UCS548 DASHBOARD PROJECT
IMDb Dataset

-Arjun Khanchandani
CS1-Roll No.102017005

DESCRIPTION OF DATASET-

The dataset is the IMDb 5000 dataset, which has 5000+ records of movies from all over the world. It has 31 features (which have been increased during the process of pre-processing). I have 6 different tables which have been joined and pre-processed using R commands on R Studio. In total there are 5043 different rows of unique movies and their details.

R Commands-

- **Reading the 6 tables as dataframes from their respective csv files**

```
1  file1 <- "/Users/arjunkhanchandani/Desktop/table_content.csv"
2  table1 <- read.csv(file1)
3
4  file2 <- "/Users/arjunkhanchandani/Desktop/table_director.csv"
5  table2 <- read.csv(file2)
6
7  file3 <- "/Users/arjunkhanchandani/Desktop/table_movies.csv"
8  table3 <- read.csv(file3)
9
10 file4 <- "/Users/arjunkhanchandani/Desktop/table_finance.csv"
11 table4 <- read.csv(file4)
12
13 file5 <- "/Users/arjunkhanchandani/Desktop/table_critics.csv"
14 table5 <- read.csv(file5)
15
16 file6 <- "/Users/arjunkhanchandani/Desktop/table_aspect.csv"
17 table6 <- read.csv(file6)
18
```

Data	
table1	5043 obs. of 3 variables
table2	5043 obs. of 2 variables
table3	5043 obs. of 10 variables
table4	5043 obs. of 3 variables
table5	5043 obs. of 13 variables
table6	5043 obs. of 3 variables

- **Displaying the first 5 rows of each dataframe**

```
19  head(table1)
20  head(table2)
21  head(table3)
22  head(table4)
23  head(table5)
24  head(table6)
```

```
> head(table1)
                                      movie_title                                          plot_keywords content_rating
1                                         Avatar¬†                      avatar|future|marine|native|paraplegic          PG-13
2          Pirates of the Caribbean: At World's End¬†  goddess|marriage ceremony|marriage proposal|pirate|singapore          PG-13
3                                        Spectre¬†                          bomb|espionage|sequel|spy|terrorist          PG-13
4                          The Dark Knight Rises¬† deception|imprisonment|lawlessness|police officer|terrorist plot          PG-13
5 Star Wars: Episode VII - The Force Awakens¬†
6                                      John Carter¬†           alien|american civil war|male nipple|mars|princess          PG-13
>
```

```
> head(table2)
      director_name director_facebook_likes
1     James Cameron                     0
2    Gore Verbinski                   563
3        Sam Mendes                     0
4 Christopher Nolan                 22000
5       Doug Walker                   131
6    Andrew Stanton                   475
>
```

```
> head(table3)
  color     director_name                                        movie_title duration                          genres
1 Color     James Cameron                                          Avatar¬†      178 Action|Adventure|Fantasy|Sci-Fi
2 Color    Gore Verbinski          Pirates of the Caribbean: At World's End¬†      169        Action|Adventure|Fantasy
3 Color        Sam Mendes                                        Spectre¬†      148       Action|Adventure|Thriller
4 Color Christopher Nolan                              The Dark Knight Rises¬†      164                 Action|Thriller
5           Doug Walker Star Wars: Episode VII - The Force Awakens¬†               NA                     Documentary
6 Color    Andrew Stanton                                      John Carter¬†      132         Action|Adventure|Sci-Fi
                          movie_imdb_link language country title_year imdb_score
1 http://www.imdb.com/title/tt0499549/?ref_=fn_tt_tt_1  English     USA       2009        7.9
2 http://www.imdb.com/title/tt0449088/?ref_=fn_tt_tt_1  English     USA       2007        7.1
3 http://www.imdb.com/title/tt2379713/?ref_=fn_tt_tt_1  English      UK       2015        6.8
4 http://www.imdb.com/title/tt1345836/?ref_=fn_tt_tt_1  English     USA       2012        8.5
5 http://www.imdb.com/title/tt5289954/?ref_=fn_tt_tt_1                          NA        7.1
6 http://www.imdb.com/title/tt0401729/?ref_=fn_tt_tt_1  English     USA       2012        6.6
>
```

```
> head(table4)
                                      movie_title    budget     gross
1                                         Avatar¬† 237000000 760505847
2          Pirates of the Caribbean: At World's End¬† 300000000 309404152
3                                        Spectre¬† 245000000 200074175
4                          The Dark Knight Rises¬† 250000000 448130642
5 Star Wars: Episode VII - The Force Awakens¬†                NA        NA
6                                      John Carter¬† 263700000  73058679
>
```

```
> head(table5)
                                      movie_title num_voted_users num_critic_for_reviews    actor_1_name actor_1_facebook_likes
1                                          Avatar¬†         886204                  723     CCH Pounder                  1000
2              Pirates of the Caribbean: At World's End¬†         471220                  302     Johnny Depp                 40000
3                                        Spectre¬†         275868                  602 Christoph Waltz                 11000
4                                The Dark Knight Rises¬†        1144337                  813       Tom Hardy                 27000
5 Star Wars: Episode VII - The Force Awakens¬†                      8                   NA     Doug Walker                   131
6                                     John Carter¬†         212204                  462    Daryl Sabara                   640
      actor_2_name actor_2_facebook_likes         actor_3_name actor_3_facebook_likes cast_total_facebook_likes num_critic_for_reviews.1
1 Joel David Moore                   936            Wes Studi                   855                     4834                     723
2    Orlando Bloom                  5000       Jack Davenport                  1000                    48350                     302
3     Rory Kinnear                   393     Stephanie Sigman                   161                    11700                     602
4   Christian Bale                 23000 Joseph Gordon-Levitt                 23000                   106759                     813
5      Rob Walker                    12                                          NA                      143                      NA
6  Samantha Morton                   632         Polly Walker                   530                     1873                     462
  num_user_for_reviews movie_facebook_likes
1                 3054                33000
2                 1238                    0
3                  994                85000
4                 2701               164000
5                   NA                    0
6                  738                24000
>
```

```
> head(table6)
                                      movie_title aspect_ratio facenumber_in_poster
1                                          Avatar¬†         1.78                   0
2              Pirates of the Caribbean: At World's End¬†         2.35                   0
3                                        Spectre¬†         2.35                   1
4                                The Dark Knight Rises¬†         2.35                   0
5 Star Wars: Episode VII - The Force Awakens¬†                   NA                   0
6                                     John Carter¬†         2.35                   1
>
```

- **Joining the 6 dataframes together to form one single dataframe and its summary**

```
25
26  final_table <- cbind(table1, table2, table3, table4, table5, table6)
27  summary(final_table)
28
```

```
> final_table <- cbind(table1, table2, table3, table4, table5, table6)
> summary(final_table)
 movie_title        plot_keywords      content_rating     director_name      director_facebook_likes
 Length:5043        Length:5043        Length:5043        Length:5043        Min.   :    0.0
 Class :character   Class :character   Class :character   Class :character   1st Qu.:    7.0
 Mode  :character   Mode  :character   Mode  :character   Mode  :character   Median :   49.0
                                                                            Mean   :  686.5
                                                                            3rd Qu.:  194.5
                                                                            Max.   :23000.0
                                                                            NA's   :104
    color           director_name      movie_title          duration        genres          movie_imdb_link
 Length:5043        Length:5043        Length:5043        Min.   :  7.0   Length:5043        Length:5043
 Class :character   Class :character   Class :character   1st Qu.: 93.0   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Median :103.0   Mode  :character   Mode  :character
                                                          Mean   :107.2
                                                          3rd Qu.:118.0
                                                          Max.   :511.0
                                                          NA's   :15
   language           country            title_year     imdb_score    movie_title           budget
 Length:5043        Length:5043        Min.   :1916   Min.   :1.600   Length:5043        Min.   :2.180e+02
 Class :character   Class :character   1st Qu.:1999   1st Qu.:5.800   Class :character   1st Qu.:6.000e+06
 Mode  :character   Mode  :character   Median :2005   Median :6.600   Mode  :character   Median :2.000e+07
                                       Mean   :2002   Mean   :6.442                      Mean   :3.975e+07
                                       3rd Qu.:2011   3rd Qu.:7.200                      3rd Qu.:4.500e+07
                                       Max.   :2016   Max.   :9.500                      Max.   :1.222e+10
                                       NA's   :108                                       NA's   :492
```

```
     gross           movie_title        num_voted_users   num_critic_for_reviews actor_1_name
 Min.   :      162   Length:5043        Min.   :      5   Min.   :  1.0          Length:5043
 1st Qu.:  5340988   Class :character   1st Qu.:   8594   1st Qu.: 50.0          Class :character
 Median : 25517500   Mode  :character   Median :  34359   Median :110.0          Mode  :character
 Mean   : 48468408                      Mean   :  83668   Mean   :140.2
 3rd Qu.: 62309438                      3rd Qu.:  96309   3rd Qu.:195.0
 Max.   :760505847                      Max.   :1689764   Max.   :813.0
 NA's   :884                                              NA's   :50
 actor_1_facebook_likes actor_2_name       actor_2_facebook_likes actor_3_name       actor_3_facebook_likes
 Min.   :     0         Length:5043        Min.   :     0         Length:5043        Min.   :    0.0
 1st Qu.:   614         Class :character   1st Qu.:   281         Class :character   1st Qu.:  133.0
 Median :   988         Mode  :character   Median :   595         Mode  :character   Median :  371.5
 Mean   :  6560                            Mean   :  1652                            Mean   :  645.0
 3rd Qu.: 11000                            3rd Qu.:   918                            3rd Qu.:  636.0
 Max.   :640000                            Max.   :137000                            Max.   :23000.0
 NA's   :7                                 NA's   :13                                NA's   :23
 cast_total_facebook_likes num_critic_for_reviews.1 num_user_for_reviews movie_facebook_likes movie_title
 Min.   :     0            Min.   :  1.0           Min.   :   1.0       Min.   :     0       Length:5043
 1st Qu.:  1411            1st Qu.: 50.0           1st Qu.:  65.0       1st Qu.:     0       Class :character
 Median :  3090            Median :110.0           Median : 156.0       Median :   166       Mode  :character
 Mean   :  9699            Mean   :140.2           Mean   : 272.8       Mean   :  7526
 3rd Qu.: 13756            3rd Qu.:195.0           3rd Qu.: 326.0       3rd Qu.:  3000
 Max.   :656730            Max.   :813.0           Max.   :5060.0       Max.   :349000
                           NA's   :50              NA's   :21
  aspect_ratio   facenumber_in_poster
 Min.   : 1.18   Min.   : 0.000
 1st Qu.: 1.85   1st Qu.: 0.000
 Median : 2.35   Median : 1.000
 Mean   : 2.22   Mean   : 1.371
 3rd Qu.: 2.35   3rd Qu.: 2.000
 Max.   :16.00   Max.   :43.000
 NA's   :329     NA's   :13
```

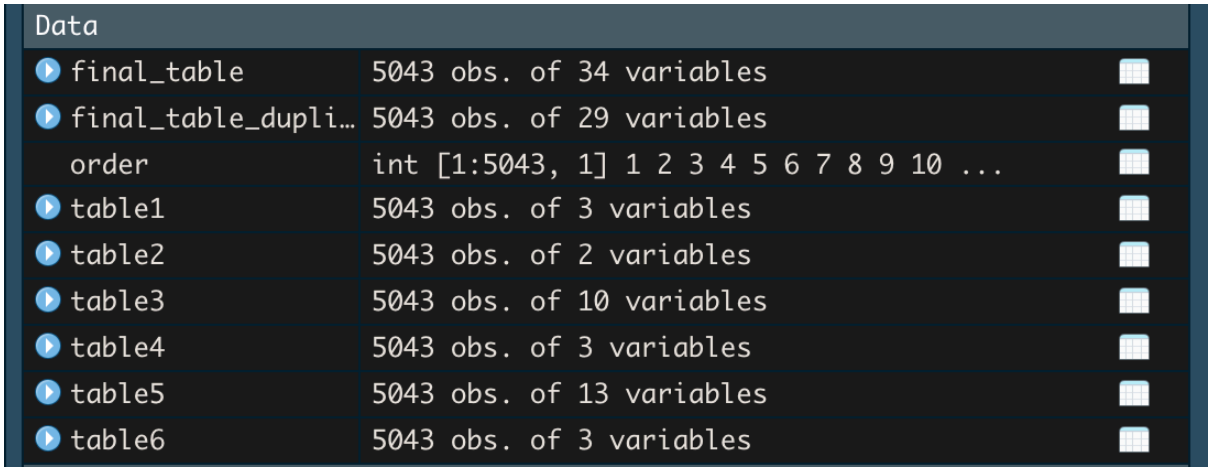
- **Delete duplicate columns from the dataframe**

```
28
29  final_table_duplicates <- final_table[!duplicated(as.list(final_table))]
30
```

Data	
final_table	5043 obs. of 34 variables
final_table_dupli…	5043 obs. of 28 variables
order	int [1:5043, 1] 1 2 3 4 5 6 7 8 9 10 ...
table1	5043 obs. of 3 variables
table2	5043 obs. of 2 variables
table3	5043 obs. of 10 variables
table4	5043 obs. of 3 variables
table5	5043 obs. of 13 variables
table6	5043 obs. of 3 variables

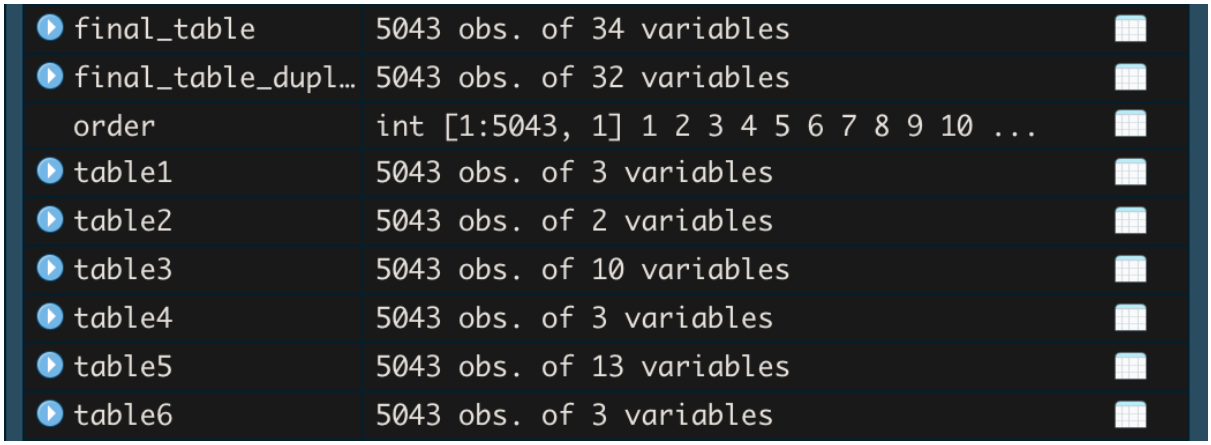
- **Adding a column to give the order number of the records**

```
31  no <- nrow(final_table_duplicates)
32  order <- seq.int(1,no)
33  order <- matrix(order)
34  final_table_duplicates <- cbind(order,final_table_duplicates)
```



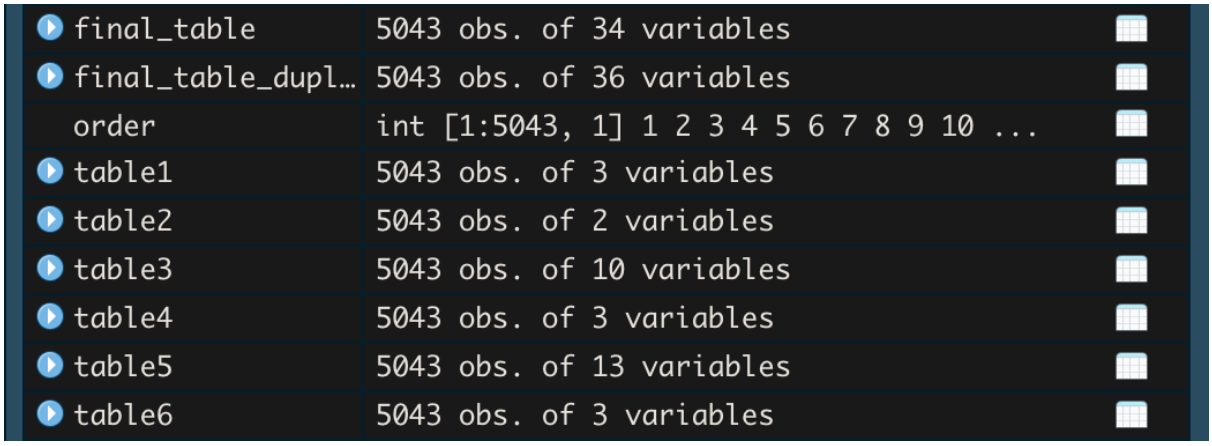
- **Splitting the "genres" column into three using dplyr and tidyr library**

```
36 library(dplyr)
37 library(tidyr)
38
39 final_table_duplicates <- final_table_duplicates %>% separate(genres, c('genre1', 'genre2', 'genre3'))
40 final_table_duplicates <- cbind(final_table_duplicates,final_table$genres)
```



- **Splitting the "plot_keywords" column into three using dplyr and tidyr library**

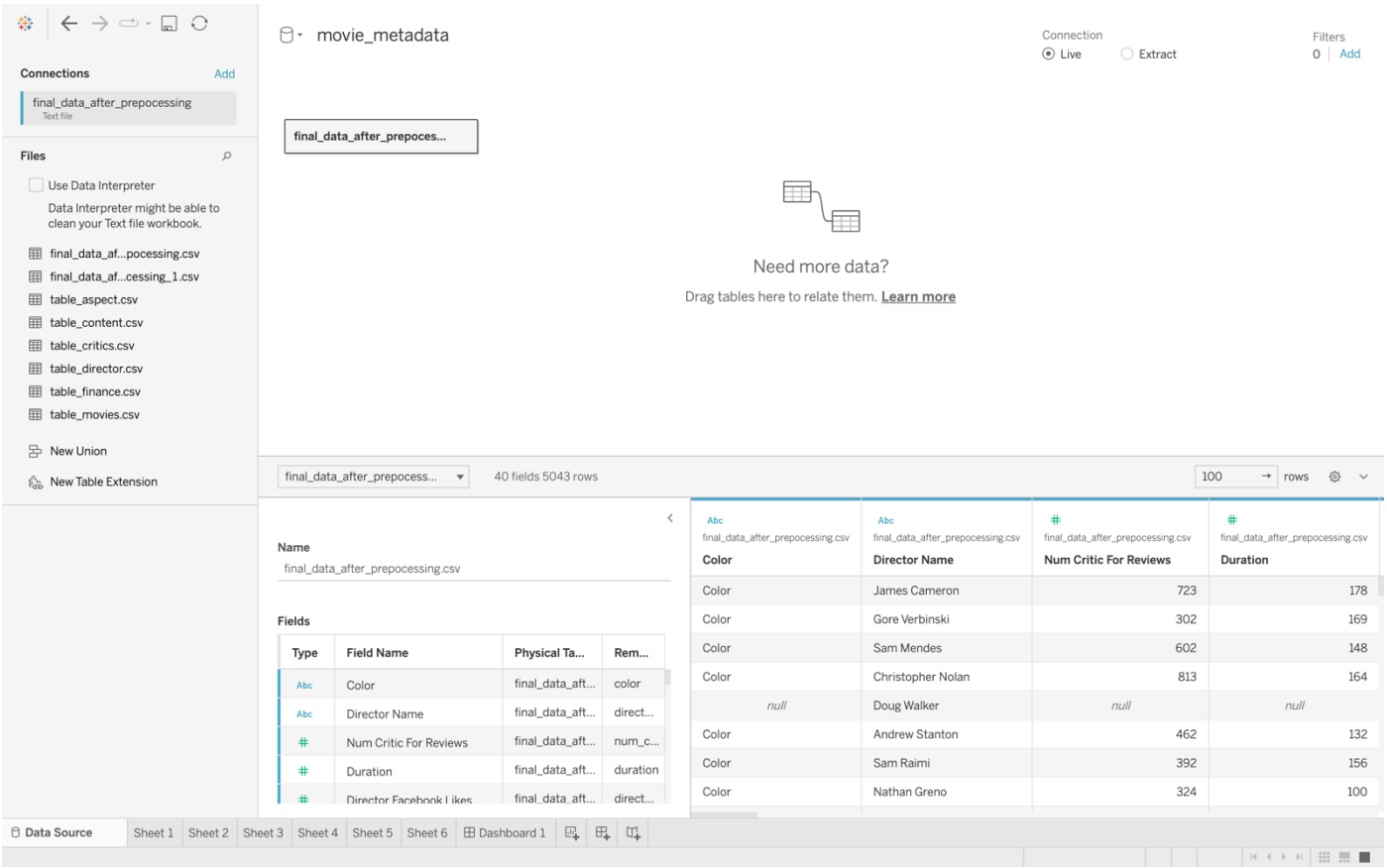
```
42 final_table_duplicates <- final_table_duplicates %>% separate(plot_keywords,
43                                                              c('plot_keyword1', 'plot_keyword2', 'plot_keyword3', 'plot_keyword4'))
44 final_table_duplicates <- cbind(final_table_duplicates,final_table$plot_keywords)
```



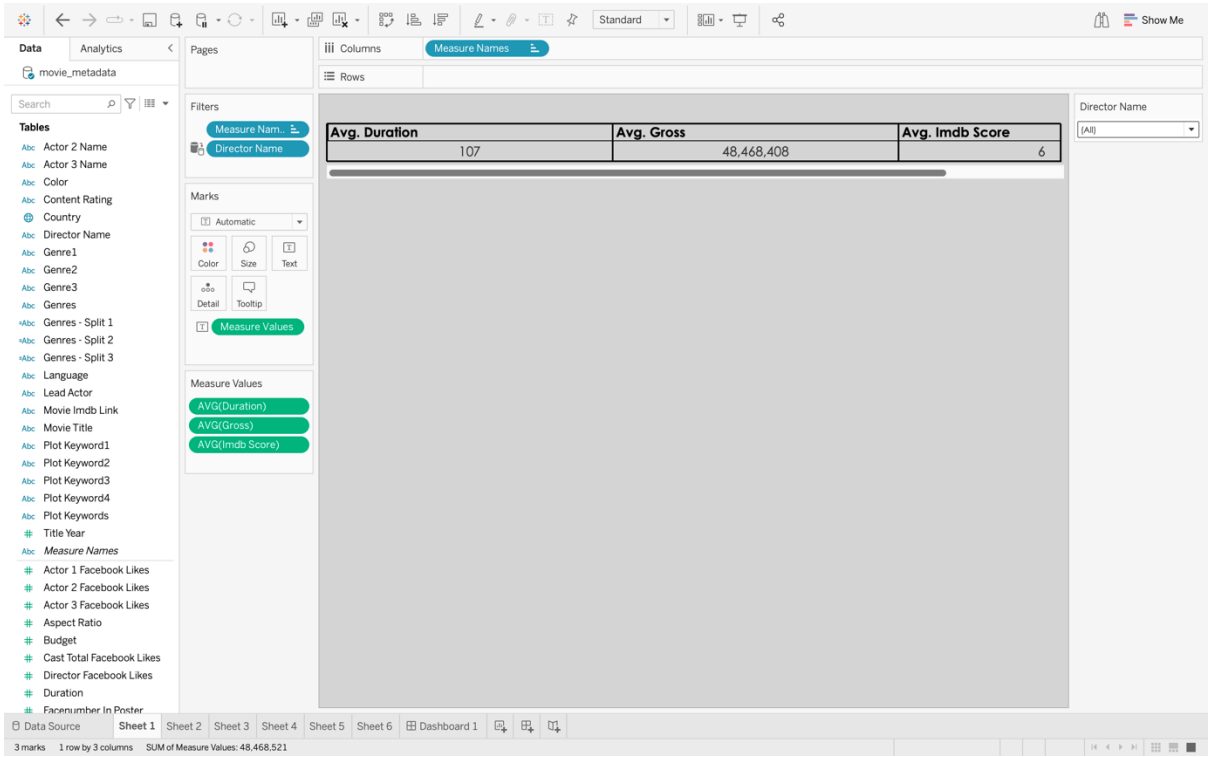
- **Write the dataframe to a csv file to use it in tableau**

```
45
46 write.csv(final_table_duplicates,"/Users/arjunkhanchandani/Desktop/final_data_after_prepocessing.csv")
47
48
```

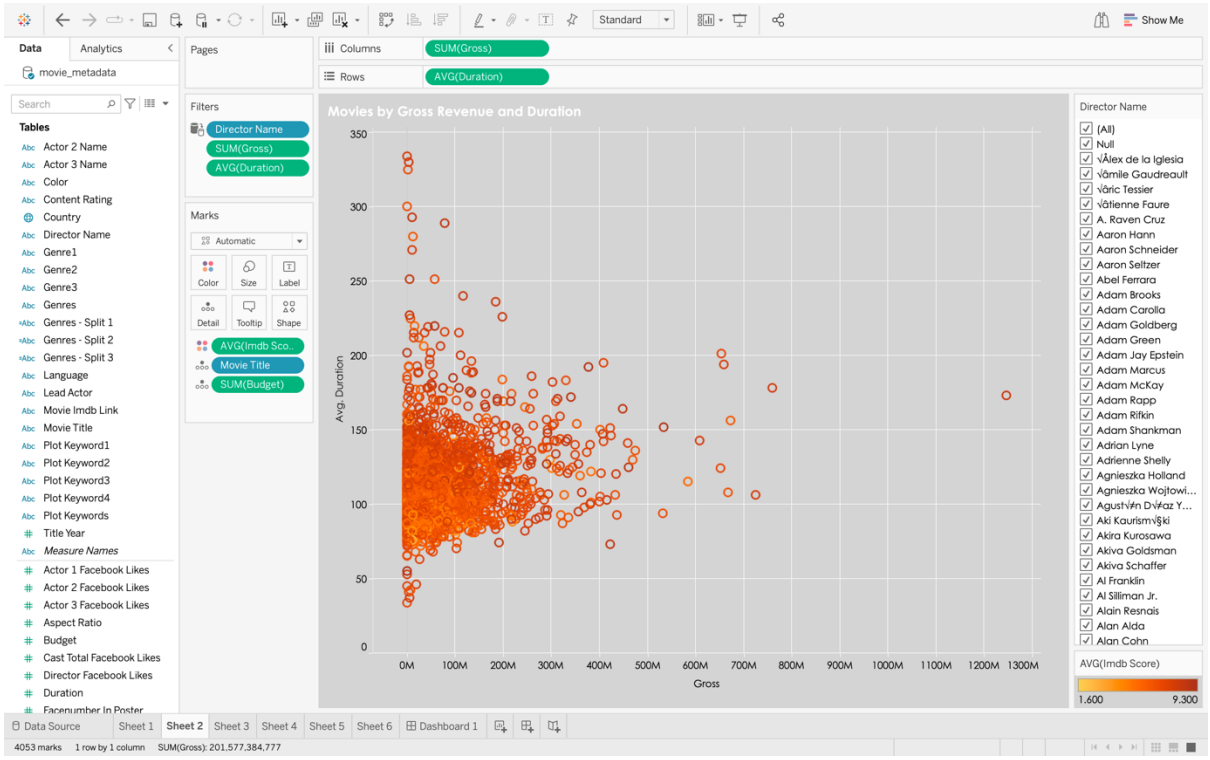
CONNECTING DATA SOURCE FILE TO TABLEAU WORKSPACE-



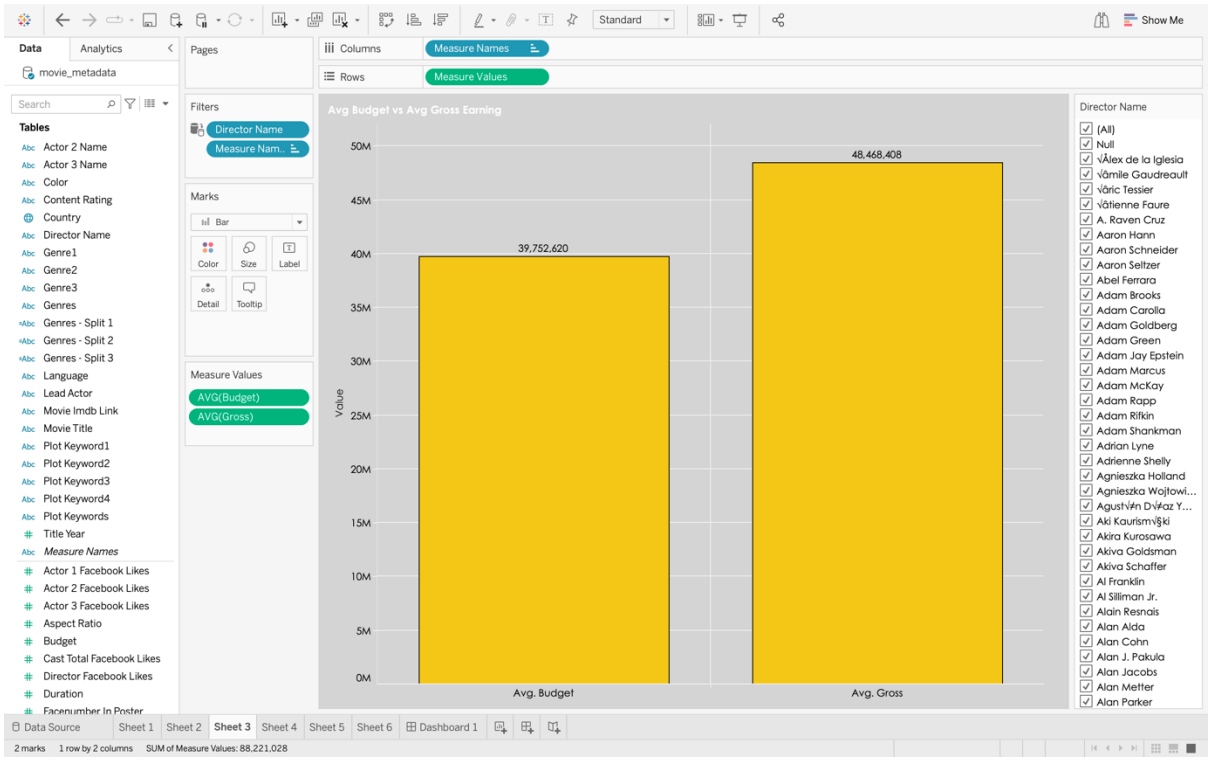
SHEET 1 DATA FIELDS-



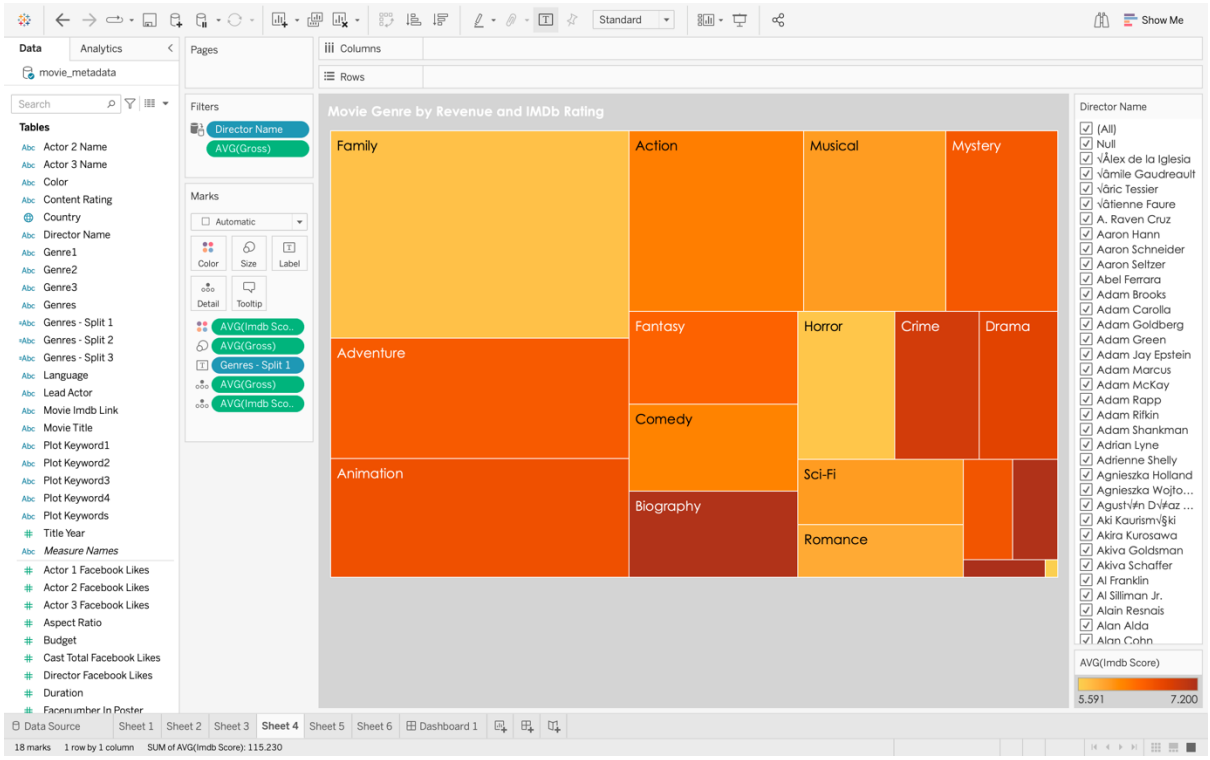
SHEET 2 DATA FIELDS-



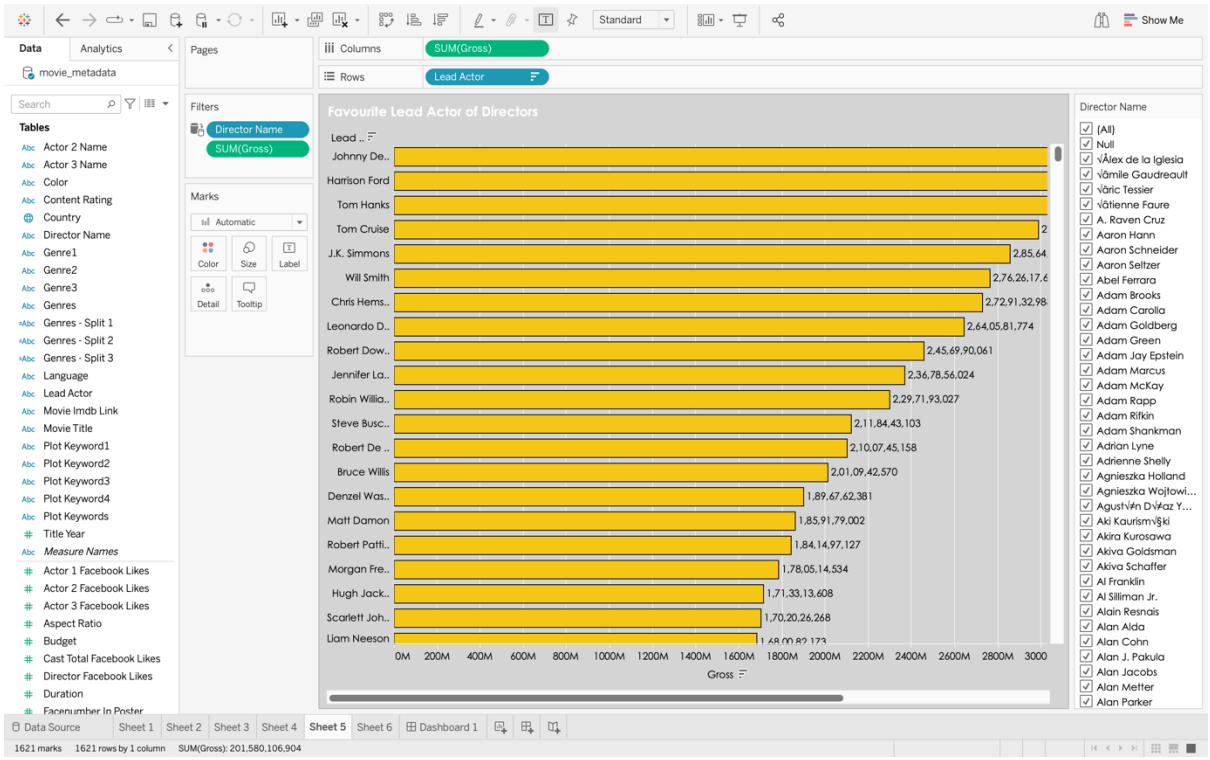
SHEET 3 DATA FIELDS-



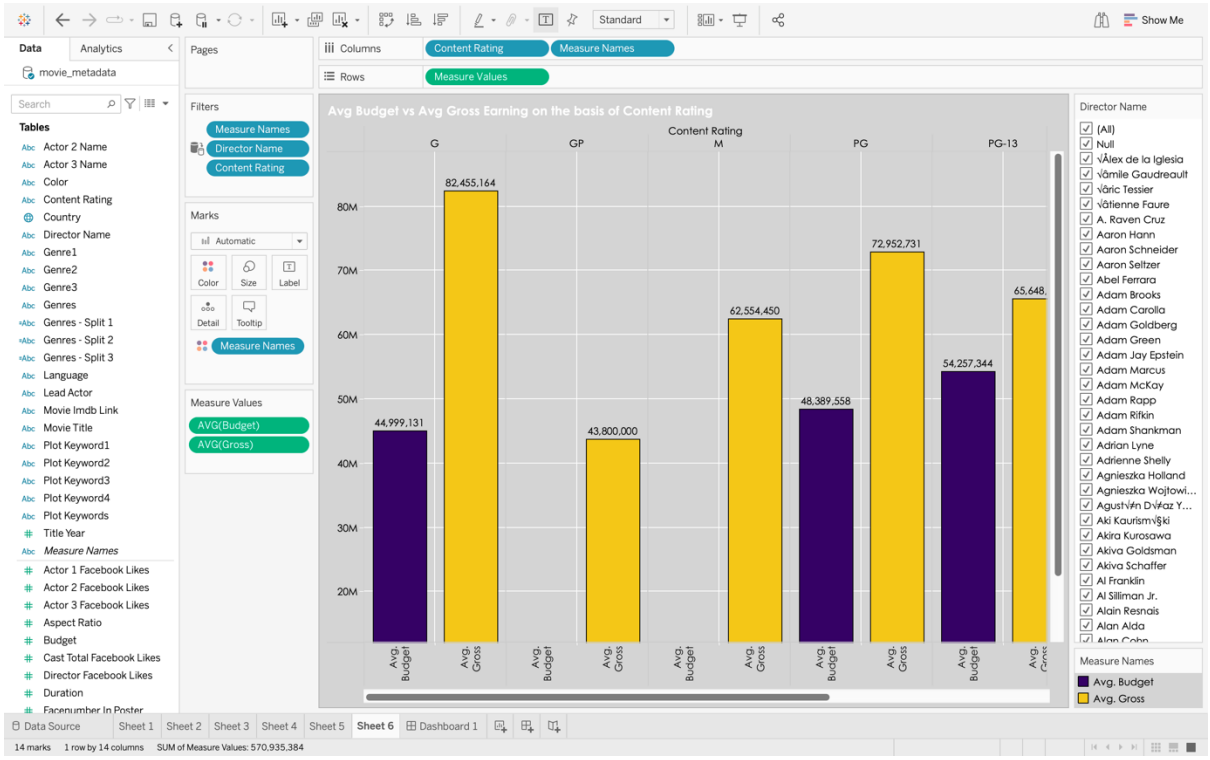
SHEET 4 DATA FIELDS-



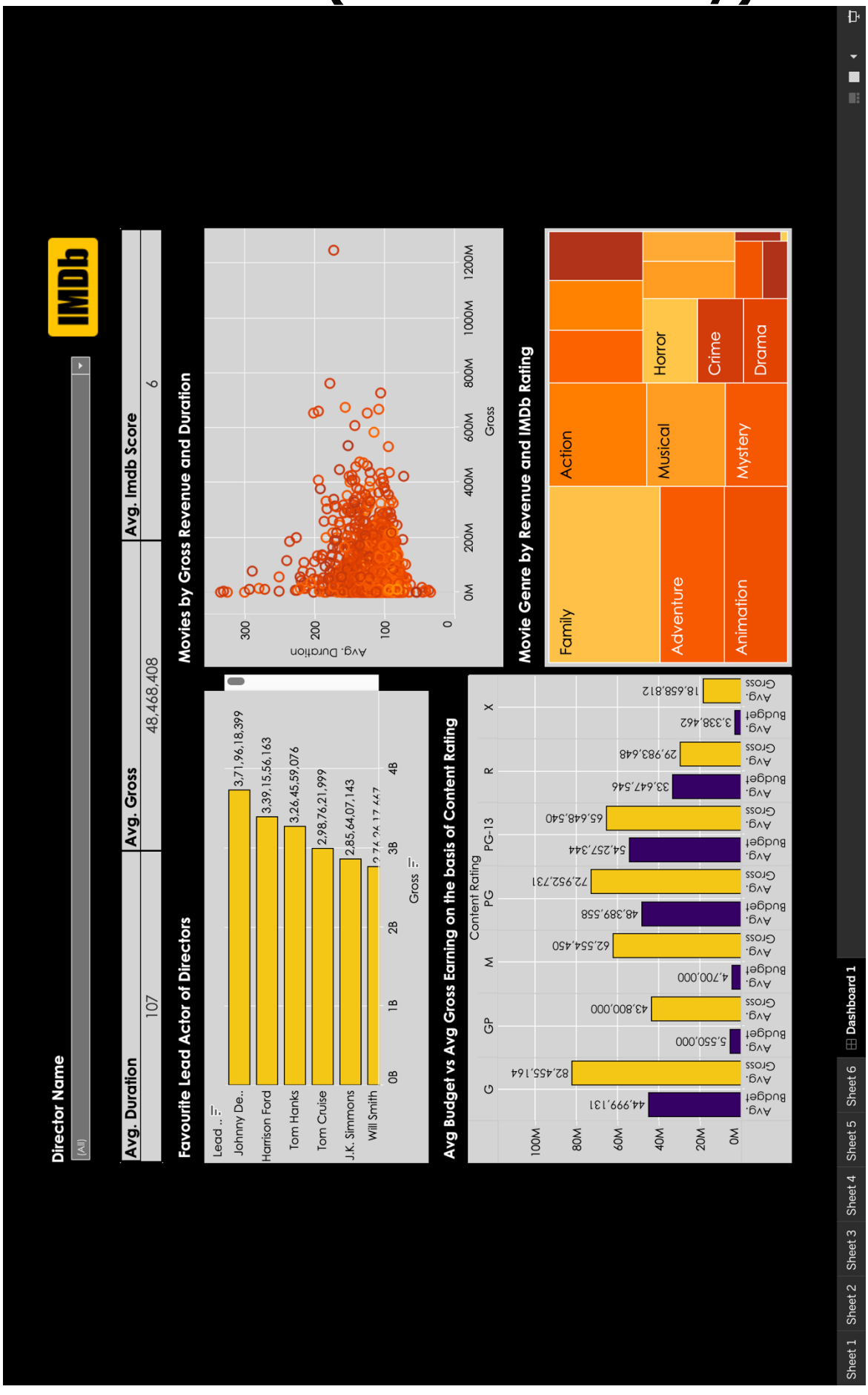
SHEET 5 DATA FIELDS-



SHEET 6 DATA FIELDS-

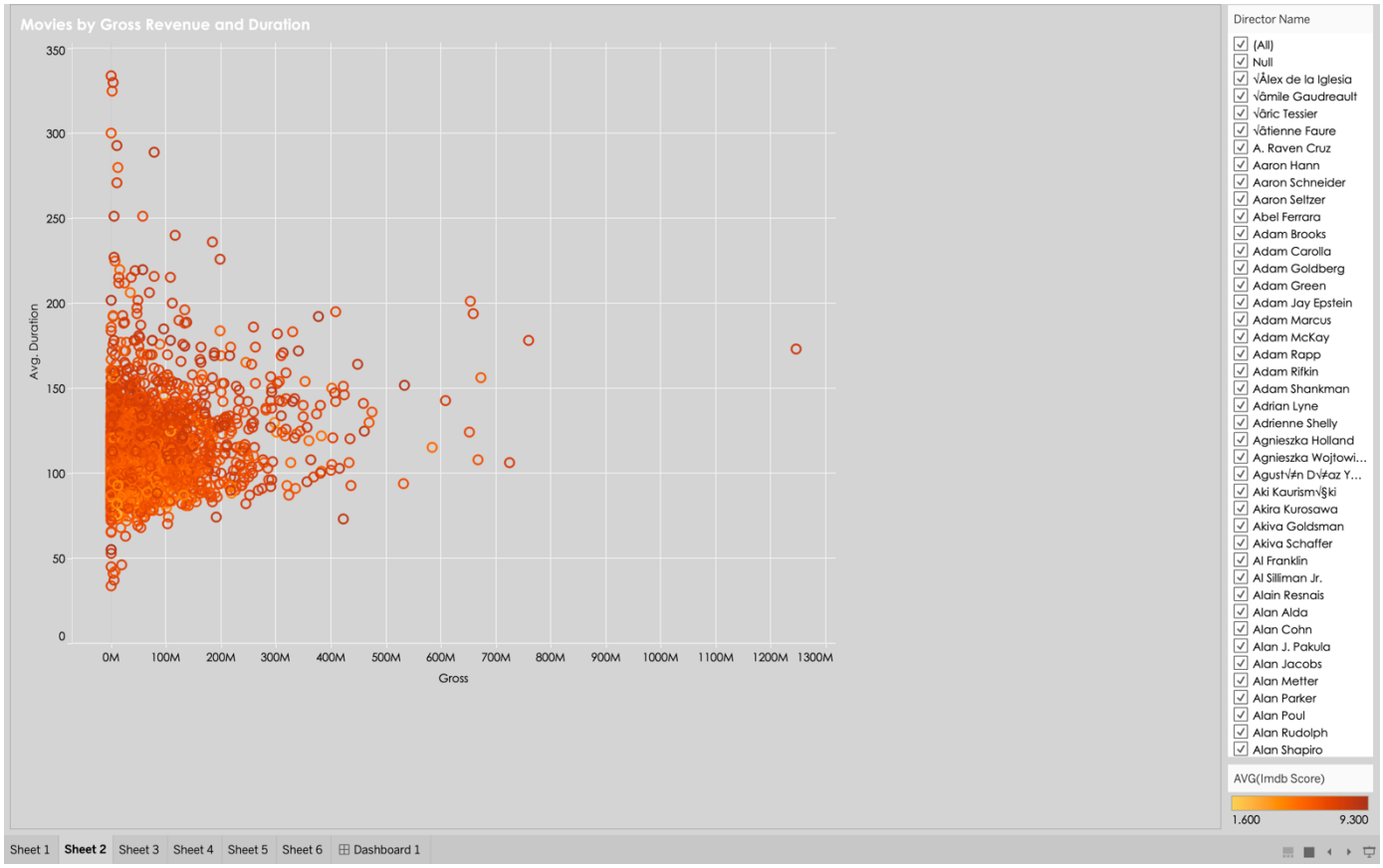


DASHBOARD (All Data Query)

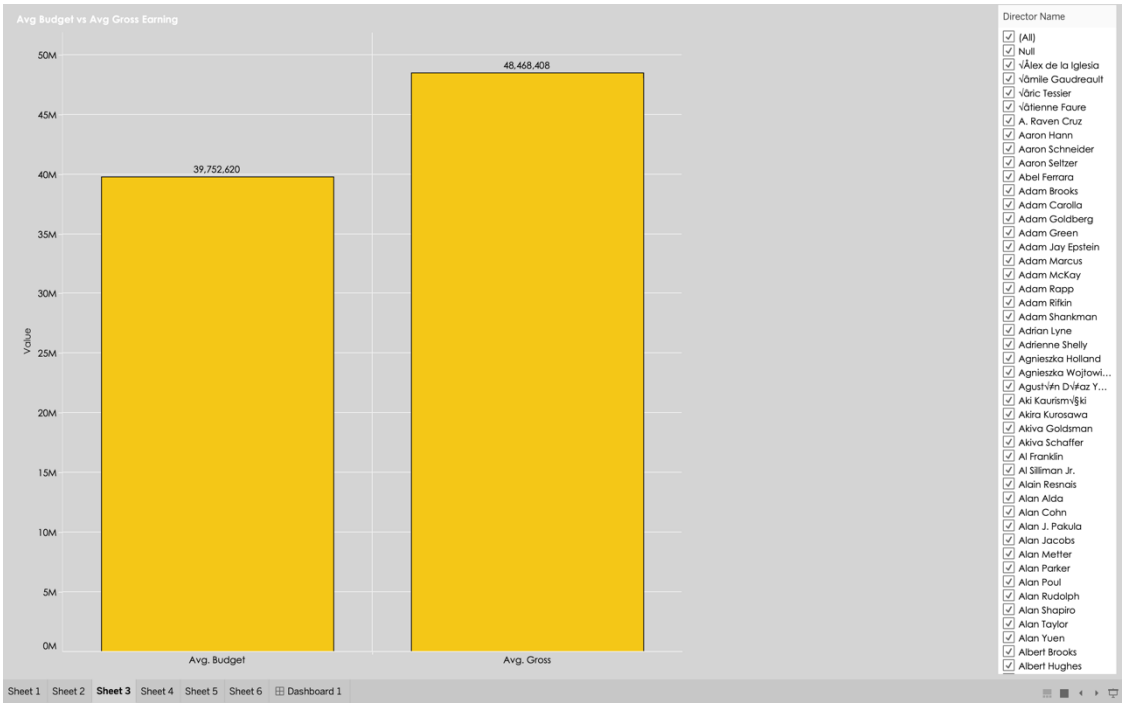


Sheet 1 (With All The Directors' Data Query)

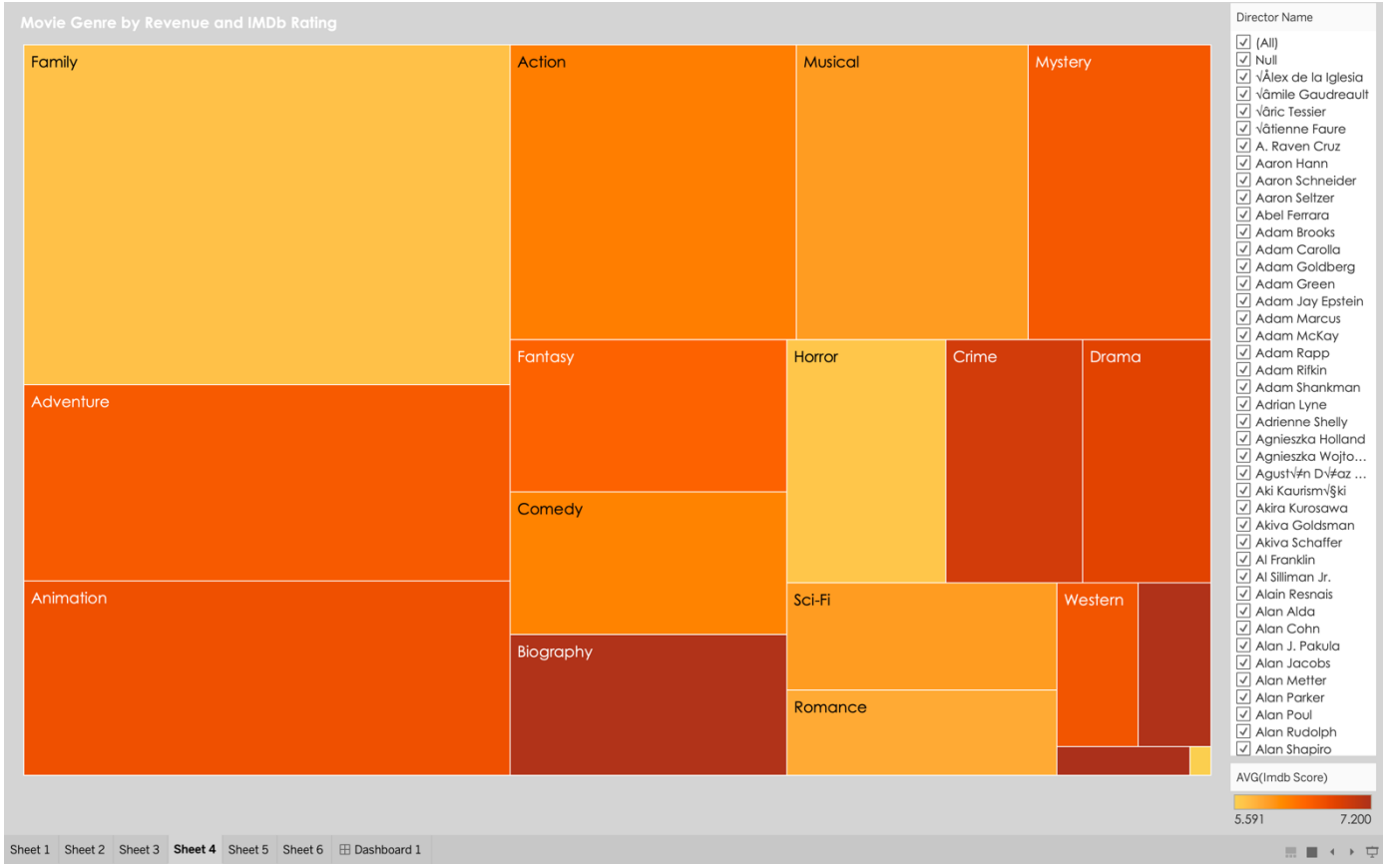
Sheet 2 (With All The Directors' Data Query)



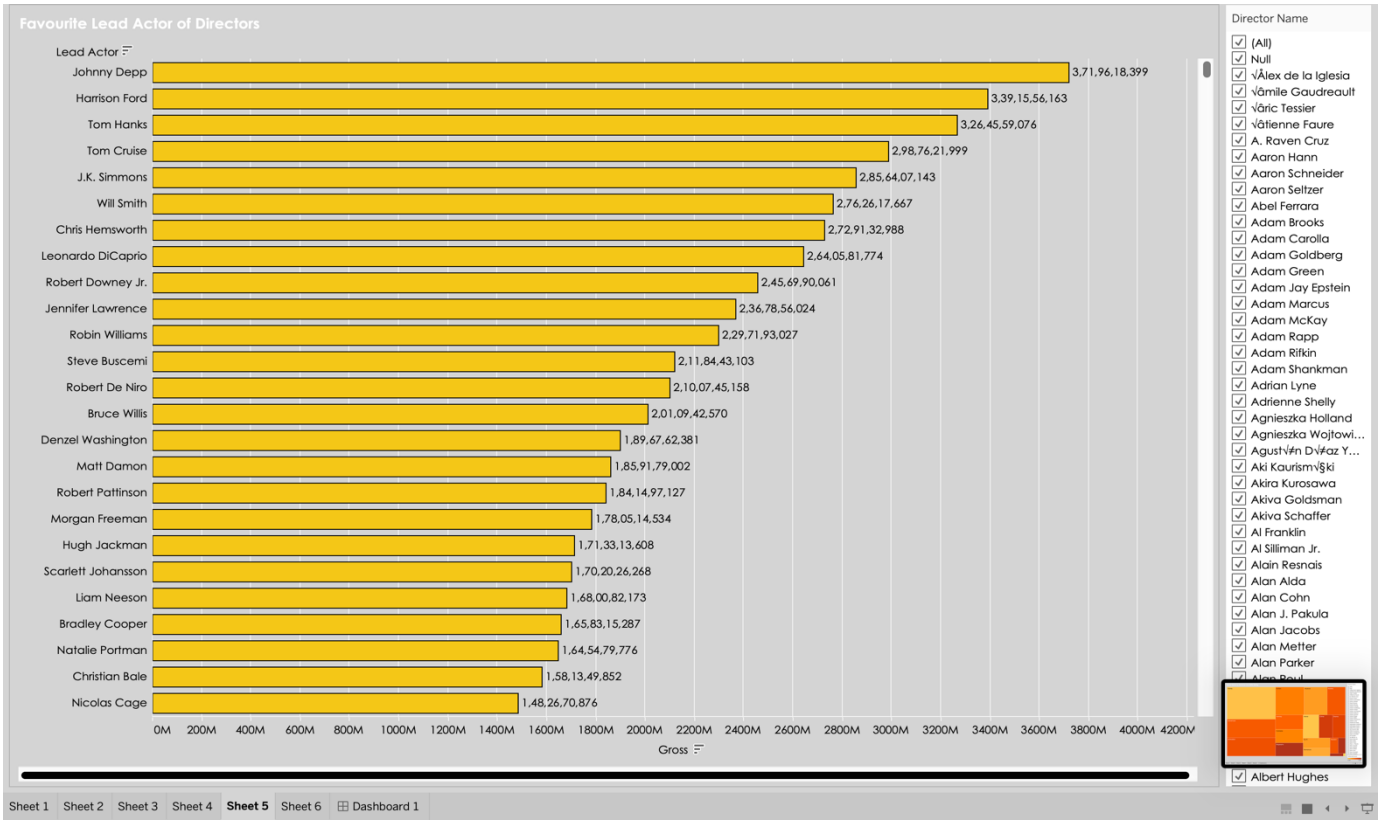
Sheet 3 (With All The Directors' Data Query)



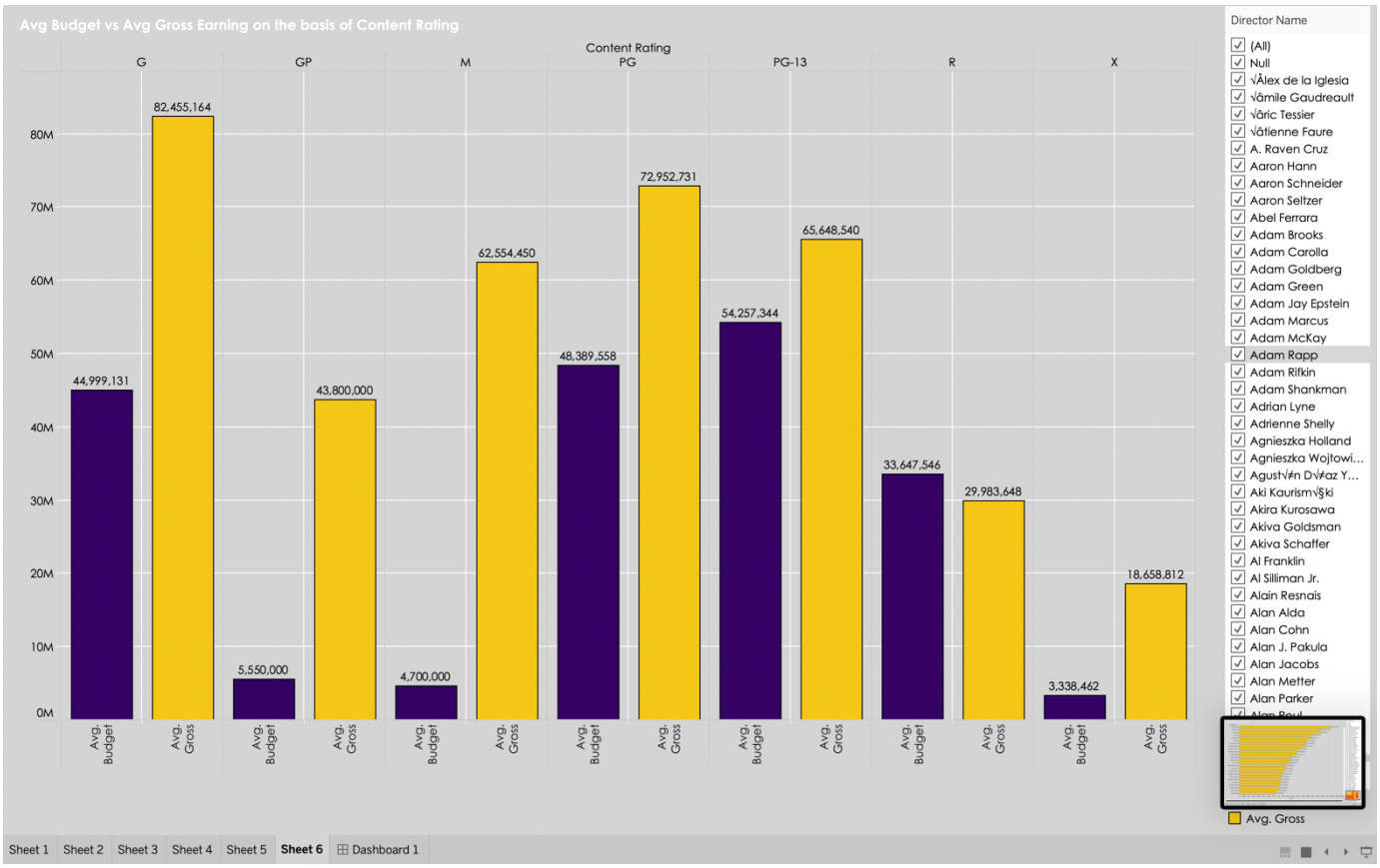
Sheet 4 (With All The Directors' Data Query)



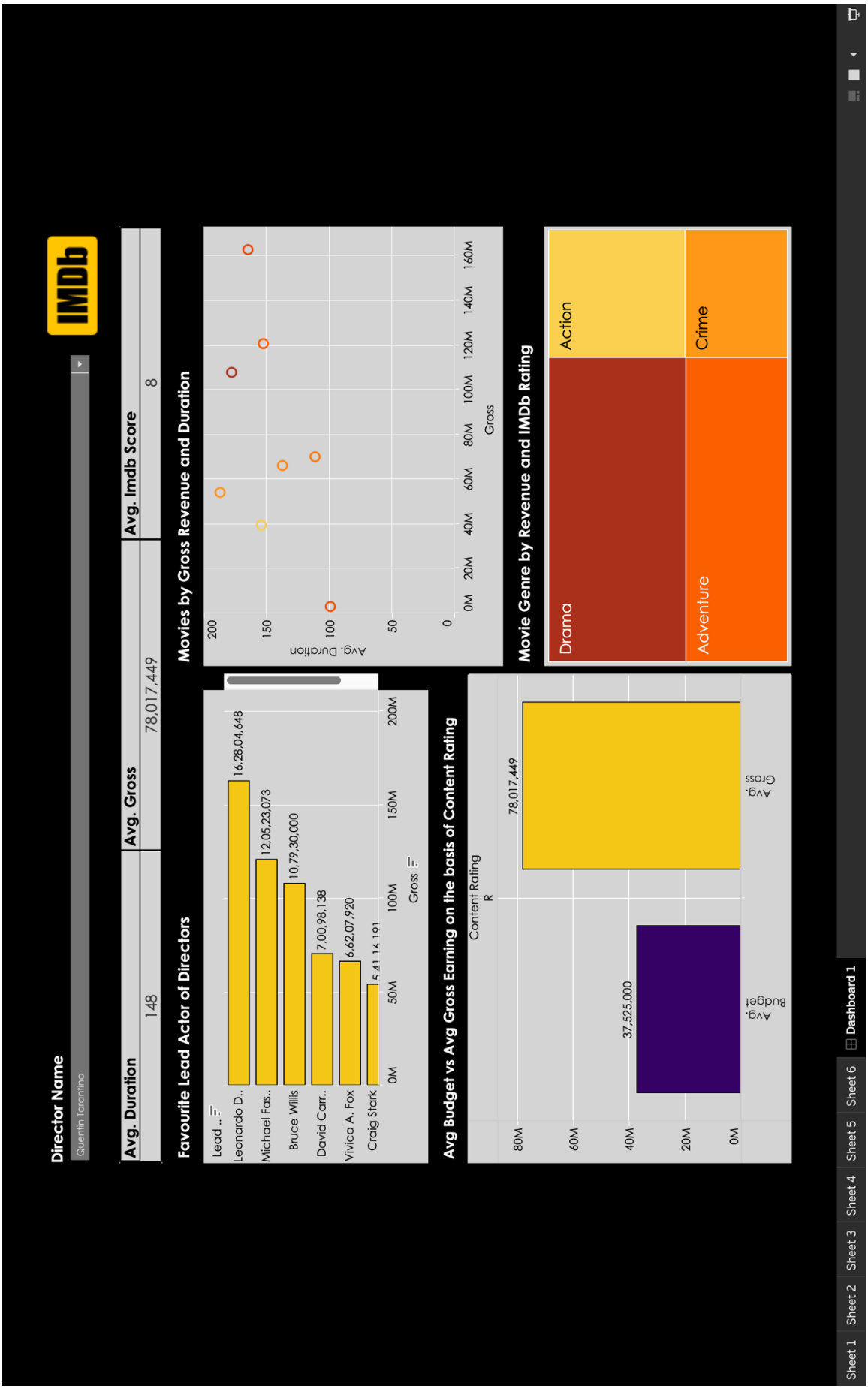
Sheet 5 (With All The Directors' Data Query)



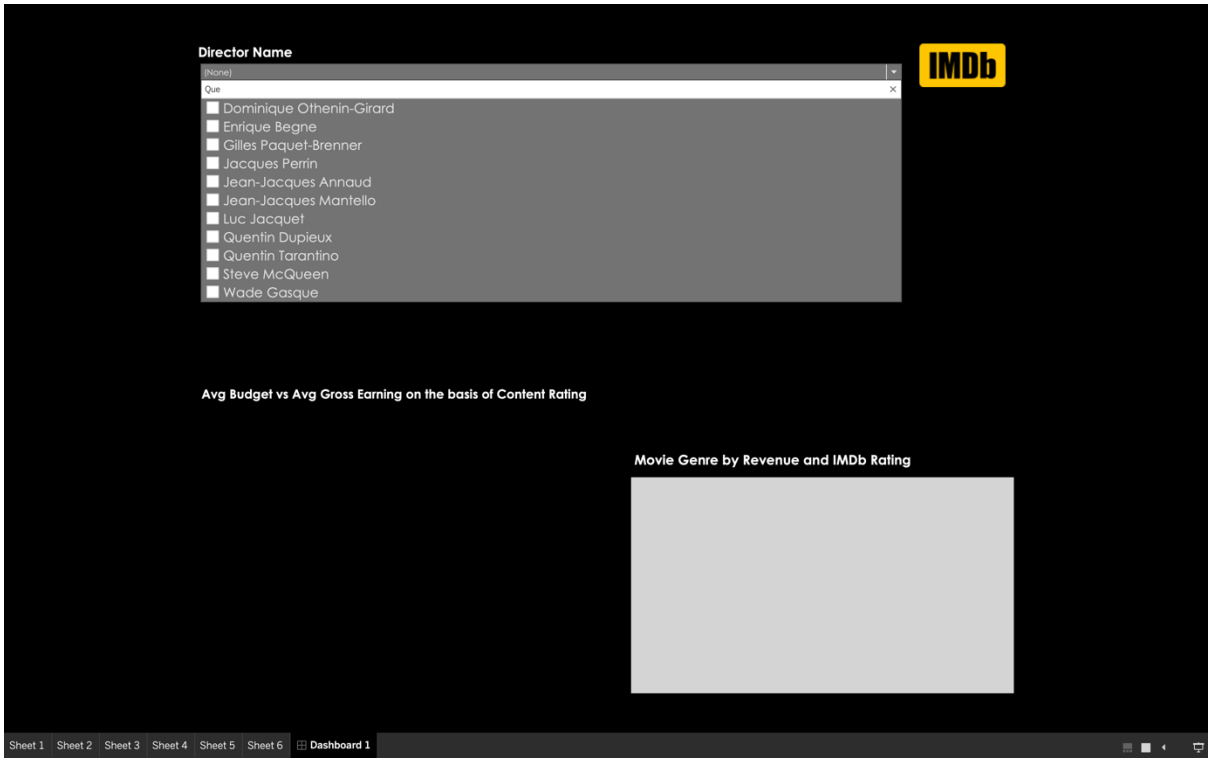
Sheet 6 (With All The Directors' Data Query)



DASHBOARD (Filtered Query)

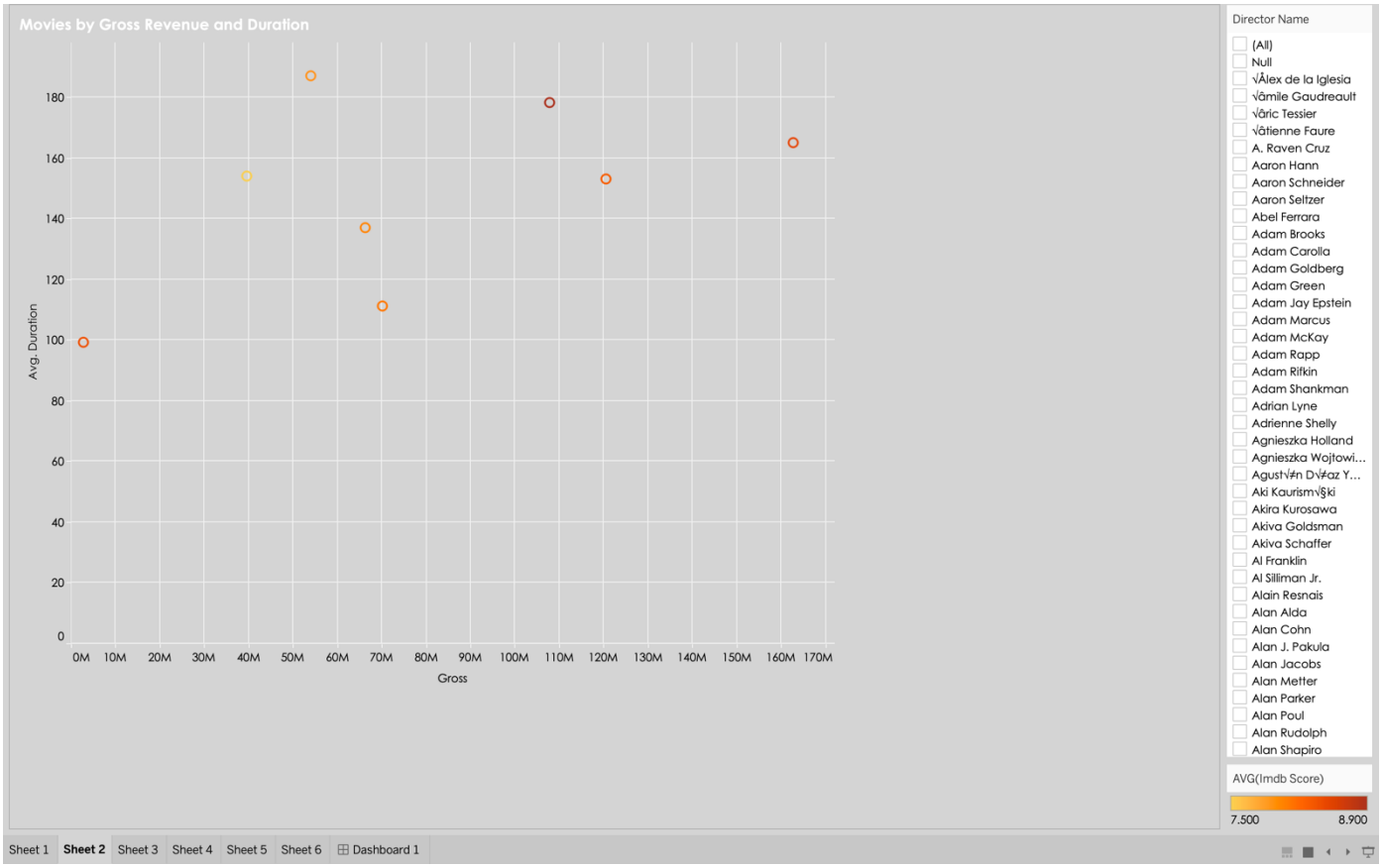


Search Function (filter)

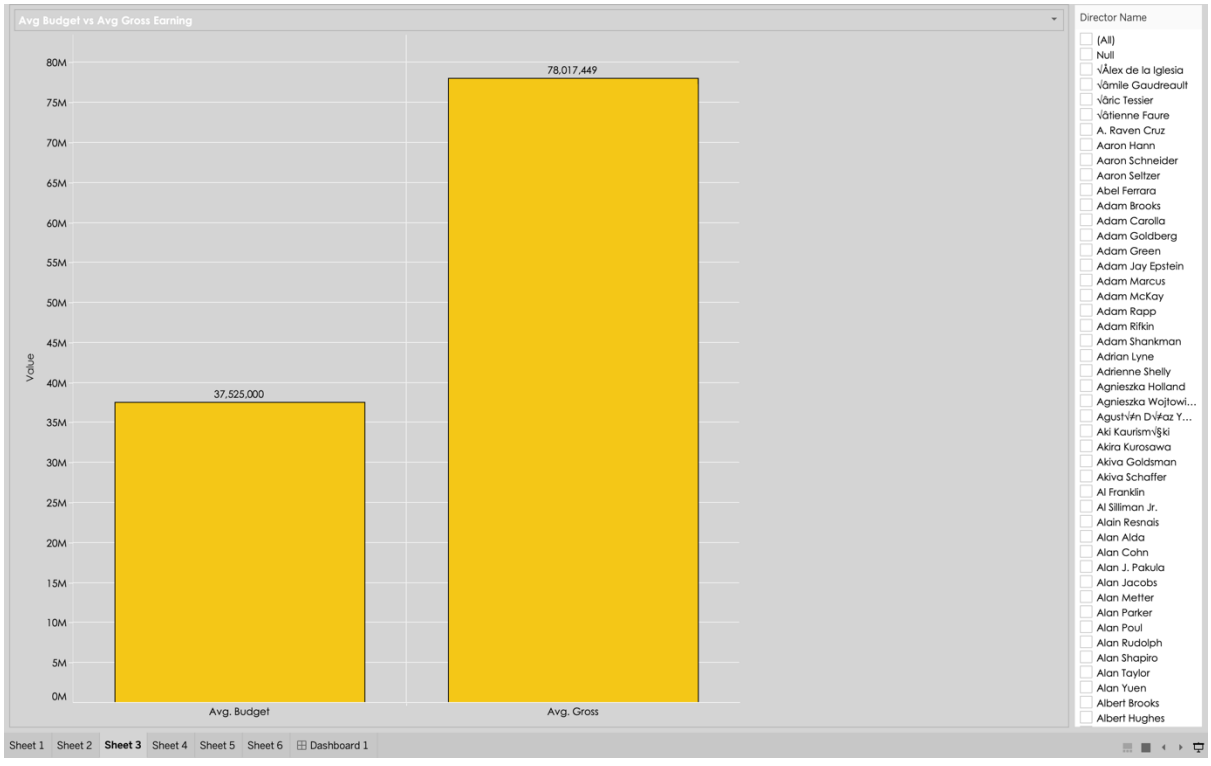


Sheet 1 (Filtered Query)

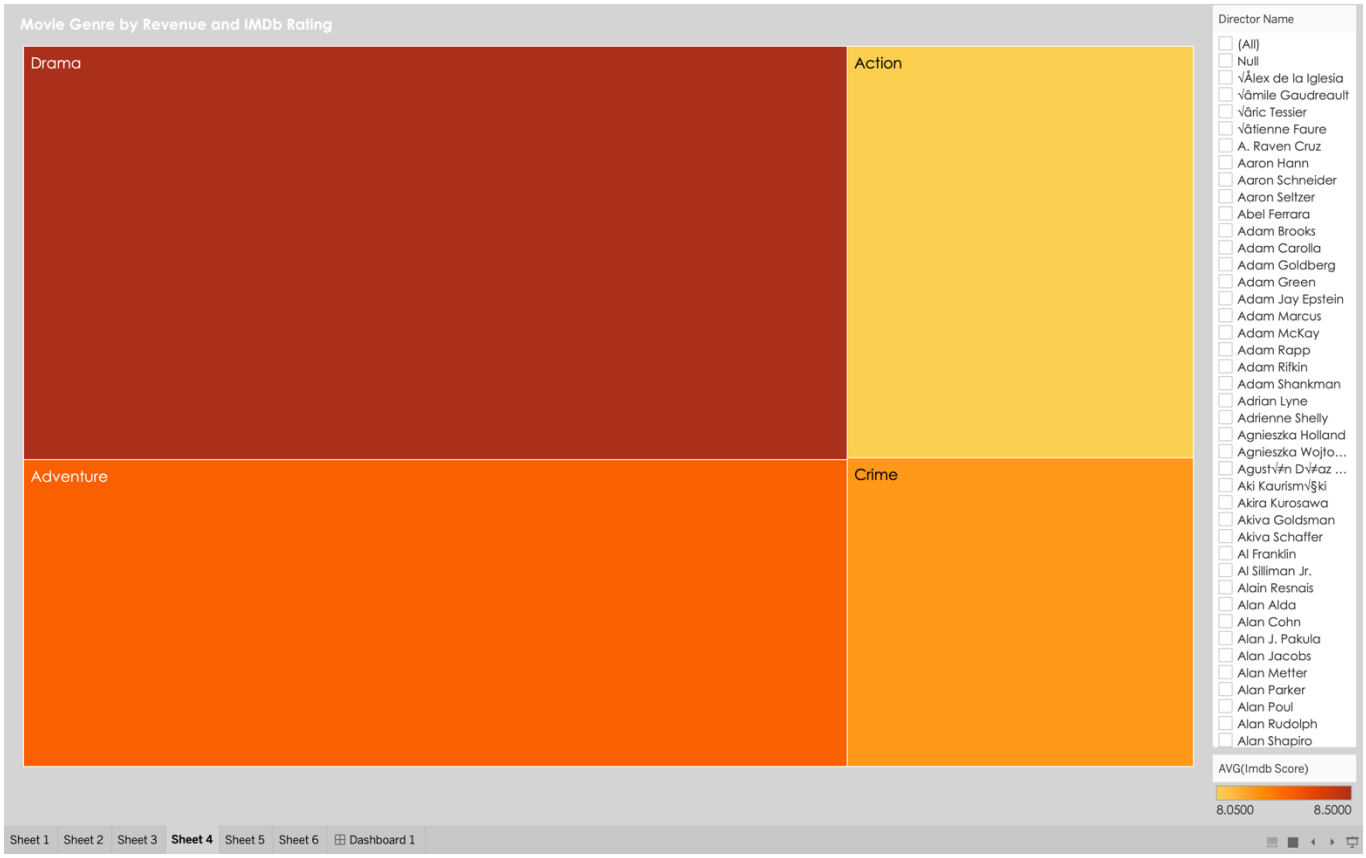
Sheet 2 (Filtered Query)



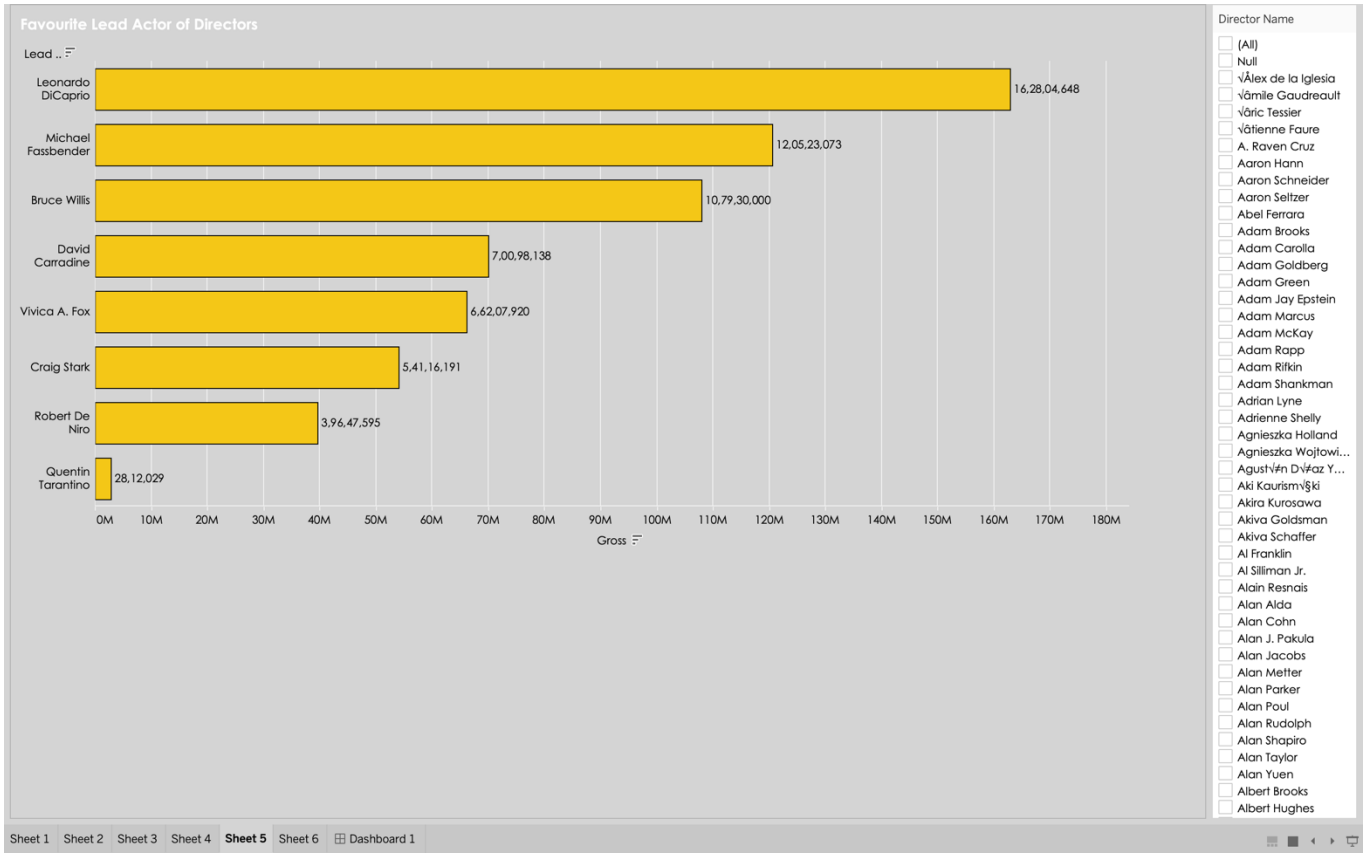
Sheet 3 (Filtered Query)



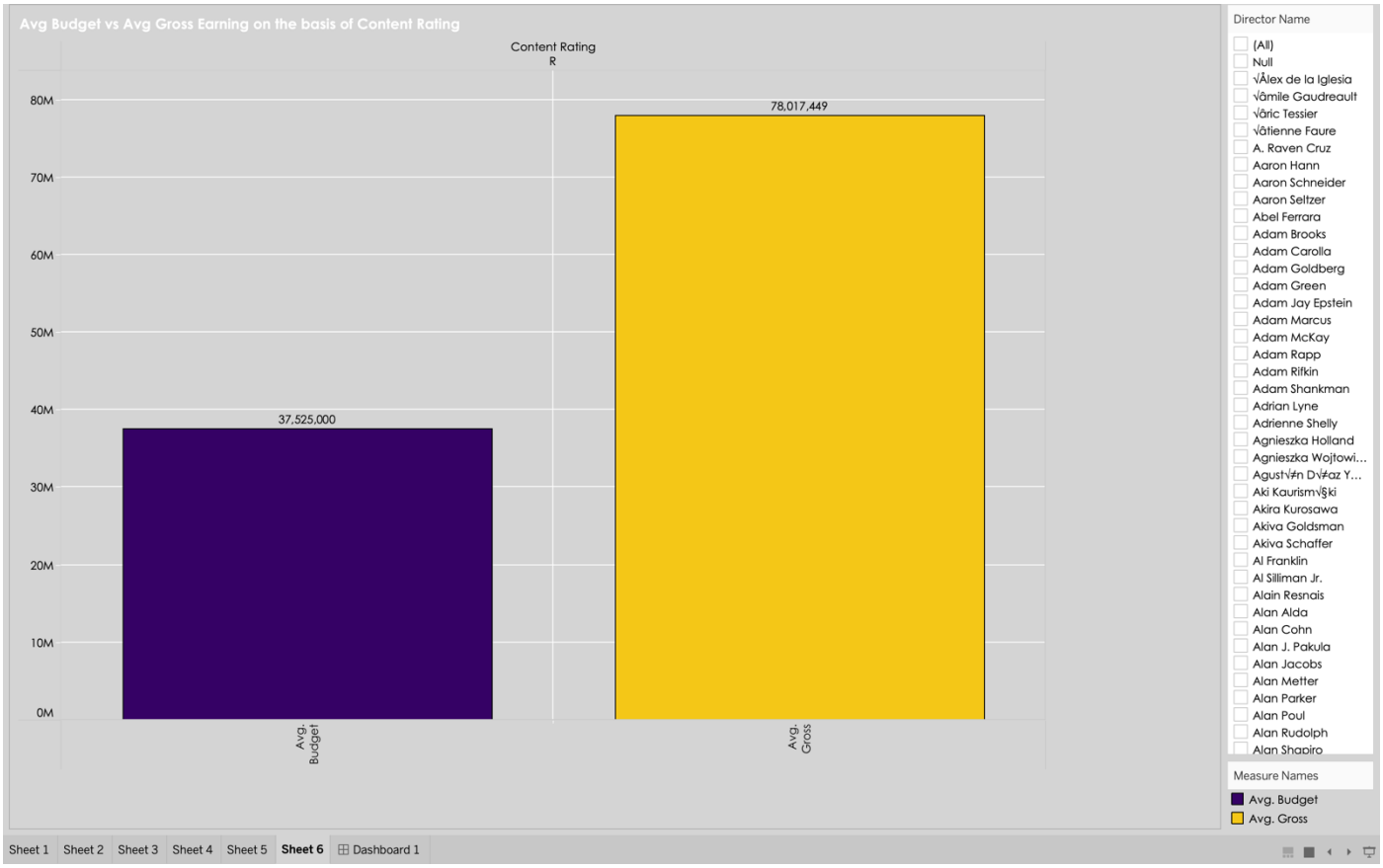
Sheet 4 (Filtered Query)



Sheet 5 (Filtered Query)



Sheet 6 (Filtered Query)



-Arjun Khanchandani
CS1-Roll No.102017005