Project Proposal

due February 25, 2022 by 11:59 PM

Matthew Distell, Bo Aldridge, Haris Adnan, Lucan Franzblau, Devin Obee

2-25-2022

Load Packages

```
library(tidyverse)
install.packages("readxl")
library("readxl")
```

Load Data

```
myMovieData <- read_csv("IMDB-Movie-Data.csv")
myNBAData <- read_csv("nba.games.stats.csv")
myMusicData <- read_excel("Hot 100 Audio Features.xlsx")
```

Introduction and Data

Please introduce and discuss your data here.

For our three datasets, we are using data called myMovieData, myMusicData, and myNBAData. My-MovieData is collected from IMDB, or the International Movie Database. On that website, there is data listed for 1000 movies about 12 different variables, including genre, director, year released, runtime, and audience rating. MyMusicData is collected from the Billboard100 Hot Charts that features 29503 songs featured on the chart. The variables in the set include the runtime, loudness, performer, and genre of a given song. Finally, our myNbaData set analyzes the results of 9840 NBA games played from 2014-2018. This dataset includes statistics such as Field Goal Percentage, Three Point Percentage, who the home team was, and rebounding.

Research Questions

Please discuss your research questions here.

For each of these three datasets, we have specific research questions designed to answer interesting questions. For the myMovieData set, we plan on analyzing the differences between critic reviews and fan reviews, and identify any trends within that framework. More generally, our question is posed as: What do critics prefer in a movie as compared to fans? Our brief hypothesis is that fans prefer more action/adventure movies with shorter runtimes, while critics prefer longer dramas. For the myNBAData, we plan on figuring out what factors contribute most to winning. We will look at the relationship between home/away, field goal percentage, and rebounding numbers to find specific correlative properties with winning. Our hypothesis is that home/away will matter less than many think, and stats such as field goals attempted will be key as more shots and possessions leads to more opportunities to score. Finally, for the myMusicData, we will answer this

question: What type of song historically becomes a hit? We hypothesize that shorter songs oftentimes rise to the top, while the specific genre might change. Louder, more energetic songs should also perform better under this evaluation.

Glimpse

Please use `glimpse` for your datasets here.

```
glimpse(myMovieData)
```

```
## Rows: 1,000
## Columns: 12
## $ Rank                <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15~
## $ Title               <chr> "Guardians of the Galaxy", "Prometheus", "Split",~
## $ Genre               <chr> "Action,Adventure,Sci-Fi", "Adventure,Mystery,Sci~
## $ Description         <chr> "A group of intergalactic criminals are forced to~
## $ Director            <chr> "James Gunn", "Ridley Scott", "M. Night Shyamalan~
## $ Actors              <chr> "Chris Pratt, Vin Diesel, Bradley Cooper, Zoe Sal~
## $ Year                <dbl> 2014, 2012, 2016, 2016, 2016, 2016, 2016, 2016, 2~
## $ `Runtime (Minutes)` <dbl> 121, 124, 117, 108, 123, 103, 128, 89, 141, 116, ~
## $ Rating              <dbl> 8.1, 7.0, 7.3, 7.2, 6.2, 6.1, 8.3, 6.4, 7.1, 7.0,~
## $ Votes               <dbl> 757074, 485820, 157606, 60545, 393727, 56036, 258~
## $ `Revenue (Millions)` <dbl> 333.13, 126.46, 138.12, 270.32, 325.02, 45.13, 15~
## $ Metascore           <dbl> 76, 65, 62, 59, 40, 42, 93, 71, 78, 41, 66, 74, 6~
```

```
glimpse(myNBAData)
```

```
## Rows: 9,840
## Columns: 41
## $ ...1                <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14~
## $ Team                <chr> "ATL", "ATL", "ATL", "ATL", "ATL", "ATL", "AT~
## $ Game                <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14~
## $ Date                <date> 2014-10-29, 2014-11-01, 2014-11-05, 2014-11-~
## $ Home                <chr> "Away", "Home", "Away", "Away", "Home", "Away~
## $ Opponent            <chr> "TOR", "IND", "SAS", "CHO", "NYK", "NYK", "UT~
## $ WINorLOSS           <chr> "L", "W", "L", "L", "W", "W", "W", "W", "L", ~
## $ TeamPoints          <dbl> 102, 102, 92, 119, 103, 91, 100, 114, 94, 109~
## $ OpponentPoints      <dbl> 109, 92, 94, 122, 96, 85, 97, 103, 127, 114, ~
## $ FieldGoals          <dbl> 40, 35, 38, 43, 33, 27, 39, 42, 40, 41, 38, 3~
## $ FieldGoalsAttempted <dbl> 80, 69, 92, 93, 81, 71, 76, 75, 90, 85, 80, 8~
## $ FieldGoals.         <dbl> 0.500, 0.507, 0.413, 0.462, 0.407, 0.380, 0.5~
## $ X3PointShots        <dbl> 13, 7, 8, 13, 9, 10, 9, 11, 3, 9, 5, 7, 10, 5~
## $ X3PointShotsAttempted <dbl> 22, 20, 25, 33, 22, 27, 20, 28, 22, 27, 24, 2~
## $ X3PointShots.       <dbl> 0.591, 0.350, 0.320, 0.394, 0.409, 0.370, 0.4~
## $ FreeThrows          <dbl> 9, 25, 8, 20, 28, 27, 13, 19, 11, 18, 18, 27,~
## $ FreeThrowsAttempted <dbl> 17, 33, 11, 26, 36, 28, 18, 23, 13, 23, 26, 3~
## $ FreeThrows.         <dbl> 0.529, 0.758, 0.727, 0.769, 0.778, 0.964, 0.7~
## $ OffRebounds         <dbl> 10, 3, 10, 7, 12, 9, 13, 3, 11, 13, 8, 12, 9,~
## $ TotalRebounds       <dbl> 42, 37, 37, 38, 41, 38, 46, 36, 37, 38, 46, 4~
## $ Assists             <dbl> 26, 26, 26, 28, 18, 20, 23, 33, 26, 22, 20, 1~
## $ Steals              <dbl> 6, 10, 14, 8, 10, 7, 8, 10, 6, 7, 7, 16, 2, 6~
## $ Blocks              <dbl> 8, 6, 5, 3, 5, 3, 4, 5, 8, 3, 5, 6, 4, 7, 5, ~
## $ Turnovers           <dbl> 17, 12, 13, 19, 8, 15, 18, 13, 18, 10, 9, 10,~
## $ TotalFouls          <dbl> 24, 20, 25, 33, 17, 16, 12, 20, 12, 17, 18, 2~
## $ Opp.FieldGoals      <dbl> 37, 31, 31, 48, 40, 36, 43, 35, 49, 47, 31, 3~
```

```
## $ Opp.FieldGoalsAttempted  <dbl> 90, 81, 69, 97, 84, 83, 86, 74, 90, 87, 83, 8~
## $ Opp.FieldGoals.          <dbl> 0.411, 0.383, 0.449, 0.495, 0.476, 0.434, 0.5~
## $ Opp.3PointShots          <dbl> 8, 12, 5, 6, 8, 6, 5, 10, 19, 6, 4, 9, 14, 7,~
## $ Opp.3PointShotsAttempted <dbl> 26, 32, 17, 21, 21, 26, 23, 21, 31, 17, 21, 2~
## $ Opp.3PointShots.         <dbl> 0.308, 0.375, 0.294, 0.286, 0.381, 0.231, 0.2~
## $ Opp.FreeThrows           <dbl> 27, 18, 27, 20, 8, 7, 6, 23, 10, 14, 23, 21, ~
## $ Opp.FreeThrowsAttempted  <dbl> 33, 21, 38, 27, 11, 12, 12, 25, 13, 22, 24, 2~
## $ Opp.FreeThrows.          <dbl> 0.818, 0.857, 0.711, 0.741, 0.727, 0.583, 0.5~
## $ Opp.OffRebounds          <dbl> 16, 11, 11, 11, 13, 11, 8, 5, 12, 13, 13, 16,~
## $ Opp.TotalRebounds        <dbl> 48, 44, 50, 51, 44, 40, 30, 32, 47, 44, 45, 4~
## $ Opp.Assists              <dbl> 26, 25, 25, 31, 26, 23, 28, 27, 39, 24, 12, 2~
## $ Opp.Steals               <dbl> 13, 5, 7, 6, 2, 4, 12, 10, 12, 7, 6, 2, 7, 8,~
## $ Opp.Blocks               <dbl> 9, 5, 9, 7, 6, 2, 8, 3, 2, 0, 3, 3, 7, 4, 4, ~
## $ Opp.Turnovers            <dbl> 9, 18, 19, 19, 15, 15, 11, 14, 13, 11, 12, 20~
## $ Opp.TotalFouls           <dbl> 22, 26, 15, 30, 29, 26, 17, 20, 14, 24, 20, 2~
```

```
glimpse(myMusicData)
```

```
## Rows: 29,503
## Columns: 22
## $ SongID                   <chr> "-twistin'-White Silver SandsBill Black's Co~
## $ Performer                <chr> "Bill Black's Combo", "Augie Rios", "Andy Wi~
## $ Song                     <chr> "-twistin'-White Silver Sands", "¿Dònde Està~
## $ spotify_genre            <chr> "[]", "['novelty']", "['adult standards', 'b~
## $ spotify_track_id         <chr> NA, NA, "3tvqPPpXyIgKrm4PR9HCf0", "1fHHq3qHU~
## $ spotify_track_preview_url <chr> NA, NA, "https://p.scdn.co/mp3-preview/cef48~
## $ spotify_track_duration_ms <dbl> NA, NA, 166106, 172066, 211066, 208186, 2055~
## $ spotify_track_explicit   <lgl> NA, NA, FALSE, FALSE, FALSE, FALSE, TRUE, FA~
## $ spotify_track_album      <chr> NA, NA, "The Essential Andy Williams", "Comp~
## $ danceability             <dbl> NA, NA, 0.154, 0.588, 0.759, 0.613, NA, 0.64~
## $ energy                   <dbl> NA, NA, 0.185, 0.672, 0.699, 0.764, NA, 0.68~
## $ key                      <dbl> NA, NA, 5, 11, 0, 2, NA, 2, NA, NA, 7, NA, 1~
## $ loudness                 <dbl> NA, NA, -14.063, -17.278, -5.745, -6.509, NA~
## $ mode                     <dbl> NA, NA, 1, 0, 0, 1, NA, 0, NA, NA, 1, NA, 0,~
## $ speechiness              <dbl> NA, NA, 0.0315, 0.0361, 0.0307, 0.1360, NA, ~
## $ acousticness             <dbl> NA, NA, 0.91100, 0.00256, 0.20200, 0.05270, ~
## $ instrumentalness         <dbl> NA, NA, 2.67e-04, 7.45e-01, 1.31e-04, 0.00e+~
## $ liveness                 <dbl> NA, NA, 0.1120, 0.1450, 0.4430, 0.1970, NA, ~
## $ valence                  <dbl> NA, NA, 0.150, 0.801, 0.907, 0.417, NA, 0.95~
## $ tempo                    <dbl> NA, NA, 83.969, 121.962, 92.960, 160.015, NA~
## $ time_signature           <dbl> NA, NA, 4, 4, 4, 4, NA, 4, NA, NA, 4, NA, 4,~
## $ spotify_track_popularity <dbl> NA, NA, 38, 11, 77, 73, 61, 40, NA, NA, 31, ~
```