18F-0240_7A_MovieRatingAnalysis

18F-0240

12/19/2021

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Google Trend Analysis

Github link: "https://github.com/HuzaifahZia/Google-Search-Analytics-/blob/main/18F-0240_7A_GoogleTrendsAnalytics.Rmd"

Installing Libraries

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.2
```

```
library(ggthemes)
```

```
## Warning: package 'ggthemes' was built under R version 4.1.2
```

```
library(scales)
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
```

```
library(lessR)
```

```
## Warning: package 'lessR' was built under R version 4.1.2
```

```
##
## lessR 4.1.3  feedback: gerbing@pdx.edu  web: lessRstats.com/new
## ------------------------------------------------------------
## > d <- Read("")   Read text, Excel, SPSS, SAS, or R data file
##   d is default data frame, data= in analysis routines optional
##
## Learn about reading, writing, and manipulating data, graphics,
## testing means and proportions, regression, factor analysis,
## customization, and descriptive statistics from pivot tables.
##   Enter:  browseVignettes("lessR")
##
## View changes in this new version of lessR.
##   Enter: help(package=lessR)  Click: Package NEWS
##   Enter: interact()  for access to interactive graphics
```

```
##
## Attaching package: 'lessR'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:scales':
##
##     rescale
```

-extracting data from .dat file where seperator is '::' -making extracted data a dataframe -displaying its top 6 rows

```
movies = do.call(rbind,strsplit(readLines('movies.dat'),'::',fixed=T))
movies <- as.data.frame(movies)
head(movies)
```

```
##        V1                                                 V2
## 1 0000008        Edison Kinetoscopic Record of a Sneeze (1894)
## 2 0000010                 La sortie des usines LumiÃ¨re (1895)
## 3 0000012                       The Arrival of a Train (1896)
## 4      25 The Oxford and Cambridge University Boat Race (1895)
## 5 0000091                           Le manoir du diable (1896)
## 6 0000131                             Une nuit terrible (1896)
##                    V3
## 1   Documentary|Short
## 2   Documentary|Short
## 3   Documentary|Short
## 4                  25
## 5        Short|Horror
## 6 Short|Comedy|Horror
```

-assigning column names to dataframe -making ID column numeric -displaying its top 6 rows

```
colnames(movies) <- c("ID","Title","Genre")
movies$ID <- as.numeric(movies$ID)
head(movies)
```

```
##    ID                                             Title               Genre
## 1   8        Edison Kinetoscopic Record of a Sneeze (1894)   Documentary|Short
## 2  10                 La sortie des usines LumiÃ¨re (1895)   Documentary|Short
## 3  12                        The Arrival of a Train (1896)   Documentary|Short
## 4  25 The Oxford and Cambridge University Boat Race (1895)                  25
## 5  91                            Le manoir du diable (1896)        Short|Horror
## 6 131                              Une nuit terrible (1896) Short|Comedy|Horror
```

-extracting data from .dat file where seperator is '::' -displaying its top 6 rows

```
ratings = read.delim("ratings.dat", header= FALSE ,sep = ':', colClasses = c(NA, "NULL"))
head(ratings)
```

```
##   V1      V3 V5         V7
## 1  1  114508  8 1381006850
## 2  2  499549  9 1376753198
## 3  2 1305591  8 1376742507
## 4  2 1428538  1 1371307089
## 5  3   75314  1 1595468524
## 6  3  102926  9 1590148016
```

-assigning column names to dataframe -making ID column numeric -displaying its top 6 rows

```
colnames(ratings) <- c("User","ID","Ratings","Timestamp")
head(ratings)
```

```
##   User      ID Ratings  Timestamp
## 1    1  114508       8 1381006850
## 2    2  499549       9 1376753198
## 3    2 1305591       8 1376742507
## 4    2 1428538       1 1371307089
## 5    3   75314       1 1595468524
## 6    3  102926       9 1590148016
```

-merging dataframes based on ID -displaying its top 6 rows

```
data = merge(movies,ratings,by = "ID")
data = data[with(data, order(ID)),]
head(data)
```

```
##   ID                                             Title             Genre
## 1  8        Edison Kinetoscopic Record of a Sneeze (1894) Documentary|Short
## 2 10                 La sortie des usines LumiÃ¨re (1895) Documentary|Short
## 3 12                        The Arrival of a Train (1896) Documentary|Short
## 4 25 The Oxford and Cambridge University Boat Race (1895)                25
```

```
## 5 91                          Le manoir du diable (1896)      Short|Horror
## 6 91                          Le manoir du diable (1896)      Short|Horror
##    User Ratings  Timestamp
## 1 42898       5 1396981211
## 2 70577      10 1412878553
## 3 69535      10 1439248579
## 4 37628       8 1488189899
## 5 54465       7 1562928526
## 6 37239       5 1532347349
```

-ploting pie chart

```
ratings<- dplyr::count(data, Ratings, sort = TRUE) %>%
  mutate(perc = `n` / sum(`n`)) %>%
  arrange(perc) %>%
  mutate(labels = scales::percent(perc))

ggplot(ratings, aes(x = "", y = perc, fill = factor(Ratings)),alpha = 0.8) +
  geom_col() +
  geom_text(aes(label = labels),
            position = position_stack(vjust = 0.5)) +
  coord_polar(theta = "y")
```