DAS Group Project 2

Group 7

1 Introduction

Introduction paragraph

2 Exploratory Data Analysis

```
film_id    year  length  budget   votes   genre  rating
      0       0      92       0       0       0       0
```

```
Rows: 2,387
Columns: 8
$ film_id <int> 39891, 33810, 20282, 33131, 50633, 37020, 55337, 28037, 13291,~
$ year    <int> 2003, 2004, 1941, 1959, 1917, 1934, 2003, 1988, 1981, 1935, 19~
$ length  <int> 75, 120, 78, 106, 70, 64, 91, 101, 78, 7, 21, 90, 99, 101, 110~
$ budget  <dbl> 10.9, 19.6, 11.7, 12.0, 14.8, 11.6, 12.6, 10.1, 14.2, 6.6, 10.~
$ votes   <int> 17, 21, 14, 14, 9, 8, 182, 274, 61, 10, 5, 8, 349, 24, 20168, ~
$ genre   <chr> "Action", "Documentary", "Action", "Drama", "Drama", "Drama", ~
$ rating  <dbl> 4.4, 7.3, 2.7, 4.9, 5.6, 4.7, 4.4, 4.3, 4.3, 8.8, 7.3, 8.3, 7.~
$ above_7 <dbl> 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1,~
```

```
                          length      budget     votes
Proportion of Outliers 0.1805614 0.004608295 0.1625471
```

```
                       length_log      budget  votes_log
Proportion of Outliers  0.1818182 0.004608295 0.03812317
```

```
::: {.cell}
::: {.cell-output .cell-output-stdout}
```

$length [1] 7 21 2 153 4 4 30 7 5 10 22 6 7 14 7 15 8 9 [19] 25 18 8 21 147 17 19 141 170 10 4 146 23 10 12 154 141 6 [37] 11 8 170 399 8 3 24 156 18 12 7 15 30 7 8 17 7 15 [55] 7 7 11 17 7 185 7 17 6 18 1 7 15 19 28 1 8 218 [73] 20 30 17 10 21 3 15 187 8 235 17 6 9 30 17 15 9 10 [91] 10 13 10 7 311 4 24 8 8 34 20 12 7 5 7 14 7 7 [109] 7 20 27 7 1 9 7 7 3 242 31 30 13 10 8 7 9 17 [127] 6 6 16 5 9 7 13 3 5 13 13 6 19 16 20 7 7 8 [145] 1 8 7 6 19 192 269 6 8 18 19 7 6 29 7 15 32 25 [163] 11 14 7 7 17 6 7 18 155 7 6 10 141 2 160 8 6 33 [181] 30 30 10 7 1 3 7 7 17 145 7 6 7 5 10 7 27 7 [199] 13 8 9 6 8 29 11 10 10 155 3 13 13 142 15 6 15 8 [217] 168 6 3 17 7 10 7 6 20 6 13 210 28 141 140 9 27 15 [235] 6 22 149 181 5 24 7 3 7 152 14 7 7 28 11 179 12 7 [253] 3 30 7 10 7 17 7 5 7 16 30 170 13 3 200 29 10 12 [271] 20 6 20 15 20 160 9 4 175 12 27 4 7 8 20 185 29 7 [289] 13 7 6 178 4 16 8 25 145 15 6 5 8 18 5 185 12 20 [307] 21 10 18 34 188 9 6 7 12 26 152 155 13 156 6 8 240 19 [325] 6 9 8 14 33 7 26 6 27 14 7 6 9 180 2 161 7 9 [343] 9 140 169 142 4 20 14 15 7 7 16 9 25 17 173 140 1 19 [361] 15 29 147 5 5 5 8 12 9 10 10 11 5 170 180 3 17 14 [379] 16 12 30 10 20 10 14 7 10 11 9 2 231 13 155 10 8 2 [397] 9 320 6 20 5 7 11 14 5 3 11 6 28 27 8 11 11 180 [415] 165 26 14 5 9 27 1 157 1 7 145 175 17 15 30 8 25

$budget [1] 2.1 22.1 22.1 2.6 4.1 20.0 23.1 22.4 2.6 23.7 20.9

$votes [1] 349 20168 326 642 462 1113 507 2268 528 11127 [11] 22214 779 780 1098 1813 1521 575 815 470 987 [21] 862 381 5509 630 1258 808 723 649 2046 307 [31] 18169 1234 7184 1475 312 5008 5605 1737 1771 585 [41] 4389 20966 640 305 1596 3468 5128 986 849 2904 [51] 1222 590 906 350 2541 5440 811 4050 563 323 [61] 1430 385 342 362 319 7728 648 344 304 2505 [71] 2542 307 969 1863 4235 3656 475 1218 5022 1274 [81] 810 6938 315 5846 407 529 1243 2572 310 4231 [91] 5152 2111 13765 431 405 10065 3651 8340 340 2928 [101] 779 3533 2102 1040 19100 1087 602 2697 1612 325 [111] 731 522 2023 3423 2283 1506 370 5652 4027 668 [121] 1084 462 399 1350 1489 287 458 5774 400 338 [131] 1062 300 662 446 369 579 2379 2216 15565 9196 [141] 2873 14575 1047 2004 455 908 332 298 8371 280 [151] 291 406 1037 2755 416 1601 890 330 5640 339 [161] 6892 6058 2495 51961 1966 282 329 1884 2068 33188 [171] 590 598 980 999 466 9592 1024 9324 8738 1120 [181] 2723 11887 593 459 1742 18277 3530 354 14997 592 [191] 2269 793 7916 1452 308 103854 642 444 577 21462 [201] 466 836 4954 1860 900 2065 998 597 8526 446 [211] 418 565 1364 459 5893 917 1585 322 69600 15539 [221] 12923 2633 959 449 2844 4231 285 1172 295 654 [231] 1010 751 333 377 2291 2492 478 343 2566 619 [241] 1728 411 2336 307 299 1355 1640 400 372 297 [251] 461 4339 1397 7593 743 544 690 5066 1407 755 [261] 3735 322 787 764 1803 1103 10797 360 548 322 [271] 7279 1106 1322 406 2942 4190 9155 1280 627 352 [281] 280 2458 513 2541 391 411 920 316 343 7933 [291] 2778 292 799 2221 283 401 5020 421 384 707 [301] 1373 499 497 288 951 352 2670 7123 325 319 [311] 1718 11483 281 817 4316 13989 3694 449 316 462 [321] 9038 100267 366 334 581 464 2662 5039 1165 1564 [331] 1065 3288 20690 368 407 2019 2250 1252 877 970 [341] 3563 496 5811 288 1567 1509 1211 4294 352 396 [351] 4590 496 7247 603 2612 757 17521 582 529 385 [361] 1757 616 497 448 84488 735 1323 1306 301 618 [371] 1195 1775 391 3794 7771 297 8830 1327 1257 35648 [381] 1112 733 5044 306 17166 773 1386 4646

:::
:::

```
::: {.cell}
::: {.cell-output-display}
![](Group_07_Analysis_files/figure-pdf/unnamed-chunk-10-1.pdf)
:::
:::

::: {.cell}
::: {.cell-output-display}
![](Group_07_Analysis_files/figure-pdf/unnamed-chunk-11-1.pdf)
:::
:::

::: {.cell}
::: {.cell-output-display}
![](Group_07_Analysis_files/figure-pdf/unnamed-chunk-12-1.pdf)
:::
:::

::: {.cell}
::: {.cell-output-display}
![](Group_07_Analysis_files/figure-pdf/unnamed-chunk-13-1.pdf)
:::
:::

::: {.cell}
::: {.cell-output-display}
![](Group_07_Analysis_files/figure-pdf/unnamed-chunk-14-1.pdf)
:::
:::

::: {.cell}
::: {.cell-output-display}
![](Group_07_Analysis_files/figure-pdf/unnamed-chunk-15-1.pdf)
:::
:::

::: {.cell}
::: {.cell-output-display}
![](Group_07_Analysis_files/figure-pdf/unnamed-chunk-16-1.pdf)
```

```
:::
:::



# Formal Data Analysis {#sec-FDA}


::: {.cell}

:::

::: {.cell}

:::

::: {.cell}

:::

::: {.cell}
::: {.cell-output .cell-output-stdout}

                         Accuracy Sensitivity Specificity    AUC      BIC
```

Full Model 0.8659 0.8624 0.8734 0.9350 992.8782 Full model with Log 0.8869 0.8912 0.8777 0.9451 956.8562 Model without Year 0.8855 0.8871 0.8821 0.9457 950.4586 Model without Year and Votes 0.8883 0.8871 0.8908 0.9450 950.8248 “‘ ::: :::

3 Conclusions

4 Reference