

Kuis IF-C

123200102_Muhammad_Dzaki_KUIS_IF-C

2022-10-05

##Intro

1. Boleh menggunakan PC lab / Laptop pribadi
2. Pengumpulan berupa hasil knit Rmd ke pdf dengan nama NIM_NAMA_KUIS_IF-C.pdf.
3. Durasi 2 Jam + 5 menit submit, > tidak bisa mengumpul.
4. Pengerjaan offline, pengumpulan di Spada (online).
5. Tidak boleh buka modul.
6. Tidak boleh membuka internet (googling, WhatsApp, ig, sosmed, dan media komunikasi lain).
7. Boleh bawa catatan 1 lembar A4.
8. Izin keluar maks. 1x dengan durasi 2 menit.
9. Tas dan HP diletakkan di depan.

Selamat mengerjakan, semoga sukses!

1. Load library apa saja yang kira-kira digunakan! Lalu gunakan data 'movielens'! **point 2**

```
library(dslabs)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr 0.3.4
## v tibble 3.1.8       v dplyr 1.0.10
## v tidyr 1.2.1        v stringr 1.4.1
## v readr 2.1.2        v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()      masks stats::lag()
```

```
data('movielens')
```

2. Tampilkan struktur dari data Movielens! **point 5**

```
str(movielens)
```

```
## 'data.frame': 100004 obs. of 7 variables:
## $ movieId : int 31 1029 1061 1129 1172 1263 1287 1293 1339 1343 ...
## $ title : chr "Dangerous Minds" "Dumbo" "Sleepers" "Escape from New York" ...
## $ year : int 1995 1941 1996 1981 1989 1978 1959 1982 1992 1991 ...
## $ genres : Factor w/ 901 levels "(no genres listed)",...: 762 510 899 120 762 836 81 762 844 899 .
## $ userId : int 1 1 1 1 1 1 1 1 1 1 ...
## $ rating : num 2.5 3 3 2 4 2 2 2 3.5 2 ...
## $ timestamp: int 1260759144 1260759179 1260759182 1260759185 1260759205 1260759151 1260759187 1260759187 1260759187 1260759187 ...
```

Jawablah pertanyaan berikut a. Class dari object movielens: data.frame b. Jumlah baris: 100004 c. Jumlah kolom: 7

3. Implementasikan fungsi yang dapat digunakan jika hanya ingin menampilkan nama-nama variabel dari object movielens! **point 5**

```
names(movielens)
```

```
## [1] "movieId" "title" "year" "genres" "userId" "rating"
## [7] "timestamp"
```

4. Tuliskan fungsi untuk mengetahui tipe data dari variabel genres, dan jelaskan pada bawah chunk kenapa variabel genres menggunakan tipe data tersebut! **point 5**

```
class(movielens$genres)
```

```
## [1] "factor"
```

menggunakan factor dikarenakan tiap karakter angka merepresentasikan value genre yang berbeda, semisal karakter angka 1 bisa merepresentasikan sebagai genre drama

5. Tampilkan 5 observasi terakhir dari data Movielens! **point 5**

```
tail(movielens, n = 5)
```

```
##      movieId      title year      genres
## 100000    6268 Raising Victor Vargas 2002 Comedy|Drama|Romance
## 100001    6269      Stevie 2002      Documentary
## 100002    6365 Matrix Reloaded, The 2003 Action|Adventure|Sci-Fi|Thriller|IMAX
## 100003    6385      Whale Rider 2002      Drama
## 100004    6565      Seabiscuit 2003      Drama
##      userId rating  timestamp
## 100000    671    2.5 1065579370
## 100001    671    4.0 1065149201
## 100002    671    4.0 1070940363
## 100003    671    2.5 1070979663
## 100004    671    3.5 1074784724
```

6. Tampilkan 10 observasi teratas movielens tanpa variabel movieId & userId, simpan ke dalam object bernama "tiny_df"! **point 15**

```
tiny_df = tail(movielens %>% select(title,year,genres,rating,timestamp), n=10)
tiny_df
```

```
##      title year
## 99995 Lord of the Rings: The Two Towers, The 2002
## 99996      Catch Me If You Can 2002
## 99997      Chicago 2002
## 99998      Pianist, The 2002
## 99999      Bringing Down the House 2003
```

```
## 100000          Raising Victor Vargas 2002
## 100001          Stevie 2002
## 100002          Matrix Reloaded, The 2003
## 100003          Whale Rider 2002
## 100004          Seabiscuit 2003
##
##          genres rating  timestamp
## 99995          Adventure|Fantasy      5.0 1063502716
## 99996          Crime|Drama           4.0 1064890625
## 99997          Comedy|Crime|Drama|Musical 4.5 1064245387
## 99998          Drama|War             4.0 1066793014
## 99999          Comedy                2.5 1065149436
## 100000          Comedy|Drama|Romance    2.5 1065579370
## 100001          Documentary           4.0 1065149201
## 100002 Action|Adventure|Sci-Fi|Thriller|IMAX 4.0 1070940363
## 100003          Drama                2.5 1070979663
## 100004          Drama                3.5 1074784724
```

7. Tampilkan 20 data teratas dengan genre hanya romance dengan diurutkan rating dari yang tertinggi. Tampilkan data dengan variabel title, year, genres, dan rating saja! **point 15** Hilangkan data yang dengan judul berulang(Poin plus) **point 5**

```
movielens %>% select(title,year,genres,rating) %>% filter(genres == "Romance",!duplicated(title)) %>%
```

```
##
##          title year  genres rating
## 1          Against All Odds 1984 Romance    4.0
## 2          Everybody's All-American 1988 Romance    4.0
## 3          Dirty Dancing: Havana Nights 2004 Romance    4.0
## 4          Hustler White 1996 Romance    4.0
## 5          Paris, I Love You (Paris, je t'aime) 2006 Romance    4.0
## 6          Tango Lesson, The 1997 Romance    4.0
## 7          Love and Other Catastrophes 1996 Romance    4.0
## 8          Autumn Tale, An (Conte d'automne) 1998 Romance    4.0
## 9          Jack and Sarah 1995 Romance    4.0
## 10         Kama Sutra: A Tale of Love 1996 Romance    3.5
## 11         Message in a Bottle 1999 Romance    3.0
## 12         Raise Your Voice 2004 Romance    2.5
## 13         House of Mirth, The 2000 Romance    2.5
## 14         Meet Joe Black 1998 Romance    2.0
```

8. Buatlah histogram dengan sumbu x tahun pembuatan film dan sumbu y frekuensi banyaknya film yang ada di tahun tersebut. Berikan warna kuning dan garis batas biru pada histogram dan beri judul yang sesuai. Di bawah chunk tuliskan kesimpulan yang dapat diperoleh dari histogram yang anda buat! **point 13**

```
"ndak paham ingfo caranya mas"
```

```
## [1] "ndak paham ingfo caranya mas"
```

Kesimpulan: Tahun dengan jumlah film terbanyak dalam data movielens adalah sekitar tahun...

9. Klasifikasikan data dari object "tiny_df" dari soal no 4 berdasarkan rating dengan kondisi : -Rating kurang dari 2 dikategorikan sebagai "Negative" -Rating lebih dari 3 dikategorikan sebagai "Positive" -Rating antara kedua kondisi diatas dikategorikan sebagai "Mixed" **point 15**

```
dummy1 = ifelse(tiny_df$rating<2,"Negative",ifelse(tiny_df$rating>3,"Positive","Mixed"))
dummy1
```

```
## [1] "Positive" "Positive" "Positive" "Positive" "Mixed"      "Mixed"
## [7] "Positive" "Positive" "Mixed"      "Positive"
```

10. Tambahkan variabel baru dalam data “tiny_df” dengan tambahan kolom baru dengan nama “sentiment” yang isinya merupakan implementasi nomor 9! **point 20**

```
tiny_df = mutate(tiny_df, sentiment = dummy1)
tiny_df
```

```
##                                     title year
## 99995 Lord of the Rings: The Two Towers, The 2002
## 99996          Catch Me If You Can 2002
## 99997                  Chicago 2002
## 99998          Pianist, The 2002
## 99999      Bringing Down the House 2003
## 100000      Raising Victor Vargas 2002
## 100001          Stevie 2002
## 100002      Matrix Reloaded, The 2003
## 100003          Whale Rider 2002
## 100004          Seabiscuit 2003
##                                     genres rating timestamp sentiment
## 99995          Adventure|Fantasy      5.0 1063502716 Positive
## 99996          Crime|Drama      4.0 1064890625 Positive
## 99997      Comedy|Crime|Drama|Musical      4.5 1064245387 Positive
## 99998          Drama|War      4.0 1066793014 Positive
## 99999          Comedy      2.5 1065149436 Mixed
## 100000      Comedy|Drama|Romance      2.5 1065579370 Mixed
## 100001          Documentary      4.0 1065149201 Positive
## 100002      Action|Adventure|Sci-Fi|Thriller|IMAX      4.0 1070940363 Positive
## 100003          Drama      2.5 1070979663 Mixed
## 100004          Drama      3.5 1074784724 Positive
```