

Analisis: Spotify Weekly Chart

Data yang digunakan untuk analisis dari kaagle

Data ini terdiri dari beberapa kolom, berikut ini penjelasan dari masing-masing kolom: 1. Pos :position of the song in the chart

2. P. /IncreasedPos :number of position increased since last week

3. Artist :name of the artist

4. Song :title of the song

5. Wks :number of weeks spent on the chart

6. Pk :peak position of the song

7. x. /TimesPeak :number of times the peak position has been attained

8. Streams :number of streams of the song for one we

9. Streams. /IncreasedStreams :number of streams increased or decreased since last week

10.Total :total number of streams

Pada kolom P. /IncreasedPos ini bersisi angka, NEW, dan RE. Jika berisi angka maka yang menunjukkan kenaikan dari minggu lalu.

Jika berisi RE atau RE-ENTRY menunjukkan lagu lama yang masuk kembali ke chart.

Jika berisi NEW menunjukkan lagu baru.

```
data <- read.csv("spotify_weekly_chart.csv")
head(data)
```

##	Pos	P.	Artist	Title	Wks	Pk	x	Streams	Streams.
## 1	1	4	SZA	Kill Bill	4	1	(x1)	41474366	1578141
## 2	2	9	Sam Smith	Unholy	15	1	(x7)	33132175	2872164
## 3	3	12	Harry Styles	As It Was	40	1	(x11)	30998084	3982809
## 4	4	16	David Guetta	I'm Good (Blue)	19	2	(x4)	30704256	7144547
## 5	5	8	Metro Boomin	Creepin'	5	3	(x1)	28092232	381484
## 6	6	10	Manuel Turizo	La Bachata	27	4		27357095	727736
##			Total						
## 1	160755788								
## 2	642800599								
## 3	1719435686								
## 4	541637842								
## 5	153469486								
## 6	667941457								

```
tail(data,22)
```

##	Pos	P.	Artist	Title	Wks	Pk	x
## 179	179	RE	Mora	LA INOCENTE	24	74	
## 180	180	NEW	RAYE	Escapism.	1	180	
## 181	181	RE	50 Cent	In Da Club	13	77	
## 182	182	RE	Taylor Swift	You're On Your Own, Kid	10	6	
## 183	183	RE	Kenshi Yonezu	KICK BACK	9	66	
## 184	184	RE	Eminem	Lose Yourself	207	44	
## 185	185	RE	Billie Eilish	TV	11	27	

##	186	186	RE	Camila Cabello	Bam Bam	41	5
##	187	187	RE	Keane	Somewhere Only We Know	13	120
##	188	188	RE	KAROL G	GATÚBELA	16	33
##	189	189	RE	Pritam	Kesariya	10	71
##	190	190	RE	The Police	Every Breath You Take	74	109
##	191	191	RE	Seafret	Atlantis	21	63
##	192	192	RE	Guilherme & Benuto	Haja Colírio	2	187
##	193	193	RE	Future	WAIT FOR U	34	10
##	194	194	RE	Tears For Fears	Everybody Wants To Rule The World	23	105
##	195	195	RE	Post Malone	Sunflower	215	1 (x2)
##	196	196	RE	Bad Bunny	Party	33	6
##	197	197	RE	Freddie Dredd	Limbo	7	158
##	198	198	RE	Bruno Mars	Locked Out Of Heaven	82	45
##	199	199	RE	MEDUZA	Bad Memories	2	199
##	200	200	RE	TWISTED	WORTH NOTHING	5	181
##				Streams	Streams.	Total	
##	179	7077846		NA		188339481	
##	180	7048604		NA		7048604	
##	181	6975892		NA		80329014	
##	182	6955426		NA		174292913	
##	183	6949309		NA		76744573	
##	184	6932729		NA		703684632	
##	185	6928883		NA		82967316	
##	186	6888725		NA		571221486	
##	187	6876234		NA		90755157	
##	188	6842540		NA		162007366	
##	189	6840658		NA		77686725	
##	190	6820435		NA		424204320	
##	191	6815573		NA		175770219	
##	192	6807866		NA		13580265	
##	193	6805425		NA		349455284	
##	194	6803992		NA		159047407	
##	195	6801735		NA		2425355542	
##	196	6782777		NA		479414025	
##	197	6775767		NA		51408912	
##	198	6758678		NA		335046675	
##	199	6748442		NA		13244341	
##	200	6735908		NA		34008257	

Ask

1. Lagu apa yang paling populer menempati posisi paling tinggi di chart?
2. Lagu apa yang paling populer menempati posisi paling terendah di chart?
3. Lagu apa dengan kategori NEW yang paling sering diputar pendengar?
4. Lagu apa dengan kategori RE (Re-Entry) yang paling sering diputar pendengar?
5. Lagu apa yang paling lama menduduki chart?
6. Lagu apa yang paling sedikit menduduki chart?

Prepare

Kita akan mengubah nama kolom: P+ menjadi IncreasedPos (x?) menjadi TimesPeak Streams+ menjadi IncreasedStreams

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.5    v purrr 0.3.4
## v tibble 3.1.1     v dplyr 1.0.6
## v tidyr 1.1.3      v stringr 1.4.0
## v readr 1.4.0      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

names(data)[names(data)=='P.'] <- 'IncreasedPos'
names(data)[names(data)=='x'] <- 'TimesPeak'
names(data)[names(data)=='Streams.'] <- 'IncreasedStreams'
```

Kita akan mengubah data kolom IncreasedPos dari NEW ke bentuk 0 (NEW=0) dari RE ke bentuk 1 (RE=1)
Ini bertujuan untuk menyamakan tipe data

```
data['IncreasedPos'][data['IncreasedPos']=='NEW'] <- 0
data['IncreasedPos'][data['IncreasedPos']=='RE'] <- 1
tail(data,22)
```

##	Pos	IncreasedPos	Artist	Title	Wks
## 179	179	1	Mora	LA INOCENTE	24
## 180	180	0	RAYE	Escapism.	1
## 181	181	1	50 Cent	In Da Club	13
## 182	182	1	Taylor Swift	You're On Your Own, Kid	10
## 183	183	1	Kenshi Yonezu	KICK BACK	9
## 184	184	1	Eminem	Lose Yourself	207
## 185	185	1	Billie Eilish	TV	11
## 186	186	1	Camila Cabello	Bam Bam	41
## 187	187	1	Keane	Somewhere Only We Know	13
## 188	188	1	KAROL G	GATÚBELA	16
## 189	189	1	Pritam	Kesariya	10
## 190	190	1	The Police	Every Breath You Take	74
## 191	191	1	Seafret	Atlantis	21
## 192	192	1	Guilherme & Benuto	Haja Colirio	2
## 193	193	1	Future	WAIT FOR U	34
## 194	194	1	Tears For Fears	Everybody Wants To Rule The World	23
## 195	195	1	Post Malone	Sunflower	215
## 196	196	1	Bad Bunny	Party	33
## 197	197	1	Freddie Dredd	Limbo	7
## 198	198	1	Bruno Mars	Locked Out Of Heaven	82
## 199	199	1	MEDUZA	Bad Memories	2
## 200	200	1	TWISTED	WORTH NOTHING	5
##	Pk	TimesPeak	Streams	IncreasedStreams	Total
## 179	74		7077846	NA	188339481
## 180	180		7048604	NA	7048604
## 181	77		6975892	NA	80329014
## 182	6		6955426	NA	174292913
## 183	66		6949309	NA	76744573
## 184	44		6932729	NA	703684632
## 185	27		6928883	NA	82967316
## 186	5		6888725	NA	571221486
## 187	120		6876234	NA	90755157
## 188	33		6842540	NA	162007366
## 189	71		6840658	NA	77686725

```
## 190 109          6820435          NA 424204320
## 191 63           6815573          NA 175770219
## 192 187          6807866          NA 13580265
## 193 10           6805425          NA 349455284
## 194 105          6803992          NA 159047407
## 195 1      (x2) 6801735          NA 2425355542
## 196 6           6782777          NA 479414025
## 197 158          6775767          NA 51408912
## 198 45           6758678          NA 335046675
## 199 199          6748442          NA 13244341
## 200 181          6735908          NA 34008257
```

Process

Mengganti format penulisan pada kolom TimesPeak

```
pattern = '\\(x'
pattern1 <- ')'
data$TimesPeak <-gsub(pattern,"",data$TimesPeak)
data$TimesPeak <-gsub(pattern1,'',data$TimesPeak)
head(data)
```

##	Pos	IncreasedPos	Artist	Title	Wks	Pk	TimesPeak	Streams
## 1	1	4	SZA	Kill Bill	4	1	1	41474366
## 2	2	9	Sam Smith	Unholy	15	1	7	33132175
## 3	3	12	Harry Styles	As It Was	40	1	11	30998084
## 4	4	16	David Guetta	I'm Good (Blue)	19	2	4	30704256
## 5	5	8	Metro Boomin	Creepin'	5	3	1	28092232
## 6	6	10	Manuel Turizo	La Bachata	27	4		27357095
##	IncreasedStreams		Total					
## 1	1578141		160755788					
## 2	2872164		642800599					
## 3	3982809		1719435686					
## 4	7144547		541637842					
## 5	381484		153469486					
## 6	727736		667941457					

Mengubah data tipe TimePeak dari chr menjadi numeric

```
data$TimesPeak <-as.integer(data$TimesPeak)
head(data)
```

##	Pos	IncreasedPos	Artist	Title	Wks	Pk	TimesPeak	Streams
## 1	1	4	SZA	Kill Bill	4	1	1	41474366
## 2	2	9	Sam Smith	Unholy	15	1	7	33132175
## 3	3	12	Harry Styles	As It Was	40	1	11	30998084
## 4	4	16	David Guetta	I'm Good (Blue)	19	2	4	30704256
## 5	5	8	Metro Boomin	Creepin'	5	3	1	28092232
## 6	6	10	Manuel Turizo	La Bachata	27	4	NA	27357095
##	IncreasedStreams		Total					
## 1	1578141		160755788					
## 2	2872164		642800599					
## 3	3982809		1719435686					
## 4	7144547		541637842					
## 5	381484		153469486					
## 6	727736		667941457					

Mengecek apakah dalam kolom yang ada di data terdapat missing value

```
colSums(sapply(data, is.na))
```

```
##           Pos      IncreasedPos      Artist      Title
##           0           0           0           0
##           Wks           Pk      TimesPeak      Streams
##           0           0           163           0
## IncreasedStreams      Total
##           71           0
```

Mengganti missing value dengan 0

```
data$TimesPeak<-ifelse(is.na(data$TimesPeak),0,
                        data$IncreasedPos)
data$IncreasedStreams<-ifelse(is.na(data$IncreasedStreams),0,
                              data$IncreasedStreams)
tail(data,22)
```

```
##      Pos IncreasedPos      Artist      Title Wks
## 179 179           1      Mora      LA INOCENTE 24
## 180 180           0      RAYE      Escapism.   1
## 181 181           1    50 Cent      In Da Club 13
## 182 182           1 Taylor Swift  You're On Your Own, Kid 10
## 183 183           1 Kenshi Yonezu KICK BACK   9
## 184 184           1    Eminem      Lose Yourself 207
## 185 185           1 Billie Eilish      TV        11
## 186 186           1 Camila Cabello      Bam Bam   41
## 187 187           1    Keane      Somewhere Only We Know 13
## 188 188           1    KAROL G      GATÚBELA   16
## 189 189           1    Pritam      Kesariya   10
## 190 190           1    The Police      Every Breath You Take 74
## 191 191           1    Seafret      Atlantis   21
## 192 192           1 Guilherme & Benuto      Haja Colírio   2
## 193 193           1    Future      WAIT FOR U   34
## 194 194           1 Tears For Fears Everybody Wants To Rule The World 23
## 195 195           1    Post Malone      Sunflower 215
## 196 196           1    Bad Bunny      Party      33
## 197 197           1    Freddie Dredd      Limbo    7
## 198 198           1    Bruno Mars      Locked Out Of Heaven 82
## 199 199           1    MEDUZA      Bad Memories   2
## 200 200           1    TWISTED      WORTH NOTHING  5
##      Pk TimesPeak Streams IncreasedStreams      Total
## 179  74           0 7077846           0 188339481
## 180 180           0 7048604           0   7048604
## 181  77           0 6975892           0  80329014
## 182   6           0 6955426           0 174292913
## 183  66           0 6949309           0   76744573
## 184  44           0 6932729           0 703684632
## 185  27           0 6928883           0   82967316
## 186   5           0 6888725           0 571221486
## 187 120           0 6876234           0   90755157
## 188  33           0 6842540           0 162007366
## 189  71           0 6840658           0   77686725
## 190 109           0 6820435           0 424204320
## 191  63           0 6815573           0 175770219
```

```
## 192 187      0 6807866      0 13580265
## 193 10      0 6805425      0 349455284
## 194 105      0 6803992      0 159047407
## 195 1      1 6801735      0 2425355542
## 196 6      0 6782777      0 479414025
## 197 158      0 6775767      0 51408912
## 198 45      0 6758678      0 335046675
## 199 199      0 6748442      0 13244341
## 200 181      0 6735908      0 34008257
```

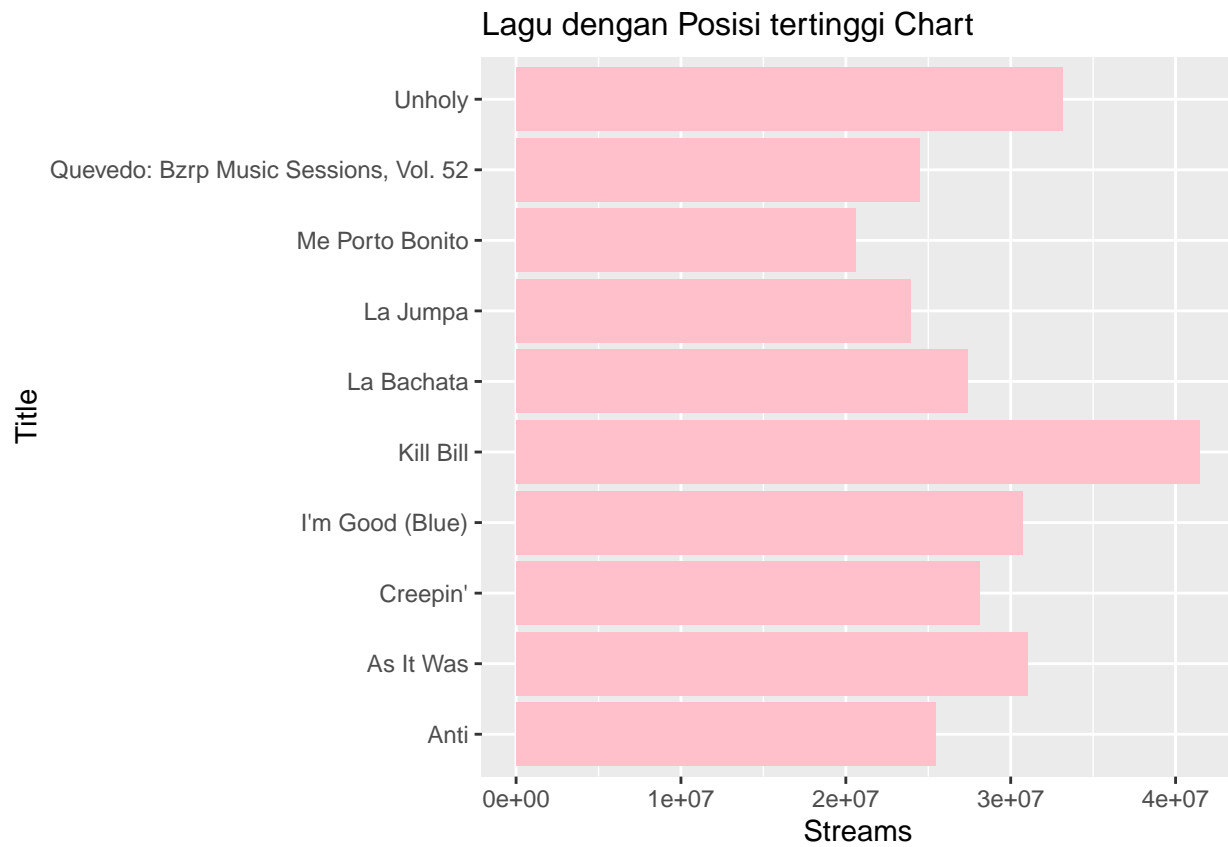
Mengecek kembali apakah dalam kolom yang ada di data terdapat missing value

```
colSums(sapply(data, is.na))
```

```
##          Pos      IncreasedPos      Artist      Title
##          0          0          0          0
##          Wks          Pk      TimesPeak      Streams
##          0          0          0          0
## IncreasedStreams      Total
##          0          0
```

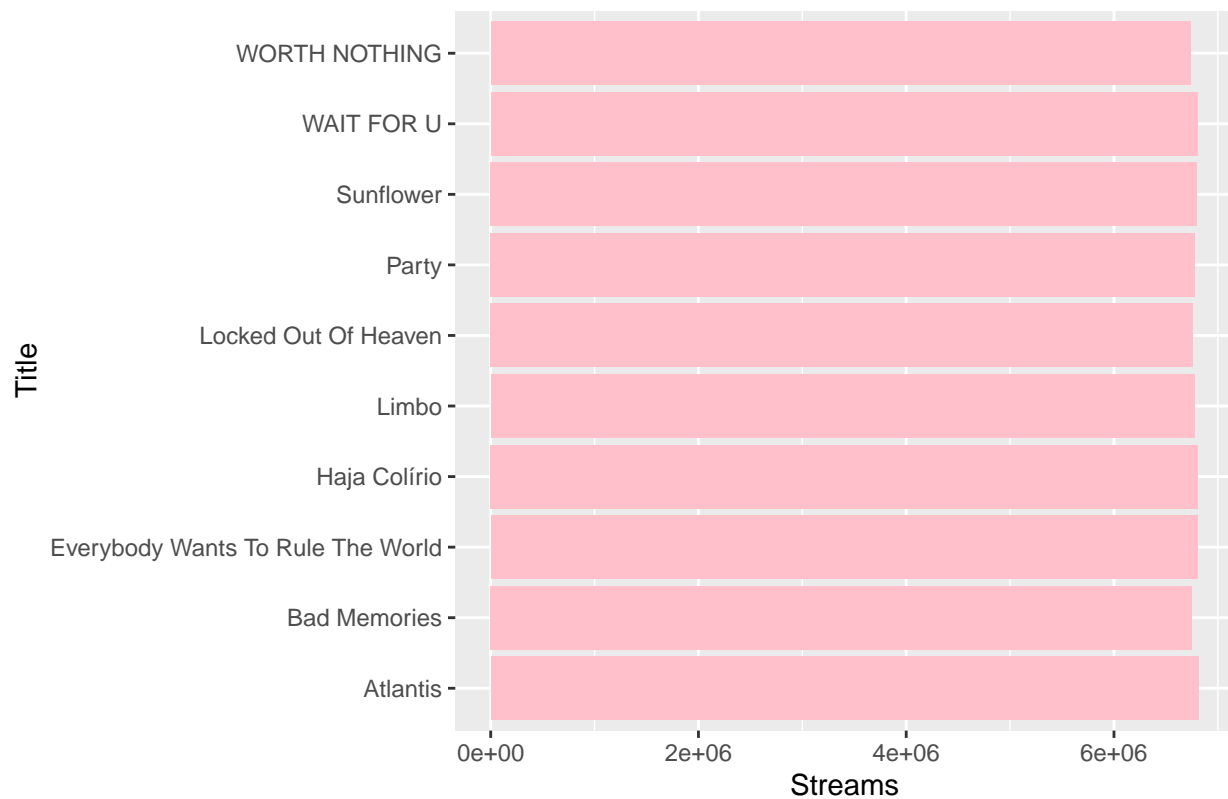
Analyze &Share

```
library(dplyr)
data%>%top_n(10, Streams) %>%
  ggplot()+
  aes(x=Streams,y=Title)+
  geom_bar(stat="identity",fill="pink")+
  ggtitle('Lagu dengan Posisi tertinggi Chart')+
  theme(plot.title = element_text( size=12))
```



```
data%>%top_n(-10, Streams) %>%
  ggplot()+
  aes(x=Streams,y=Title)+
  geom_bar(stat="identity",fill="pink")+
  ggtitle('Lagu dengan Posisi terendah Chart')+
  theme(plot.title = element_text( size=12))
```

Lagu dengan Posisi terendah Chart

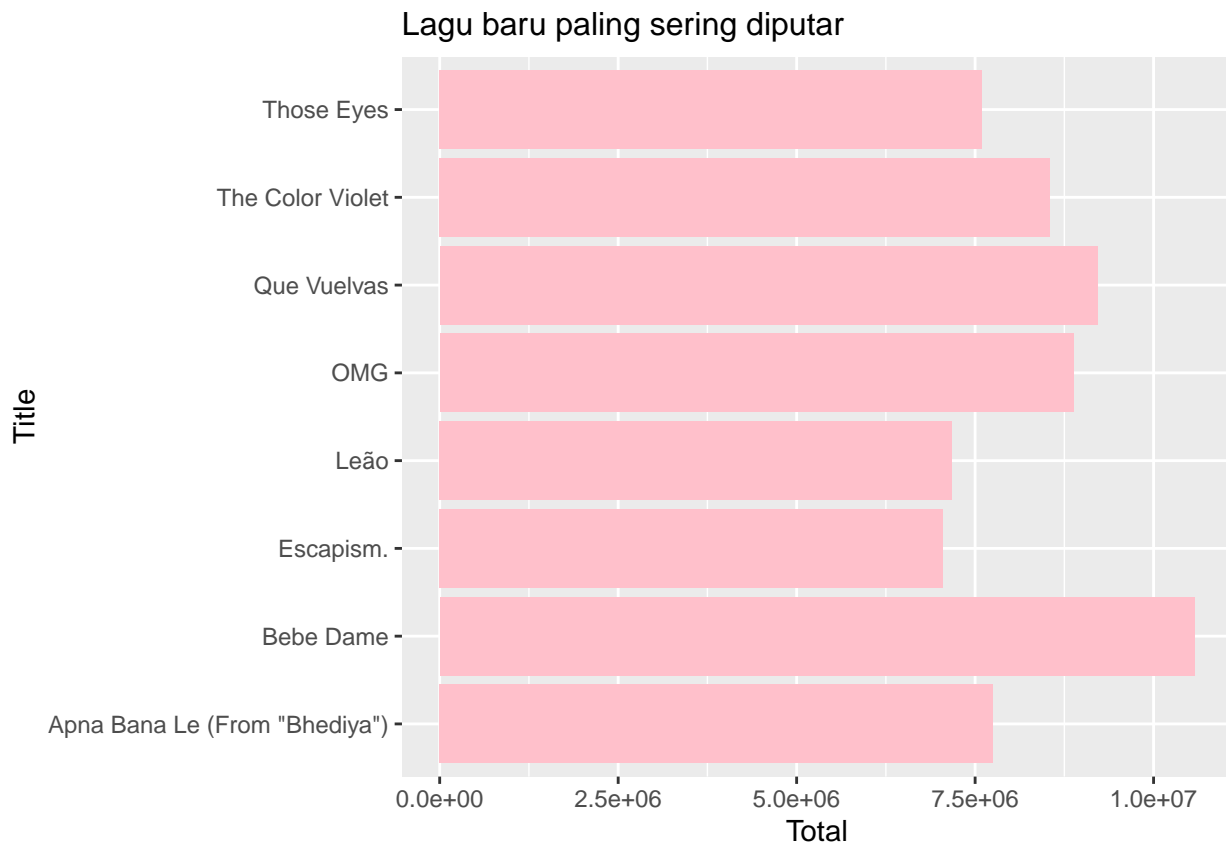


```
dataplot1<-data[data$IncreasedPos==0,]
head(dataplot1)
```

##	Pos	IncreasedPos	Artist	Title	Wks	Pk
## 59	59	0	Fuerza Regida	Bebe Dame	1	59
## 79	79	0	Carin Leon	Que Vuelvas	1	79
## 95	95	0	NewJeans	OMG	1	95
## 106	106	0	Tory Lanez	The Color Violet	1	106
## 142	142	0	Arijit Singh	Apna Bana Le (From "Bhediya")	1	142
## 151	151	0	New West	Those Eyes	1	151

##	TimesPeak	Streams	IncreasedStreams	Total
## 59	0	10571178	0	10571178
## 79	0	9217131	0	9217131
## 95	0	8878049	0	8878049
## 106	0	8552377	0	8552377
## 142	0	7754469	0	7754469
## 151	0	7596355	0	7596355

```
dataplot1%>%
  ggplot()+
  aes(x=Total,y=Title)+
  geom_bar(stat="identity",fill="pink")+
  ggtitle('Lagu baru paling sering diputar')+
  theme(plot.title = element_text( size=12))
```

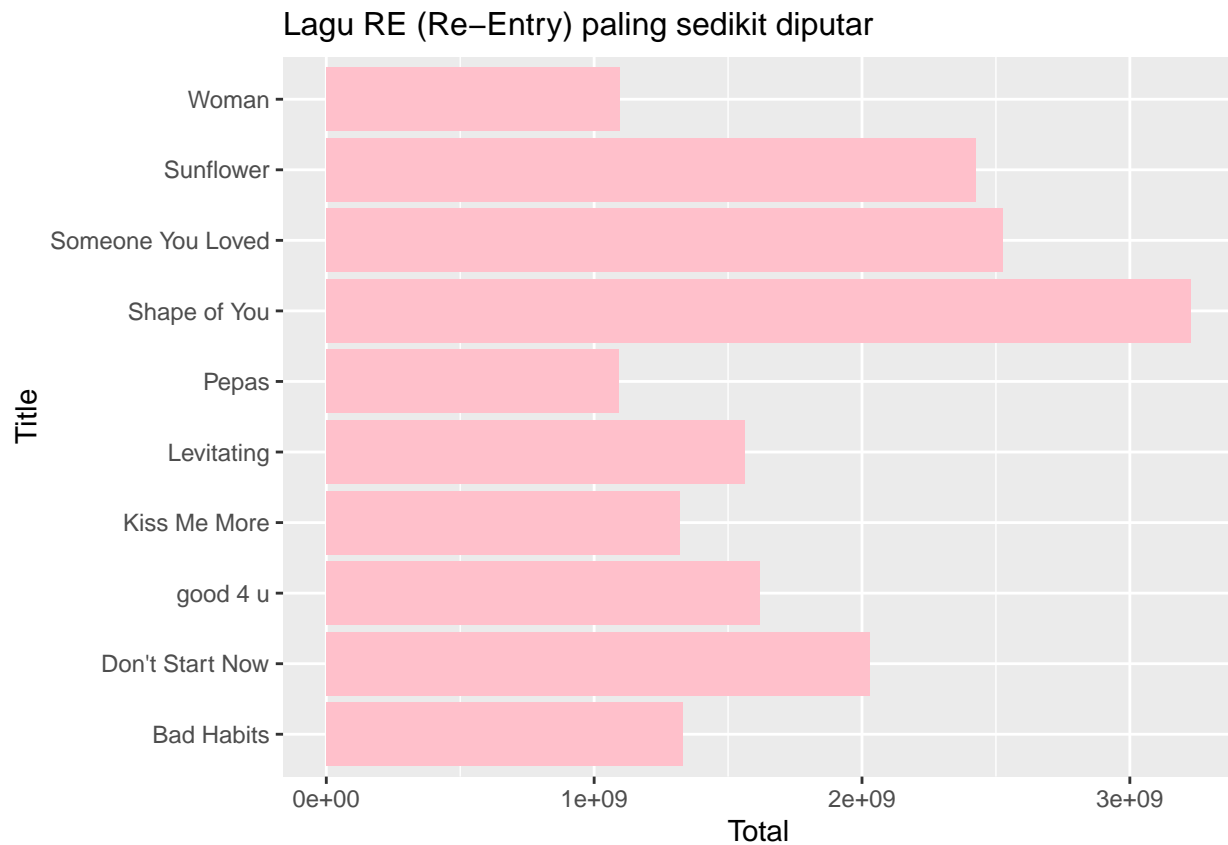



```
dataplot2<-data[data$IncreasedPos==1,]
head(dataplot1)
```

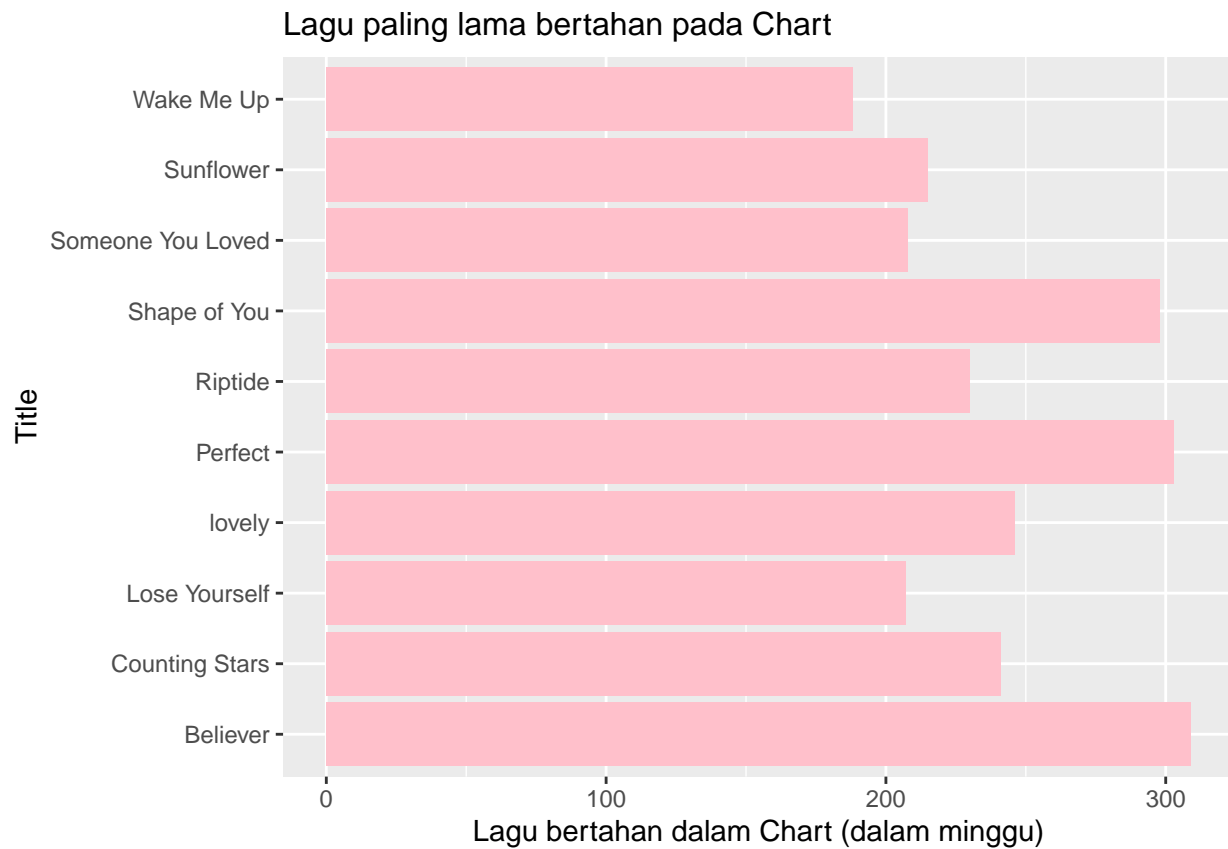
##	Pos	IncreasedPos	Artist	Title	Wks	Pk
## 59	59	0	Fuerza Regida	Bebe Dame	1	59
## 79	79	0	Carin Leon	Que Vuelvas	1	79
## 95	95	0	NewJeans	OMG	1	95
## 106	106	0	Tory Lanez	The Color Violet	1	106
## 142	142	0	Arijit Singh	Apna Bana Le (From "Bhediya")	1	142
## 151	151	0	New West	Those Eyes	1	151

##	TimesPeak	Streams	IncreasedStreams	Total
## 59	0	10571178	0	10571178
## 79	0	9217131	0	9217131
## 95	0	8878049	0	8878049
## 106	0	8552377	0	8552377
## 142	0	7754469	0	7754469
## 151	0	7596355	0	7596355

```
dataplot2%>%top_n(10, Total) %>%
  ggplot()+
  aes(x=Total,y=Title)+
  geom_bar(stat="identity",fill="pink")+
  ggtitle('Lagu RE (Re-Entry) paling sedikit diputar')+
  theme(plot.title = element_text( size=12))
```



```
data%>%top_n(10, Wks) %>%
  ggplot()+
  aes(x=Wks,y=Title)+
  geom_bar(stat="identity",fill="pink")+
  ggtitle('Lagu paling lama bertahan pada Chart')+
  xlab('Lagu bertahan dalam Chart (dalam minggu)')+
  theme(plot.title = element_text( size=12))
```



```
data%>%top_n(-10, Wks) %>%  
  ggplot()+  
  aes(x=Wks,y=Title)+  
  geom_bar(stat="identity",fill="pink")+  
  ggtitle('Lagu dengan waktu singkat bertahan pada Chart')+  
  xlab('Lagu bertahan dalam Chart (dalam minggu)')+  
  theme(plot.title = element_text( size=12))
```

