# RESPONSI_DS-(C)

Yohanes Ageng H.Z.P._123190075

## Intro

1. Kerjakan soal-soal yang ada! Jangan lupa AUTHOR diberi NAMA dan NIM (pada bagian atas soal ini)
2. Perhatikan instruksi soal dan jawab dengan mengisi chunk dibawah soal!
3. Waktu pengerjaan mulai pukul 10.30 hingga 12.30 dan batas submit di Spada sampai 12.35 (2 Jam + 5 menit)
4. Tidak ada toleransi kerja sama saat responsi.
5. Telat pengumpulan tiap 3 menit akan ada pengurangan nilai 10 point dengan maksimal pengurangan 50 point. Telat lebih dari 15 menit atau melebihi pukul 12.50 dianggap **GUGUR**.
6. Soal yang rancu bisa ditanyakan ke asisten.
7. Pengumpulan hanya dalam bentuk **WORD Document atau PDF**. Jika pengumpulan dalam bentuk **Rmd** akan dianggap tidak mengumpulkan jawaban.
8. Pastikan jawaban dapat dijalankan dengan baik sebelum submit. Jika masih ada error, cukup dicomment codenya tanpa perlu dijalankan/RUN/keluar outputnya.
9. Jawaban dikumpulkan dengan format file NIM_Nama_Responsi dalam bentuk PDF atau WORD Document.

##Soal

1. Load library apa saja yang kira-kira akan digunakan! Jika terjadi error pastikan library terinstall. Lalu load dataset 'googleplay.csv' dan 'googleplay_user_review.csv'! Lalu tampilkan TOP 6 data yang dimiliki 'googleplay.csv'

```
library(tidyverse)

## -- Attaching packages ------------------------------------- tidyverse 1.
3.1 --

## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.6      v dplyr   1.0.7
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.0.2      v forcats 0.5.1

## Warning: package 'tibble' was built under R version 4.1.2

## -- Conflicts --------------------------------------------- tidyverse_conflict
s() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(dslabs)
library(dslabs)
library(tm)
```

```
## Warning: package 'tm' was built under R version 4.1.2
```

```
## Loading required package: NLP
```

```
##
## Attaching package: 'NLP'
```

```
## The following object is masked from 'package:ggplot2':
##
##     annotate
```

```
library(vroom)
library(here)
```

```
## Warning: package 'here' was built under R version 4.1.2
```

```
## here() starts at D:/Ageng/Kuliah/Semester 7/Prak DS/Responsi
```

```
library(RColorBrewer)
```

```
data_google_play_user <- vroom(here('./googleplaystore_user_reviews.csv'))
```

```
## Rows: 64295 Columns: 5
```

```
## -- Column specification -------------------------------------------------
------
## Delimiter: ","
## chr (3): App, Translated_Review, Sentiment
## dbl (2): Sentiment_Polarity, Sentiment_Subjectivity
```

```
##
## 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 m
essage.
```

```
data_google_play <- vroom(here('./googleplaystore.csv'))
```

```
## Rows: 8196 Columns: 13
```

```
## -- Column specification -------------------------------------------------
------
## Delimiter: ","
## chr (11): App, Category, Size, Installs, Type, Price, Content Rating, Genr
es...
## dbl  (2): Rating, Reviews
```

```
##
## 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 m
essage.

head(data_google_play, 6)

## # A tibble: 6 x 13
##    App         Category Rating Reviews Size  Installs Type  Price `Content R
ating`
##    <chr>       <chr>     <dbl>   <dbl> <chr> <chr>    <chr> <chr> <chr>
## 1 Photo Edi~ ART_AND~    4.1     159 19M   10,000+  Free  0     Everyone
## 2 Coloring ~ ART_AND~    3.9     967 14M   500,000+ Free  0     Everyone
## 3 U Launche~ ART_AND~    4.7   87510 8.7M  5,000,0~ Free  0     Everyone
## 4 Sketch - ~ ART_AND~    4.5  215644 25M   50,000,~ Free  0     Teen
## 5 Pixel Dra~ ART_AND~    4.3     967 2.8M  100,000+ Free  0     Everyone
## 6 Paper flo~ ART_AND~    4.4     167 5.6M  50,000+  Free  0     Everyone
## # ... with 4 more variables: Genres <chr>, Last Updated <chr>,
## #   Current Ver <chr>, Android Ver <chr>

head(data_google_play_user, 6)

## # A tibble: 6 x 5
##    App                  Translated_Revi~ Sentiment Sentiment_Polar~ Sentim
ent_Subje~
##    <chr>                <chr>            <chr>                <dbl>
<dbl>
## 1 10 Best Foods for You "I like eat del~ Positive                 1
0.533
## 2 10 Best Foods for You "This help eati~ Positive                 0.25
0.288
## 3 10 Best Foods for You "nan"            nan                     NaN
NaN
## 4 10 Best Foods for You "Works great es~ Positive                 0.4
0.875
## 5 10 Best Foods for You "Best idea us"   Positive                 1
0.3
## 6 10 Best Foods for You "Best way"       Positive                 1
0.3
```

**Pada soal nomor 2 s/d 7 hanya menggunakan dataset 'googleplay.csv'**

2. Tampilkan TOP 10 Aplikasi berdasarkan kategori 'GAME' dan banyaknya REVIEWS dari yang terbesar!

```
data_game <- data_google_play %>%
  filter(Category == "GAME") %>%
  arrange(desc(Reviews))
head(data_game, 10)

## # A tibble: 10 x 13
##    App         Category Rating  Reviews Size  Installs Type  Price `Content R
ating`
##    <chr>       <chr>     <dbl>    <dbl> <chr> <chr>    <chr> <chr> <chr>
```

```
##  1 Clash o~ GAME          4.6 44891723 98M    100,000~ Free  0      Everyone 1
0+
##  2 Subway ~ GAME          4.5 27722264 76M    1,000,0~ Free  0      Everyone 1
0+
##  3 Clash R~ GAME          4.6 23133508 97M    100,000~ Free  0      Everyone 1
0+
##  4 Candy C~ GAME          4.4 22426677 74M    500,000~ Free  0      Everyone
##  5 My Talk~ GAME          4.5 14891223 Vari~ 500,000~ Free  0      Everyone
##  6 8 Ball ~ GAME          4.5 14198297 52M    100,000~ Free  0      Everyone
##  7 Shadow ~ GAME          4.6 10979062 88M    100,000~ Free  0      Everyone 1
0+
##  8 Pou      GAME          4.3 10485308 24M    500,000~ Free  0      Everyone
##  9 Pok<e9>~ GAME          4.1 10424925 85M    100,000~ Free  0      Everyone
## 10 Minion ~ GAME          4.5 10216538 Vari~ 100,000~ Free  0      Everyone 1
0+
## # ... with 4 more variables: Genres <chr>, Last Updated <chr>,
## #   Current Ver <chr>, Android Ver <chr>
```

3. Tampilkan TOP 10 Aplikasi berdasarkan banyaknya unduhan, dan tampilkan secara urut berdasarkan rating! Perhatikan clue, karena berkaitan dengan data preprocessing. Clue : Kolom (Variabel) banyaknya unduhan memiliki tipe data yang tidak sesuai

```
data_google_play$Installs =
as.numeric(gsub(',|\\+','',data_google_play$Installs))
data_google_play %>% top_n(n = 10, Installs) %>% arrange(desc(Rating)) %>%
 head(10)
```

```
## # A tibble: 10 x 13
##    App      Category  Rating Reviews Size   Installs Type  Price `Content R
ating`
##    <chr>    <chr>      <dbl>   <dbl> <chr>     <dbl> <chr> <chr> <chr>
##  1 Subway~ GAME        4.5  2.77e7 76M        1e9 Free  0      Everyone 1
0+
##  2 Instag~ SOCIAL      4.5  6.66e7 Varie~     1e9 Free  0      Teen
##  3 Google~ PHOTOGRA~   4.5  1.09e7 Varie~     1e9 Free  0      Everyone
##  4 WhatsA~ COMMUNIC~   4.4  6.91e7 Varie~     1e9 Free  0      Everyone
##  5 Google  TOOLS       4.4  8.03e6 Varie~     1e9 Free  0      Everyone
##  6 Google~ PRODUCTI~   4.4  2.73e6 Varie~     1e9 Free  0      Everyone
##  7 Google~ COMMUNIC~   4.3  9.64e6 Varie~     1e9 Free  0      Everyone
##  8 Gmail   COMMUNIC~   4.3  4.60e6 Varie~     1e9 Free  0      Everyone
##  9 Google~ ENTERTAI~   4.3  7.17e6 Varie~     1e9 Free  0      Teen
## 10 Maps -~ TRAVEL_A~   4.3  9.24e6 Varie~     1e9 Free  0      Everyone
## # ... with 4 more variables: Genres <chr>, Last Updated <chr>,
## #   Current Ver <chr>, Android Ver <chr>
```

4. Tampilkan rata-rata RATING untuk setiap kategori aplikasi!

```
mean.rating = data_google_play %>% group_by(Category) %>% summarize(Mean =
mean(Rating, na.rm=TRUE)) %>% head(10)

mean.rating
```
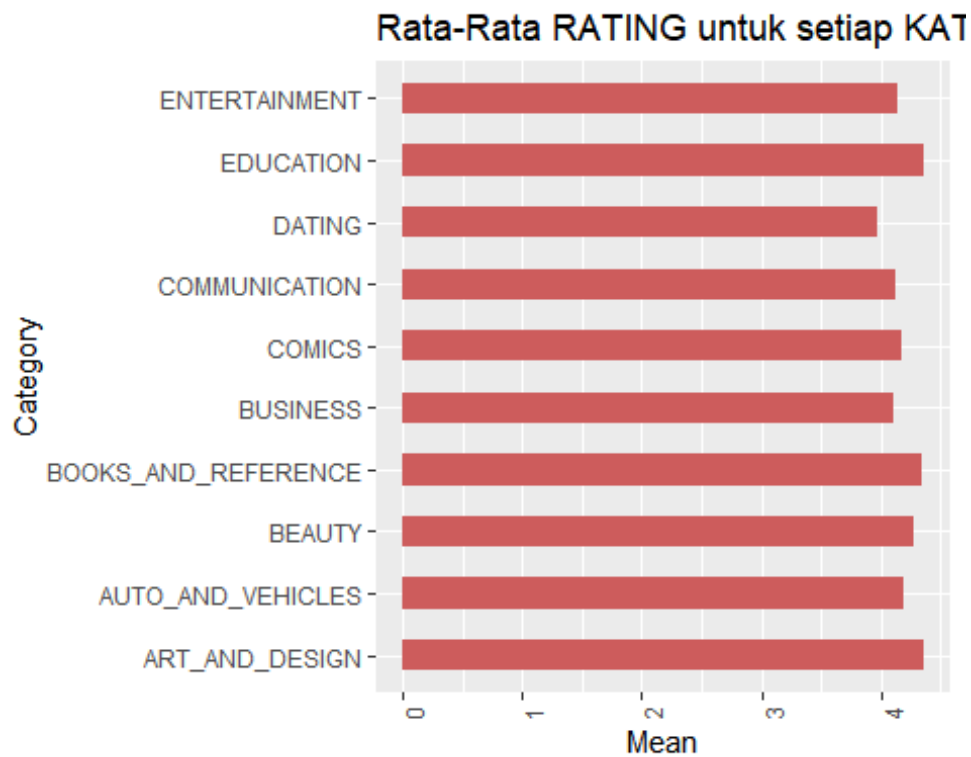
```
## # A tibble: 10 x 2
##    Category            Mean
##    <chr>              <dbl>
##  1 ART_AND_DESIGN      4.36
##  2 AUTO_AND_VEHICLES   4.19
##  3 BEAUTY              4.28
##  4 BOOKS_AND_REFERENCE 4.34
##  5 BUSINESS            4.10
##  6 COMICS              4.18
##  7 COMMUNICATION       4.12
##  8 DATING              3.97
##  9 EDUCATION           4.36
## 10 ENTERTAINMENT       4.14
```

5. Berdasarkan soal nomor 4, buat plot untuk memvisualisasikan hasilnya! (Bentuk plot bebas)

```r
mean.plot <- mean.rating

mean.plot <- data.frame(Category= mean.plot$Category, Mean = mean.plot$Mean)

ggplot(mean.plot, aes(x = Category, y = Mean)) +
  geom_bar(stat = "identity", width = 0.5, fill = "indianred") +
  coord_flip() +
  labs(title = "Rata-Rata RATING untuk setiap KATEGORI") +
  theme(axis.text.x = element_text(angle = 90))
```

6. Tampilkan rata-rata Intalls untuk setiap kategori aplikasi!

```
rata_install = data_google_play %>% group_by(Category) %>% summarize(Mean =
mean(Installs, na.rm=TRUE)) %>% head(10)

rata_install

## # A tibble: 10 x 2
##    Category                 Mean
##    <chr>                   <dbl>
##  1 ART_AND_DESIGN       1872674.
##  2 AUTO_AND_VEHICLES     727805.
##  3 BEAUTY                640862.
##  4 BOOKS_AND_REFERENCE  9856755.
##  5 BUSINESS             2650259.
##  6 COMICS                832057.
##  7 COMMUNICATION       43118131.
##  8 DATING               1051585.
##  9 EDUCATION            2990271.
## 10 ENTERTAINMENT       20722157.
```

7. Berdasarkan soal nomor 6, buat plot untuk memvisualisasikan hasilnya! (Bentuk plot bebas)

```
install_plot <- rata_install

install_plot <- data.frame(Category= install_plot$Category, Average = install
_plot$Mean)

ggplot(install_plot, aes(x = Category, y = Average)) +
  geom_bar(stat = "identity", width = 0.5, fill = "indianred") +
  coord_flip() +
  labs(title = "Total App Reviews for Each Category") +
  theme(axis.text.x = element_text(angle = 90))
```

## Total App Reviews for Each Categor
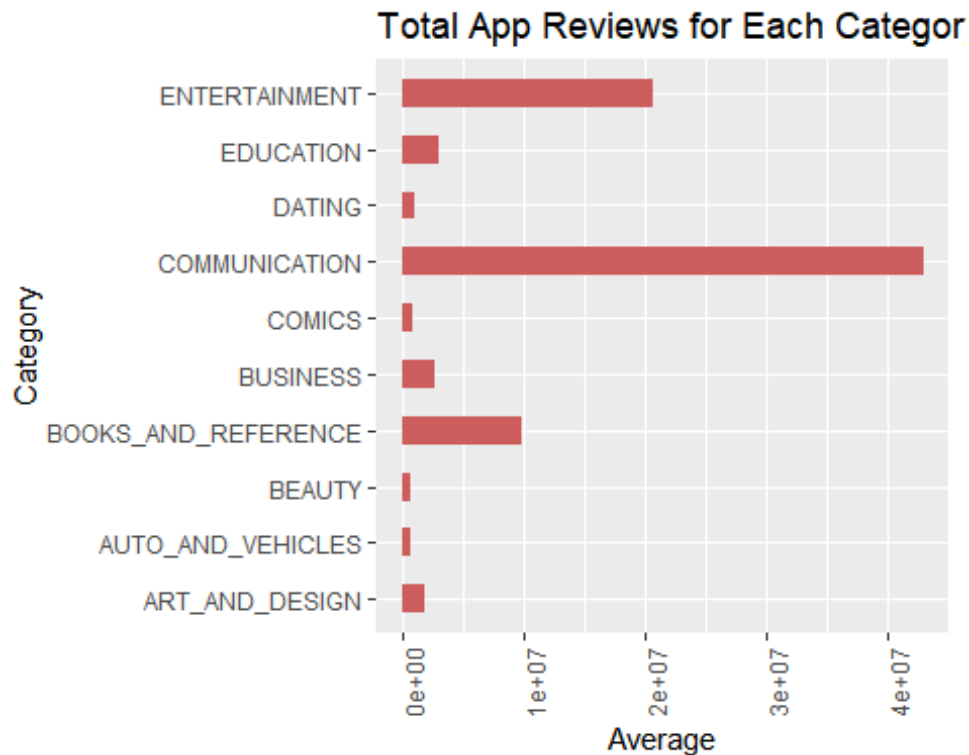


Import library 'textdata' dan 'tidytext' untuk nomor 8 s/d 10. Jika eror pastikan kedua library tersebut sudah terinstall.

```
library(tidytext)

## Warning: package 'tidytext' was built under R version 4.1.2

library(textdata)

## Warning: package 'textdata' was built under R version 4.1.2
```

8. Gabungkan data googleplaystore.csv dan googleplaystore_user_reviews.csv menggunakan inner join. Tampilkan TOP 6 datanya!

```
gabungan_data <- inner_join(data_google_play, data_google_play_user)

## Joining, by = "App"

head(gabungan_data, 6)

## # A tibble: 6 x 17
##    App                 Category Rating Reviews Size  Installs Type  Price `
Content Rating`
##    <chr>               <chr>     <dbl>   <dbl> <chr>    <dbl> <chr> <chr> <
chr>
## 1 Coloring book moana ART_AND~    3.9     967 14M     500000 Free  0     E
veryone
## 2 Coloring book moana ART_AND~    3.9     967 14M     500000 Free  0     E
veryone
```

```
## 3 Coloring book moana ART_AND~     3.9     967 14M      500000 Free  0     E
veryone
## 4 Coloring book moana ART_AND~     3.9     967 14M      500000 Free  0     E
veryone
## 5 Coloring book moana ART_AND~     3.9     967 14M      500000 Free  0     E
veryone
## 6 Coloring book moana ART_AND~     3.9     967 14M      500000 Free  0     E
veryone
## # ... with 8 more variables: Genres <chr>, Last Updated <chr>,
## #   Current Ver <chr>, Android Ver <chr>, Translated_Review <chr>,
## #   Sentiment <chr>, Sentiment_Polarity <dbl>, Sentiment_Subjectivity <dbl
>
```

9. Dari data pada nomor 8, tampilkan nama aplikasi, translated_review, sentiment, dan reviews berdasarkan sentiment negatif dan reviews terbanyak!

```
tampilan_gabungan <- gabungan_data %>%
  filter(Sentiment == "Negative") %>%
  arrange(desc(Reviews)) %>%
  select(App, Translated_Review, Sentiment, Reviews)

tail(tampilan_gabungan, 10)
```

```
## # A tibble: 10 x 4
##    App                                    Translated_Review Sentiment R
eviews
##    <chr>                                  <chr>             <chr>
<dbl>
##  1 Azpen eReader                          My attempt unins~ Negative
156
##  2 Airway Ex - Intubate. Anesthetize. Train. So I thought gam~ Negative
123
##  3 Airway Ex - Intubate. Anesthetize. Train. In tutorial won'~ Negative
123
##  4 Airway Ex - Intubate. Anesthetize. Train. THEY HAVE TO PAY~ Negative
123
##  5 Airway Ex - Intubate. Anesthetize. Train. Fricken SCREW DO~ Negative
123
##  6 Airway Ex - Intubate. Anesthetize. Train. Controls f******~ Negative
123
##  7 Airway Ex - Intubate. Anesthetize. Train. Great game! Can'~ Negative
123
##  8 Airway Ex - Intubate. Anesthetize. Train. Bad controls.The~ Negative
123
##  9 Airway Ex - Intubate. Anesthetize. Train. I'M A DOCTOR NOW~ Negative
123
## 10 DIY On A Budget                        Useless app, alo~ Negative
114
```

10. Dalam tahap pre-processing data, ketika ingin melakukan analisis sentimen beberapa hal harus dilakukan sebelum data dapat digunakan. Bersihkan dan

rapikan data dengan membuang data yang "nan" di bagian Translated_Review. Setelah itu, data juga harus dibersihkan dari kata-kata yang mengandung stop_words (seperti: a, a's, after, dll). Tampilkan 6 data teratas dari data yang memiliki sentiment 'Positive' dengan rating lebih dari sama dengan 4.5

```
tidy_user_reviews <- gabungan_data %>%
  filter(Translated_Review != "nan") %>%
  unnest_tokens(word, Translated_Review) %>%
  anti_join(stop_words)

## Joining, by = "word"

tidy_user_reviews

## # A tibble: 362,048 x 17
##    App                  Category Rating Reviews Size  Installs Type  Price
`Content Rating`
##    <chr>                <chr>     <dbl>   <dbl> <chr>    <dbl> <chr> <chr>
<chr>
##  1 Coloring book moana ART_AND~    3.9     967 14M     500000 Free  0
Everyone
##  2 Coloring book moana ART_AND~    3.9     967 14M     500000 Free  0
Everyone
##  3 Coloring book moana ART_AND~    3.9     967 14M     500000 Free  0
Everyone
##  4 Coloring book moana ART_AND~    3.9     967 14M     500000 Free  0
Everyone
##  5 Coloring book moana ART_AND~    3.9     967 14M     500000 Free  0
Everyone
##  6 Coloring book moana ART_AND~    3.9     967 14M     500000 Free  0
Everyone
##  7 Coloring book moana ART_AND~    3.9     967 14M     500000 Free  0
Everyone
##  8 Coloring book moana ART_AND~    3.9     967 14M     500000 Free  0
Everyone
##  9 Coloring book moana ART_AND~    3.9     967 14M     500000 Free  0
Everyone
## 10 Coloring book moana ART_AND~    3.9     967 14M     500000 Free  0
Everyone
## # ... with 362,038 more rows, and 8 more variables: Genres <chr>,
## #   Last Updated <chr>, Current Ver <chr>, Android Ver <chr>, Sentiment <c
hr>,
## #   Sentiment_Polarity <dbl>, Sentiment_Subjectivity <dbl>, word <chr>
```

*=== Selamat Mengerjakan ===*