

# **LAPORAN HASIL ANALISIS**

## **MATA KULIAH STATISTIKA KOMPUTASI**

### **KELAS B**



### **“Analisis Pengaruh Jumlah Subscribers terhadap Total Engagement Tiktokers”**

#### **DISUSUN OLEH KELOMPOK I:**

- |                            |                           |
|----------------------------|---------------------------|
| 1. ANGELA LISANTHONI       | ( 21083010032 ) - KETUA   |
| 2. DENDY ARIZKI KUSWARDANA | ( 21083010006 ) - ANGGOTA |
| 3. NAOMI DWI ANGGRAINI     | ( 21083010010 ) - ANGGOTA |
| 4. ALYA SETYA PARAMITA     | ( 21083010046 ) - ANGGOTA |
| 5. MUHIMMATUL AROFAH       | ( 21083010055 ) - ANGGOTA |
| 6. DIVAYANTI FEBRI SAKINA  | ( 21083010099 ) - ANGGOTA |
| 7. YAYANG DIMAS SAPUTRA    | ( 21083010102 ) - ANGGOTA |
| 8. ZULFAZ REFIE ABABIL     | ( 21083010122 ) - ANGGOTA |

#### **DOSEN PENGAMPU:**

AMRI MUHAIMIN, S.STAT, M.STAT, MS  
DR. ENG. DWI ARMAN PRASETYA, ST., MT., IPU. ASEAN. ENG

PROGRAM STUDI SAINS DATA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”  
JAWA TIMUR  
2022

## DAFTAR ISI

|                                             |           |
|---------------------------------------------|-----------|
| <b>1.BAB 1: PENDAHULUAN .....</b>           | <b>3</b>  |
| 1.1 LATAR BELAKANG .....                    | 3         |
| 1.2 PERMASALAHAN .....                      | 4         |
| 1.3 TUJUAN .....                            | 4         |
| 1.4 MANFAAT .....                           | 4         |
| <b>2.BAB II: HASIL DAN PEMBAHASAN .....</b> | <b>4</b>  |
| 2.1 DATASET .....                           | 4         |
| 2.2 STATISTIKA DESKRIPTIF .....             | 9         |
| 2.3 STATISTIKA INFERENSIAL .....            | 17        |
| <b>3.BAB III: KESIMPULAN .....</b>          | <b>21</b> |
| <b>4.LAMPIRAN .....</b>                     | <b>22</b> |
| 4.1 PEMBAGIAN TUGAS .....                   | 22        |

## 1. BAB 1: PENDAHULUAN

### 1.1 Latar Belakang

Di era modern seperti sekarang, penjualan serta promosi produk bukan hanya berbasis *offline* tetapi, lebih mengarah ke *online*. Salah satu strategi yang paling sering digunakan adalah melakukan *endorsement* atau membayar *influencer* untuk mempromosikan produk yang dijual oleh perusahaan. Pengaruh *Influencer* menjadi komponen yang sangat penting untuk meningkatkan *branding* suatu produk. *Influencer* bisa mempromosikan produk *endorsement* melalui sosial media seperti Instagram, youtube, tiktok, dan lain sebagainya sebagai media penyebaran pengenalan produk ke masyarakat.

Salah satu platform media sosial yang terkenal adalah tiktok. Tiktok bukan hanya terkenal di Indonesia, tapi hampir di seluruh penjuru dunia sehingga jumlah akun penggunaanya sangat banyak serta beragam. Data terbaru 26 Juli 2021 dari Statista menyebutkan total akun aktif pengguna tiktok mencapai lebih dari 150 juta pengguna. Amerika Serikat menduduki peringkat teratas dengan total 65.9 juta akun aktif pengguna tiktok dan Indonesia menduduki peringkat kedua dengan total 22.2 juta akun aktif pengguna tiktok. Semakin banyak akun aktif, akan ada lebih banyak kemungkinan orang – orang mengenal produk yang dipasarkan.

Namun kebanyakan perusahaan berfokus pada jumlah *subscribers*, padahal masih banyak pertimbangan lain dalam memilih *influencer* yang tepat. Perusahaan juga harus memperhatikan jumlah *views*, *likes*, *comments*, dan *shares* dari setiap postingan untuk memilih *influencer* yang tepat agar meminimalisir kerugian. Sangatlah penting bagi perusahaan untuk memilih *influencer* yang sesuai dengan produk yang dijualnya agar mendapatkan keuntungan maksimal serta memperhatikan jumlah *reached accounts* dan *engagement* dari sosial media *influencer* tersebut.

Ditinjau dari kesalahan ini, hal khusus yang mendasari laporan analisis berikut adalah untuk membuktikan adanya hubungan antara jumlah *subscribers* dengan *views*, *likes*, *comment*, dan *shares*. Sehingga analisis ini dilakukan untuk memberikan wawasan dan pembuktian bahwasannya ada pengaruh atau korelasi antara jumlah *subscribers* dengan *views*, *likes*, *comment*, dan *shares* dengan analisis sederhana serta pembuktian bahwa apakah jumlah *subscribers* tinggi menjadi indikator bahwa *engagement*nya tinggi atau bisa sebaliknya.

## 1.2 Permasalahan

Adapun permasalahan yang dibahas dalam laporan yang **Analisis Pengaruh Jumlah Subscribers terhadap Total Engagement Tiktokers** ini adalah:

- Apakah ada korelasi antara jumlah *subscribers* dengan rata – rata *views*, *like*, *comments*, dan *share* yang didapatkan?
- Apakah jumlah *subscribers* tinggi menjadi indikator bahwa tingkat *engagement*nya tinggi?

## 1.3 Tujuan

Adapun tujuan dari dibuatnya proyek yang berjudul **Analisis Pengaruh Jumlah Subscribers terhadap Total Engagement Tiktokers** ini adalah:

- Untuk mengetahui pengaruh jumlah *subscribers* terhadap rata – rata *views*, *like*, *comments*, dan *share* yang didapatkan.
- Untuk mengetahui apakah jumlah *subscribers* yang tinggi bisa menjadi indikator tingkat *engagement*.

## 1.4 Manfaat

Adapun manfaat dari dibuatnya proyek yang berjudul **Analisis Pengaruh Jumlah Subscribers terhadap Total Engagement Tiktokers** ini adalah:

- Dapat menambah wawasan akan adanya korelasi pengaruh jumlah *subscribers* terhadap rata – rata *views*, *like*, *comments*, dan *share* yang didapatkan.
- Dapat menambah wawasan mengenai apakah jumlah *subscribers* yang tinggi bisa menjadi indikator tingkat *engagement*.

# 2. BAB II: HASIL DAN PEMBAHASAN

## 2.1 Dataset

Dataset yang digunakan dalam analisis ini adalah daftar nama *influencer* dalam media sosial tiktok beserta nama akun, jumlah *subscribers*, rata - rata *views*, rata - rata *likes*, rata - rata *comment*, dan rata - rata *shares*. Ini adalah data jeda 6 bulan, mulai dari Maret 2022 hingga September 2022.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Gambar 1. Import Library

### a. Data Awal

```
df = pd.read_csv('data_tiktok.csv')
df
```

|     | S.no | Tiktoker name | Tiktok name   | Subscribers | Views avg. | Likes avg. | Comments avg. | Shares avg. |
|-----|------|---------------|---------------|-------------|------------|------------|---------------|-------------|
| 0   | 1    | jypestraykids | Stray Kids    | 13.8M       | 6.4M       | 2.3M       | 50.2K         | 34.2K       |
| 1   | 2    | khaby.lame    | Khabane lame  | 149.2M      | 17.3M      | 2.3M       | 15.2K         | 8.7K        |
| 2   | 3    | scarlettspam2 | scarlett      | 2.1M        | 17.9M      | 845.8K     | 53.9K         | 6.3K        |
| 3   | 4    | addisonre     | Addison Rae   | 88.7M       | 22M        | 906.6K     | 7.6K          | 26.2K       |
| 4   | 5    | belindatok    | Belinda       | 4.8M        | 14.2M      | 1.5M       | 14.5K         | 15.3K       |
| ... | ...  | ...           | ...           | ...         | ...        | ...        | ...           | ...         |
| 995 | 996  | brendadialoy  | Brenda Dialoy | 371.9K      | 1.2M       | 187.5K     | 626           | 2K          |
| 996 | 997  | jujuftcats    | Juju Fitcats  | 4.3M        | 2.2M       | 280.8K     | 380           | 558         |
| 997 | 998  | xoteam        | XO Team       | 37.8M       | 2.5M       | 180.2K     | 859           | 450         |
| 998 | 999  | kimsnwuo      |               | 100.9K      | 957.4K     | 226.8K     | 908           | 1.7K        |
| 999 | 1000 | jass_mcg      | jass_mcg      | 52.2K       | 503.8K     | 116K       | 8.1K          | 1.8K        |

1000 rows × 8 columns

Gambar 2. Memunculkan isi dataset

Pada gambar 2 menunjukkan isi *dataset* awal yang memiliki 1000 baris dan 8 kolom. Kolom tersebut adalah nomor, tiktoker name, tiktok name, subscribers, views average, likes average, comments average, dan shares averages.

### b. Pengecekan data null dan menghapus data null

```
df.isna().sum() |
```

|               |       |
|---------------|-------|
| S.no          | 0     |
| Tiktoker name | 0     |
| Tiktok name   | 1     |
| Subscribers   | 0     |
| Views avg.    | 0     |
| Likes avg.    | 0     |
| Comments avg. | 0     |
| Shares avg.   | 0     |
| dtype:        | int64 |

Gambar 3. Pengecekan data null

Pada gambar 3, dilakukan pengecekan apakah ada data null dari setiap kolom. Terlihat ada satu data null pada kolom Tiktok Name sehingga data null ini perlu dihapus untuk mengoptimalkan hasil analisis. Fungsi `isna()` digunakan untuk mengecek apakah ada data null dan fungsi `sum()` digunakan untuk menjumlahkan banyak data null pada setiap kolom.

```
df = df.dropna(subset=["Tiktok name"], axis=0) |
df
```

|     | S.no | Tiktoker name | Tiktok name   | Subscribers | Views avg. | Likes avg. | Comments avg. | Shares avg. |
|-----|------|---------------|---------------|-------------|------------|------------|---------------|-------------|
| 0   | 1    | jypestraykids | Stray Kids    | 13.8M       | 6.4M       | 2.3M       | 50.2K         | 34.2K       |
| 1   | 2    | khaby.lame    | Khabane lame  | 149.2M      | 17.3M      | 2.3M       | 15.2K         | 8.7K        |
| 2   | 3    | scarlettspam2 | scarlett      | 2.1M        | 17.9M      | 845.8K     | 53.9K         | 6.3K        |
| 3   | 4    | addisonre     | Addison Rae   | 88.7M       | 22M        | 906.6K     | 7.6K          | 26.2K       |
| 4   | 5    | belindatok    | Belinda       | 4.8M        | 14.2M      | 1.5M       | 14.5K         | 15.3K       |
| ... | ...  | ...           | ...           | ...         | ...        | ...        | ...           | ...         |
| 995 | 996  | brendadialoy  | Brenda Dialoy | 371.9K      | 1.2M       | 187.5K     | 626           | 2K          |
| 996 | 997  | jujufitcats   | Juju Fitcats  | 4.3M        | 2.2M       | 280.8K     | 380           | 558         |
| 997 | 998  | xoteam        | XO Team       | 37.8M       | 2.5M       | 180.2K     | 859           | 450         |
| 998 | 999  | kimsnwuo      |               | 100.9K      | 957.4K     | 226.8K     | 908           | 1.7K        |
| 999 | 1000 | jass_mcg      | jass_mcg      | 52.2K       | 503.8K     | 116K       | 8.1K          | 1.8K        |

999 rows x 8 columns

Gambar 4. Menghapus baris yang terdapat data null

Pada gambar 4, dilakukan penghapusan baris terhadap baris di kolom 'Tiktok name' yang memuat data null. Berdasarkan pengecekan pada gambar 3, hanya ada satu data null di kolom 'Tiktok name' maka hanya ada satu baris yang dihapus. Oleh sebab itu data sebelumnya yang 1000 rows x 8 columns berubah menjadi 999 rows x 8 columns.

c. Pengecekan duplikasi data

```
df.duplicated().any()
```

False

Gambar 5. Pengecekan duplikasi data

Pada gambar 5, dilakukan pengecekan apakah adanya data yang terduplikasi dalam dataset. Dari pengecekan diatas, ternyata tidak ada data yang terduplikat sehingga tidak perlu adanya penghapusan baris. Sehingga *dataset* terakhir adalah 999 rows x 8 columns dan data sudah dilakukan *cleansing*.

d. Pengecekan tipe data

```
df.dtypes

S.no          int64
Tiktoker name object
Tiktok name   object
Subscribers   object
Views avg.    object
Likes avg.    object
Comments avg. object
Shares avg.   object
dtype: object
```

Gambar 6. Pengecekan tipe data setiap kolom

Pada gambar 6 menunjukkan bahwa tipe data kolom Subscribers, Views avg., Likes avg., Comments avg., dan Shares avg., masih berubah object atau string sehingga akan menimbulkan error untuk melakukan analisis kemudiannya. Sehingga perlu dilakukan perubahan tipe data

```
df['Subscribers'] = (df['Subscribers'].replace(r'[KM]+$', '', regex=True).astype(float) * \
df['Subscribers'].str.extract(r'[\d\.]+([KM]+)', expand=False)
.fillna(1)
.replace(['K', 'M'], [10**3, 10**6])).astype(int))

df['Views avg.'] = (df['Views avg.'].replace(r'[KM]+$', '', regex=True).astype(float) * \
df['Views avg.'].str.extract(r'[\d\.]+([KM]+)', expand=False)
.fillna(1)
.replace(['K', 'M'], [10**3, 10**6])).astype(int))

df['Likes avg.'] = (df['Likes avg.'].replace(r'[KM]+$', '', regex=True).astype(float) * \
df['Likes avg.'].str.extract(r'[\d\.]+([KM]+)', expand=False)
.fillna(1)
.replace(['K', 'M'], [10**3, 10**6])).astype(int))

df['Comments avg.'] = (df['Comments avg.'].replace(r'[KM]+$', '', regex=True).astype(float) * \
df['Comments avg.'].str.extract(r'[\d\.]+([KM]+)', expand=False)
.fillna(1)
.replace(['K', 'M'], [10**3, 10**6])).astype(int))

df['Shares avg.'] = (df['Shares avg.'].replace(r'[KM]+$', '', regex=True).astype(float) * \
df['Shares avg.'].str.extract(r'[\d\.]+([KM]+)', expand=False)
.fillna(1)
.replace(['K', 'M'], [10**3, 10**6])).astype(int))
```

Gambar 7. Mengubah Tipe Data dan nilai ribuan serta nilai jutaan

Pada gambar 7 menunjukkan perubahan tipe data kolom Subscribers, Views avg., Likes avg., Comments avg., dan Shares avg., menjadi tipe data float. Nilai yang terdapat 'K' melambangkan nominal ribuan sehingga harus diubah dengan mengalikan 1000 sedangkan nilai yang terdapat 'M' melambangkan nominal juta sehingga harus diubah dengan mengalikan 1000000.

```
df.dtypes
```

```
S.no          int64
Tiktoker name object
Tiktok name   object
Subscribers   float64
Views avg.    float64
Likes avg.    float64
Comments avg. float64
Shares avg.   float64
dtype: object
```

Gambar 8. Pengecekan tipe data setiap kolom setelah dirubah

Pada gambar 8 menunjukkan hasil perubahan tipe data kolom Subscribers, Views avg., Likes avg., Comments avg., dan Shares avg. menjadi tipe data float. Hal ini menunjukkan bahwa perubahan berhasil dilakukan dan data dapat digunakan untuk perhitungan statistik kemudiannya.

| df  |      |               |               |             |            |            |               |             |
|-----|------|---------------|---------------|-------------|------------|------------|---------------|-------------|
|     | S.no | Tiktoker name | Tiktok name   | Subscribers | Views avg. | Likes avg. | Comments avg. | Shares avg. |
| 0   | 1    | jypestraykids | Stray Kids    | 13800000.0  | 6400000.0  | 2300000.0  | 50200.0       | 34200.0     |
| 1   | 2    | khaby.lame    | Khabane lame  | 149200000.0 | 17300000.0 | 2300000.0  | 15200.0       | 8700.0      |
| 2   | 3    | scarlettspam2 | scarlett      | 2100000.0   | 17900000.0 | 845800.0   | 53900.0       | 6300.0      |
| 3   | 4    | addisonre     | Addison Rae   | 88700000.0  | 22000000.0 | 906600.0   | 7600.0        | 26200.0     |
| 4   | 5    | belindatok    | Belinda       | 4800000.0   | 14200000.0 | 1500000.0  | 14500.0       | 15300.0     |
| ... | ...  | ...           | ...           | ...         | ...        | ...        | ...           | ...         |
| 995 | 996  | brendadialoy  | Brenda Dialoy | 371900.0    | 1200000.0  | 187500.0   | 626.0         | 2000.0      |
| 996 | 997  | jujuftcats    | Juju Fitcats  | 4300000.0   | 2200000.0  | 280800.0   | 380.0         | 558.0       |
| 997 | 998  | xoteam        | XO Team       | 37800000.0  | 2500000.0  | 180200.0   | 859.0         | 450.0       |
| 998 | 999  | kimsnwuo      |               | 100900.0    | 957400.0   | 226800.0   | 908.0         | 1700.0      |
| 999 | 1000 | jass_mcg      | jass_mcg      | 52200.0     | 503800.0   | 116000.0   | 8100.0        | 1800.0      |

999 rows x 8 columns

Gambar 9. Hasil dataframe

Pada gambar 9 menunjukkan hasil datafram setelah dilakukam perubahan tipe data kolom Subscribers, Views avg., Likes avg., Comments avg., dan Shares avg. Terlihat nilai yang memiliki 'K' dan 'M' juga sudah berubah sesuai nominal semestinya.

#### e. Menghapus Kolom yang Kurang Memuat Informasi

```
del df['S.no']
```

Gambar 10. Menghapus Kolom S.no



| df  |               |               |             |            |            |               |             |
|-----|---------------|---------------|-------------|------------|------------|---------------|-------------|
|     | Tiktoker name | Tiktok name   | Subscribers | Views avg. | Likes avg. | Comments avg. | Shares avg. |
| 0   | jypestraykids | Stray Kids    | 13800000.0  | 6400000.0  | 2300000.0  | 50200.0       | 34200.0     |
| 1   | khaby.lame    | Khabane lame  | 149200000.0 | 17300000.0 | 2300000.0  | 15200.0       | 8700.0      |
| 2   | scarlettsspm2 | scarlett      | 2100000.0   | 17900000.0 | 845800.0   | 53900.0       | 6300.0      |
| 3   | addisonre     | Addison Rae   | 88700000.0  | 22000000.0 | 906600.0   | 7600.0        | 26200.0     |
| 4   | belindatok    | Belinda       | 4800000.0   | 14200000.0 | 1500000.0  | 14500.0       | 15300.0     |
| ... | ...           | ...           | ...         | ...        | ...        | ...           | ...         |
| 995 | brendadialoy  | Brenda Dialoy | 371900.0    | 1200000.0  | 187500.0   | 626.0         | 2000.0      |
| 996 | jujufitcats   | Juju Fitcats  | 4300000.0   | 2200000.0  | 280800.0   | 380.0         | 558.0       |
| 997 | xoteam        | XO Team       | 37800000.0  | 2500000.0  | 180200.0   | 859.0         | 450.0       |
| 998 | kimsnwuo      |               | 100900.0    | 957400.0   | 226800.0   | 908.0         | 1700.0      |
| 999 | jass_mcg      | jass_mcg      | 52200.0     | 503800.0   | 116000.0   | 8100.0        | 1800.0      |

999 rows × 7 columns

Gambar 11. Hasil Dataframe

## 2.2 Statistika Deskriptif

### a. Ukuran Pemusatan Data

- Mean (rata – rata)

```
print(df['Subscribers'].mean())
print(df['Views avg.'].mean())
print(df['Likes avg.'].mean())
print(df['Comments avg.'].mean())
print(df['Shares avg.'].mean())
```

7090693.793793794  
2856087.4874874875  
352627.82782782783  
2537.766766766767  
3263.4464464464463

Gambar 12. Hasil Mean

- Median

```
print(df['Subscribers'].median())
print(df['Views avg.'].median())
print(df['Likes avg.'].median())
print(df['Comments avg.'].median())
print(df['Shares avg.'].median())
```

3500000.0  
2200000.0  
274900.0  
1600.0  
1800.0

Gambar 13. Hasil Median

- Mode (modus)

```
print(df['Subscribers'].mode())
print(df['Views avg.'].mode())
print(df['Likes avg.'].mode())
print(df['Comments avg.'].mode())
print(df['Shares avg.'].mode())
```

```
0    1100000.0
Name: Subscribers, dtype: float64
0    1600000.0
Name: Views avg., dtype: float64
0    1100000.0
Name: Likes avg., dtype: float64
0     1400.0
Name: Comments avg., dtype: float64
0     1300.0
Name: Shares avg., dtype: float64
```

Gambar 14. Hasil Mode

## b. Ukuran Penyebaran Data

- Standar Deviasi

```
print(df['Subscribers'].std())
print(df['Views avg.'].std())
print(df['Likes avg.'].std())
print(df['Comments avg.'].std())
print(df['Shares avg.'].std())
```

```
11609832.80580642
2216568.628674682
286269.1873073131
3692.961026187326
4780.954926047169
```

Gambar 15. Hasil Standar Deviasi

- Varians

```
print(df['Subscribers'].var())
print(df['Views avg.'].var())
print(df['Likes avg.'].var())
print(df['Comments avg.'].var())
print(df['Shares avg.'].var())
```

```
134788217778779.0
4913176485624.761
81950047601.5895
13637961.140938548
22857530.00489469
```

Gambar 16. Hasil Varians

- IQR (Interquartile Range)

```
np.percentile(df.Subscribers, [25, 50, 75])
```

```
array([1300000., 3500000., 8100000.])
```

```
Q1 = np.percentile(df.Subscribers,25)
Q3 = np.percentile(df.Subscribers,75)
IQR = Q3 - Q1
IQR
```

```
6800000.0
```

(a)

```
np.percentile(df['Views avg.'],[25, 50, 75])
```

```
array([1600000., 2200000., 3350000.])
```

```
Q1 = np.percentile(df['Views avg.'],25)
Q3 = np.percentile(df['Views avg.'],75)
IQR = Q3 - Q1
IQR
```

```
1750000.0
```

(b)

|                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <pre>np.percentile(df['Likes avg.'],[25, 50, 75])</pre> <pre>array([190000., 274900., 400750.])</pre> <pre>Q1 = np.percentile(df['Likes avg.'],25) Q3 = np.percentile(df['Likes avg.'],75) IQR = Q3 - Q1 IQR</pre> <p>210750.0</p> | <pre>np.percentile(df['Comments avg.'],[25, 50, 75])</pre> <pre>array([ 984., 1600., 2700.])</pre> <pre>Q1 = np.percentile(df['Comments avg.'],25) Q3 = np.percentile(df['Comments avg.'],75) IQR = Q3 - Q1 IQR</pre> <p>1716.0</p> |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

(c)

(d)

```
np.percentile(df['Shares avg.'],[25, 50, 75])
```

```
array([ 822., 1800., 3750.])
```

```
Q1 = np.percentile(df['Shares avg.'],25)
Q3 = np.percentile(df['Shares avg.'],75)
IQR = Q3 - Q1
IQR
```

2928.0

(e)

Gambar 17. Hasil IQR untuk a) kolom Subscribers b) kolom Views avg. c) kolom Likes avg. d) kolom Comments avg. e) kolom Shares avg.

### c. Korelasi

- Menghitung kovarian

```
df2.cov()
```

|               | Subscribers  | Views avg.   | Likes avg.   | Comments avg. | Shares avg.  |
|---------------|--------------|--------------|--------------|---------------|--------------|
| Subscribers   | 1.380190e+15 | 1.304547e+14 | 1.051421e+13 | 1.504512e+11  | 3.810474e+10 |
| Views avg.    | 1.304547e+14 | 3.941028e+13 | 2.498178e+12 | 2.430125e+10  | 4.335146e+10 |
| Likes avg.    | 1.051421e+13 | 2.498178e+12 | 4.400720e+11 | 2.833145e+09  | 1.552705e+09 |
| Comments avg. | 1.504512e+11 | 2.430125e+10 | 2.833145e+09 | 4.046250e+07  | 1.060692e+07 |
| Shares avg.   | 3.810474e+10 | 4.335146e+10 | 1.552705e+09 | 1.060692e+07  | 6.687138e+07 |

Gambar 18. Hasil Kovarian

- Menghitung Korelasi

```
df2.corr()
```

|               | Subscribers | Views avg. | Likes avg. | Comments avg. | Shares avg. |
|---------------|-------------|------------|------------|---------------|-------------|
| Subscribers   | 1.000000    | 0.559352   | 0.426624   | 0.636649      | 0.125427    |
| Views avg.    | 0.559352    | 1.000000   | 0.599869   | 0.608552      | 0.844460    |
| Likes avg.    | 0.426624    | 0.599869   | 1.000000   | 0.671399      | 0.286225    |
| Comments avg. | 0.636649    | 0.608552   | 0.671399   | 1.000000      | 0.203912    |
| Shares avg.   | 0.125427    | 0.844460   | 0.286225   | 0.203912      | 1.000000    |

Gambar 19. Hasil Korelasi

Pada gambar 19 menunjukkan tingkat korelasi antar Subscribers, Views avg., Likes avg., Comments avg., Shares avg. Terlihat korelasi antar kolom adalah positif sehingga setiap kolom saling mempengaruhi secara positif.

#### d. Data Visualization

```
subs = df.sort_values(by = ["Subscribers"], ascending = False)
subs
```

|     | Tiktok name         | Tiktok name       | Subscribers | Views avg. | Likes avg. | Comments avg. | Shares avg. |
|-----|---------------------|-------------------|-------------|------------|------------|---------------|-------------|
| 1   | khaby.lame          | Khaby Lame        | 149200000.0 | 17300000.0 | 2300000.0  | 15200.0       | 8700.0      |
| 10  | charlidamelio       | Charli D'Amelio   | 146200000.0 | 15200000.0 | 959400.0   | 20800.0       | 4600.0      |
| 74  | bellapoarch         | Bella Poarch      | 91400000.0  | 7400000.0  | 677300.0   | 4900.0        | 1100.0      |
| 3   | addisonre           | Addison Rae       | 88700000.0  | 22000000.0 | 906600.0   | 7600.0        | 26200.0     |
| 19  | kimberlyloaiza      | Kimberly Loaiza   | 67200000.0  | 8800000.0  | 1200000.0  | 12100.0       | 4300.0      |
| ... | ...                 | ...               | ...         | ...        | ...        | ...           | ...         |
| 206 | iamprincey          | Priscilla         | 13900.0     | 2300000.0  | 522900.0   | 1200.0        | 4200.0      |
| 307 | meganfitton6        | user3892932795857 | 8700.0      | 2000000.0  | 315300.0   | 1600.0        | 4400.0      |
| 190 | wilmah10            | wilmah10          | 6300.0      | 2900000.0  | 208200.0   | 2700.0        | 9900.0      |
| 244 | strange_56uuuuy     | alkihorus?        | 5900.0      | 1500000.0  | 276600.0   | 2200.0        | 10800.0     |
| 776 | soapimymouthonemore | tedwards          | 5100.0      | 943300.0   | 189600.0   | 1900.0        | 3100.0      |

Gambar 20. Mengurutkan data berdasarkan jumlah subscribers tertinggi

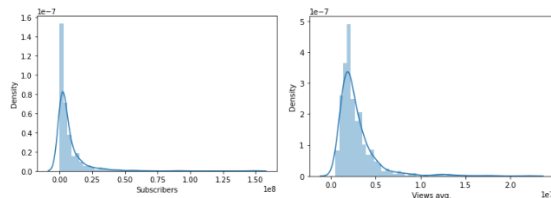
```
df1 = subs.iloc[0:9]
df1
```

|     | Tiktok name    | Tiktok name     | Subscribers | Views avg. | Likes avg. | Comments avg. | Shares avg. |
|-----|----------------|-----------------|-------------|------------|------------|---------------|-------------|
| 1   | khaby.lame     | Khaby Lame      | 149200000.0 | 17300000.0 | 2300000.0  | 15200.0       | 8700.0      |
| 10  | charlidamelio  | Charli D'Amelio | 146200000.0 | 15200000.0 | 959400.0   | 20800.0       | 4600.0      |
| 74  | bellapoarch    | Bella Poarch    | 91400000.0  | 7400000.0  | 677300.0   | 4900.0        | 1100.0      |
| 3   | addisonre      | Addison Rae     | 88700000.0  | 22000000.0 | 906600.0   | 7600.0        | 26200.0     |
| 19  | kimberlyloaiza | Kimberly Loaiza | 67200000.0  | 8800000.0  | 1200000.0  | 12100.0       | 4300.0      |
| 326 | cznburak       | cznburak        | 63600000.0  | 4600000.0  | 228400.0   | 2300.0        | 978.0       |
| 16  | domelipa       | domelipa        | 58100000.0  | 10400000.0 | 1500000.0  | 9100.0        | 3500.0      |
| 630 | dixiedamelio   | dixie           | 57400000.0  | 2400000.0  | 220500.0   | 1900.0        | 424.0       |
| 6   | therock        | The Rock        | 55400000.0  | 13100000.0 | 1500000.0  | 14300.0       | 11900.0     |

Gambar 21. Mengambil 10 data teratas dengan subscribers tertinggi

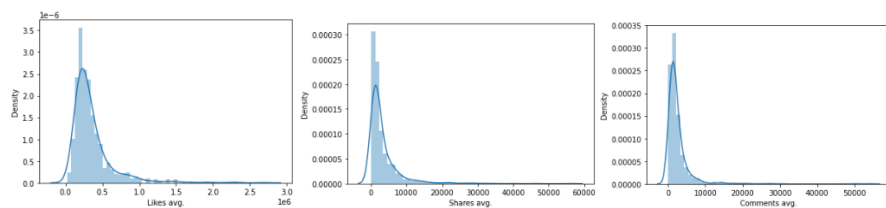
Pada gambar 20 dan 21 menunjukkan proses pengambilan 10 data teratas berdasarkan subscribers tertinggi. Hal ini adalah proses pengalambihan sampel dari populasi sehingga adanya fokus analisis utama untuk data visualisasi. Berikut adalah jenis – jenis data visualisasi:

- Distplot



(a)

(b)



(c)

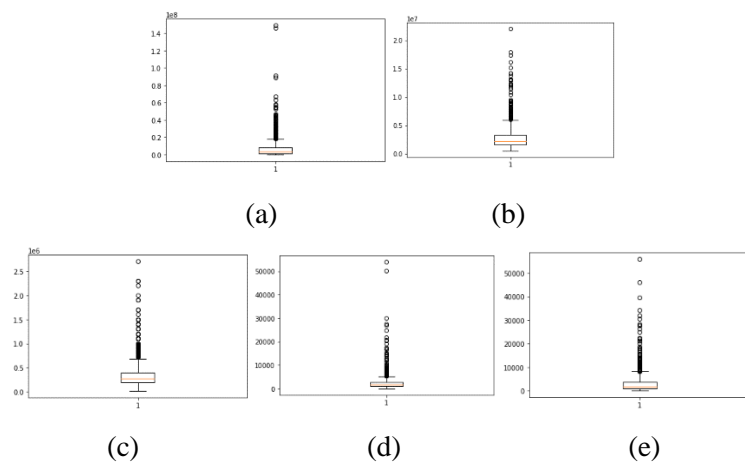
(d)

(e)

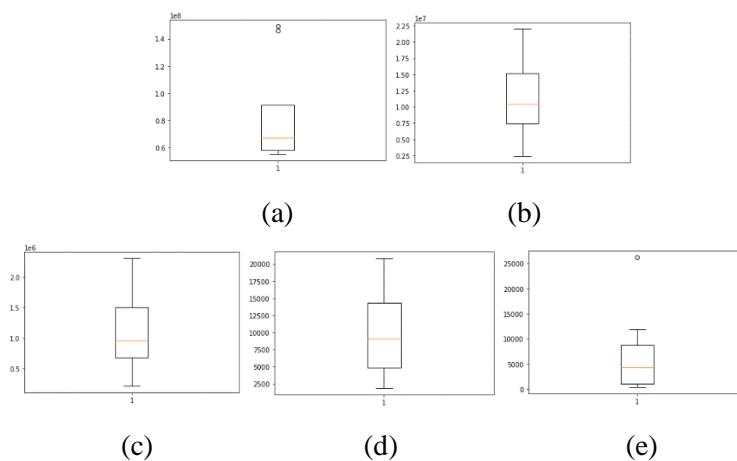
Gambar 22. Visualiasi penyebaran data dengan distplot a) kolom Subscribers b) kolom Views avg. c) kolom Likes avg. d) kolom Shares avg. e) kolom Comments avg.

Pada gambar 22 menunjukkan visualisasi penyebaran data dengan distplot. Data yang digunakan masih berupa data awal untuk melihat seberapa menyebar data yang digunakan dan ternyata ditinjau dari subscribers, Views avg., Likes avg., Shares avg., dan Comments avg., data berupa distribusi miring ke kanan karena setiap puncak berada di sisi kanan dan semakin menurun di sisi kiri.

- Boxplot



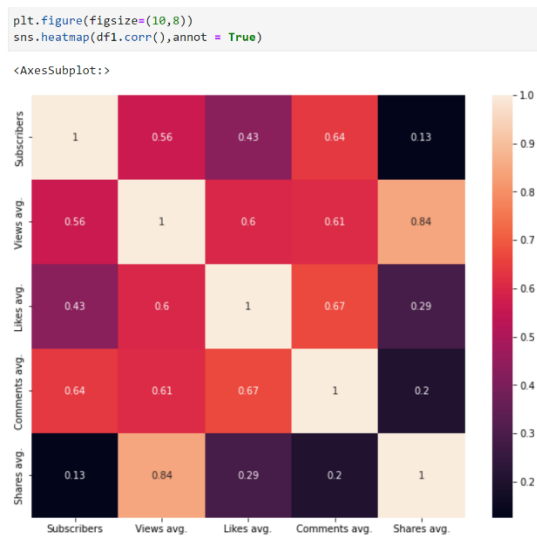
Gambar 23. Visualisasi penyebaran data awal dengan boxplot a) Kolom Subscribers b) kolom Views avg. c) kolom Likes avg. d) kolom Shares avg. e) kolom comments avg.



Gambar 24. Visualisasi penyebaran 10 data teratas dengan boxplot a) Kolom Subscribers b) kolom Views avg. c) kolom Likes avg. d) kolom Shares avg. e) kolom comments avg.

Perbandingan gambar 23 dan 24 menjadi salah satu alasan untuk menggunakan 10 data teratas sebagai sampel analisis. Terlihat pada gambar 23, banyak outlier disetiap kolom dan untuk meminimalisir outlier diambil 10 data teratas menurut subscribers dengan penyebarannya terlihat pada gambar 24.

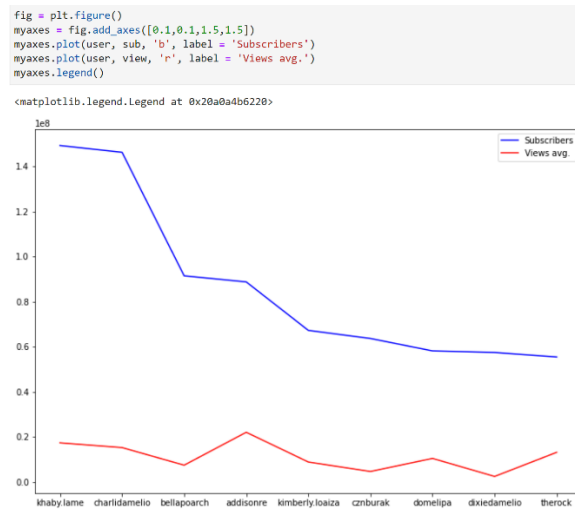
- Heatmap



Gambar 25. Visualisasi korelasi dengan heatmap

Jumlah *subscribers* dengan *view*, *like*, *comment*, dan *share* memiliki korelasi positif yang artinya setiap jumlah *subscribers* naik, akan ada kenaikan pada *view*, *like*, *comment*, dan *share*. Korelasi paling tinggi adalah antara *subscribers* dengan *comment* sebesar 0.64 dan semakin besar *comment*, maka juga akan mempengaruhi *like*, *comment*, dan *share* karena tingkat korelasinya yang cukup tinggi. Jumlah *subscribers* memang mempengaruhi tingkat *engagement* tetapi, perlu diperhatikan berapa banyak jumlah *comment* agar jumlah *view*, *like*, *share* ikut menaik dan semakin lebih banyak orang untuk mengenal produk yang dipasarkan.

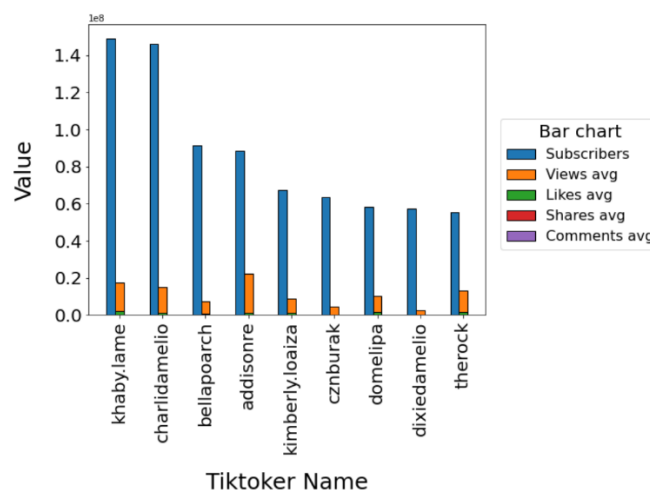
- Line Chart



Gambar 26. Visualisasi antara Subscribers dan Views avg.

Pada gambar 26 terlihat perbandingan *subscribers* dengan *views avg.* Disini terlihat meskipun, jumlah *subscribers* itu tinggi, belum tentu *views* yang didapatkan juga tinggi dan begitupula sebaliknya. Sehingga pentingnya untuk melihat jumlah *views* sebelum menentukan *endorsement* terhadap influencer. Jumlah *subscribers* yang tinggi bukan menjadi acuan untuk mendapatkan *views* tertinggi.

- Bar Chart

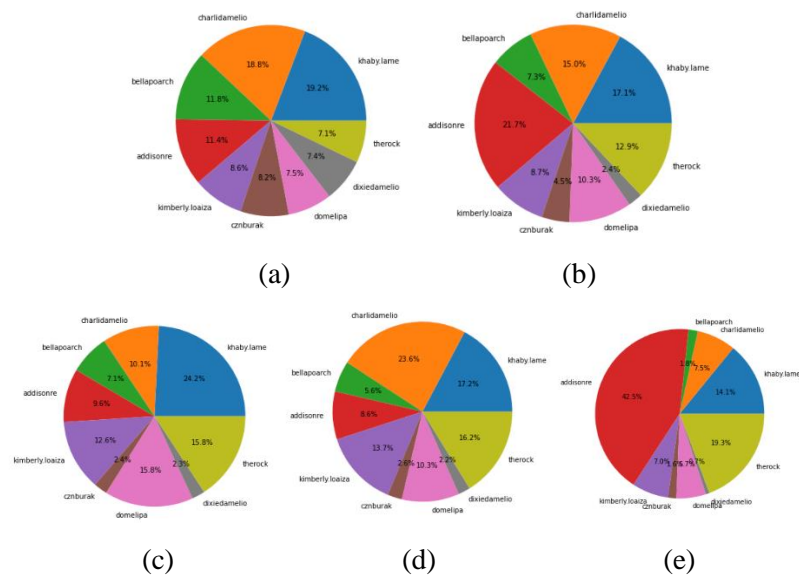


Gambar 27. Visualisasi antara Subscribers dan total engagement

Pada gambar 27 terlihat perbandingan subscribers dengan total engagement. Total *engagement* disini adalah total keseluruhan dari *Views*

*avg.*, *Likes avg.*, *Comments avg.*, dan *shares avg.* Disini terlihat meskipun, jumlah *subscribers* itu tinggi, belum tentu tingkat *engagement* yang didapatkan juga tinggi dan begitupula sebaliknya. Sehingga pentingnya untuk melihat tingkat *engagement* sebelum menentukan *endorsement* terhadap *influencer*.

- Pie chart

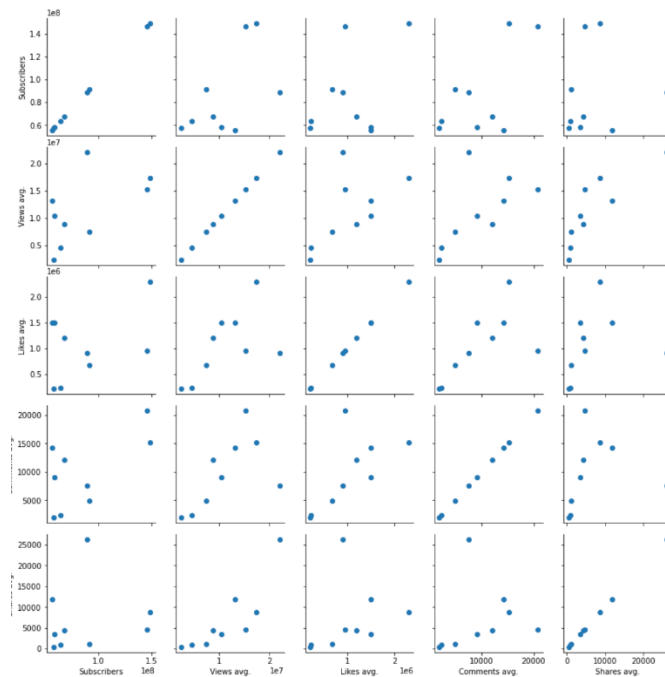


Gambar 28. Visualisasi dengan pie chart a) Kolom Subscribers b) kolom Views avg. c) kolom Likes avg. d) kolom Shares avg. e) kolom comments avg.

Pada gambar 28 terlihat perbandingan *subscribers*, *views avg*, *likes avg*, *shares avg*, dan *comments avg*. Visualisasi ini menunjukkan bahwa tiktokers yang memiliki jumlah *subscribers* tertinggi, bukan berarti memiliki *views*, *likes*, *shares*, dan *comments* yang tertinggi, begitupula sebaliknya. Perlunya ada pertimbangan lebih lanjut terkait tingkat *engagement* sebelum melakukan *endorsement*.



- Scatter Plot



Gambar 29. Visualisasi scatter plot secara keseluruhan

Pada gambar 29 menunjukkan hubungan antar setiap kolom. Terlihat pada bagian views sangat mempengaruhi views, comments, dan shares serta selalu menunjukkan korelasi positif. Sebelumnya, pada gambar 25 yaitu visualiasi heatmap menunjukkan views memang berpengaruh positif dengan views, comments, dan shares. Serta jumlah subscribers berpengaruh 0.56 terhadap views meskipun, ada beberapa pertimbangan lebih lanjut.

## 2.3 Statistika Inferensial

### a. Uji Z

- Uji Hipotesis 1

Data yang digunakan pada uji Z adalah 60 data teratas berdasarkan Subscribers.

```
import pandas as pd
from statsmodels.stats import weightstats as stests

ztest, p = stests.ztest(df2['Views avg.'], value=5700000)
print("Pvalue:", str(p))
print("Zvalue:", str(ztest))

if p < 0.05:
    print("tolak H0")
else:
    print("gagal tolak H0")

Pvalue: 0.9759547039712174
Zvalue: 0.03014087251775373
gagal tolak H0
```

Gambar 30. Hasil Uji Hipotesis 1 dengan uji Z

$H_0 = \text{rata - rata views avg.} = 5700000$

$H_1 = \text{rata - rata views avg.} \neq 5700000$

Karena hasil yang didapatkan adalah gagal tolak  $H_0$ . Maka hipotesis bahwa rata – rata views avg. adalah 5700000 diterima.

- Uji Hipotesis 2

Data yang digunakan pada uji Z hipotesis 2 adalah 60 data teratas berdasarkan Subscribers. Yang kemudian dipisahkan menjadi 30 data – 30 data secara berurutan.

```
tval, pval = stats.ztest(df3['Views avg.'], df4['Views avg.'])

if pval < 0.05:
    print("Tolak H0")
else:
    print("gagal tolak H0")

print("Pvalue:", str(pval))
print("Tvalue:", str(tval))

Tolak H0
Pvalue: 0.007874277962319844
Tvalue: 2.6574135651446382
```

Gambar 31. Hasil Uji Hipotesis 2 dengan uji Z

$H_0 = \mu \text{ Views avg. 1} = \mu \text{ Views avg. 2}$

$H_1 = \mu \text{ Views avg. 1} \neq \mu \text{ Views avg. 2}$

Karena hasil yang didapatkan adalah tolak  $H_0$ . Maka hipotesis bahwa  $\mu \text{ Views avg. 1} = \mu \text{ Views avg. 2}$  ditolak.

b. Uji T

- Uji Hipotesis 1

Data yang digunakan pada uji F hipotesis 1 adalah 2 dataframe berbeda yang setiapnya terdiri dari 10 data mengenai jumlah comments secara random dari data awal yang sudah diurutkan berdasarkan jumlah subscribers

```
tval, pval = stats.ttest_ind(df3['Comments avg.'], df4['Comments avg.'])

print("Pvalue:", str(pval))
print("Tvalue:", str(tval))

if pval/2 < 0.05:
    print("Tolak H0")
else:
    print("gagal tolak H0")

Pvalue: 0.0013467431451231312
Tvalue: 3.7880625433317237
Tolak H0
```

Gambar 32. Hasil Uji Hipotesis 1 dengan uji T

$$H_0 = \mu_{df3} \leq \mu_{df4}$$

$$H_1 = \mu_{df3} > \mu_{df4}$$

Karena hasil yang didapatkan adalah gagal tolak  $H_0$ . Maka hipotesis bahwa  $\mu_{df3} \leq \mu_{df4}$  ditolak. Hal ini berarti menunjukkan bahwa nilai varians pada dataframe3 lebih tinggi daripada nilai varians dataframe4.

- Uji Hipotesis 2

Data yang digunakan pada uji F hipotesis 1 adalah 2 dataframe berbeda yang setiapnya terdiri dari 10 data mengenai jumlah comments secara random dari data awal yang sudah diurutkan berdasarkan jumlah subscribers

```
# Uji beda varians
group1 = df3['Comments avg.']
group2 = df4['Comments avg.']

# converting the list to array
x = np.array(group1)
y = np.array(group2)

def t_test(group1, group2):
    t = np.var(group1, ddof=1)/np.var(group2, ddof=1)
    num = x.size-1
    den = y.size-1
    p_value = 1 - scipy.stats.t.cdf(t, num, den)
    return t, p_value

# perform T-test
t, pvalue = t_test(x, y)

print("Pvalue:", str(pvalue))
print("Tvalue:", str(t))

if pvalue < 0.05:
    print("Tolak H0")
else:
    print("gagal tolak H0")

Pvalue: 1.1712852989795482e-13
Tvalue: 74.2817810281781
Tolak H0
```

Gambar 33. Hasil Uji Hipotesis 2 dengan uji T

$$H_0 = \sigma_1 = \sigma_2$$

$$H_1 = \sigma_1 \neq \sigma_2$$

Karena hasil yang didapatkan adalah tolak  $H_0$ . Maka hipotesis bahwa  $\sigma_1 = \sigma_2$  ditolak. Berarti ada perbedaan varians antara dua sampel yang diujikan.

### c. Uji F

- Uji Hipotesis 1

Data yang digunakan pada uji F hipotesis 1 adalah 4 dataframe berbeda yang setiapnya terdiri dari 5 data mengenai jumlah share secara random dari data awal yang sudah diurutkan berdasarkan jumlah subscribers.

```

fvalue ,p = scipy.stats.f_oneway(dataframe1['Shares avg.'],
                                dataframe2['Shares avg.'],
                                dataframe3['Shares avg.'],
                                dataframe4['Shares avg.'])

print("Pvalue:", str(p))
print("Fvalue:", str(fvalue))

if p < 0.05:
    print("Tolak H0")
else:
    print("gagal tolak H0")

# H0 :  $\mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$  (Ini menyiratkan bahwa rata-rata semua populasi adalah sama)
# H1 : Ini menyatakan bahwa akan ada setidaknya satu rata-rata populasi yang berbeda dari yang lain

Pvalue: 0.5747466558947758
Fvalue: 0.6839460423974559
gagal tolak H0

```

Gambar 34. Hasil Uji Hipotesis 1 dengan Uji F

$$H_0 = \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_1 = \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$$

Karena hasil yang didapatkan adalah gagal tolak  $H_0$ . Maka hipotesis bahwa  $\mu_1 = \mu_2 = \mu_3 = \mu_4$  diterima.

- Uji Hipotesis 2

Data yang digunakan pada uji F hipotesis 2 adalah 5 data teratas pada dataframe awal bagian kolom Subscribers dan 5 data terbawah pada dataframe awal bagian kolom Subscribers.

```

# Create data
group1 = dfhead['Subscribers']
group2 = dfetail['Subscribers']

# converting the list to array
x = np.array(group1)
y = np.array(group2)

def f_test(group1, group2):
    f = np.var(group1, ddof=1)/np.var(group2, ddof=1)
    num = x.size-1
    den = y.size-1
    p_value = 1-scipy.stats.f.cdf(f, num, den)
    return f, p_value

# perform f-test
fval,p = f_test(x, y)

print("Pvalue:", str(p))
print("Zvalue:", str(fval))

if p < 0.05:
    print("tolak H0")
else:
    print("gagal tolak H0")

Pvalue: 0.010382459250767195
Zvalue: 15.654810093687447
tolak H0

```

Gambar 35. Hasil Uji Hipotesis 2 dengan Uji F

$$H_0 = \sigma_1 = \sigma_2$$

$$H_1 = \sigma_1 \neq \sigma_2$$

Karena hasil yang didapatkan adalah tolak  $H_0$ . Maka hipotesis bahwa  $\sigma_1 = \sigma_2$  ditolak. Berarti ada perbedaan varians antara dua sampel yang diujikan.

### 3. BAB III: KESIMPULAN

Salah satu strategi marketing produk yang paling sering digunakan adalah melakukan *endorsement* terhadap *influencer* di berbagai *platform* termasuk tiktok. Jumlah *subscribers* memang menjadi pengaruh penting dalam pemilihan *tiktokers* yang tepat karena memilih korelasi positif dengan tingkat *engagement*. Setiap jenis *engagement* juga saling mempengaruhi satu sama lain secara positif. Namun, perlu ditinjau lebih lanjut bahwa jumlah *subscribers* yang tinggi bukan menjadi indikator tingkat *engagement* yang tinggi maupun sebaliknya. Masih banyak berbagai faktor lain yang mempengaruhi sehingga dalam pemilihan *tiktokers* yang tepat, masih harus dilakukan analisis mendalam terkait tingkat *engagement* yang didapatkan oleh *tiktokers* tersebut.

#### 4. LAMPIRAN

##### 4.1 Pembagian Tugas

| Nama                    | NPM         | Jabatan | Tugas                                                                                                                              |
|-------------------------|-------------|---------|------------------------------------------------------------------------------------------------------------------------------------|
| Angela Lisanthoni       | 21083010032 | Ketua   | Mencari Dataset, Melakukan Data Cleansing, Pembuatan coding untuk melakukan hipotesis uji Z, Penyusunan Laporan, Pembuatan PPT     |
| Dendy Arizki Kuswardana | 21083010006 | Anggota | Pembuatan coding untuk melakukan hipotesis uji T                                                                                   |
| Naomi Dwi Anggraini     | 21083010010 | Anggota | Pembuatan coding untuk mencari mean, median, mode, sum, dan count                                                                  |
| Alya Setya Paramita     | 21083010046 | Anggota | Pembuatan coding Data Visualization yang terdiri dari line chart, bar chart, boxplot, piechart, scatter plot, heatmap, dan displot |
| Muhimmatul Arofah       | 21083010055 | Anggota | Pembuatan coding Data Visualization yang terdiri dari line chart, bar chart, boxplot, piechart, scatter plot, heatmap, dan displot |
| Divayanti Febri Sakina  | 21083010099 | Anggota | Pembuatan coding untuk standar deviasi, varians, dan IQR.                                                                          |
| Yayang Dimas Saputra    | 21083010102 | Anggota | Pembuatan coding untuk mencari kovarian dan korelasi                                                                               |
| Zulfaz Refie Ababil     | 21083010122 | Anggota | Pembuatan coding untuk melakukan hipotesis uji F                                                                                   |