la documentación del mismo puede encontrarse en https://www.kaggle.com/kemical/kickstarter-projects?select=ks-projects-201801.csv

Este conjunto de datos se compone de dos ficheros CSV, siendo uno de ellos (*ks-projects-201612.csv*) un subconjunto de los registros del otro (ks-projects-201801.csv), por lo que solo se tratará este último fichero, más completo. Este *dataset* cuenta con un total de 378661 proyectos, descritos con 15 variables diferentes:

- **ID**: identificador numérico y único del proyecto en la plataforma Kickstarter.
- **Name**: Nombre del proyecto.
- **Category**: categoría de clasificación del proyecto, por ejemplo, poesía.
- **main_category**: categoría principal para la clasificación del proyecto en clases más generales, por ejemplo, arte.
- **Currency**: acrónimo de la moneda de la financiación del proyecto, por ejemplo, USD.
- **Deadline**: fecha de finalización de la campaña de financiación del proyecto.
- **Goal**: objetivo de financiación propuesto por los autores del proyecto. Cantidad indicada en la moneda especificada por el campo *currency*.
- **Launched**: fecha y hora de lanzamiento de la campaña de financiación del proyecto.
- **Pledged**: cantidad recaudada por el proyecto, expresada en la moneda especificada por el campo *currency*.
- **State**: estado del proyecto. Posee varias categorías posibles, como cancelado, fallido, o existoso.
- **Backers**: número de patrocinadores o mecenas del proyecto.
- **Country**: país de origen del proyecto, descrito mediante el acrónimo de dos letras, por ejemplo, US.
- **usd_pledged**: cantidad recaudada por el proyecto, expresada en dólares americanos (USD), siendo la conversión entre la moneda original realizada por Kickstarter.
- **usd_pledged_real**: cantidad recaudada por el proyecto, expresada en dólares

americanos (USD), siendo la conversión entre la moneda original realizada por la plataforma Fixer (https://fixer.io).

- **usd_goal_real**: objetivo de financiación propuesto por los autores del proyecto. Cantidad indicada en dólares americanos (USD) siendo la conversión entre la moneda original realizada por la plataforma Fixer (https://fixer.io).

El presente documento supone una memoria con la descripción de los pasos y conclusiones extraídas del preprocesado y análisis de estos datos. Dichos pasos se han realizado con el código que se adjunta en este proyecto.

2 Lectura y selección de los datos de interés

En el código adjunto a este proyecto se ha realizado, tras la carga del fichero, un resumen de todas las variables que posee el conjunto de datos donde, entre otros, se muestra se muestra la codificación en R de cada variable, que deberá ser subsanada para que considere como variables categóricas a todos aquellos campos que toman para cada registro un valor entre una serie de clases concretas; y también para las variables con fechas (*launched* y *deadline*), que deberán codificarse como datos temporales. Se ha comprobado también que no existen duplicados de registros que deban borrarse para no considerar el mismo proyecto más de una vez. Tras esta correcta lectura de los datos del fichero se procede a preprocesar los datos.

En primer lugar, se procede a una selección de las variables a analizar. Recordemos que se dispone de los datos objetivo y recaudación sobre la financiación del proyecto, tanto en dólares americanos como en la moneda original. Esta información es redundante y, considerando que los datos en dólares americanos resultan una estandarización que facilita la comparación entre diferentes proyectos, las variables que referencian la financiación en la moneda original serán eliminadas. Además, la conversión a dólares realizada por Kickstarter (*usd pledged*) no resulta del todo satisfactoria ya que, como se aprecia en el resumen de las variables, presenta datos perdidos a pesar de que la variable original *pledged* no posee datos perdidos. La conversión realizada por Fixer (*usd_pledged_real*) resulta más completa y satisfactoria, luego será esta variable la que represente la cantidad recuadada en dólares.

Por tanto, se seleccionarán todas las variables excepto *goal*, *pledged*, *currency* y *usd pledged*.

3 Limpieza de los datos

Continuando con una limpieza de datos más exhaustiva se han realizado acciones de reducción del volumen de datos, conversión y creación de nuevas variables, tratamiento de datos perdidos y análisis de los valores extremos. A continuación, se describen estos procesos.

3.1 Reducción de casos

A parte de la selección de variables realizada y comentada anteriormente, se realiza una reducción del número de registros, bastante elevado. En concreto, teniendo en cuenta que uno de los intereses del análisis de estos datos es conocer el estado de los mismos para poder distinguir las características clave de los proyectos exitosos, los proyectos con estado indefinido (valor *undefined* en el campo *state*) serán eliminados.

3.2 Conversión y creación de nuevas variables

Otro proceso habitual en el preprocesado consiste en la conversión de variables y la creación de nuevos campos que permitan un análisis posterior más fácil de interpretar.

Se poseen dos campos, *launched* y *deadline*, que representan la fecha de lanzamiento y de final de la campaña de recaudación de los proyectos, respectivamente. Sin embargo, más que las fechas concretas de estos periodos, interesa una localización más general del lanzamiento del proyecto en el tiempo, como pueda ser el año en que este comenzó. Además, la fecha concreta de finalización de la recaudación puede ser sustituida por una variable más explicativa que represente la duración, en días, de la campaña de financiación que, sin duda, influirá en la cantidad recaudada. Por tanto, estas dos variables se sustituirán por dos nuevas variables *launch_year*, con el año de lanzamiento del proyecto, y *days_of_campaign*, con el número de días para la financiación desde el lanzamiento del proyecto.

3.3 Datos perdidos

En el resumen de las variables realizado durante la carga de los datos, entre otros, se apreciaban la cantidad de datos perdidos que posee cada variable. La única variable con datos perdidos, como ya se comentó, era *usd_pledged*, que fue descartada durante la etapa de selección de variables. Sin embargo, no es cierto que no se presenten más valores perdidos ya que, en la variable *country* se encuentran valores 'N, 0"', que no se corresponden con un país si no que codifican valores perdidos para el país de origen del proyecto. Si bien resulta imposible y carente de sentido realizar una imputación de los países, sí que se puede proceder a codificarlos de una forma más representativa con la palabra 'unknown'.

3.4 Valores extremos

Se ha realizado un análisis de los valores extremos de las variables numéricas del conjunto de datos: *backers*, *usd_goal_real*, *usd_pledged_real*, *days_of_campaign* (Tabla 1). Se encuentra una gran multitud de valores atípicamente altos en comparación con el resto de valores suelen ser cercanos a cero. Estos valores atípicos modifican notablemente las estadísticas y distribuciones de las variables.

Además, se realiza una comparación gráfica y numérica de la distribución de cada una de estas variables considerando los valores atípicos y una corrección de los mismos. Las variables sin valores atípicos resultan más contundentes y poseen una distribución con más sentido, lo que podría indicar que todos estos valores atípicos suponen errores de codificación que no representan realmente los valores correctos. Sin embargo, basta buscar algunos de los proyectos que presentan estos valores atípicos en Kickstarter para darse cuenta de que no se trata de errores en la creación del *dataset*, si no que realmente estos proyectos cuentan con valores tan altos en su número de patrocinadores y respecto a su financiación.

Por tanto, estos valores extremos no serán tratados ya que se asume, y será parte del análisis, que hay proyectos que no consiguen nada o muy poca financiación mientras que existen proyectos que sobrepasan con mucha diferencia los objetivos de financiación asumidos.

4 Análisis de los datos

Se procede a realizar un análisis de datos que determine, entre otros, la normalidad de las variables, la homogeneidad de la varianza entre diferentes grupos, la correlación entre variables, y una serie de pruebas estadísticas sobre la media. Este análisis de datos se centrará en la variable que cuantifica el dinero total recaudado por los proyectos (*usd_pledged_real*) ya que se considera la variable de mayor interés en este estudio.

4.1 Selección de los grupos de datos que se quieren analizar/comparar

Uno de los estudios entre grupos de datos que se va a realizar será el contraste de hipótesis entre las medias de dinero total recaudado por los proyectos estadounidenses y los proyectos británicos. Por tanto, se han separado ambos grupos de datos.

4.2 Comprobación de la normalidad y homogeneidad de la varianza

Para la comprobación de que los valores que toman nuestras variables cuantitativas provienende una población distribuida normalmente, se utilizará la prueba de normalidad de Anderson-Darling. Si en la prueba se obtiene un p-valor superior al nivel de significación prefijado **α= 0,05** (nivel de confianza del 95%), se considera que dicha variable sigue una distribución normal. Esto es particularmente útil porque la distribución normal tiene ciertas propiedades matemáticas de simetría que permiten estudiar los fenómenos que modela de acuerdo con unos patrones ya documentados.

Recordemos que, gracias a las propiedades simétricas, cuando la distribución es normal el 68%de las observaciones se encuentran en torno a la media en un radio marcado por la desviación estándar, el 95% entredos desviaciones estándar y el 99,7% entre tres. Es una regla empírica.

Las variables que no siguen una distribución normal son: *id*, *backers*, *usd_pledged_real*, *usd_goal_real*, *launch_year*, *days_of_campaign*. Es decir, ninguna da las variables continuas del conjunto de datos sigue una distribución normal (ver Gráfico 7).

En segundo lugar, se estudia la homogeneidad respecto a la recaudación final de los proyectos (*usd_pledged_real*) según los grupos conformados por la categoría principal del proyecto (*main_category*). En la siguiente prueba, la hipótesis nula consiste en que ambas varianzas son iguales.

```
        Fligner-Killeen test of homogeneity of variances

data:  usd_pledged_real by main_category
Fligner-Killeen:med chi-squared = 40053, df = 14, p-value < 2.2e-16
```

Dado que se obtiene un p-valor inferior a 0.05, rechazamos la hipótesis de que las varianzas deque las muestras por categorías son homogéneas. Hay diferencias significativas entre la variación de la recaudación según la categoría del proyecto. Luego, entre proyectos de la misma categoría hay diferencias que permiten recaudar más a unos que a otros.

4.3 Aplicación de pruebas estadísticas para comparar los grupos de datos

4.3.1 Correlación

Se procede a realizar un análisis de correlación entre las distintas variables numéricas (Gráfico 5) para determinar cuáles de ellas ejercen una mayor influencia sobre la suma total recaudada por el proyecto (*usd_pledged_real*). Para ello, se utilizará el coeficiente de correlación de Spearman puesto que, según hemos visto anteriormente, tenemos datos que no siguen una distribución normal.

```
                    estimate      p-value
id                -0.001723277 2.912314e-01
backers            0.958755792 0.000000e+00
usd_pledged_real   1.000000000 0.000000e+00
usd_goal_real      0.180384012 0.000000e+00
launch_year       -0.067902966 0.000000e+00
days_of_campaign   0.017190732 6.332847e-26
```

Con base en los resultados, se puede identificar cuáles son las variables más correlacionadas con la suma recaudada por la campaña en función de su aporte entre los valores -1 y +1. Para cada coeficiente de correlación se muestra también su p-valor asociado, puesto que éste puede dar información acerca del peso estadístico de la correlación obtenida.

Como cabía esperar, la variable *id* no está relacionada con la variable de estudio, ya que solo contiene el identificador del proyecto y no aporta ningún otro tipo de información. El año de lanzamiento (*launch_year*) y la duración de la campaña (*days_of_campaign*) no parecen tampoco ser muy relevantes en la cantidad de dinero que un proyecto puede recibir, señalando que la línea temporal no es tan importante como pudiera parecer.

La variable más correlacionada es el número de patrocinadores (*backers*), lo cual resulta lógico ya que cuantas más personas quieran invertir en un proyecto, más dinero conseguirá este, indicando que las aportaciones de los proyectos suelen estar distribuidas en forma de muchos patrocinadores que invierten pequeñas cantidades, en lugar de pocos patrocinadores con grandes sumas de dinero. Seguidamente, la segunda variable más destacable, es el objetivo de financiación marcado por los autores del proyecto (*usd_goal_real*) que, aun con menor importancia que el número de patrocinadores, indica que, a mayor objetivo fijado, mayor suele ser la cantidad recaudada, lo cual también es lógico.

4.3.2 Contraste de hipótesis sobre las medias

Esta prueba estadística se usa para realizar un contraste de hipótesis sobre dos muestras para determinar si el dinero total recaudado por los proyectos es superior dependiendo si el país es Estados Unidos o Gran Bretaña, los dos países más frecuentes en el conjunto de datos del que disponemos, como se aprecia en la Gráfica 1. Se construyen por tanto dos muestras: la primera de ellas corresponde a los valores de dinero recaudado en dólares por los proyectos estadounidenses, la segunda, para los valores de dinero recaudado por proyectos británicos. Dado que ambas muestras tienen tamaños n>30, por el Teorema del Límite Central podemos considerar que provienen de poblaciones normales y, por tanto, usar la prueba.

Se plantean dos hipótesis:

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 < 0$$

Donde **μ1** es la media de la población de la que se extrae la primera muestra y **μ2** es la media de la población de la que se extrae la segunda. Tomamos un nivel de confianza del 95%.

```
        One Sample t-test

data:  goal_usd_us
t = 52.664, df = 292626, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
  9310.302 10030.084
sample estimates:
mean of x
 9670.193
```

Puesto que obtenemos un p-valor menor que el valor de significación fijado, rechazamos la hipótesis nula. Por tanto, podemos concluir, con un nivel de confianza del 95%, que la media del dinero recaudado por los proyectos americanos no es inferior a la media de los proyectos británicos.

4.3.3 Regresión lineal

Otra forma de estudiar la relación entre variables es construir un modelo de regresión lineal que prediga la cantidad de dinero que recauda cada proyecto en función de diferentes variables. En este caso, se ha definido el modelo de regresión lineal sobre la variable *usd_pledged_year* considerado como explicativas las variables *usd_goal_real*, *backers*, *launch_year* (las tres más correlacionadas con la cantidad recaudada, como ya se vio anteriormente), además de la categoría principal de los proyectos, *main_category*. Los resultados muestran que el objetivo de recaudación no es realmente significativo para la cantidad final recaudada, mientras que sí lo son el número de patrocinadores (cuantos más obtenga un proyecto mayor será la cantidad recaudada) y el año de lanzamiento (los proyectos de los últimos años son capaces de recaudar más dinero). Por otro lado, en cuanto a la categoría de los proyectos parece que solo influyen significativamente en la cantidad de dinero recaudada si el proyecto es de índole tecnológica, publicitaria, de videojuegos, audiovisual, de diseño o de cómics, donde los proyectos de tecnología, diseño o películas afectan positivamente a la cantidad total recaudada, mientras que los proyectos de publicidad, videojuegos o cómics suelen recaudar menos dinero.

El coeficiente de determinación, R^2 que se consigue es 0.56, lo que indica que con este modelo se consigue explicar el 56% de la variabilidad de los datos de recaudación de los proyectos. Esto no constituye un modelo muy adecuado para realzar predicciones, pero sí que ha servido para realizar un estudio desde otro enfoque sobre las variables que más afectan a la cantidad de dinero que recauda cada proyecto.

```
Call:
lm(formula = usd_pledged_real ~ usd_goal_real + main_category +
    backers + launch_year, data = ks_df)

Residuals:
     Min       1Q   Median       3Q      Max
-7761401    -2053     -622      808 14413896

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             -5.977e+05  1.068e+05  -5.597 2.18e-08 ***
usd_goal_real            1.294e-04  8.475e-05   1.527  0.12675
main_categoryComics     -3.698e+03  6.802e+02  -5.436 5.44e-08 ***
main_categoryCrafts     -7.339e+02  7.354e+02  -0.998  0.31830
main_categoryDance       2.927e+02  1.043e+03   0.281  0.77898
main_categoryDesign      5.955e+03  5.005e+02  11.897  < 2e-16 ***
main_categoryFashion     8.389e+02  5.367e+02   1.563  0.11800
main_categoryFilm & Video 1.290e+03  4.322e+02   2.985  0.00283 **
main_categoryFood        8.443e+02  5.252e+02   1.608  0.10792
main_categoryGames      -3.481e+03  4.830e+02  -7.209 5.66e-13 ***
main_categoryJournalism -4.623e+02  9.428e+02  -0.490  0.62388
main_categoryMusic      -7.107e+01  4.494e+02  -0.158  0.87433
main_categoryPhotography 5.839e+02  6.809e+02   0.858  0.39110
main_categoryPublishing -9.447e+02  4.691e+02  -2.014  0.04401 *
main_categoryTechnology  8.404e+03  4.922e+02  17.074  < 2e-16 ***
main_categoryTheater     5.999e+02  6.785e+02   0.884  0.37665
backers                  7.543e+01  1.082e-01 697.058  < 2e-16 ***
launch_year              2.968e+02  5.302e+01   5.597 2.18e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 60110 on 375081 degrees of freedom
Multiple R-squared:  0.5676,    Adjusted R-squared:  0.5676
F-statistic: 2.896e+04 on 17 and 375081 DF,  p-value: < 2.2e-16
```

5 Representación de los resultados a partir de tablas y gráficas

A continuación, se muestran y comentan las visualizaciones de algunas características relevantes en este estudio, como la representación de las distribuciones de las variables o el gráfico de correlación.

5.1 Variables discretas

Con un simple diagrama de barras se representan visualmente las variables categóricas. Como vemos, los proyectos más comunes son los de naturaleza audiovisual o musical. Además, lo más común es que los proyectos fracasen en su intento de conseguir la financiación marcada como objetivo, aunque también existe una gran cantidad de proyectos que consiguen rebasar su objetivo. Por último, los proyectos provienen, en su gran mayoría, de Estados Unidos.

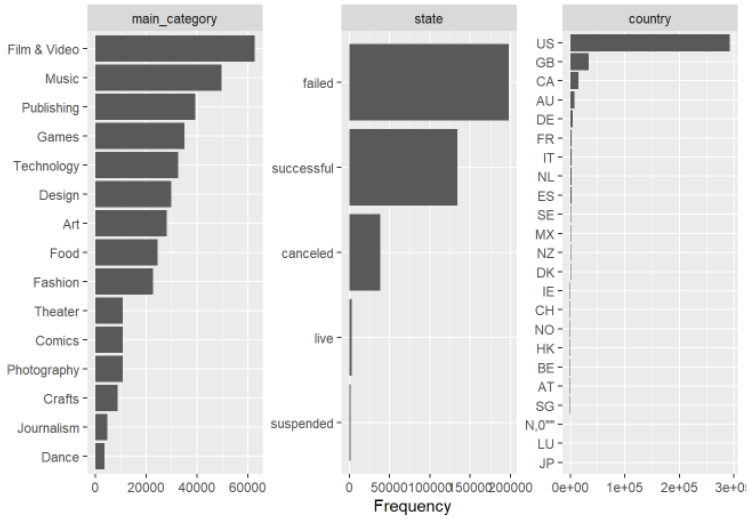

Gráfico 1. Diagrama de barras de las variables categóricas.

5.2 Detección de valores anómalos

Tal como se comentó, las variables numéricas presentan numerosos valores atípicos que provocan que sus parámetros estadísticos, como la media, aumenten notablemente, tal como vemos en la tabla y gráficas siguientes. Sin embargo, tal como se dijo, no se deben tratar estos valores atípicos ya que se corresponden con datos reales de proyectos, y hacerlo supondría una pérdida de información muy sensible, ya que resulta de interés esa distinción de los proyectos que consiguen valores extremadamente altos de dinero recaudado.

variables <chr>	outliers_cnt <int>	outliers_ratio <dbl>	outliers_mean <dbl>	with_mean <dbl>	without_mean <dbl>
backers	44713	11.92779	722.38823	106.68728	23.30168
usd_goal_real	44700	11.92432	327236.95622	45863.17875	7768.75340
usd_pledged_real	50558	13.48702	58217.57049	9120.80292	1466.81448
days_of_campaign	76235	20.33671	46.37488	34.49768	31.46562

Tabla 1. Diagnóstico de valores extremos.

Se presentan además las distribuciones de estas variables si se corrigieran los valores extremos, aunque, como ya se ha dicho, no conviene realizarlo. Tal como se aprecia en las siguientes gráficas la distribución del número de patrocinadores, del objetivo de recaudación y de la suma recaudada tras tratar los valores atípicos son distribuciones asimétricas a la izquierda, que se

asemejan en forma a la distribución de Poisson, aunque habría que comprobar si realmente siguen dicha distribución con pruebas estadísticas. Sin embargo, con los datos atípicos sin tratar estas variables tienen valores mayoritariamente nulos o cercanos a 0, y algunos valores excepcionales mucho más altos.

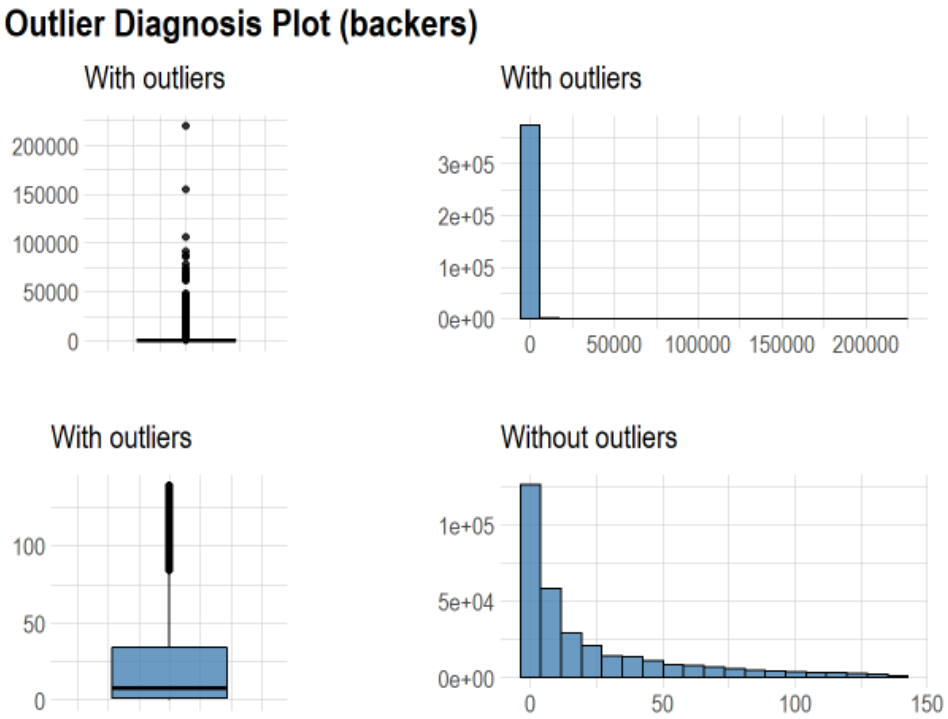

Gráfico 2. Diagnóstico de valores atípicos del número de patrocinadores.

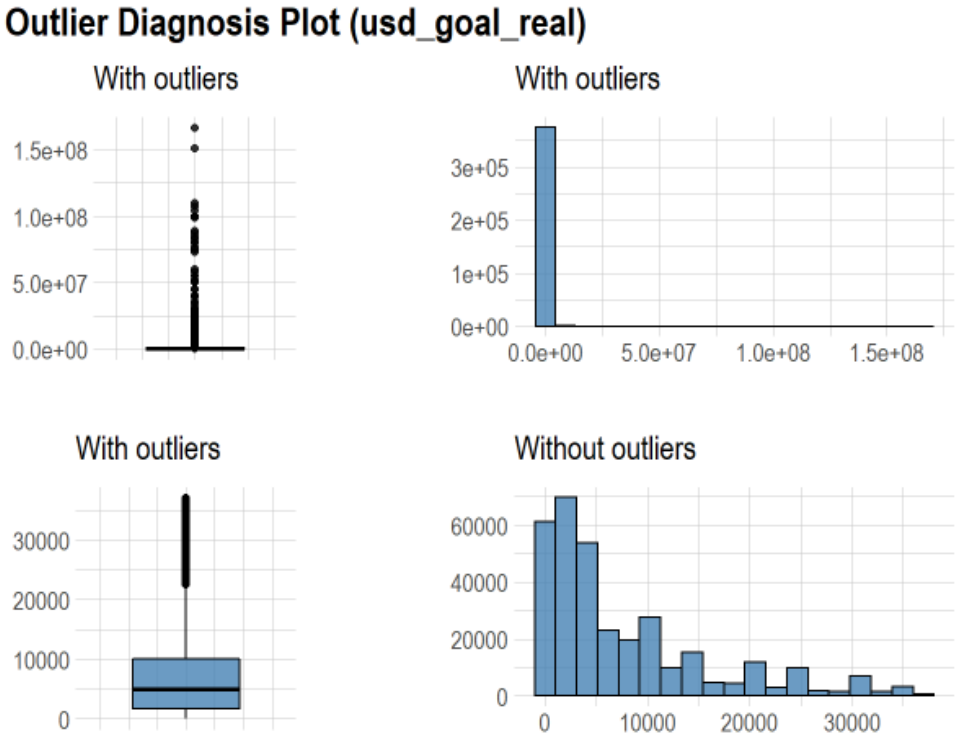

Gráfico 3. Diagnóstico de valores atípicos del objetivo de financiación de los proyectos.

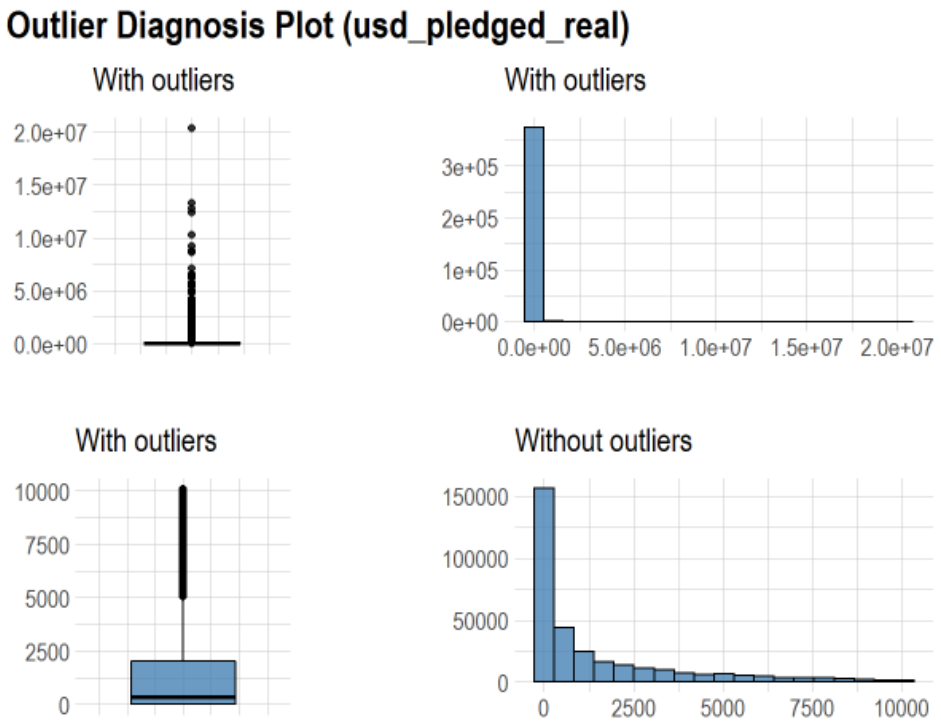

Gráfico 4. Diagnóstico de valores atípicos de la cantidad de dinero recaudada por los proyectos.

Por otro lado, el Gráfico 5 con los valores atípicos corregidos del periodo de campaña muestran una distribución con un pico central de gran densidad, lo que podría asemejarse a una distribución normal con muy baja varianza. Sin embargo, hay proyectos excepcionalmente largos que se deben tener en cuenta.

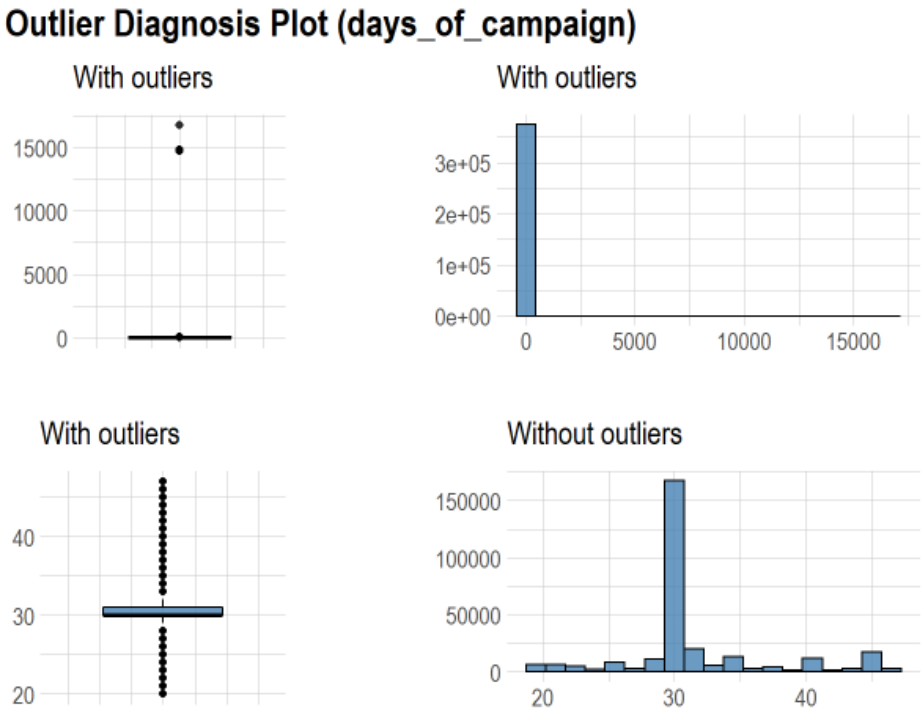

Gráfico 5. Diagnóstico de valores atípicos de la duración del periodo de campaña de financiación.

5.3 Correlación de variables numéricas

La matriz de correlación de variables numéricas muestra únicamente dos relaciones importantes. En primer lugar, la más fuerte de todas, la ya comentada anteriormente entre el número de patrocinadores y la cantidad de dinero recaudada por un proyecto, que indica que, a mayor número de patrocinadores, más dinero se consigue. En segundo lugar, la relación entre la duración de la campaña de financiación y el año de lanzamiento del proyecto, que indica que, conforme avanzan los años las campañas tienden a ser más cortas.

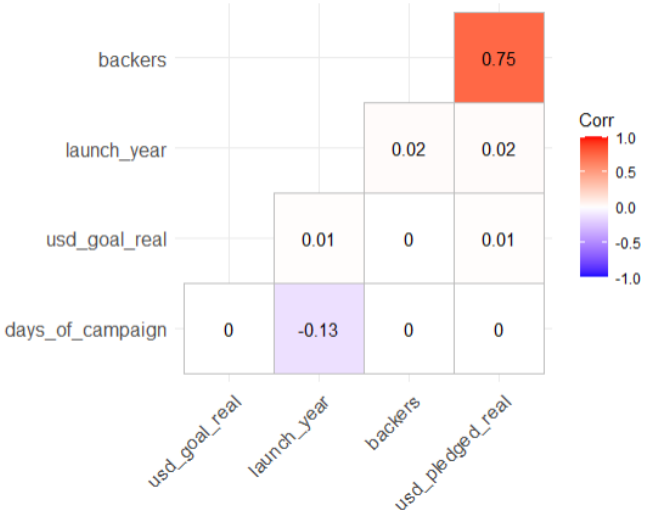

Gráfico 5. Matriz de correlación entre las variables numéricas.

5.4 Pruebas de normalidad

Los gráficos cuantil-cuantil son una forma visual de inspeccionar si una variable sigue o no una distribución normal. Una variable que sigue una distribución normal tendrá sus datos representados sobre la línea que también se dibuja, que representa el cuantil de una distribución normal. Como vemos, se confirman visualmente las conclusiones obtenidas con los tests estadísticos de normalidad: ninguna de las variables numéricas sigue una distribución normal.

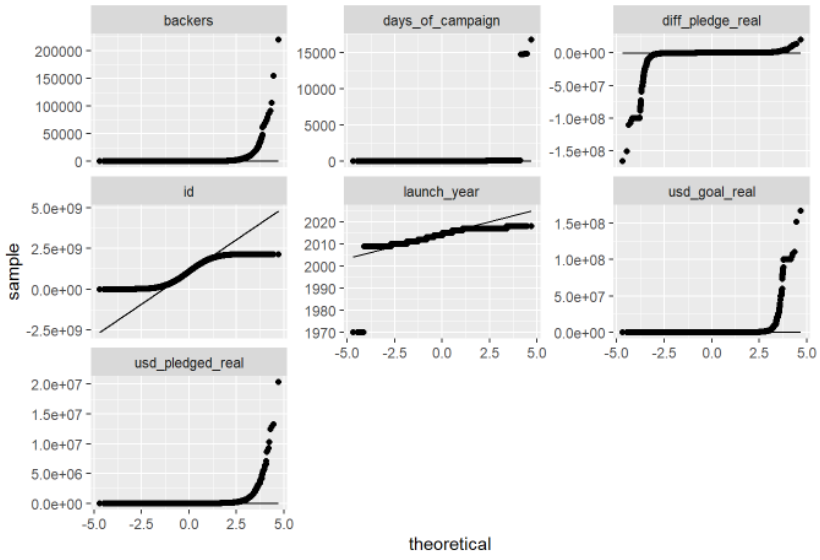

Gráfico 6. Diagramas cuantil-cuantil para comprobar la normalidad de las variables.

6 Conclusiones

Hemos podido apreciar que el conjunto de datos de Kickstarter tiene una historia y cantidad relevante de registros que permiten realizar diversos análisis y de esta manera tener conclusiones relevantes acerca de cómo se comportan los proyectos de *crowdfunding*.

Tras realizar la etapa de limpieza de los datos, que ha contado con etapas como la selección de datos, el filtrado de registros, la creación de nuevas variables, el tratamiento de los datos perdidos y el análisis de los valores extremos, se ha conseguido un *dataset* limpio y procesado que se ha guardado en un nuevo fichero CSV.

En cuanto al análisis de datos, en primer lugar, se evaluó la normalidad de los datos encontrando que la mayoría no tienen una distribución normal, presentando un sesgo hacia la izquierda (ver imágenes de apoyo). También se analizó la homogeneidad de la varianza de la recaudación total de los proyectos según la categoría principal de los mismos, encontrando que dichas varianzas no son homogéneas de acuerdo con la prueba realizada.

Seguidamente se han aplicado un total de tres pruebas estadísticas para comparar distintas variables y grupos de datos. En primer lugar, con el análisis de los coeficientes de correlación se ha determinado que la variable más relacionada con la cantidad de dinero recaudado es el número de patrocinadores, siendo la relación entre ambas positiva.

La segunda prueba estadística realizada fue el contraste de hipótesis para comparar la media de dinero recaudado entre proyectos de Estados Unidos y proyectos de Gran Bretaña, donde se ha encontrado que la cantidad media de dinero recaudada por los proyectos de Estados Unidos es significativamente mayor que la de los británicos.

Finalmente, un modelo de regresión lineal con la recaudación total de cada proyecto como variable dependiente confirmaba la relación ya predicha por el análisis de correlación con respecto a los patrocinadores, así como que los proyectos más recientes tienden a recaudar más dinero. También se concluye a partir de este modelo que los proyectos sobre tecnología, cine o diseño suelen recaudar más dinero, mientras que los proyectos de videojuegos, comics o publicidad presentan más dificultades.

Para un siguiente análisis sería prometedor combinar estos datos con otras fuentes de información como valores demográficos y detalles concretos de los proyectos financiados.

7 Bibliografía

Kickstarter - Wikipedia, la enciclopedia libre. (s. f.). Recuperado 30 de mayo de 2021, de https://es.wikipedia.org/wiki/Kickstarter

Kickstarter e impuestos — Kickstarter. (s. f.). Recuperado 30 de mayo de 2021, de https://www.kickstarter.com/help/taxes?lang=es

ANEXO: Contribuciones

Contribuciones	Firma
Investigación previa	ILR JSR
Redacción de las respuestas	ILR JSR
Desarrollo de código	ILR JSR