

Data Analysis Using R

By Egi Fermana Putra



Table of contents



01 Case Study

02 Demografi
Konsumen

03 Uji Validitas &
Reliabilitas

04 Tingkat Kepuasan
Konsumen



Case Study

Seorang Marketing Manager sebuah perusahaan ingin mengetahui profil demografi konsumen pengguna mereknya, karena itu dilakukan survei terhadap 300 responden pengguna mereknya yang tersebar di 6 kota. Keluarkan output demografi (kota, usia, jenis kelamin, pendidikan dan status perkawinan).

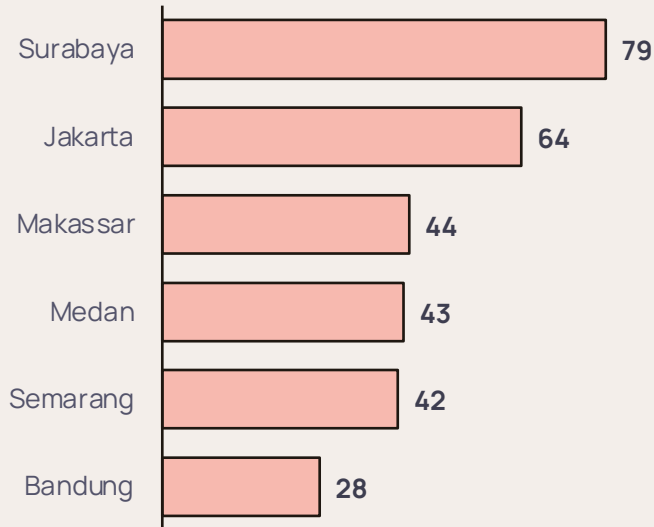
Tingkat kepuasan konsumen terhadap Brand A pada data, diukur dari tingkat kepuasan dalam hal pelayanan, kualitas produk dan harga. Lakukan uji validitas dan reliabilitas pada data, apakah variabel tingkat kepuasan dalam hal pelayanan, kualitas produk dan harga dapat digunakan untuk mengukur tingkat kepuasan secara keseluruhan dengan baik! ($\alpha = 0.6$).

Marketing Manager tersebut juga ingin mengetahui tingkat kepuasan konsumennya di masing-masing kota. Ada 4 tingkat kepuasan yang diukur yaitu tingkat kepuasan konsumen secara keseluruhan, tingkat kepuasan konsumen terhadap pelayanan yang diberikan, tingkat kepuasan konsumen terhadap kualitas produk dan tingkat kepuasan konsumen terhadap harga produk. Tingkat kepuasan didefinisikan sebagai rata-rata dari tingkat kepuasan semua responden.

- Keluarkan tingkat kepuasan konsumen secara keseluruhan untuk masing-masing kota dan secara total (gabungan semua kota).
- Apakah ada perbedaan tingkat kepuasan untuk masing-masing kota? Jika ada, kota mana yang berbeda? Analisa pada $\alpha = 5.0$.

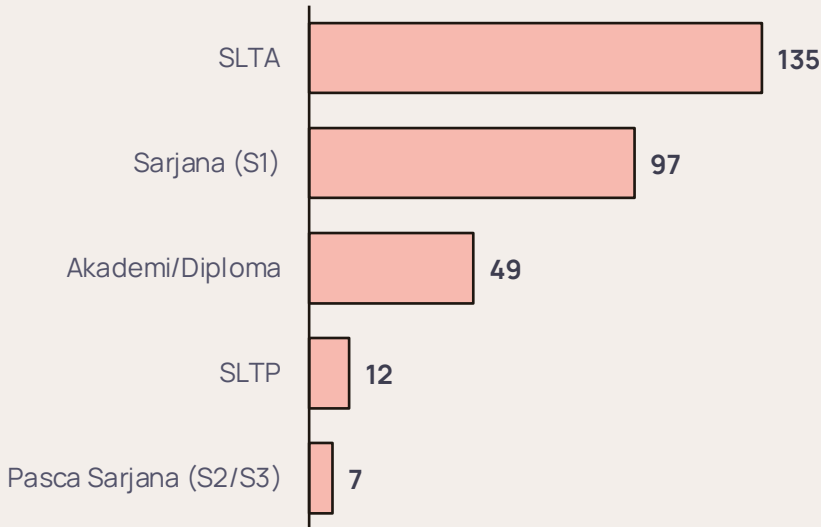
Demografi

Distribusi Responden di Berbagai Kota:
Surabaya Mendominasi dengan 79 Responden.



Mayoritas kota asal responden adalah **Surabaya** sebanyak **79 responden**, dan kota kedua **Jakarta** sebanyak **64 responden**, Makassar, Medan, dan Semarang rata-rata sebanyak 43 responden, sedangkan Bandung dengan jumlah sedikit yaitu 28 responden.

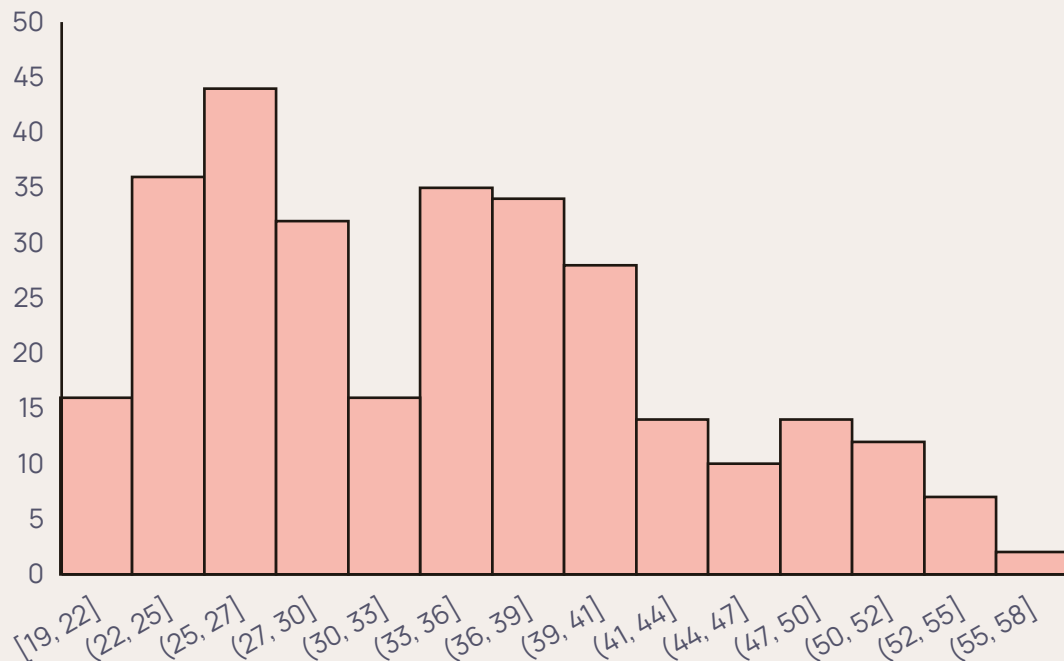
Mayoritas Tingkat Pendidikan Responden adalah
SLTA dengan 135 Responden.



Sebagian besar responden pada tingkat Pendidikan **SLTA 135** responden, diikuti **Sarjana (S1) 97 responden**, kemudian **Akademi/Diploma 49 responden**, **SLTP sebanyak 12 responden**, dan **Pasca Sarjana (S2/S3) sebanyak 7 responden**.

Demografi

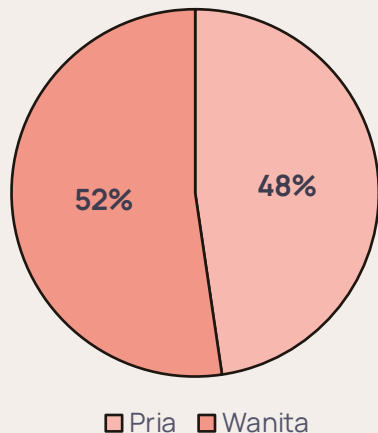
Distribusi Usia Responden, Rata-rata Usia Responden yaitu 36.63.



- **Kelompok Usia yang Paling Umum:**
Kelompok usia dengan frekuensi tinggi adalah usia 25, 26, 30, 33, 35, dan 40, menunjukkan jumlah yang signifikan dalam rentang usia ini.
- **Kelompok Usia yang Jarang:**
Beberapa kelompok usia, seperti 29, 41, 43, 45, 46, 51, 53, 54, 55, dan 58, memiliki frekuensi yang lebih rendah, menunjukkan representasi lebih kecil dari individu dalam rentang usia ini.
- **Rata-rata usia** responden yaitu **36.63**.

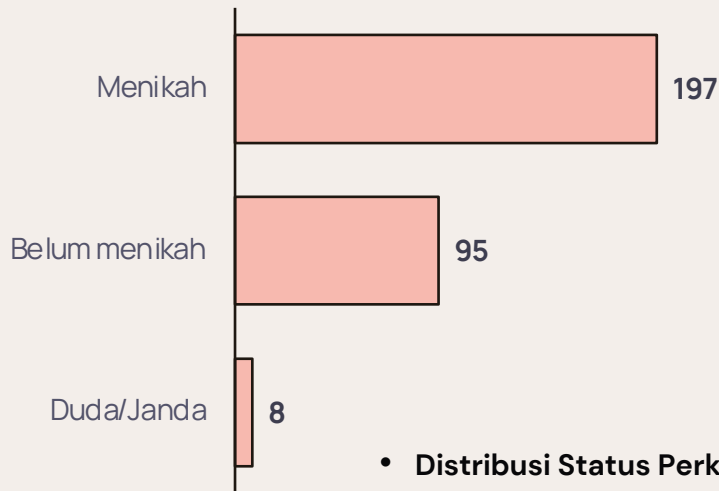
Demografi

Jenis Kelamin Responden



Representasi jenis kelamin cukup seimbang, dengan **143 responden** dikategorikan sebagai **Pria**, dan **157 responden** kategori **Wanita**.

Distribusi Status Perkawinan, Sebagian Besar Responden Sudah Menikah.



- **Distribusi Status Perkawinan:** Ada 197 responden yang sudah menikah, 95 responden yang belum menikah, dan 8 responden duda/janda.
- **Mayoritas Status Perkawinan:** Sebagian besar responden dikategorikan sudah menikah.



Uji Validitas & Reliabilitas

Reliabilitas Skala Kepuasan: analisis reliabilitas untuk mengukur sejauh mana tiga pertanyaan terkait tingkat kepuasan konsumen terhadap pelayanan, tingkat kepuasan konsumen terhadap kualitas produk, tingkat kepuasan konsumen terhadap harga produk cukup reliabel sebagai satu kesatuan.

Hasil Analisis:

- Reliabilitas keseluruhan **alpha** adalah **0.77**, menunjukkan **tingkat konsistensi yang cukup**.
 - Jika satu pertanyaan dihapus, reliabilitas bisa meningkat (misalnya, menghapus tingkat kepuasan konsumen terhadap pelayanan dapat meningkatkan reliabilitas menjadi 0.86).
-

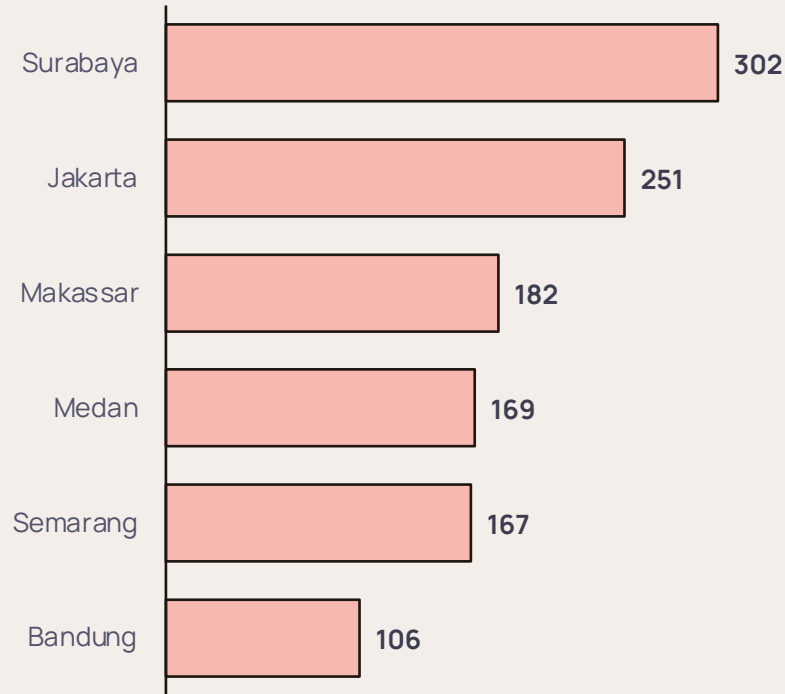
Variabel **tingkat kepuasan konsumen terhadap pelayanan** memiliki **dampak signifikan pada reliabilitas**, menunjukkan pentingnya variabel tersebut.

Kesimpulan: apakah reliabilitas keseluruhan dapat diterima berdasarkan alpha 0.6?.
Hasilnya reliabilitas saat ini tidak memenuhi kriteria yang ditetapkan.

Tingkat Kepuasan Konsumen



Perbandingan Kepuasan Pelanggan di Seluruh Kota: Surabaya Mendominasi dengan 302, Bandung di Posisi Terbawah dengan 106



- **Total Kepuasan per Kota:** kota Surabaya memiliki total kepuasan sebanyak 302, Jakarta sebanyak 251, Makassar sebanyak 182, Medan sebanyak 169, Semarang sebanyak 167, dan Bandung sebanyak 106.
- **Rata-rata skor kepuasan konsumen keseluruhan** adalah sekitar 3.92.
- **Skor kepuasan konsumen total bervariasi di berbagai kota**, menunjukkan perbedaan pengalaman atau persepsi konsumen.
- **Kota Surabaya memiliki skor kepuasan konsumen tertinggi** yaitu 302, sedangkan kota **Bandung memiliki skor terendah** yaitu 106.
- Kota dengan skor kepuasan konsumen yang lebih rendah seperti Bandung, mungkin memerlukan perhatian lebih untuk mengidentifikasi masalah spesifik yang mempengaruhi kepuasan konsumen.



Perbedaan Tingkat Kepuasan Tiap Kota

Hasil menunjukkan bahwa tidak ada perbedaan signifikan dalam tingkat kepuasan di semua kota ($p > 0.05$).

Ada perbedaan signifikan antara Kota Makassar dan Kota Bandung ($p = 0.028$), menunjukkan bahwa kedua kota ini memiliki tingkat kepuasan yang berbeda.

Beberapa kota memiliki nilai p yang lebih besar dari 0.05, menunjukkan perbedaan yang tidak signifikan. Seperti Kota Surabaya dan Kota Bandung ($p = 0.999$) tidak menunjukkan perbedaan signifikan dalam tingkat kepuasan.

Secara keseluruhan, analisis ini mengidentifikasi perbedaan signifikan dalam tingkat kepuasan antara beberapa pasangan kota, memberikan wawasan yang dapat dijadikan dasar untuk perbaikan yang ditargetkan dalam kepuasan pelanggan.

R Syntax

Syntax

```
1 # Read CSV file
2 data <- read.csv("data.csv", header=TRUE)
3
4 # No. 1, mean dan standar deviasi dari demografi
5 demographic <- aggregate(cbind(usia, sex, didik, status) ~ kota, data = data,
6 FUN = function(x) c(mean = mean(x), sd = sd(x)))
7 print(demographic)
8
9 # No. 1, frekuensi demografi kota, usia, sex, pendidikan, dan status pernikahan
10 kota_freq <- table(data$kota)
11 usia_freq <- table(data$usia)
12 sex_freq <- table(data$sex)
13 didik_freq <- table(data$didik)
14 status_freq <- table(data$status)
15
16 # Print demografi kota, usia, sex, pendidikan, dan status pernikahan
17 print("Frekuensi Kota:")
18 print(kota_freq)
19 print("Frekuensi Usia:")
20 print(usia_freq)
21 print("Frekuensi Jenis Kelamin:")
22 print(sex_freq)
23 print("Frekuensi Pendidikan:")
24 print(didik_freq)
25 print("Frekuensi Status:")
26 print(status_freq)
27
```

```
> demographic <- aggregate(cbind(usia, sex, didik, status) ~ kota, data = data,
+ FUN = function(x) c(mean = mean(x), sd = sd(x)))
> print(demographic)
  kota usia.mean  usia.sd  sex.mean  sex.sd didik.mean didik.sd status.mean status.sd
1 1 34.515625 10.056968 1.5156250 0.5037065 3.6718750 0.9929064 1.7500000 0.4364358
2 2 34.678571 9.076031 1.3571429 0.4879500 3.6785714 1.0202967 1.8571429 0.4483951
3 3 33.642857 8.260460 1.5000000 0.5060608 3.8333333 0.9084070 1.6904762 0.5625766
4 4 33.784810 8.259681 1.4810127 0.5028320 3.9240506 1.0472449 1.7974684 0.4349992
5 6 35.488372 8.740633 1.6744186 0.4741373 3.9534884 0.9500248 1.6744186 0.6063525
6 7 30.113636 9.638141 1.5909091 0.4973503 3.9318182 1.0652596 1.4545455 0.5479155
> |
```

Output

```
> print("Frekuensi Kota:")
[1] "Frekuensi Kota:"
> print(kota_freq)

 1  2  3  4  6  7
64 28 42 79 43 44
> print("Frekuensi Usia:")
[1] "Frekuensi Usia:"
> print(usia_freq)

19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52
 3  5  8 10 14 12 17 20 7 13 1 18 10 6 11 7 17 9 13 12 10 15 3 5 6 3 6 4 8 5 1 6 4 2
53 64 55 66 58
 3  1  3  1  1
> print("Frekuensi Jenis Kelamin:")
[1] "Frekuensi Jenis Kelamin:"
> print(sex_freq)

 1  2
143 157
> print("Frekuensi Pendidikan:")
[1] "Frekuensi Pendidikan:"
> print(didik_freq)

 2  3  4  5  6
12 135 49 97 7
> print("Frekuensi Status:")
[1] "Frekuensi Status:"
> print(status_freq)

 1  2  3
95 197 8
> |
```

R Syntax

Output

Syntax

```
27
28 # No. 2, uji validitas dan reliabilitas, dengan alpha = 0.6
29 library(psych)
30 satisfaction <- data[, c("puas_lay", "puas_kua", "puas_pri")]
31 reliability_result <- alpha(satisfaction)
32 print(reliability_result)
33
34 ▾ if(all(reliability_result$total > 0.6)) {
35   print("Reliabilitas dapat diterima.")
36 ▾ } else {
37   print("Reliabilitas tidak dapat diterima.")
38 ▲ }
39
```

```
> library(psych)
> satisfaction <- data[, c("puas_lay", "puas_kua", "puas_pri")]
> reliability_result <- alpha(satisfaction)
> print(reliability_result)

Reliability analysis
Call: alpha(x = satisfaction)

      raw_alpha std.alpha G6(smc) average_r S/N ase mean sd median_r
      0.77      0.81      0.77      0.58 4.2 0.023 3.8 0.44 0.54

      95% confidence boundaries
              lower alpha upper
Feldt      0.73 0.77 0.81
Duhachek 0.73 0.77 0.82

Reliability if an item is dropped:
      raw_alpha std.alpha G6(smc) average_r S/N alpha se var.r med.r
puas_lay 0.86      0.86      0.75      0.75 6.0 0.017 NA 0.75
puas_kua 0.67      0.70      0.54      0.54 2.4 0.034 NA 0.54
puas_pri 0.59      0.62      0.45      0.45 1.7 0.043 NA 0.45

Item statistics
      n raw.r std.r r.cor r.drop mean sd
puas_lay 300 0.84 0.78 0.58 0.53 3.6 0.65
puas_kua 300 0.82 0.86 0.79 0.65 3.9 0.44
puas_pri 300 0.87 0.90 0.85 0.73 3.9 0.45

Non missing response frequency for each item
      1 2 3 4 5 miss
puas_lay 0 0.04 0.33 0.58 0.05 0
puas_kua 0 0.00 0.16 0.80 0.04 0
puas_pri 0 0.00 0.17 0.78 0.05 0
>
> if(all(reliability_result$total > 0.6)) {
+   print("Reliabilitas dapat diterima.")
+ } else {
+   print("Reliabilitas tidak dapat diterima.")
+ }
[1] "Reliabilitas tidak dapat diterima."
>
```

R Syntax

Syntax

```
39
40 # No. 3a, tingkat kepuasan konsumen masing-masing kota
41 total_satisfaction_by_city <- aggregate(cbind(puas_all) ~ kota,
42                                       data = data, FUN = sum)
43 total_overall_satisfaction <- mean(data$puas_all)
44 print(total_satisfaction_by_city)
45 print(paste("Total kepuasan konsumen keseluruhan: ",
46            total_overall_satisfaction))
47
```

Output

```
> total_satisfaction_by_city <- aggregate(cbind(puas_all) ~ kota,
+                                       data = data, FUN = sum)
> total_overall_satisfaction <- mean(data$puas_all)
> print(total_satisfaction_by_city)
  kota puas_all
1     1      251
2     2      106
3     3      167
4     4      302
5     6      169
6     7      182
> print(paste("Total kepuasan konsumen keseluruhan: ",
+            total_overall_satisfaction))
[1] "Total kepuasan konsumen keseluruhan:  3.92333333333333"
> |
```

R Syntax

Output

Syntax

```
47  
48 # No. 3b, tes apakah ada perbedaan tingkat kepuasan  
49 # masing-masing kota, dengan alpha = 5.0  
50 library(stats)  
51 data$kota <- as.factor(data$kota)  
52 anova_result <- aov(puas_all ~ kota, data = data)  
53 tukey_result <- TukeyHSD(anova_result)  
54 print(tukey_result)  
55 significant_differences <- tukey_result$kota[, 4]  
56 print("Perbedaan signifikan:")  
57 print(significant_differences)  
58
```

```
> library(stats)  
> data$kota <- as.factor(data$kota)  
> anova_result <- aov(puas_all ~ kota, data = data)  
> tukey_result <- TukeyHSD(anova_result)  
> print(tukey_result)  
Tukey multiple comparisons of means  
95% family-wise confidence level
```

```
Fit: aov(formula = puas_all ~ kota, data = data)
```

\$kota		diff	lwr	upr	p adj
2-1	-0.136160714	-0.44363440	0.1713130	0.8009782	
3-1	0.054315476	-0.21516130	0.3237923	0.9923953	
4-1	-0.099090190	-0.32730695	0.1291266	0.8139501	
6-1	0.008357558	-0.25922064	0.2759358	0.9999992	
7-1	0.214488636	-0.05126462	0.4802419	0.1912012	
3-2	0.190476190	-0.14060035	0.5215527	0.5657977	
4-2	0.037070524	-0.26138689	0.3355279	0.9992418	
6-2	0.144518272	-0.18501477	0.4740513	0.8074563	
7-2	0.350649351	0.02259641	0.6787023	0.0283716	
4-3	-0.153405666	-0.41254751	0.1057362	0.5340965	
6-3	-0.045957918	-0.34035509	0.2484393	0.9977177	
7-3	0.160173160	-0.13256631	0.4529126	0.6192885	
6-4	0.107447748	-0.14971922	0.3646147	0.8373642	
7-4	0.313578826	0.05831121	0.5688464	0.0064928	
7-6	0.206131078	-0.08486163	0.4971238	0.3267121	

```
> significant_differences <- tukey_result$kota[, 4]  
> print("Perbedaan signifikan:")  
[1] "Perbedaan signifikan:"  
> print(significant_differences)  
      2-1      3-1      4-1      6-1      7-1      3-2      4-2      6-2  
0.800978235 0.992395341 0.813950130 0.999999185 0.191201170 0.565797709 0.999241762 0.807456323  
      7-2      4-3      6-3      7-3      6-4      7-4      7-6  
0.028371647 0.534096502 0.997717706 0.619288546 0.837364208 0.006492782 0.326712097  
> |
```

Coding Variable

kota	
1	Jakarta
2	Bandung
3	Semarang
4	Surabaya
5	Palembang
6	Medan
7	Makassar
8	Banjarmasin

usia_responden	
1	19-25 thn
2	26-31 thn
3	32-37 thn
4	38-43 thn
5	44-49 thn
6	50-55 thn
7	56-61 thn

sex	
1	Pria
2	Wanita

status	
1	Belum menikah
2	Menikah
3	Duda/Janda

pendidikan	
1	SD
2	SLTP
3	SLTA
4	Akademi/Diploma
5	Sarjana (S1)
6	Pasca Sarjana (S2/S3)

puas_all, puas_kua, puas_lay, puas_pri	
1	Very Much Dissatisfied
2	Dissatisfied
3	Neutral
4	Satisfied
5	Very Much Satisfied

media	
1	Televisi
2	Radio
3	Koran
4	Majalah
5	Tabloid
6	Komunitas
7	Event/Pameran
8	Internet



Thanks!

Do you have any questions?

egifermn@gmail.com

<https://www.linkedin.com/in/egifermana/>

