importing required libraries

```
library(ggplot2)
library(dplyr)
library(tidyr)
library(stringr)
library(lubrid)
library(ggthemes)
```

```
C:/Users/ajayk/Anaconda3/lib/site-packages/numpy/core/umath.pyd: error: (126) The specified module could not be found.
error: cannot open shared object file: No such file or directory
```

```
C:/Users/ajayk/Anaconda3/lib/site-packages/numpy/core/umath.pyd: error: (126) The specified module could not be found.
error: cannot open shared object file: No such file or directory
```

import dataset

```
netflix_data <- read.csv("C:/Users/ajayk/Desktop/netflix_data.csv")
```

```
netflix_data <- read.csv("C:/Users/ajayk/Desktop/netflix_data.csv")
```

```
show_id, type, title, director, cast, country, date_added, release_year, rating, duration, listed_in, description
```

```
'data.frame':	8807 obs. of  12 variables:
 $ show_id     : chr  "s1" "s2" "s3" "s4" ...
 $ type        : chr  "Movie" "TV Show" "TV Show" "TV Show" ...
 $ title       : chr  "Dick Johnson Is Dead" "Blood & Water" "Ganglands" "Jailbirds New Orleans" ...
 $ director    : chr  "Kirsten Johnson" "" "Julien Leclercq" "" ...
 $ cast        : chr  "" "Ama Qamata, Khosi Ngema, Gail Mabalane, Thabang Molaba, Dinh"| __truncated__ "Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nicolas Duvauchelle, Nicolas Wanczak, Zi"| __truncated__ "Mel Davis, Santiago Cabrera, Jillian Harris, Erica Ash, Derek Luke, Michael Kenneth Wi"| __truncated__ "Tahar Rahim, Céline Sallette, Philippe Rebbot, Christophe Montenery, Jérôme Kircher, "| __truncated__
 $ country     : chr  "United States" "South Africa" "" "" ...
 $ date_added  : chr  "September 25, 2021" "September 24, 2021" "September 23, 2021" "September 22, 2021" ...
 $ release_year: int  2020 2021 2021 2021 2021 2021 2021 2021 2021 2021 ...
 $ rating      : chr  "PG-13" "TV-MA" "TV-MA" "TV-MA" ...
 $ duration    : chr  "90 min" "2 Seasons" "1 Season" "1 Season" ...
 $ listed_in   : chr  "Documentaries" "International TV Shows, TV Dramas, TV Mysteries" "Crime TV Shows, International TV Shows, TV Action & Adventure" "Docuseries, Reality TV" ...
 $ description : chr  "As her father nears the end of his life, filmmaker Kirsten Johnson stumbles upon moments she"| __truncated__ "After crossing paths at a party, a Cape Town teen sets off on a journey to find her biological"| __truncated__ "To protect his family from a powerful drug lord, skilled assassin Yuri Zhirkov has to sacrifice"| __truncated__ "Feuds, flirtations and toilet paper are the order of the day when a wildlife photographer and"| __truncated__
```

Data Exploration

```
dim(netflix_data)
```

```
[1] 8807   12
```

```
head(netflix_data,10)
```

```
show_id, type, title, director, cast, country, date_added, release_year, rating, duration, listed_in, description
```

```
'data.frame':	8807 obs. of  12 variables:
 $ show_id     : chr  "s1" "s2" "s3" "s4" ...
 $ type        : chr  "Movie" "TV Show" "TV Show" "TV Show" ...
 $ title       : chr  "Dick Johnson Is Dead" "Blood & Water" "Ganglands" "Jailbirds New Orleans" ...
 $ director    : chr  "Kirsten Johnson" "" "Julien Leclercq" "" ...
 $ cast        : chr  "" "Ama Qamata, Khosi Ngema, Gail Mabalane, Thabang Molaba, Dinh"| __truncated__ "Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nicolas Duvauchelle, Nicolas Wanczak, Zi"| __truncated__ "Mel Davis, Santiago Cabrera, Jillian Harris, Erica Ash, Derek Luke, Michael Kenneth Wi"| __truncated__ "Tahar Rahim, Céline Sallette, Philippe Rebbot, Christophe Montenery, Jérôme Kircher, "| __truncated__
 $ country     : chr  "United States" "South Africa" "" "" ...
 $ date_added  : chr  "September 25, 2021" "September 24, 2021" "September 23, 2021" "September 22, 2021" ...
 $ release_year: int  2020 2021 2021 2021 2021 2021 2021 2021 2021 2021 ...
 $ rating      : chr  "PG-13" "TV-MA" "TV-MA" "TV-MA" ...
 $ duration    : chr  "90 min" "2 Seasons" "1 Season" "1 Season" ...
 $ listed_in   : chr  "Documentaries" "International TV Shows, TV Dramas, TV Mysteries" "Crime TV Shows, International TV Shows, TV Action & Adventure" "Docuseries, Reality TV" ...
 $ description : chr  "As her father nears the end of his life, filmmaker Kirsten Johnson stumbles upon moments she"| __truncated__ "After crossing paths at a party, a Cape Town teen sets off on a journey to find her biological"| __truncated__ "To protect his family from a powerful drug lord, skilled assassin Yuri Zhirkov has to sacrifice"| __truncated__ "Feuds, flirtations and toilet paper are the order of the day when a wildlife photographer and"| __truncated__
```

```
dim(netflix_data)
```

```
[1] 8807   12
```

```
colnames(netflix_data)
```

```
 [1] "show_id"     "type"        "title"       "director"    "cast"        "country"     "date_added"  "release_year" "rating"      "duration"
[11] "listed_in"   "description"
```

```
str(netflix_data)
```

```
'data.frame':	8807 obs. of  12 variables:
 $ show_id     : chr  "s1" "s2" "s3" "s4" ...
 $ type        : chr  "Movie" "TV Show" "TV Show" "TV Show" ...
 $ title       : chr  "Dick Johnson Is Dead" "Blood & Water" "Ganglands" "Jailbirds New Orleans" ...
 $ director    : chr  "Kirsten Johnson" "" "Julien Leclercq" "" ...
 $ cast        : chr  "" "Ama Qamata, Khosi Ngema, Gail Mabalane, Thabang Molaba, Dinh"| __truncated__ "Sami Bouajila, Tracy Gotoas, Samuel Jouy, Nicolas Duvauchelle, Nicolas Wanczak, Zi"| __truncated__ "Mel Davis, Santiago Cabrera, Jillian Harris, Erica Ash, Derek Luke, Michael Kenneth Wi"| __truncated__ "Tahar Rahim, Céline Sallette, Philippe Rebbot, Christophe Montenery, Jérôme Kircher, "| __truncated__
 $ country     : chr  "United States" "South Africa" "" "" ...
 $ date_added  : chr  "September 25, 2021" "September 24, 2021" "September 23, 2021" "September 22, 2021" ...
 $ release_year: int  2020 2021 2021 2021 2021 2021 2021 2021 2021 2021 ...
 $ rating      : chr  "PG-13" "TV-MA" "TV-MA" "TV-MA" ...
 $ duration    : chr  "90 min" "2 Seasons" "1 Season" "1 Season" ...
 $ listed_in   : chr  "Documentaries" "International TV Shows, TV Dramas, TV Mysteries" "Crime TV Shows, International TV Shows, TV Action & Adventure" "Docuseries, Reality TV" ...
 $ description : chr  "As her father nears the end of his life, filmmaker Kirsten Johnson stumbles upon moments she"| __truncated__ "After crossing paths at a party, a Cape Town teen sets off on a journey to find her biological"| __truncated__ "To protect his family from a powerful drug lord, skilled assassin Yuri Zhirkov has to sacrifice"| __truncated__ "Feuds, flirtations and toilet paper are the order of the day when a wildlife photographer and"| __truncated__
```

```
colnames(netflix_data)
```

```
 [1] "show_id"     "type"        "title"       "director"    "cast"        "country"     "date_added"  "release_year" "rating"      "duration"
[11] "listed_in"   "description"
```

```
sapply(netflix_data, class)
```

```
show_id     type    title 
"character" "character" "character" 
director     cast  country 
"character" "character" "character" 
date_added release_year     rating 
"character"   "integer" "character" 
duration  listed_in description 
"character" "character" "character" 
```

```
sapply(netflix_data, function(x) length(unique(x)))
```

```
show_id     type    title 
    8807       2    8807 
director     cast  country 
    4528    7692     748 
date_added release_year     rating 
    1767      74      14 
duration  listed_in description 
     220     514    8807 
```

```
sapply(netflix_data, function(x) sum(is.na))
```

```
show_id     type    title 
      0       0       0 
director     cast  country 
      0       0       0 
date_added release_year     rating 
      0       0       0 
duration  listed_in description 
      0       0       0 
```

```
sapply(netflix_data, function(x) sum(is.na(x)))
```

```
show_id     type    title 
      0       0       0 
director     cast  country 
      0       0       0 
date_added release_year     rating 
      0       0       0 
duration  listed_in description 
      0       0       0 
```

Data cleaning

```
sapply(netflix_data, function(x) sum(is.na(x)))
```

```
show_id     type    title 
      0       0       0 
director     cast  country 
      0       0       0 
date_added release_year     rating 
      0       0       0 
duration  listed_in description 
      0       0       0 
```

```
sapply(netflix_data, function(x) sum(is.na(x)))
```

```
show_id     type    title 
      0       0       0 
director     cast  country 
      0       0       0 
date_added release_year     rating 
      0       0       0 
duration  listed_in description 
      0       0       0 
```

```
dim(netflix_data)
```

```
[1] 8807   12
```

```
sapply(netflix_data, function(x) sum(is.na(x)))
```

```
show_id     type    title 
      0       0       0 
director     cast  country 
      0       0       0 
date_added release_year     rating 
      0       0       0 
duration  listed_in description 
      0       0       0 
```

```
dim(netflix_data)
```

```
[1] 8807   12
```

```
sapply(netflix_data, function(x) sum(is.na(x)))
```

```
show_id     type    title 
      0       0       0 
director     cast  country 
      0       0       0 
date_added release_year     rating 
      0       0       0 
duration  listed_in description 
      0       0       0 
```