

# Operazioni su dataframe

## Pulizia dataframe

```
dati <- read.csv("C:/Users/anton/Downloads/movies.csv", sep = ",")
dati <- na.omit(dati)
colonne <- colnames(dati)
dati <- subset.data.frame(dati, mpa <= "PG", select = colonne )
```

## Qualche operazione base

### Funzioni apply

```
colonne_numeriche <- function(df) {
  if(!is.data.frame(df)) {stop("L'argomento non è un dataframe")}
  new_df <- data.frame(id = 1:nrow(df))
  i <- 1
  while(i < ncol(df)) {
    if(is.numeric(df[,i])) {
      new_df <- data.frame(new_df, df[,i])
    }
    i <- i + 1
  }
  new_df <- new_df[,-1]
  return(new_df)
}

nome_colonne_numeriche <- function(df) {
  if(!is.data.frame(df)) {stop("L'argomento non è un dataframe")}
  nomi <- c()
  i <- 1
  while(i < ncol(df)) {
    if(is.numeric(df[,i])) {
      nomi <- c(nomi, names(df) [i])
    }
    i <- i + 1
  }
  return(nomi)
}

dati_numeric <- colonne_numeriche(dati)
nome_colonne_numeriche_dati <- nome_colonne_numeriche(dati)
names(dati_numeric)[1:length(nome_colonne_numeriche_dati)] <- nome_colonne_numeriche_dati [1: length(nome_colonne_numeriche_dati)]
round(apply(dati_numeric, 2, mean), 1)
```

##	rownames	year	length	budget	rating	votes
##	28559.7	1999.1	106.9	28026051.6	5.9	9313.7
##	r1	r2	r3	r4	r5	r6
##	7.9	5.1	5.5	6.6	9.5	13.7
##	r7	r8	r9	r10	Action	Animation
##	16.9	14.3	8.5	12.1	0.2	0.0
##	Comedy	Drama	Documentary	Romance		
##	0.4	0.5	0.0	0.2		

## dplyr

```
library(dplyr)
```

```
## Warning: il pacchetto 'dplyr' è stato creato con R versione 4.3.3
```

```
##
```

```
## Caricamento pacchetto: 'dplyr'
```

```
## I seguenti oggetti sono mascherati da 'package:stats':
```

```
##
```

```
##     filter, lag
```

```
## I seguenti oggetti sono mascherati da 'package:base':
```

```
##
```

```
##     intersect, setdiff, setequal, union
```

```
medie_per_anno <- dati %>%  
  group_by(year) %>%  
  summarise(budget = mean(budget), rating = mean(rating))
```

```
fasce_di_rating <- dati %>%  
  group_by(cut(rating, breaks = 5)) %>%  
  summarise(budget = mean(budget))
```

```
film_999 <- dati %>%  
  filter(year == 1999) %>%  
  summarise(film = title, rating = rating)
```

```
## Warning: Returning more (or less) than 1 row per 'summarise()' group was deprecated in
```

```
## dplyr 1.1.0.
```

```
## i Please use 'reframe()' instead.
```

```
## i When switching from 'summarise()' to 'reframe()', remember that 'reframe()'
```

```
## always returns an ungrouped data frame and adjust accordingly.
```

```
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
```

```
## generated.
```

```
Gli_anni_migliori <- dati %>%  
  group_by(year) %>%  
  summarise(rating = mean(rating)) %>%  
  arrange(desc(rating))
```