Deliverable 4

2025-09-18

Research Motivation

In evaluating audience reception of films, factors such as actor performance, genre, and budget have been extensively studied as significant predictors of individual ratings (Wallace et al., 1993). Runtime, frequently included as a control variable in prior research, also carries meaningful implications for audience perception (Ashari et al., 2022). A longer duration of the movie often reflects higher production value and suggests a narrative depth that justifies viewers' time investment. However, excessive runtime may adversely affect enjoyment due to decreased audience attention and potential fatigue.

Empirical evidence on the relationship between movie duration and audience ratings remains inconclusive. Some studies identify a positive association, whereas others report a non-linear or genre specific effect. To address this gap, the current study investigates the influence of movie duration on audience ratings.

Given that rating behaviour and audience preferences evolve over time (Amendola et al., 2015), we incorporate release year as a control variable to enhance the internal validity of our analysis and account for temporal dynamics in consumer behaviour.

Research Question

To what extent does movie duration influence audience ratings, controlling for release year?

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:

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

```
## Classes 'data.table' and 'data.frame':   1614145 obs. of  3 variables:
##  $ tconst       : chr  "tt0000001" "tt0000002" "tt0000003" "tt0000004" ...
##  $ averageRating: num  5.7 5.5 6.4 5.2 6.2 5 5.3 5.3 5.3 6.8 ...
##  $ numVotes     : int  2178 299 2244 193 2989 218 927 2335 228 8071 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

Table 1: Data snapshots

dataset	n_rows	n_cols
title.basics	11912581	9
title.ratings	1614145	3

Table 1. Variable Explanation

Variable	Type	Definition	Role
runtimeMinutes	integer	Duration of the movie in minutes	Independ
averageRating	double	Average IMDb user rating (0–10 scale, aggregated from user votes)	Depende
startYear	integer	Year the movie was released	Control v

```
## [1] 1614145       3
```

```
## [1] 1614145
```

```
## [1] 3
```

```
## [1] "tconst"        "averageRating" "numVotes"
```

```
## Classes 'data.table' and 'data.frame':   1614145 obs. of  3 variables:
##  $ tconst       : chr  "tt0000001" "tt0000002" "tt0000003" "tt0000004" ...
##  $ averageRating: num  5.7 5.5 6.4 5.2 6.2 5 5.3 5.3 5.3 6.8 ...
##  $ numVotes     : int  2178 299 2244 193 2989 218 927 2335 228 8071 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

```
##     tconst          averageRating       numVotes
##  Length:1614145     Min.   : 1.000   Min.   :      5
##  Class :character   1st Qu.: 6.200   1st Qu.:     12
##  Mode  :character   Median : 7.100   Median :     26
##                     Mean   : 6.942   Mean   :   1020
##                     3rd Qu.: 7.900   3rd Qu.:    100
##                     Max.   :10.000   Max.   :3097823
```

```
##        tconst averageRating numVotes
##        <char>         <num>    <int>
##  1: tt0000001           5.7     2178
##  2: tt0000002           5.5      299
##  3: tt0000003           6.4     2244
##  4: tt0000004           5.2      193
##  5: tt0000005           6.2     2989
##  6: tt0000006           5.0      218
##  7: tt0000007           5.3      927
##  8: tt0000008           5.3     2335
##  9: tt0000009           5.3      228
## 10: tt0000010           6.8     8071
```

```
## $n_rows
```

Table 2. Descriptive Statistics

Variable	N	Missing	Mean	SD	Min	Max
runtimeMinutes	1137850	476295	58.195385	3462.858836	0	3692080
averageRating	1614145	0	6.941712	1.392415	1	10
startYear	1613889	256	2004.383463	21.194358	1874	2026

Table 2: Quick data checks

metric	value
rows	301682
min_runtime	1
max_runtime	300
min_year	1894
max_year	2025
any_na	0

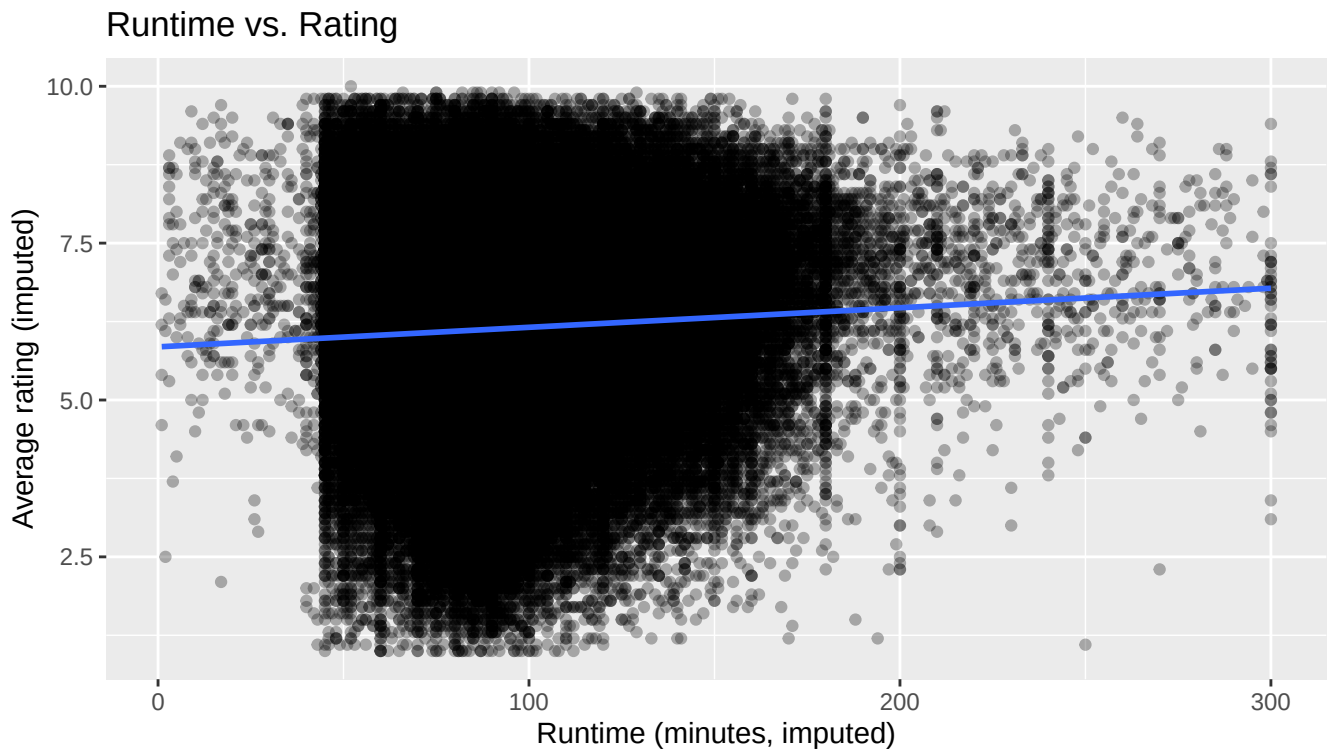
```
## [1] 301682
##
## $min_runtime
## [1] 1
##
## $max_runtime
## [1] 300
##
## $min_year
## [1] 1894
##
## $max_year
## [1] 2025
##
## $any_na
## [1] FALSE
```

Table 3: Preview: yearly_dt (first 10)

start_year	n_movies	avg_runtime	avg_rating	med_votes
1894	1	45.00000	5.30	228.0
1896	1	61.00000	3.20	25.0
1897	1	100.00000	5.20	568.0
1899	1	135.00000	3.90	86.0
1900	2	59.50000	5.00	35.5
1903	2	52.50000	5.20	407.5
1904	1	68.00000	5.30	378.0
1905	3	65.66667	5.20	37.0
1906	2	87.50000	4.45	544.0
1907	2	67.50000	6.00	71.5

Table 4: Preview: runtime_summary_dt (first 10)

runtime_bin	n	wmean_rating
[1,90]	149794	6.346994
(90,120]	120788	6.802389
(120,150]	23611	7.363439
(150,180]	6057	7.815277
(180,300]	1432	8.150471



#Add interpretation here

Table 5: Number of movies by release year

start_year	n_movies
1894	1
1896	1
1897	1
1899	1
1900	2
1903	2
1904	1
1905	3
1906	2
1907	2
1908	2
1909	4
1910	11
1911	18
1912	28
1913	75
1914	109
1915	155
1916	228
1917	259
1918	265
1919	292
1920	252
1921	252
1922	246
1923	276
1924	296
1925	370
1926	414
1927	426
1928	486
1929	519
1930	651
1931	748
1932	834
1933	787
1934	883
1935	999
1936	1041
1937	1074
1938	1052
1939	1012
1940	975
1941	951
1942	973
1943	936
1944	780
1945	694
1946	773
1947	822
1948	876
1949	992
1950	1058
1951	1097
1952	1088
1953	1175

Table 6: Movies per rating bucket by year (wide)

start_year	4–6	<4	6–7	>=8	7–8
1894	1	0	0	0	0
1896	0	1	0	0	0
1897	1	0	0	0	0
1899	0	1	0	0	0
1900	2	0	0	0	0
1903	0	1	1	0	0
1904	1	0	0	0	0
1905	1	1	1	0	0
1906	1	1	0	0	0
1907	1	0	1	0	0

Table 7: OLS: rating on runtime + year fixed effects

term	estimate	std.error	statistic	p.value
(Intercept)	5.1595696	1.3417508	3.8454009	0.0001204
runtime_imp	0.0031207	0.0001042	29.9526330	0.0000000
factor(start_year)1896	-2.1499308	1.8975113	-1.1330266	0.2572040
factor(start_year)1897	-0.2716372	1.8975192	-0.1431539	0.8861688
factor(start_year)1899	-1.6808609	1.8975337	-0.8858134	0.3757188
factor(start_year)1900	-0.3452498	1.6432930	-0.2100963	0.8335926
factor(start_year)1903	-0.1234051	1.6432925	-0.0750962	0.9401382
factor(start_year)1904	-0.0717756	1.8975121	-0.0378261	0.9698263
factor(start_year)1905	-0.1644940	1.5493124	-0.1061722	0.9154458
factor(start_year)1906	-0.9826287	1.6432983	-0.5979613	0.5498662
factor(start_year)1907	0.6297848	1.6432940	0.3832453	0.7015381
factor(start_year)1908	-0.3701460	1.6432972	-0.2252459	0.8217881
factor(start_year)1909	-0.6468101	1.5001146	-0.4311738	0.6663423
factor(start_year)1910	-0.7905248	1.4014065	-0.5640938	0.5726907
factor(start_year)1911	0.2476112	1.3785098	0.1796224	0.8574491
factor(start_year)1912	0.2449522	1.3654937	0.1793873	0.8576337
factor(start_year)1913	0.5107089	1.3506613	0.3781177	0.7053435
factor(start_year)1914	0.4110013	1.3478858	0.3049229	0.7604251
factor(start_year)1915	0.5479489	1.3460648	0.4070747	0.6839534
factor(start_year)1916	0.7210245	1.3446832	0.5362040	0.5918180
factor(start_year)1917	0.5833504	1.3443315	0.4339334	0.6643371
factor(start_year)1918	0.7305119	1.3442734	0.5434251	0.5868376
factor(start_year)1919	0.5925702	1.3440410	0.4408870	0.6592951
factor(start_year)1920	0.6249928	1.3444057	0.4648841	0.6420148
factor(start_year)1921	0.6907571	1.3444059	0.5138010	0.6073915
factor(start_year)1922	0.6234221	1.3444710	0.4636933	0.6428679
factor(start_year)1923	0.6603385	1.3441758	0.4912591	0.6232436
factor(start_year)1924	0.7916055	1.3440118	0.5889870	0.5558704
factor(start_year)1925	0.7579469	1.3435589	0.5641338	0.5726635
factor(start_year)1926	0.8363857	1.3433662	0.6226044	0.5335450
factor(start_year)1927	0.8165029	1.3433201	0.6078245	0.5433043
factor(start_year)1928	0.5172422	1.3431272	0.3851029	0.7001614
factor(start_year)1929	0.7793694	1.3430397	0.5803026	0.5617110
factor(start_year)1930	0.6309273	1.3427771	0.4698675	0.6384500
factor(start_year)1931	0.8067179	1.3426441	0.6008427	0.5479452
factor(start_year)1932	0.8471760	1.3425512	0.6310195	0.5280282
factor(start_year)1933	0.6492267	1.3425992	0.4835596	0.6286988
factor(start_year)1934	0.6017113	1.3425070	0.4481998	0.6540093
factor(start_year)1935	0.6255466	1.3424184	0.4659848	0.6412267
factor(start_year)1936	0.6325925	1.3423916	0.4712429	0.6374677
factor(start_year)1937	0.6799064	1.3423723	0.5064962	0.6125087
factor(start_year)1938	0.6682663	1.3423857	0.4978199	0.6186114
factor(start_year)1939	0.7113143	1.3424113	0.5298780	0.5961969
factor(start_year)1940	0.7241587	1.3424366	0.5394361	0.5895864
factor(start_year)1941	0.6749474	1.3424544	0.5027712	0.6151255
factor(start_year)1942	0.6500929	1.3424381	0.4842628	0.6281997
factor(start_year)1943	0.7190643	1.3424654	0.5356297	0.5922149
factor(start_year)1944	0.6970675	1.3426088	0.5191888	0.6036295
factor(start_year)1945	0.7579615	1.3427156	0.5644989	0.5724151
factor(start_year)1946	0.7771923	1.3426168	0.5788638	0.5626815
factor(start_year)1947	0.8167757	1.3425659	0.6083692	0.5429431
factor(start_year)1948	0.8206969	1.3425153	0.6113129	0.5409929
factor(start_year)1949	0.7896682	1.3424265	0.5882394	0.5563720
factor(start_year)1950	0.6379193	1.3423845	0.4752135	0.6346351
factor(start_year)1951	0.7382942	1.3423621	0.5499963	0.5823223
factor(start_year)1952	0.7240012	1.3423676	0.5393464	0.5896482