

Лабораторна робота №1

Виконали: Кузьменко Юрій, Болотов Єгор

Завдання 1: Знайомство з R Studio

(A) Створити проєкт (New Project) під назвою “Topic 1”;

 Topic 1.Rproj

218 B

Feb 14, 2023, 10:31 PM

(B) Створити R Script під назвою “lab 1”;


 lab_1.R

0 B

Feb 14, 2023, 10:33 PM

Завдання 2: Знайомство з базою даних

(A) Приєднати dataset до проєкту.

Data	
 Spotify_final_dataset	11084 obs. of 9 variables

(B) Вивантажити dataset з сайту та описати його:

a. Описати кількість змінних (факторів) та їх довжину;

```
> df <- read.csv('Spotify_final_dataset.csv')
> dim(df)
[1] 11084      9
```

Кількість записів – 11084, кількість змінних - 9

b. Описати що означає кожна змінна;

- **Position** - Spotify Ranking
- **Artist Name** - Artist Name
- **Song Name** - Song Name
- **Days** - No of days since the release of the song (Updated)
- **Top 10 (xTimes)** - No of times inside top 10
- **Peak Position** - Peak position attained
- **Peak Position (xTimes)** - No of times Peak position attained
- **Peak Streams** - Total no of streams during Peak position
- **Total Streams** - Total song streams

с. Описати який тип у кожної змінної;

```
> sapply(df, class)
      Position      Artist.Name      Song.Name      Days      Top.10..xTimes.
      "integer"      "character"      "character"      "integer"      "numeric"
Peak.Position Peak.Position..xTimes. Peak.Streams      Total.Streams
      "integer"      "character"      "integer"      "integer"
```

Завдання 3: Побудова різноманітних моделей

(A) Визначити вхідні фактори для моделей;

```
library(readr)
dt_frame <- read_csv("../Spotify_final_dataset.csv")

## Rows: 11084 Columns: 9
## — Column specification
## Delimiter: ","
## chr (3): Artist Name, Song Name, Peak Position (xTimes)
## dbl (6): Position, Days, Top 10 (xTimes), Peak Position, Peak Streams,
Total...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this
message.

colnames(dt_frame)[2]='Artist_name'
colnames(dt_frame)[3]='Song_name'
colnames(dt_frame)[5]='Top_ten_times'
colnames(dt_frame)[6]='Peak_position'
colnames(dt_frame)[7]='Peak_position_times'
colnames(dt_frame)[8]='Peak_streams'
colnames(dt_frame)[9]='Total_streams'
```

(B) Побудувати моделі;

- $\text{lm}(y \sim x)$
- $\text{lm}(y \sim x + I(x^2))$
- $\text{lm}(y \sim x + I(x^2) + I(x^3))$

(C) Накласти на графік розсіювання криві кожної моделі, використовуючи різний колір.

У ролі залежної змінної будемо використовувати Position, саме за допомогою неї зможемо дізнатись вплив змінних на рейтинг пісні

First

x - Position (Позиція у рейтингу) y - Days (К-сть днів після виходу)

```
x <- dt_frame$Days
y <- dt_frame$Position

lm(y ~ x )

##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)          x
##    6276.13       -13.75

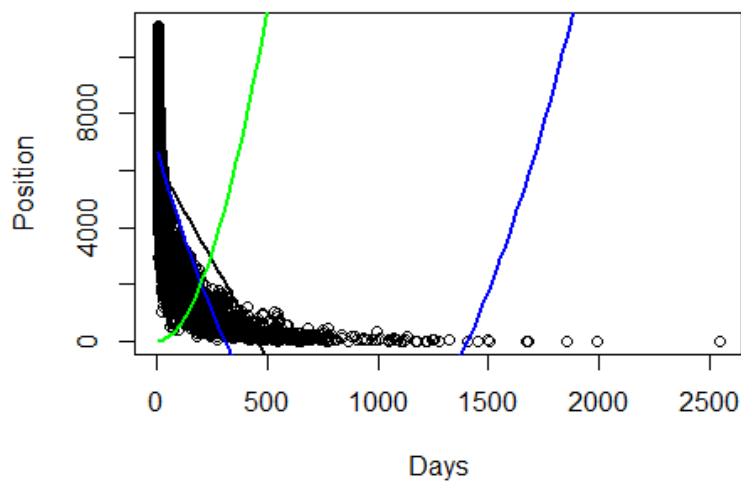
lm(y ~ x + I(x^2))

##
## Call:
## lm(formula = y ~ x + I(x^2))
##
## Coefficients:
## (Intercept)          x      I(x^2)
##    6646.9371    -26.3736     0.0154

lm(y ~ x + I(x^2) + I(x^3))

##
## Call:
## lm(formula = y ~ x + I(x^2) + I(x^3))
##
## Coefficients:
## (Intercept)          x      I(x^2)      I(x^3)
##    6.952e+03   -4.200e+01    5.541e-02   -1.738e-05

plot(x, y, xlab = "Days", ylab = "Position")
curve( (6276.13 - 13.75 * x), add=TRUE, col="black", lwd = 2)
curve( (6646.9371 - 26.3736 * x + 0.0154 * x^2), add=TRUE, col="blue", lwd = 2)
curve( (0.0154 + 0.0154 * x + 5.541e-02 * x^2 - 1.738e-05 * x^3),
      add=TRUE, col="green", lwd = 2)
```



Second

x - Position (Позиція у рейтингу) y - Top 10 xTimes (К-сть перебування у топ-10)

```
x <- dt_frame$Top_ten_times
y <- dt_frame$Position

lm(y ~ x )

##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)          x
##    5698.70       -57.56

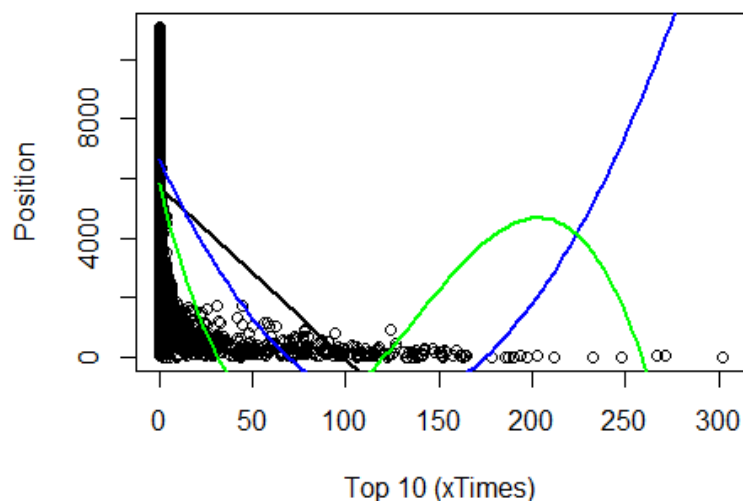
lm(y ~ x + I(x^2))

##
## Call:
## lm(formula = y ~ x + I(x^2))
##
## Coefficients:
## (Intercept)          x      I(x^2)
##    5768.5988   -134.0563     0.5489

lm(y ~ x + I(x^2) + I(x^3))

##
## Call:
## lm(formula = y ~ x + I(x^2) + I(x^3))
##
## Coefficients:
## (Intercept)          x      I(x^2)      I(x^3)
##    5.830e+03   -2.552e+02    2.429e+00   -5.911e-03

plot(x, y, xlab = "Top 10 (xTimes)", ylab = "Position")
curve( (5698.70 -57.56 * x), add=TRUE, col="black", lwd = 2)
curve( (6646.9371 -134.0563 * x + 0.5489 * x^2), add=TRUE, col="blue", lwd = 2)
curve( (5.830e+03 + -2.552e+02 * x + 2.429e+00 * x^2 -5.911e-03 * x^3),
      add=TRUE, col="green", lwd = 2)
```



Third

x - Position (Позиція у рейтингу) y - Peak Position (Пікова позиція)

```
x <- dt_frame$Peak_position
y <- dt_frame$Position

lm(y ~ x )

##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)          x
##    1906.00         43.78

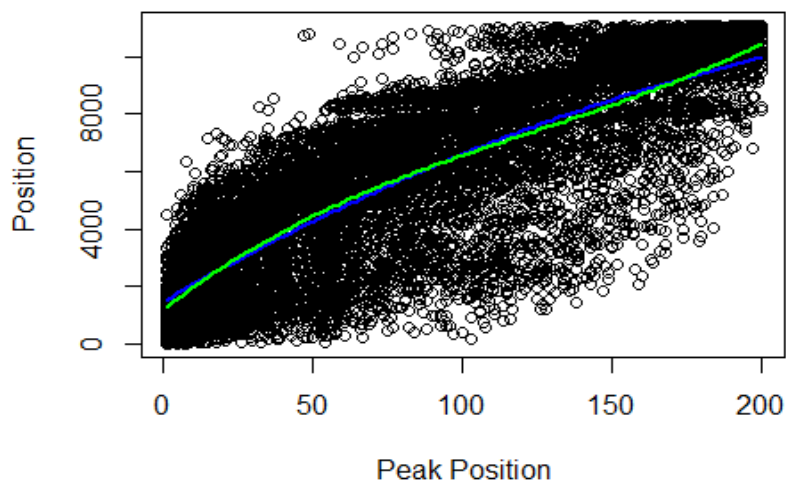
lm(y ~ x + I(x^2))

##
## Call:
## lm(formula = y ~ x + I(x^2))
##
## Coefficients:
## (Intercept)          x      I(x^2)
##  1458.34956    60.32582   -0.08938

lm(y ~ x + I(x^2) + I(x^3))

##
## Call:
## lm(formula = y ~ x + I(x^2) + I(x^3))
##
## Coefficients:
## (Intercept)          x      I(x^2)      I(x^3)
##  1.171e+03    8.171e+01   -3.806e-01    1.019e-03

plot(x, y, xlab = "Peak Position", ylab = "Position")
curve( (1906.00 + 42.78 * x), add=TRUE, col="black", lwd = 2)
curve( (1458.34956 + 60.32582 * x - 0.08938 * x^2), add=TRUE, col="blue", lwd = 2)
curve( (1.171e+03 + 8.171e+01 * x - 3.806e-01 * x^2 + 1.019e-03 * x^3),
      add=TRUE, col="green", lwd = 2)
```



Fourth

x - Position (Позиція у рейтингу) y - Peak Position (xTimes) (К-сть разів у піковій позиції)

```
dt_frame$Peak_position_times <- gsub("[ (x)]", "", dt_frame$Peak_position_times, ignore.case = TRUE)
dt_frame$Peak_position_times <- as.numeric(dt_frame$Peak_position_times)
x <- dt_frame$Peak_position_times
y <- dt_frame$Position

lm(y ~ x )

##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)          x
##    5602.6        -155.4

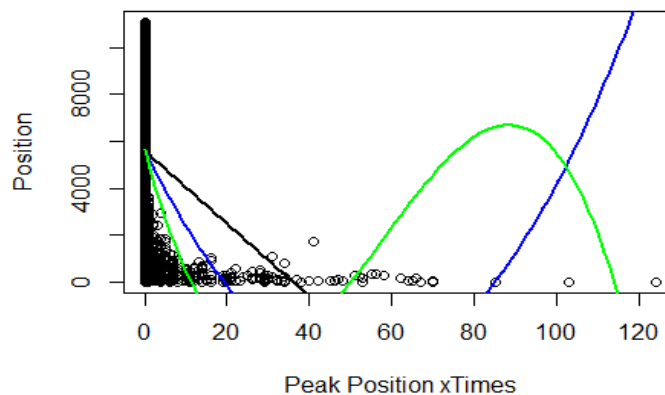
lm(y ~ x + I(x^2))

##
## Call:
## lm(formula = y ~ x + I(x^2))
##
## Coefficients:
## (Intercept)          x      I(x^2)
##    5635.851    -357.808      3.433

lm(y ~ x + I(x^2) + I(x^3))

##
## Call:
## lm(formula = y ~ x + I(x^2) + I(x^3))
##
## Coefficients:
## (Intercept)          x      I(x^2)      I(x^3)
##    5664.89543    -665.55237     15.46570     -0.08829

plot(x, y, xlab = "Peak Position xTimes", ylab = "Position")
curve( (5602.6 -155.4 * x), add=TRUE, col="black", lwd = 2)
curve( (5635.851 -357.808 * x +3.433 * x^2), add=TRUE, col="blue", lwd = 2)
curve( (5664.89543 -665.55237 * x +15.46570 * x^2 -0.08829 * x^3),
      add=TRUE, col="green", lwd = 2)
```



Fifth

x - Position (Позиція у рейтингу) y - Peak Steams (К-сть переслухувань під час піку)

```
x <- dt_frame$Peak_streams
y <- dt_frame$Position

lm(y ~ x )

##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)          x
##  7.027e+03   -2.685e-03

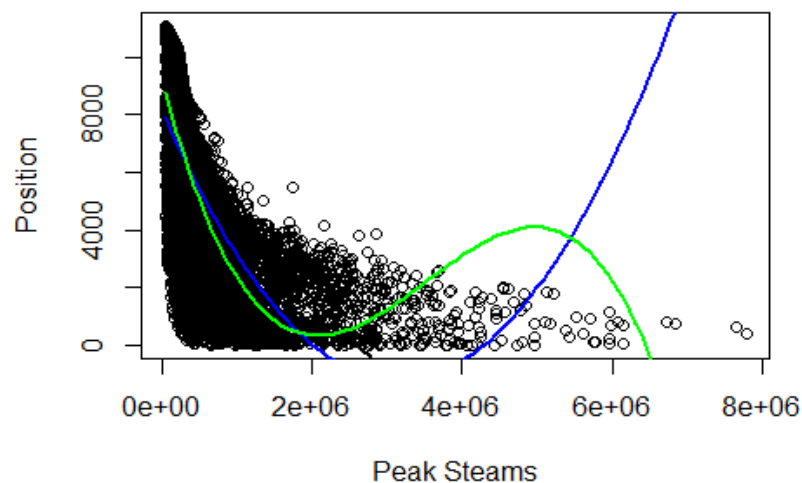
lm(y ~ x + I(x^2))

##
## Call:
## lm(formula = y ~ x + I(x^2))
##
## Coefficients:
## (Intercept)          x      I(x^2)
##  8.176e+03   -5.960e-03   9.448e-10

lm(y ~ x + I(x^2) + I(x^3))

##
## Call:
## lm(formula = y ~ x + I(x^2) + I(x^3))
##
## Coefficients:
## (Intercept)          x      I(x^2)      I(x^3)
##  9.213e+03   -9.890e-03   3.366e-09   -3.186e-16

plot(x, y, xlab = "Peak Steams", ylab = "Position")
curve( (7.027e+03 -2.685e-03 * x), add=TRUE, col="black", lwd = 2)
curve( (8.176e+03 -5.960e-03 * x +9.448e-10 * x^2), add=TRUE, col="blue", lwd = 2)
curve( (9.213e+03 -9.890e-03 * x +3.366e-09 * x^2 -3.186e-16 * x^3),
      add=TRUE, col="green", lwd = 2)
```



Sixth

x - Position (Позиція у рейтингу) y - Total Steams (Загальна к-сть переслухувань)

```
x <- dt_frame$Total_streams
y <- dt_frame$Position

lm(y ~ x )

##
## Call:
## lm(formula = y ~ x)
##
## Coefficients:
## (Intercept)          x
##  6.109e+03   -3.096e-05

lm(y ~ x + I(x^2))

##
## Call:
## lm(formula = y ~ x + I(x^2))
##
## Coefficients:
## (Intercept)          x      I(x^2)
##  6.474e+03   -6.786e-05   1.017e-13

lm(y ~ x + I(x^2) + I(x^3))

##
## Call:
## lm(formula = y ~ x + I(x^2) + I(x^3))
##
## Coefficients:
## (Intercept)          x      I(x^2)      I(x^3)
##  6.806e+03   -1.196e-04   4.317e-13   -3.783e-22

plot(x, y, xlab = "Total Steams", ylab = "Position")
curve( (6.109e+03 -3.096e-05 * x), add=TRUE, col="black", lwd = 2)
curve( (6.474e+03 -6.786e-05 * x +1.017e-13 * x^2), add=TRUE, col="blue", lwd = 2)
curve( (6.806e+03 -1.196e-04 * x +4.317e-13 * x^2 -3.783e-22 * x^3),
      add=TRUE, col="green", lwd = 2)
```

