TECNOLÓGICO NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE TIJUANA

Subdirección Académica

Departamento de sistemas y computación

Semestre:
Agosto - Diciembre 2021

Materia:
Minería de Datos

Nombre del trabajo:
Práctica Evaluatoria - Unidad 2

UNIDAD A EVALUAR:
Unidad I

NOMBRE Y NÚMERO DE CONTROL DEL ALUMNO:
Victor Fabian Hernandez Lamarque | 16212536
Cristopher Arael Garcia Torres | 17210892

NOMBRE DEL MAESTRO (A):
JOSE CHRISTIAN ROMERO

```
# Especificamos nuestro lugar de trabajo
getwd()
setwd("D:/Documentos/DataMining/Evaluacion U2")
getwd()

# Cargamos los datos con los que vamos a trabajar
movies <- read.csv("Project-Data.csv")

str(movies)
summary(movies)


#Filtramos los datos y descartamos los datos los cuales no vamos a
utilizar con “filterin Genre”
filtGenre <- movies$Genre %in%
c("action","adventure","animation","comedy","drama")


# Creamos un dataframe con los géneros limpios
filtermovies <- movies[filtGenre,]

#Filtramos los datos de nuestro nuevo dataframe
filterStudio <- filtermovies$Studio %in% c("Buena Vista
Studios","Fox","Paramount Pictures","Sony","Universal","WB")

# Creamos un nuevo dataframe con todos los datos pedidos anteriormente
Resultmovies <- filtermovies[filterStudio,]


#Renombramos las columnas para un mejor sentido
colnames(Resultmovies) <- c("Dayofweek", "Director", "Genre",
"MovieTitle", "RealseDate", "Studio","AdjustedGroosMill","BudgetMill",
"GrossMill","IMDbRating", "MovieLensRating", "OverseasMill","Overseas",
"ProfitMill", "Profit", "RuntimeMin", "UsMill", "GrossUS")

#Agregamos valores para hacer más visible los datos (Create transparent
color)
mycol <- rgb(0, 0, 255, max = 255, alpha = 125, names = "blue50")
```

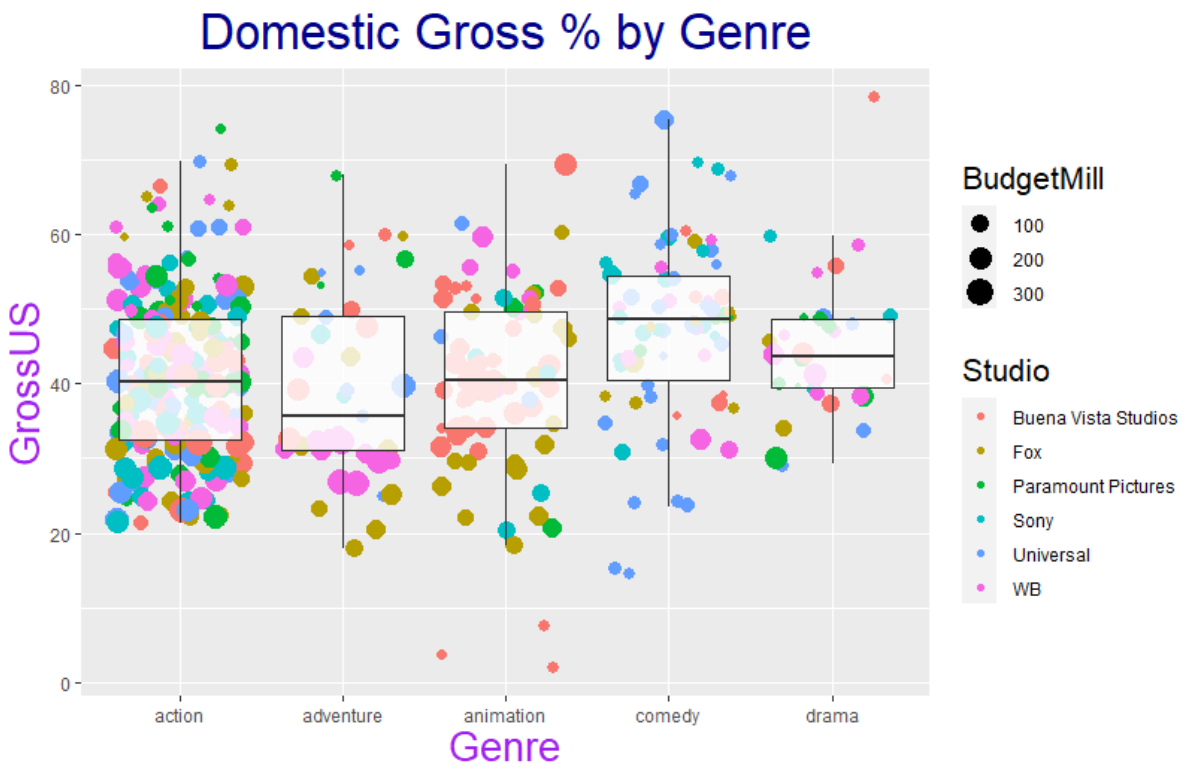
```
# instalamos e importamos la librería ggplot2
install.packages('ggplot2')

library(ggplot2)


#Creamos diagrama de dispersión
Graphic <- ggplot(data = Resultmovies,mapping = aes(x=Genre, y=GrossUS))
+
  labs(title='Domestic Gross % by Genre') +
  theme(axis.title.x = element_text(color = "Purple", size=20),
        axis.title.y = element_text(color = "Purple", size=20),
        legend.title = element_text(color="Black",size=16 ),
        plot.title =  element_text(color = "DarkBlue", size = 25, hjust
= 0.5 )
        )

Graphic + geom_boxplot()

Graphic + geom_jitter(aes( size = BudgetMill, color=Studio )) +
geom_boxplot(aes(),alpha=0.8, outlier.colour = NA )
```



Project.Data | creating a scatter plot.R

Filter

	Day.of.Week	Director	Genre	Movie.Title
1	Friday	Brad Bird	action	Tomorrowland
2	Friday	Scott Waugh	action	Need for Speed
3	Friday	Patrick Hughes	action	The Expendables 3
4	Friday	Phil Lord, Chris Miller	comedy	21 Jump Street
5	Friday	Roland Emmerich	action	White House Down
6	Friday	David Ayer	action	Fury
7	Thursday	Rob Marshall	adventure	Into the Woods
8	Friday	Daniel Espinosa	action	Safe House
9	Friday	Gary Shore	action	Dracula Untold
10	Friday	Henry Joost, Ariel Schulman	horror	Paranormal Activity 3
11	Friday	Eric Brevig	animation	Yogi Bear
12	Friday	Morten Tyldum	biography	The Imitation Game
13	Thursday	Bryan Singer	biography	Valkyrie
14	Friday	Seth Gordon	comedy	Horrible Bosses

Showing 1 to 14 of 608 entries, 18 total columns