



Universitat
de les Illes Balears

21746 - Data Mining

Final Project

Steam Successful Indie Games Study

Iván Pulgar Rodas
Jordi Sevilla Marí
Nahuel Vazquez
Yelyzaveta Denysova
Xiaozhe Cheng
Gabriel Oliver Artigues
Arturo Mus Mejías

2026-01-10

Contents

1	Introduction	1
1.1	Explanation of the Attributes	1
1.2	Objectives	1
2	Procesing de data	1
2.1	Handling of NA values	4
3	Requeriments for Successful Indie Games	6
3.1	Obtain Unique values for different columns	7

1 Introduction

1.1 Explanation of the Attributes

The dataset we will be working with contains a total number of 94948 observations and 47 columns or variables. The columns that will be used are described below:

- **appid**: Unique identifier of the game on Steam. [num]
- **name**: Name of the game. [text]
- **'released_date'**: Represents the date where the game was released. [time]
- **'required_age'**: Corresponds to the minimum age required to play the game. [num]
- **price**: How much the game costs. If its 0 it means that the game is Free to Play. [num]
- **dlc_count**: Ammount of DLCs (Downloadable Contents) the game has. [num]
- **support_url**: URL to the support page of the game. [text]
- **windows**: Determines if the game runs in windows. [categorical]
- **mac**: Determines if the game runs in mac. [categorical]
- **linux**: Determines if the game runs in linux. [categorical]
- **metacritic_score**: Metacritic score based on critical reviews (reviews from professionalss). By performing an investigation we think that an score of 0 means that when the scraping of the data was done there where no reviews for that game yet. [num]
- **achievements**: Number of achievements the game has. [num]
- **'recommendations'**: Ammount of user recommendations. [num]
- **supported_languages**: List of languages that the game supports.[**NO SÉ QUE CATEGORÍA DARLE**]
- **packages**: Aavailable packages for the game. It contains the name and a description of the package and the names, descriptions and subprices of the subpackages. [**LO MISMO QUE ARRIBA**]
- **Developers**: List of developers associated with the game. [**LO MISMO QUE ARRIBA**]
- **publishers**: List of publishers associated with the game. [**LO MISMO QUE ARRIBA**]
- **categories**: List of categories that the game has. [**LO MISMO QUE ARRIBA**]
- **genres**: List of genres that the game belongs to. [**LO MISMO QUE ARRIBA**]
- **positive**: Ammount of positive votes the game has. [num]
- **negative**: Ammount of negative votes the game has. [num]
- **estimated_owners**: Estimated owners of the game. [text]
- **average_playtime_forever**: Average playtime since March 2009 measured in minutes. [num]
- **average_playtime_2weeks**: Average playtime in the last two weeks measured in minutes. [num]
- **median_playtime_forever**: Median playtime since March 2009 measured in minutes. [num]
- **median_playtime_2weeks**: Median playtime in the last two weeks measured in minutes.[num]
- **peak_ccu**: Number of current users playing the day before the data was scrapped. [num]
- **tags**: List of tags the game has with its name and its key. [**NO SE QUE CATEGORIA DARLE**]
- **pct_pos_total**: Percentage of all reviews that are positive. [num]
- **num_reviews_total**: Nummber of the total reviews the game has. [num]

1.2 Objectives

TODO: ADD OBJECTIVES OF STUDY

2 Procesing de data

General look of the dataset

```
summary(steam)
```

```
##      appid          name      release_date      required_age
## Min.   :      20  Length:94948  Length:94948  Min.   :-1.0000
## 1st Qu.: 887338  Class :character  Class :character  1st Qu.: 0.0000
## Median :1591145  Mode  :character  Mode  :character  Median : 0.0000
## Mean   :1707530                                     Mean   : 0.1783
## 3rd Qu.:2491702                                     3rd Qu.: 0.0000
## Max.   :3570420                                     Max.   :21.0000
##
##      price      dlc_count      detailed_description about_the_game
## Min.   :  0.000  Min.   :  0.0000  Length:94948  Length:94948
## 1st Qu.:  0.990  1st Qu.:  0.0000  Class :character  Class :character
## Median :  3.990  Median :  0.0000  Mode  :character  Mode  :character
## Mean   :  6.911  Mean   :  0.5632                                     Mean   : 0.1783
## 3rd Qu.:  9.990  3rd Qu.:  0.0000                                     3rd Qu.: 0.0000
## Max.   : 999.980  Max.   :3427.0000                                     Max.   :21.0000
##
## short_description  reviews      header_image      website
## Length:94948      Length:94948  Length:94948  Length:94948
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
##
## support_url      support_email      windows      mac
## Length:94948      Length:94948  Length:94948  Length:94948
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
##
##      linux      metacritic_score metacritic_url      achievements
## Length:94948  Min.   : 0.000  Length:94948  Min.   :  0.00
## Class :character  1st Qu.: 0.000  Class :character  1st Qu.:  0.00
## Mode  :character  Median : 0.000  Mode  :character  Median :  2.00
##                  Mean   : 2.764  Mean   : 19.54
##                  3rd Qu.: 0.000  3rd Qu.: 19.00
##                  Max.   :97.000  Max.   :9821.00
##
## recommendations  notes      supported_languages full_audio_languages
## Min.   :      0  Length:94948  Length:94948  Length:94948
## 1st Qu.:      0  Class :character  Class :character  Class :character
## Median :      0  Mode  :character  Mode  :character  Mode  :character
## Mean   :  1022                                     Mean   : 0.1783
## 3rd Qu.:      0                                     3rd Qu.: 0.0000
## Max.   :4401572                                     Max.   :21.0000
##
##      packages      developers      publishers      categories
## Length:94948      Length:94948  Length:94948  Length:94948
## Class :character  Class :character  Class :character  Class :character
```

```

## Mode :character Mode :character Mode :character Mode :character
##
##
##
##
##      genres      screenshots      movies      user_score
## Length:94948    Length:94948    Length:94948    Min. : 0.00000
## Class :character Class :character Class :character 1st Qu.: 0.00000
## Mode :character Mode :character Mode :character  Median : 0.00000
##                                     Mean  : 0.03097
##                                     3rd Qu.: 0.00000
##                                     Max.  :100.00000
##
##      score_rank      positive      negative      estimated_owners
## Min. : 98.00    Min. : 0    Min. : 0.0    Length:94948
## 1st Qu.: 99.00    1st Qu.: 0    1st Qu.: 0.0    Class :character
## Median : 99.00    Median : 8    Median : 2.0    Mode :character
## Mean : 99.13    Mean : 1218    Mean : 202.1
## 3rd Qu.:100.00    3rd Qu.: 51    3rd Qu.: 15.0
## Max. :100.00    Max. :7480813    Max. :1135108.0
## NA's :94909
## average_playtime_forever average_playtime_2weeks median_playtime_forever
## Min. : 0.0    Min. : 0.000    Min. : 0.0
## 1st Qu.: 0.0    1st Qu.: 0.000    1st Qu.: 0.0
## Median : 0.0    Median : 0.000    Median : 0.0
## Mean : 108.6    Mean : 4.757    Mean : 108.4
## 3rd Qu.: 0.0    3rd Qu.: 0.000    3rd Qu.: 0.0
## Max. :1462997.0    Max. :18568.000    Max. :1462997.0
##
## median_playtime_2weeks      discount      peak_ccu
## Min. : 0.000    Min. : 0.000    Min. :0.000e+00
## 1st Qu.: 0.000    1st Qu.: 0.000    1st Qu.:0.000e+00
## Median : 0.000    Median : 0.000    Median :0.000e+00
## Mean : 5.018    Mean : 4.307    Mean :9.285e+01
## 3rd Qu.: 0.000    3rd Qu.: 0.000    3rd Qu.:0.000e+00
## Max. :18568.000    Max. :100.000    Max. :1.212e+06
##
##      tags      pct_pos_total      num_reviews_total      pct_pos_recent
## Length:94948    Min. : -1.00    Min. : -1    Min. : -1.000
## Class :character 1st Qu.: -1.00    1st Qu.: -1    1st Qu.: -1.000
## Mode :character  Median : 58.00    Median : 15    Median : -1.000
##                                     Mean : 44.63    Mean : 1448    Mean : 5.328
##                                     3rd Qu.: 84.00    3rd Qu.: 80    3rd Qu.: -1.000
##                                     Max. :100.00    Max. :8632939    Max. :100.000
##
## num_reviews_recent
## Min. : -1.00
## 1st Qu.: -1.00
## Median : -1.00
## Mean : 16.88
## 3rd Qu.: -1.00
## Max. :96473.00
##

```

Explanation of first looks of it, bad formatting, NA's, negative values...

2.1 Handling of NA values

The attributes with missing values are:

```
na_counts <- steam %>% summarise_all(~ sum(is.na(.)))

print(na_counts)

##   appid name release_date required_age price dlc_count detailed_description
## 1      0      0           0           0      0           0
##   about_the_game short_description reviews header_image website support_url
## 1              0              0      0           0           0
##   support_email windows mac linux metacritic_score metacritic_url achievements
## 1              0      0  0  0           0           0
##   recommendations notes supported_languages full_audio_languages packages
## 1              0      0              0           0
##   developers publishers categories genres screenshots movies user_score
## 1              0      0           0      0           0
##   score_rank positive negative estimated_owners average_playtime_forever
## 1      94909      0           0           0
##   average_playtime_2weeks median_playtime_forever median_playtime_2weeks
## 1              0              0
##   discount peak_ccu tags pct_pos_total num_reviews_total pct_pos_recent
## 1      0      0  0           0           0
##   num_reviews_recent
## 1              0
```

The columns with missing values are as follows:

```
add_range_minmax <- function(df, column, min_column, max_column) {
  column_quo <- enquo(column)

  df %>%
    mutate(
      "{min_column}" := as.integer(str_extract(!column_quo, "\\s*\\d+")),
      "{max_column}" := as.integer(str_extract(!column_quo, "\\d+\\s*$"))
    )
}

cleaned_steam <- cleaned_steam %>%
  add_range_minmax(
    estimated_owners,
    "estimated_owners_min",
    "estimated_owners_max"
  ) %>%
  select(-estimated_owners)
```

LOs juegos con número de reseñas = -1 creemos que son porque el scrappear ha fallado durante su ejecución. Si vamos a trabajar con las reseñas podríamos decir en la presentación que vamos a probar de volver a intentar scrappear la información

```
## tibble [88,982 x 30] (S3: tbl_df/tbl/data.frame)
## $ appid      : int [1:88982] 2556940 449940 1287250 866510 870990 439260 388390 224356
## $ name       : chr [1:88982] "! Shakabula *" "! That Bastard Is Trying To Steal Our Go
## $ release_date : chr [1:88982] "2023-10-13" "2016-03-03" "2020-04-28" "2018-06-06" ...
## $ required_age : int [1:88982] 0 0 0 0 0 0 0 0 0 0 ...
## $ price      : num [1:88982] 14.99 2.99 19.99 1.99 0.99 ...
## $ dlc_count   : int [1:88982] 0 0 0 39 0 0 0 1 3 0 ...
## $ windows     : chr [1:88982] "True" "True" "True" "True" ...
## $ mac         : chr [1:88982] "False" "False" "False" "False" ...
## $ linux       : chr [1:88982] "False" "True" "False" "False" ...
## $ metacritic_score : int [1:88982] 0 0 0 0 0 0 0 0 0 0 ...
## $ achievements : int [1:88982] 0 0 9 4997 2021 0 0 19 13 5 ...
## $ recommendations : int [1:88982] 0 0 0 495 0 0 0 108 0 0 ...
## $ supported_languages : chr [1:88982] "["English"]" "["English"]" "["English', 'Simplified Chin
## $ packages     : chr [1:88982] "["{'title': 'Buy ! Shakabula *', 'description': '', 'subs
## $ developers    : chr [1:88982] "["Skermunkel']" "["WTFOMGames']" "["Andreev Worlds']" "["
## $ publishers    : chr [1:88982] "["Skermunkel']" "["WTFOMGames']" "["Andreev Worlds']" "["
## $ categories    : chr [1:88982] "["Single-player', 'Full controller support', 'Steam Clou
## $ genres        : chr [1:88982] "["Action', 'Indie', 'RPG', 'Early Access']" "["Action',
## $ positive      : int [1:88982] 0 57 45 410 25 83 37 126 0 0 ...
## $ negative      : int [1:88982] 4 78 34 180 32 18 102 10 0 0 ...
## $ average_playtime_forever : int [1:88982] 0 312 0 360 0 0 244 0 0 0 ...
## $ average_playtime_2weeks : int [1:88982] 0 0 0 0 0 0 0 0 0 0 ...
## $ median_playtime_forever : int [1:88982] 0 391 0 378 0 0 244 0 0 0 ...
## $ median_playtime_2weeks : int [1:88982] 0 0 0 0 0 0 0 0 0 0 ...
## $ peak_ccu      : int [1:88982] 0 0 0 6 0 0 0 0 0 0 ...
## $ tags          : chr [1:88982] "{"'Early Access': 213, 'Action': 193, 'RPG': 187, 'JRPG':
## $ pct_pos_total  : int [1:88982] -1 55 61 71 55 82 55 91 95 66 ...
## $ num_reviews_total : int [1:88982] -1 68 62 495 18 101 20 108 281 12 ...
## $ estimated_owners_min : int [1:88982] 0 50000 0 100000 0 0 100000 20000 0 0 ...
## $ estimated_owners_max : int [1:88982] 20000 100000 20000 200000 20000 20000 200000 50000 0 0 ..
```

```
##      appid      name      release_date      required_age
## Min.   :      20   Length:88982      Length:88982      Min.   :-1.0000
## 1st Qu.: 852782   Class :character   Class :character   1st Qu.: 0.0000
## Median :1522535   Mode  :character   Mode  :character   Median : 0.0000
## Mean   :1655079                                     Mean   : 0.1826
## 3rd Qu.:2429338                                     3rd Qu.: 0.0000
## Max.   :3542350                                     Max.   :21.0000
##      price      dlc_count      windows      mac
## Min.   :  0.000   Min.   :  0.0000   Length:88982      Length:88982
## 1st Qu.:  0.990   1st Qu.:  0.0000   Class :character   Class :character
## Median :  4.990   Median :  0.0000   Mode  :character   Mode  :character
## Mean   :  7.349   Mean   :  0.5957
## 3rd Qu.:  9.990   3rd Qu.:  0.0000
## Max.   : 999.980   Max.   :3427.0000
##      linux      metacritic_score      achievements      recommendations
## Length:88982   Min.   : 0.00   Min.   :  0.00   Min.   :      0
## Class :character 1st Qu.: 0.00   1st Qu.:  0.00   1st Qu.:      0
## Mode  :character Median : 0.00   Median :  5.00   Median :      0
## Mean   :  2.91   Mean   : 20.67   Mean   :  1013
## 3rd Qu.:  0.00   3rd Qu.: 20.00   3rd Qu.:      0
## Max.   :97.00   Max.   :9821.00   Max.   :4401572
## supported_languages packages      developers      publishers
```

```
## Length:88982      Length:88982      Length:88982      Length:88982
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
## categories      genres      positive      negative
## Length:88982    Length:88982    Min.   :      0    Min.   :      0.0
## Class :character Class :character  1st Qu.:      1    1st Qu.:      0.0
## Mode  :character Mode  :character  Median :     10    Median :      2.0
##                                     Mean  :    1267    Mean  :    208.3
##                                     3rd Qu.:     59    3rd Qu.:    17.0
##                                     Max.   :7480813    Max.   :1135108.0
## average_playtime_forever average_playtime_2weeks median_playtime_forever
## Min.   :      0.0      Min.   :      0.000    Min.   :      0.0
## 1st Qu.:      0.0      1st Qu.:      0.000    1st Qu.:      0.0
## Median :      0.0      Median :      0.000    Median :      0.0
## Mean   :    115.6      Mean   :      4.998    Mean   :    115.5
## 3rd Qu.:      0.0      3rd Qu.:      0.000    3rd Qu.:      0.0
## Max.   :1462997.0      Max.   :18568.000    Max.   :1462997.0
## median_playtime_2weeks peak_ccu      tags      pct_pos_total
## Min.   :      0.000    Min.   :0.000e+00    Length:88982    Min.   : -1.00
## 1st Qu.:      0.000    1st Qu.:0.000e+00    Class :character 1st Qu.: -1.00
## Median :      0.000    Median :0.000e+00    Mode  :character Median : 60.00
## Mean   :      5.277    Mean   :9.778e+01                      Mean   : 45.39
## 3rd Qu.:      0.000    3rd Qu.:0.000e+00                      3rd Qu.: 84.00
## Max.   :18568.000      Max.   :1.212e+06                      Max.   :100.00
## num_reviews_total estimated_owners_min estimated_owners_max
## Min.   :      -1      Min.   :      0      Min.   :      0
## 1st Qu.:      -1      1st Qu.:      0      1st Qu.:    20000
## Median :      15      Median :      0      Median :    20000
## Mean   :    1320      Mean   :    59038      Mean   :   143056
## 3rd Qu.:      81      3rd Qu.:      0      3rd Qu.:    20000
## Max.   :8632939      Max.   :200000000      Max.   :500000000
```

3 Requeriments for Successful Indie Games

Create helper function for extracting unique elements from entries with a list of values. All lists follow the same pattern: ['e11', 'e12',...] being [] for empty lists.

```
# Function helper
get_unique_list_values <- function(x) {
  x <- x[x != "[]"]
  x <- gsub("^\\[|\\]\\$", "", x) # quitar [ y ]
  x <- gsub("'", "", x) # quitar ' y '

  elements <- strsplit(x, ",") # dividir por comas

  vals <- trimws(unlist(elements, use.names = FALSE)) # aplanar y trim
  vals <- vals[vals != ""]

  sort(unique(vals))
}
```


Create helper function for extracting unique tags. These are formatted differently and follow dictionary-like pattern where each element is a key-value pair being [] for empty entries.

```
get_unique_dict_keys <- function(x) {
  x <- x[x != "[]"]

  keys_list <- str_extract_all(x, "(?<=')[^']+(?='\\s*:.)")

  keys <- trimws(unlist(keys_list, use.names = FALSE))
  keys <- keys[keys != ""]
  sort(unique(keys))
}
```

3.1 Obtain Unique values for different columns

```
steam_categories <- get_unique_list_values(cleaned_steam$categories)
steam_categories
```

```
## [1] "Captions available"      "Co-op"
## [3] "Commentary available"    "Cross-Platform Multiplayer"
## [5] "Family Sharing"          "Full controller support"
## [7] "HDR available"           "In-App Purchases"
## [9] "Includes level editor"   "Includes Source SDK"
## [11] "LAN Co-op"               "LAN PvP"
## [13] "MMO"                     "Multi-player"
## [15] "Online Co-op"            "Online PvP"
## [17] "Partial Controller Support" "PvP"
## [19] "Remote Play on Phone"    "Remote Play on Tablet"
## [21] "Remote Play on TV"       "Remote Play Together"
## [23] "Shared/Split Screen"     "Shared/Split Screen Co-op"
## [25] "Shared/Split Screen PvP" "Single-player"
## [27] "Stats"                   "Steam Achievements"
## [29] "Steam Cloud"             "Steam Leaderboards"
## [31] "Steam Timeline"          "Steam Trading Cards"
## [33] "Steam Turn Notifications" "Steam Workshop"
## [35] "SteamVR Collectibles"    "Tracked Controller Support"
## [37] "Valve Anti-Cheat enabled" "VR Only"
## [39] "VR Support"              "VR Supported"
```

```
steam_genres <- get_unique_list_values(cleaned_steam$genres)
steam_genres
```

```
## [1] "360 Video"                "Accounting"                "Action"
## [4] "Adventure"                "Animation & Modeling"      "Audio Production"
## [7] "Casual"                   "Design & Illustration"     "Documentary"
## [10] "Early Access"             "Education"                  "Episodic"
## [13] "Free To Play"             "Game Development"           "Gore"
## [16] "Indie"                    "Massively Multiplayer"      "Movie"
## [19] "Nudity"                   "Photo Editing"              "Racing"
## [22] "RPG"                      "Sexual Content"             "Short"
## [25] "Simulation"               "Software Training"          "Sports"
```

```
## [28] "Strategy"          "Tutorial"          "Utilities"
## [31] "Video Production"  "Violent"           "Web Publishing"
```

```
steam_tags <- get_unique_dict_keys(cleaned_steam$tags)
steam_tags
```

```
## [1] "1980s"          "2.5D"
## [3] "2D"             "2D Fighter"
## [5] "2D Platformer"  "360 Video"
## [7] "3D"             "3D Fighter"
## [9] "3D Platformer"  "3D Vision"
## [11] "4 Player Local" "4X"
## [13] "6DOF"           "8-bit Music"
## [15] "Abstract"       "Action"
## [17] "Action Roguelike" "Action RPG"
## [19] "Action RTS"     "Action-Adventure"
## [21] "Addictive"      "Adventure"
## [23] "Agricuture"     "Aliens"
## [25] "Alternate History" "Ambient"
## [27] "America"        "Animation & Modeling"
## [29] "Anime"          "Arcade"
## [31] "Archery"        "Arena Shooter"
## [33] "Artificial Intelligence" "Assassin"
## [35] "Asymmetric VR"  "Asynchronous Multiplayer"
## [37] "Atmospheric"    "ATV"
## [39] "Audio Production" "Auto Battler"
## [41] "Automation"     "Automobile Sim"
## [43] "Base-Building"  "Baseball"
## [45] "Based On A Novel" "Basketball"
## [47] "Battle Royale"  "Beautiful"
## [49] "Benchmark"      "Bikes"
## [51] "Birds"          "Blood"
## [53] "BMX"            "Board Game"
## [55] "Boomer Shooter" "Boss Rush"
## [57] "Bowling"        "Boxing"
## [59] "Building"       "Bullet Hell"
## [61] "Bullet Time"    "Capitalism"
## [63] "Card Battler"   "Card Game"
## [65] "Cartoon"        "Cartoony"
## [67] "Casual"         "Cats"
## [69] "Character Action Game" "Character Customization"
## [71] "Chess"          "Choices Matter"
## [73] "Choose Your Own Adventure" "Cinematic"
## [75] "City Builder"   "Class-Based"
## [77] "Classic"        "Clicker"
## [79] "Co-op"          "Co-op Campaign"
## [81] "Coding"         "Cold War"
## [83] "Collectathon"   "Colony Sim"
## [85] "Colorful"       "Combat"
## [87] "Combat Racing"  "Comedy"
## [89] "Comic Book"     "Competitive"
## [91] "Conspiracy"     "Controller"
## [93] "Conversation"   "Cooking"
## [95] "Cozy"           "Crafting"
```

## [97]	"Creature Collector"	"Cricket"
## [99]	"Crime"	"Crowdfunded"
## [101]	"CRPG"	"Cult Classic"
## [103]	"Cute"	"Cyberpunk"
## [105]	"Cycling"	"Dark"
## [107]	"Dark Comedy"	"Dark Fantasy"
## [109]	"Dark Humor"	"Dating Sim"
## [111]	"Deckbuilding"	"Demons"
## [113]	"Design & Illustration"	"Destruction"
## [115]	"Detective"	"Dice"
## [117]	"Difficult"	"Dinosaurs"
## [119]	"Diplomacy"	"Documentary"
## [121]	"Dog"	"Dragons"
## [123]	"Drama"	"Driving"
## [125]	"Dungeon Crawler"	"Dungeons & Dragons"
## [127]	"Dwarf"	"Dynamic Narration"
## [129]	"Dystopian"	"e-sports"
## [131]	"Early Access"	"Economy"
## [133]	"Education"	"Electronic"
## [135]	"Electronic Music"	"Elf"
## [137]	"Emotional"	"Epic"
## [139]	"Episodic"	"Escape Room"
## [141]	"Experience"	"Experimental"
## [143]	"Exploration"	"Extraction Shooter"
## [145]	"Faith"	"Family Friendly"
## [147]	"Fantasy"	"Farming"
## [149]	"Farming Sim"	"Fast-Paced"
## [151]	"Feature Film"	"Female Protagonist"
## [153]	"Fighting"	"First-Person"
## [155]	"Fishing"	"Flight"
## [157]	"FMV"	"Football"
## [159]	"Football (American)"	"Football (Soccer)"
## [161]	"Foreign"	"Fox"
## [163]	"FPS"	"Free to Play"
## [165]	"Funny"	"Futuristic"
## [167]	"Gambling"	"Game Development"
## [169]	"GameMaker"	"Games Workshop"
## [171]	"Gaming"	"God Game"
## [173]	"Golf"	"Gore"
## [175]	"Gothic"	"Grand Strategy"
## [177]	"Great Soundtrack"	"Grid-Based Movement"
## [179]	"Gun Customization"	"Hack and Slash"
## [181]	"Hacking"	"Hand-drawn"
## [183]	"Hardware"	"Heist"
## [185]	"Hentai"	"Hero Shooter"
## [187]	"Hex Grid"	"Hidden Object"
## [189]	"Historical"	"Hobby Sim"
## [191]	"Hockey"	"Horror"
## [193]	"Horses"	"Hunting"
## [195]	"Idler"	"Illuminati"
## [197]	"Immersive"	"Immersive Sim"
## [199]	"Indie"	"Instrumental Music"
## [201]	"Intentionally Awkward Controls"	"Interactive Fiction"
## [203]	"Inventory Management"	"Investigation"

## [205]	"Isometric"	"Jet"
## [207]	"Job Simulator"	"JRPG"
## [209]	"Jump Scare"	"Kickstarter"
## [211]	"LEGO"	"Lemmings"
## [213]	"Level Editor"	"LGBTQ+"
## [215]	"Life Sim"	"Linear"
## [217]	"Local Co-Op"	"Local Multiplayer"
## [219]	"Logic"	"Loot"
## [221]	"Looter Shooter"	"Lore-Rich"
## [223]	"Lovecraftian"	"Magic"
## [225]	"Mahjong"	"Management"
## [227]	"Mars"	"Martial Arts"
## [229]	"Massively Multiplayer"	"Masterpiece"
## [231]	"Match 3"	"Mature"
## [233]	"Mechs"	"Medical Sim"
## [235]	"Medieval"	"Memes"
## [237]	"Metroidvania"	"Military"
## [239]	"Mini Golf"	"Minigames"
## [241]	"Minimalist"	"Mining"
## [243]	"MMORPG"	"MOBA"
## [245]	"Mod"	"Moddable"
## [247]	"Modern"	"Motocross"
## [249]	"Motorbike"	"Mouse only"
## [251]	"Movie"	"Multiplayer"
## [253]	"Multiple Endings"	"Music"
## [255]	"Music-Based Procedural Generation"	"Musou"
## [257]	"Mystery"	"Mystery Dungeon"
## [259]	"Mythology"	"Narration"
## [261]	"Narrative"	"Nature"
## [263]	"Naval"	"Naval Combat"
## [265]	"Ninja"	"Noir"
## [267]	"Nonlinear"	"Nostalgia"
## [269]	"NSFW"	"Nudity"
## [271]	"Offroad"	"Old School"
## [273]	"On-Rails Shooter"	"Online Co-Op"
## [275]	"Open World"	"Open World Survival Craft"
## [277]	"Otome"	"Outbreak Sim"
## [279]	"Parkour"	"Parody"
## [281]	"Party"	"Party Game"
## [283]	"Party-Based RPG"	"Perma Death"
## [285]	"Philosophical"	"Photo Editing"
## [287]	"Physics"	"Pinball"
## [289]	"Pirates"	"Pixel Graphics"
## [291]	"Platformer"	"Point & Click"
## [293]	"Political"	"Political Sim"
## [295]	"Politics"	"Pool"
## [297]	"Post-apocalyptic"	"Precision Platformer"
## [299]	"Procedural Generation"	"Programming"
## [301]	"Psychedelic"	"Psychological"
## [303]	"Psychological Horror"	"Puzzle"
## [305]	"Puzzle-Platformer"	"PvE"
## [307]	"PvP"	"Quick-Time Events"
## [309]	"Racing"	"Real Time Tactics"
## [311]	"Real-Time"	"Real-Time with Pause"

## [313]	"Realistic"	"Reboot"
## [315]	"Relaxing"	"Remake"
## [317]	"Replay Value"	"Resource Management"
## [319]	"Retro"	"Rhythm"
## [321]	"Robots"	"Rock Music"
## [323]	"Rogue-like"	"Rogue-lite"
## [325]	"Roguelike Deckbuilder"	"Roguevania"
## [327]	"Romance"	"Rome"
## [329]	"RPG"	"RPGMaker"
## [331]	"RTS"	"Rugby"
## [333]	"Runner"	"Sailing"
## [335]	"Sandbox"	"Satire"
## [337]	"Sci-fi"	"Science"
## [339]	"Score Attack"	"Sequel"
## [341]	"Sexual Content"	"Shooter"
## [343]	"Shop Keeper"	"Short"
## [345]	"Side Scroller"	"Silent Protagonist"
## [347]	"Simulation"	"Singleplayer"
## [349]	"Skateboarding"	"Skating"
## [351]	"Skiing"	"Sniper"
## [353]	"Snooker"	"Snow"
## [355]	"Snowboarding"	"Soccer"
## [357]	"Social Deduction"	"Software"
## [359]	"Software Training"	"Sokoban"
## [361]	"Solitaire"	"Souls-like"
## [363]	"Soundtrack"	"Space"
## [365]	"Space Sim"	"Spaceships"
## [367]	"Spectacle fighter"	"Spelling"
## [369]	"Split Screen"	"Sports"
## [371]	"Stealth"	"Steam Machine"
## [373]	"Steampunk"	"Story Rich"
## [375]	"Strategy"	"Strategy RPG"
## [377]	"Stylized"	"Submarine"
## [379]	"Superhero"	"Supernatural"
## [381]	"Surreal"	"Survival"
## [383]	"Survival Horror"	"Swordplay"
## [385]	"Tabletop"	"Tactical"
## [387]	"Tactical RPG"	"Tanks"
## [389]	"Team-Based"	"Tennis"
## [391]	"Text-Based"	"Third Person"
## [393]	"Third-Person Shooter"	"Thriller"
## [395]	"Tile-Matching"	"Time Attack"
## [397]	"Time Management"	"Time Manipulation"
## [399]	"Time Travel"	"Top-Down"
## [401]	"Top-Down Shooter"	"Touch-Friendly"
## [403]	"Tower Defense"	"TrackIR"
## [405]	"Trading"	"Trading Card Game"
## [407]	"Traditional Roguelike"	"Trains"
## [409]	"Transhumanism"	"Transportation"
## [411]	"Trivia"	"Turn-Based"
## [413]	"Turn-Based Combat"	"Turn-Based Strategy"
## [415]	"Turn-Based Tactics"	"Tutorial"
## [417]	"Twin Stick Shooter"	"Typing"
## [419]	"Underground"	"Underwater"

## [421] "Unforgiving"	"Utilities"
## [423] "Vampire"	"Vehicular Combat"
## [425] "Video Production"	"Vikings"
## [427] "Villain Protagonist"	"Violent"
## [429] "Visual Novel"	"Voice Control"
## [431] "Volleyball"	"Voxel"
## [433] "VR"	"VR Only"
## [435] "Walking Simulator"	"War"
## [437] "Wargame"	"Warhammer 40K"
## [439] "Web Publishing"	"Well-Written"
## [441] "Werewolves"	"Western"
## [443] "Wholesome"	"Word Game"
## [445] "World War I"	"World War II"
## [447] "Wrestling"	"Zombies"

From steam's Categories, Genres and Tags, 3 subsets have been created with samples interesting for this study. Given steam's laxity at classifying these values, the 3 subsets have been created with our own interpretation. Additionally, tags that refer to non-game software have been ignored.

```
#Genres
action_genre <- c(
  # Action
  "Action", "Action-Adventure", "Action RPG", "Action Roguelike", "Action RTS",
  "Arcade", "Immersive Sim", "Metroidvania", "Musou", "Sandbox", "Survival",
  "Survival Horror", "MOBA")
shooter_genre<- c(
  # Shooter
  "Arena Shooter", "Battle Royale", "Boomer Shooter", "Bullet Hell", "Extraction Shooter",
  "FPS", "Hero Shooter", "Looter Shooter", "On-Rails Shooter", "Third-Person Shooter",
  "Top-Down Shooter", "Twin Stick Shooter", "Vehicular Combat", "Naval Combat",
  "Combat Racing")
roguelike_genre <- c(
  # Roguelike
  "Rogue-like", "Rogue-lite", "Roguevania", "Traditional Roguelike",
  "Roguelike Deckbuilder", "Mystery Dungeon")
RPG_genre <- c(
  # RPG
  "RPG", "JRPG", "CRPG", "Strategy RPG", "Tactical RPG", "Otome",
  "Dating Sim", "Visual Novel", "Interactive Fiction", "Text-Based",
  "MMORPG", "MMO")
strategy_genre <- c(
  # Strategy
  "Strategy", "4X", "Grand Strategy", "RTS", "Real Time Tactics",
  "Turn-Based Strategy", "Tactical", "Tabletop", "Wargame", "Political Sim",
  "God Game", "Management", "Colony Sim", "City Builder", "Shop Keeper",
  "Tower Defense", "Creature Collector")
simulation_genre <- c(
  # Simulation
  "Simulation", "Automobile Sim", "Farming Sim", "Medical Sim",
  "Hobby Sim", "Life Sim", "Space Sim", "Outbreak Sim")
sport_genre <- c(
  # Sports
  "Sports", "Baseball", "Basketball", "Bowling", "Boxing", "BMX", "Cricket",
  "Football", "Football (American)", "Football (Soccer)", "Soccer",
```

```

"Golf", "Hockey", "Mini Golf", "Motocross", "Pool", "Racing", "Rugby",
"Skateboarding", "Skating", "Skiing", "Snowboarding", "Snooker", "Tennis",
"Volleyball", "Wrestling", "Driving")
fighting_genre <- c(
  # Fighting
  "2D Fighter", "3D Fighter", "Fighting", "Hack and Slash", "Souls-like")
platformer_genre <- c(
  # Platformers & runners
  "Platformer", "2D Platformer", "3D Platformer", "Precision Platformer",
  "Runner", "Open World Survival Craft")
tabletop_genre <- c(
  # Puzzle, logic, board & card games
  "Puzzle", "Puzzle-Platformer", "Logic", "Mahjong",
  "Match 3", "Tile-Matching", "Sokoban", "Solitaire", "Trivia",
  "Board Game", "Card Game", "Trading Card Game", "Hidden Object",
  "Escape Room", "Chess", "Word Game", "Point & Click")
casual_genre <- c(
  # Casual / idle
  "Casual", "Idler", "Clicker", "Auto Battler", "Party Game", "Pinball")
horror_genre <- c(
  # Horror
  "Horror", "Psychological Horror")
music_genre <- c(
  # Music & rhythm
  "Music", "Rhythm", "Typing")
walking_sim_genre <- c(
  # Walking
  "Walking Simulator", "FMV")

genres <- list(
  action = action_genre,
  shooter = shooter_genre,
  roguelike = roguelike_genre,
  RPG = RPG_genre,
  strategy = strategy_genre,
  simulation = simulation_genre,
  sport = sport_genre,
  fighting = fighting_genre,
  platformer = platformer_genre,
  tabletop = tabletop_genre,
  casual = casual_genre,
  horror = horror_genre,
  music = music_genre,
  walking_sim = walking_sim_genre
)

#Mechanics
resource_management_mechanic <- c(
  # Economy, building, resources
  "Base-Building", "Building", "Resource Management", "Inventory Management",
  "Crafting", "Automation", "Mining", "Farming")

```

```

customization_mechanic <- c(
  # Customization
  "Gun Customization", "Character Customization")
movement_mechanic <- c(
  # Movement
  "Stealth", "Parkour", "Swordplay", "Archery", "Sniper")
physics_mechanic <- c(
  # Physics
  "Physics", "Destruction")
failure_mechanic <- c(
  # Failure
  "Permadeath", "Perma Death")
procedural_mechanic <- c(
  # Procedural
  "Procedural Generation", "Music-Based Procedural Generation")
turn_mechanic <- c(
  # Turns
  "Turn-Based", "Turn-Based Combat", "Turn-Based Tactics",
  "Real-Time", "Real-Time with Pause",
  "Time Manipulation", "Time Management", "Time Attack")
spatial_rules_mechanic <- c(
  # Spatial rules
  "Grid-Based Movement", "Hex Grid")
quicktime_events_mechanic <- c(
  # QTE
  "Quick-Time Events")
narrative_mechanic <- c(
  # Narrative logic
  "Choices Matter", "Multiple Endings", "Nonlinear")
exploration_mechanic <- c(
  # Exploration & structure
  "Exploration", "Dungeon Crawler", "Collectathon",
  "Boss Rush", "Side Scroller", "Loot")
social_mechanic <- c(
  # Social deduction
  "Social Deduction")
card_mechanic <- c(
  # Card/Deck mechanics
  "Deckbuilding", "Card Battler")
programming_mechanic <- c(
  # Programming
  "Programming", "Coding")
economy_mechanic <- c(
  # Economy
  "Trading", "Economy", "Capitalism",
  "Diplomacy")
modding_mechanic <- c(
  # Modding / creation
  "Level Editor", "Moddable", "Mod")
tutorial_mechanic <- c(
  # Tutorial
  "Tutorial")

```



```

mechanics <- list(
  resource_management = resource_management_mechanic,
  customization = customization_mechanic,
  movement = movement_mechanic,
  physics = physics_mechanic,
  failure = failure_mechanic,
  procedural = procedural_mechanic,
  turn = turn_mechanic,
  spatial_rules = spatial_rules_mechanic,
  quicktime_events = quicktime_events_mechanic,
  narrative = narrative_mechanic,
  exploration = exploration_mechanic,
  social = social_mechanic,
  card = card_mechanic,
  programming = programming_mechanic,
  economy = economy_mechanic,
  modding = modding_mechanic,
  tutorial = tutorial_mechanic
)

# Characteristics
camera_characteristic <- c(
  # Dimensionality & camera
  "2D", "2.5D", "3D",
  "Isometric", "Top-Down",
  "6DOF",
  "First-Person", "Third Person")

vr_characteristic <- c(
  # VR
  "VR", "VR Only", "Asymmetric VR")

players_characteristic <- c(
  # Player count / mode
  "Singleplayer",
  "Multiplayer", "Massively Multiplayer",
  "Co-op", "Local Co-Op", "Online Co-Op", "Co-op Campaign",
  "4 Player Local", "Split Screen")

team_characteristic <- c(
  # Team/role structure
  "PvP", "PvE", "Team-Based",
  "Asynchronous Multiplayer")

esports_characteristic <- c(
  # Competitive/e-sports
  "e-sports")

characteristics <- list(
  camera = camera_characteristic,
  vr = vr_characteristic,
  players = players_characteristic,
  team = team_characteristic,
  esports = esports_characteristic
)

```

Thankfully the task of filtering indie games is already provided by steam with the Tags and Genres Indie, Crowdfunded and Kickstarter. From the original dataset, only Indie games are obtained.

```
indie <- c("Crowdfunded", "Kickstarter", "Indie")
indie_pattern <- str_c(indie, collapse = "|")

indie_games <- steam %>%
  filter(
    if_any(
      c(categories, genres, tags),
      ~ str_detect(.x, indie_pattern)
    )
  )
```

Prepare data for determining successful games

```
indie_games_data <- indie_games %>% select(appid, name, price, dlc_count, windows, mac, linux, achievements)

#transform suported languages into count
add_column_count <- function(df, column, new_col) {
  column_quo <- enquo(column)

  df %>%
    mutate(
      "{new_col}" := case_when(
        !!column_quo == "[]" ~ 0L,
        TRUE ~ str_count(!!column_quo, "'") %/% 2L
      )
    )
}

add_range_minmax <- function(df, column, min_column, max_column) {
  column_quo <- enquo(column)

  df %>%
    mutate(
      "{min_column}" := as.integer(str_extract(!!column_quo, "^\\s*\\d+")),
      "{max_column}" := as.integer(str_extract(!!column_quo, "\\d+\\s*$"))
    )
}

#Limpiar esta linea con los cambios de arturo TODO
indie_games_data <- indie_games_data %>%
  add_range_minmax(estimated_owners, "estimated_owners_min", "estimated_owners_max") %>% # split owners
  mutate(estimated_revenue = ((estimated_owners_min + estimated_owners_max) / 2) * price) %>% # augment
  mutate(windows = as.integer(windows == "True")) %>% # parse logical to int
  mutate(mac = as.integer(mac == "True")) %>% # parse logical to int
  mutate(linux = as.integer(linux == "True")) %>% # parse logical to int
  select(-estimated_owners)
```

Cluster the data:

```

indie_games_cluster <- indie_games_data %>%
  add_column_count(supported_languages, "languages_count") %>% # Count languages
  select(-appid, -name, -supported_languages)

indie_games_cluster <- indie_games_cluster %>% mutate( # Avoid scaling binary columns
  across(
    where(is.numeric) & !all_of(c("windows", "mac", "linux")),
    ~ as.numeric(scale(.))
  )
)

km2 <- kmeans(indie_games_cluster, centers = 3, nstart = 50)
indie_games_data$success_cluster <- km2$cluster
indie_games$success_cluster <- km2$cluster
colnames(indie_games_data)

```

```

## [1] "appid" "name"
## [3] "price" "dlc_count"
## [5] "windows" "mac"
## [7] "linux" "achievements"
## [9] "recommendations" "supported_languages"
## [11] "positive" "negative"
## [13] "average_playtime_forever" "median_playtime_forever"
## [15] "peak_ccu" "pct_pos_total"
## [17] "num_reviews_total" "estimated_owners_min"
## [19] "estimated_owners_max" "estimated_revenue"
## [21] "success_cluster"

```