Final Project Data Prep & EDA

Gabriel Valenzuela

6/28/2020

## Prediction Analysis Data Preparation and Exploratory Data Analysis - Video Game Sales

# Data Preparation

## -- Attaching packages -------------------------------------------------------------------------------------------------------------------------- tidyverse 1.3.0 --

## <U+2713> ggplot2 3.2.1 <U+2713> purrr 0.3.3  
## <U+2713> tibble 3.0.1 <U+2713> dplyr 0.8.4  
## <U+2713> tidyr 1.1.0 <U+2713> stringr 1.4.0  
## <U+2713> readr 1.3.1 <U+2713> forcats 0.5.0

## -- Conflicts ----------------------------------------------------------------------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

## Rank Name Platform Year Genre Publisher NA\_Sales  
## 1 1 Wii Sports Wii 2006 Sports Nintendo 41.49  
## 2 2 Super Mario Bros. NES 1985 Platform Nintendo 29.08  
## 3 3 Mario Kart Wii Wii 2008 Racing Nintendo 15.85  
## 4 4 Wii Sports Resort Wii 2009 Sports Nintendo 15.75  
## 5 5 Pokemon Red/Pokemon Blue GB 1996 Role-Playing Nintendo 11.27  
## 6 6 Tetris GB 1989 Puzzle Nintendo 23.20  
## 7 7 New Super Mario Bros. DS 2006 Platform Nintendo 11.38  
## 8 8 Wii Play Wii 2006 Misc Nintendo 14.03  
## 9 9 New Super Mario Bros. Wii Wii 2009 Platform Nintendo 14.59  
## 10 10 Duck Hunt NES 1984 Shooter Nintendo 26.93  
## EU\_Sales JP\_Sales Other\_Sales Global\_Sales  
## 1 29.02 3.77 8.46 82.74  
## 2 3.58 6.81 0.77 40.24  
## 3 12.88 3.79 3.31 35.82  
## 4 11.01 3.28 2.96 33.00  
## 5 8.89 10.22 1.00 31.37  
## 6 2.26 4.22 0.58 30.26  
## 7 9.23 6.50 2.90 30.01  
## 8 9.20 2.93 2.85 29.02  
## 9 7.06 4.70 2.26 28.62  
## 10 0.63 0.28 0.47 28.31

## 'data.frame': 16598 obs. of 11 variables:  
## $ Rank : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ Name : Factor w/ 11493 levels "'98 Koshien",..: 10991 9343 5532 10993 7370 9707 6648 10989 6651 2594 ...  
## $ Platform : Factor w/ 31 levels "2600","3DO","3DS",..: 26 12 26 26 6 6 5 26 26 12 ...  
## $ Year : Factor w/ 40 levels "1980","1981",..: 27 6 29 30 17 10 27 27 30 5 ...  
## $ Genre : Factor w/ 12 levels "Action","Adventure",..: 11 5 7 11 8 6 5 4 5 9 ...  
## $ Publisher : Factor w/ 579 levels "10TACLE Studios",..: 369 369 369 369 369 369 369 369 369 369 ...  
## $ NA\_Sales : num 41.5 29.1 15.8 15.8 11.3 ...  
## $ EU\_Sales : num 29.02 3.58 12.88 11.01 8.89 ...  
## $ JP\_Sales : num 3.77 6.81 3.79 3.28 10.22 ...  
## $ Other\_Sales : num 8.46 0.77 3.31 2.96 1 0.58 2.9 2.85 2.26 0.47 ...  
## $ Global\_Sales: num 82.7 40.2 35.8 33 31.4 ...

## Rank Name Platform Year Genre Publisher   
## FALSE FALSE FALSE FALSE FALSE FALSE   
## NA\_Sales EU\_Sales JP\_Sales Other\_Sales Global\_Sales   
## FALSE FALSE FALSE FALSE FALSE

## Rank Name Platform Year Genre Publisher   
## FALSE FALSE FALSE FALSE FALSE FALSE   
## NA\_Sales EU\_Sales JP\_Sales Other\_Sales Global\_Sales   
## FALSE FALSE FALSE FALSE FALSE

# Exploratory Data Analysis

## Rank Name Platform   
## Min. : 1 Need for Speed: Most Wanted: 12 DS :2163   
## 1st Qu.: 4151 FIFA 14 : 9 PS2 :2161   
## Median : 8300 LEGO Marvel Super Heroes : 9 PS3 :1329   
## Mean : 8301 Madden NFL 07 : 9 Wii :1325   
## 3rd Qu.:12450 Ratatouille : 9 X360 :1265   
## Max. :16600 Angry Birds Star Wars : 8 PSP :1213   
## (Other) :16542 (Other):7142   
## Year Genre Publisher   
## Min. : 1.00 Length:16598 Electronic Arts : 1351   
## 1st Qu.:24.00 Class :character Activision : 975   
## Median :28.00 Mode :character Namco Bandai Games : 932   
## Mean :27.61 Ubisoft : 921   
## 3rd Qu.:32.00 Konami Digital Entertainment: 832   
## Max. :40.00 THQ : 715   
## (Other) :10872   
## NA\_Sales EU\_Sales JP\_Sales Other\_Sales   
## Min. : 0.0000 Min. : 0.0000 Min. : 0.00000 Min. : 0.00000   
## 1st Qu.: 0.0000 1st Qu.: 0.0000 1st Qu.: 0.00000 1st Qu.: 0.00000   
## Median : 0.0800 Median : 0.0200 Median : 0.00000 Median : 0.01000   
## Mean : 0.2647 Mean : 0.1467 Mean : 0.07778 Mean : 0.04806   
## 3rd Qu.: 0.2400 3rd Qu.: 0.1100 3rd Qu.: 0.04000 3rd Qu.: 0.04000   
## Max. :41.4900 Max. :29.0200 Max. :10.22000 Max. :10.57000   
##   
## Global\_Sales   
## Min. : 0.0100   
## 1st Qu.: 0.0600   
## Median : 0.1700   
## Mean : 0.5374   
## 3rd Qu.: 0.4700   
## Max. :82.7400   
##

