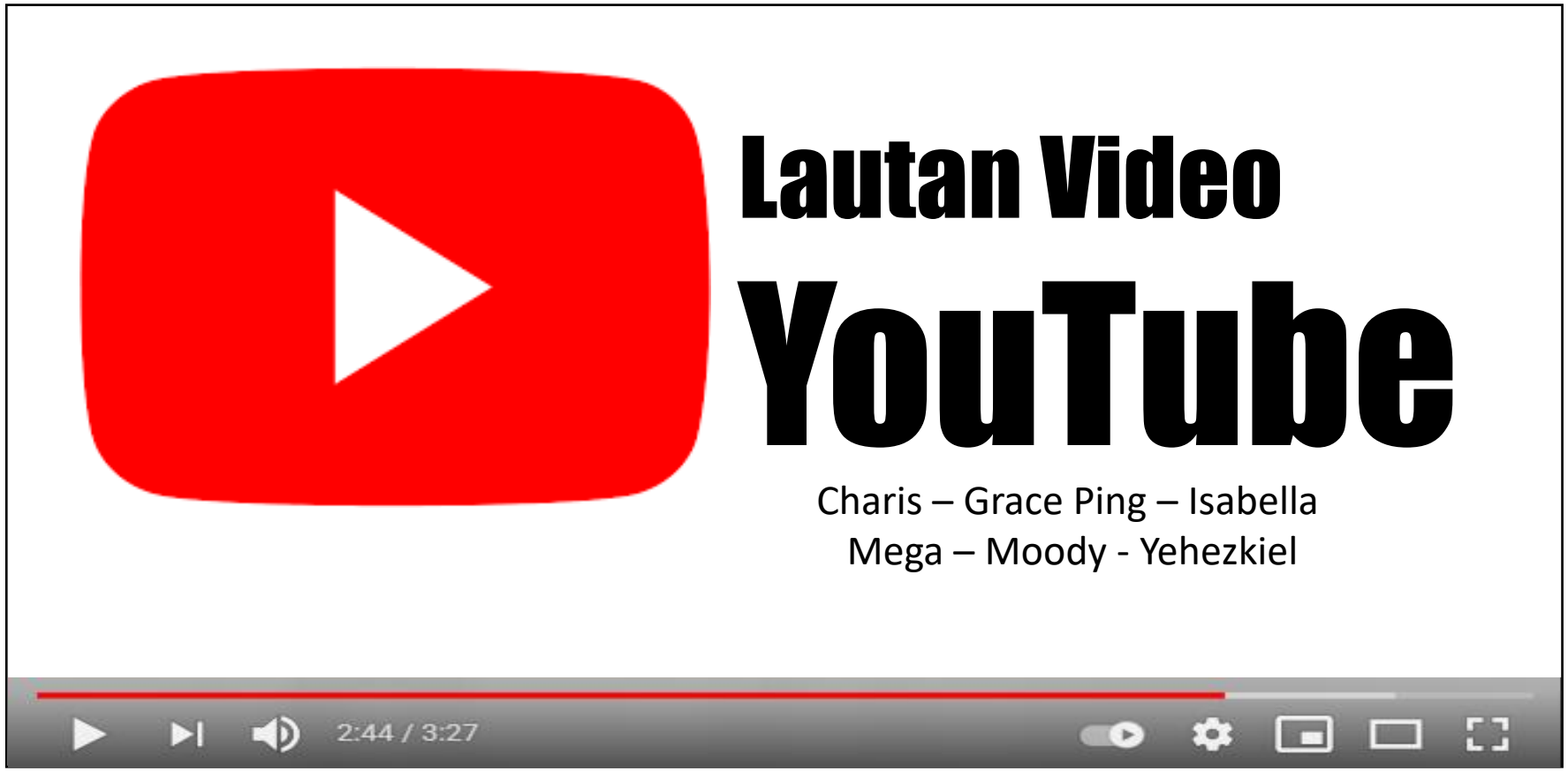
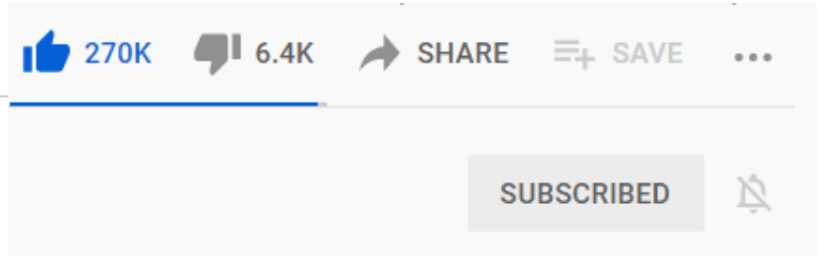


Presentasi Proyek 1A

32 views • Apr 28, 2021



KELOMPOK 1
6 subscribers

All | Contents

LATAR BELAKANG MASALAH

Mengapa dari 800 juta video Youtube, hanya sedikit yang viral?

Kualitas video ditentukan dengan 3 parameter:

TUJUAN PROYEK

Membuat rangkuman dan visualisasi data yang praktis dan nyaman dilihat

Melihat seberapa terpencilnya video viral dibanding video biasa

Menemukan informasi menarik dari data

PERALATAN STATISTIKA

- Data: dataset video Youtube yang beredar di Amerika Serikat

	views	likes	dislikes
0	748374	57527	2966
1	2418783	97185	6146
2	3191434	146033	5339
3	343168	10172	666
4	2095731	132235	1989
...	...	...	...
40944	1685609	38160	1385
40945	1064798	60008	382
40946	1066451	48068	1032
40947	5660813	192957	2846
40948	10306119	357079	212976

40949 rows × 3 columns

- Library Python:

PEMBAGIAN TUGAS

CHARIS HULU
Analisa Laporan

MOODY ASYER
Ukuran Pusat Data, Variasi, dan Lokasi

YEHEZKIEL
Histogram

MEGA dan GRACE PING ING
Boxplot

ISABELLA
Power Point

PEMROGRAMAN: UKURAN PUSAT DATA DAN VARIASI

1 RATA-RATA

```
X.mean()
```

```
views       2.360785e+06
likes       7.426670e+04
dislikes    3.711401e+03
dtype: float64
```

2 MEDIAN

```
X.median()
```

```
views       681861.0
likes        18091.0
dislikes       631.0
dtype: float64
```

3 MODUS

```
[9]: X['views'].value_counts(bins=10).sort_values(ascending=False).iloc[0]
Out[9]: 40366

In [20]: X['views'].value_counts(bins=10).sort_values(ascending=False)
Out[20]: (-224662.375, 22521686.4]     40366
         (22521686.4, 45042823.8]        368
         (45042823.8, 67563961.2]        127
```

```
In [10]: X['likes'].value_counts(bins=10).sort_values(ascending=False).iloc[0]
Out[10]: 40072

In [21]: X['likes'].value_counts(bins=10).sort_values(ascending=False)
Out[21]: (-5613.828, 561382.7]     40072
         (561382.7, 1122765.4]       566
         (1122765.4, 1684148.1]      141
```

```
In [16]: X['dislikes'].value_counts(bins=10).sort_values(ascending=False).iloc[0]
Out[16]: 40866

In [23]: X['dislikes'].value_counts(bins=10).sort_values(ascending=False)
Out[23]: (-1674.421, 167442.0]     40866
         (167442.0, 334884.0]         59
         (334884.0, 502326.0]         10
```

4 STDEV

```
[ ] X.std()
```

```
views       7.394114e+06
likes       2.288853e+05
dislikes    2.902971e+04
dtype: float64
```

PEMROGRAMAN: UKURAN PUSAT DATA DAN VARIASI

	MEAN	MEDIAN	MODUS	STANDAR DEVIASI
	2.360.785	681.861	[0, 22.521.686]	7.394.114
	74.266	18.091	[0, 561.382]	228.885
	3711	631	[0, 167.442]	29.029

PEMROGRAMAN: UKURAN LOKASI

1 Q1, Q2, Q3

```
[ ]  X.quantile([0.25,0.5,0.75])
```

	views	likes	dislikes
0.25	242329.0	5424.0	202.0
0.50	681861.0	18091.0	631.0
0.75	1823157.0	55417.0	1938.0

2 INTERKUARTIL

```
iqr = X.quantile(0.75) - X.quantile(0.25)
iqr
```

```
views       1580828.0
likes         49993.0
dislikes       1736.0
dtype: float64
```

3 BATAS ATAS DAN BAWAH

```
batas_atas = X.quantile(0.75) + (1.5*iqr)
batas_atas
```

```
views       4194399.0
likes        130406.5
dislikes       4542.0
dtype: float64
```

```
batas_bawah = X.quantile(0.25) - (1.5*iqr)
batas_bawah
```

```
views      -2128913.0
likes        -69565.5
dislikes      -2402.0
dtype: float64
```

4 JUMLAH PENCILAN

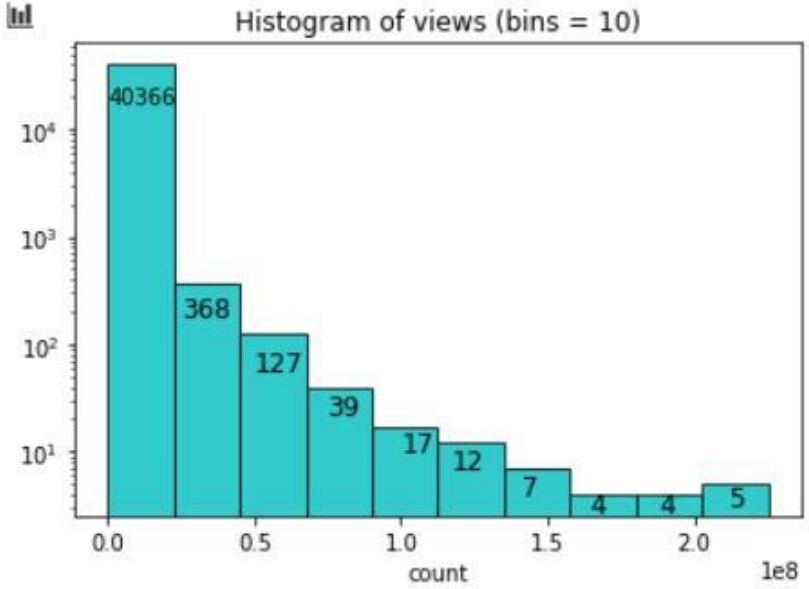
```
pencilan = ((X < batas_bawah) | (X > batas_atas)).sum()
pencilan
```

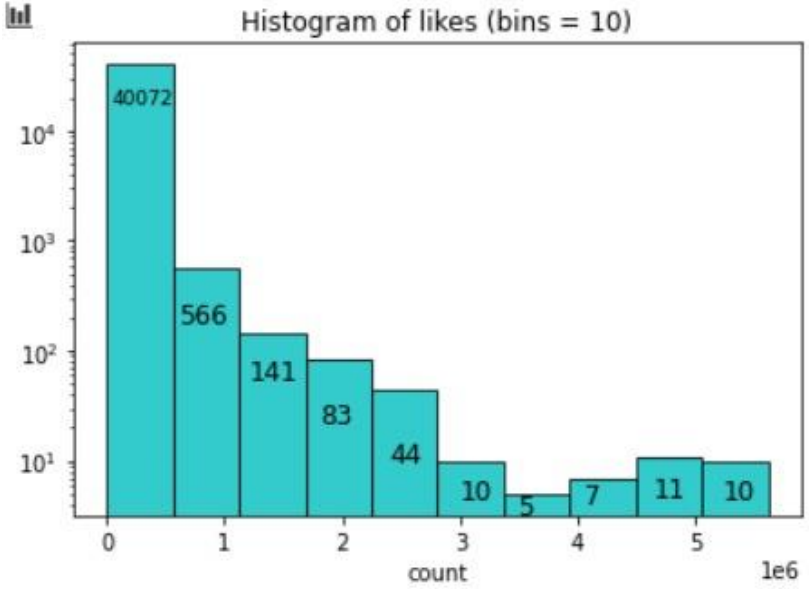
```
views       4499
likes       5136
dislikes    5288
dtype: int64
```

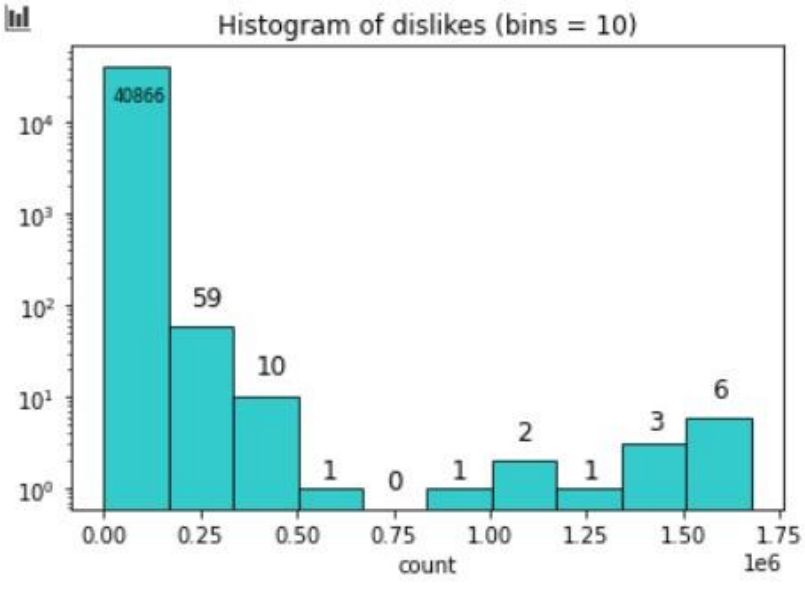
PEMROGRAMAN: UKURAN LOKASI

	KUARTIL 2	KUARTIL 3	IQR	BATAS BAWAH	BATAS ATAS	JUMLAH PENCILAN
	242.329	1.823.157	1.580.828	-2.128.913	4.194.399	4499
	5.242	55.417	49.993	-69.565,5	130.406,5	5136
	202	1.938	1.736	-2.402	4.542	5288

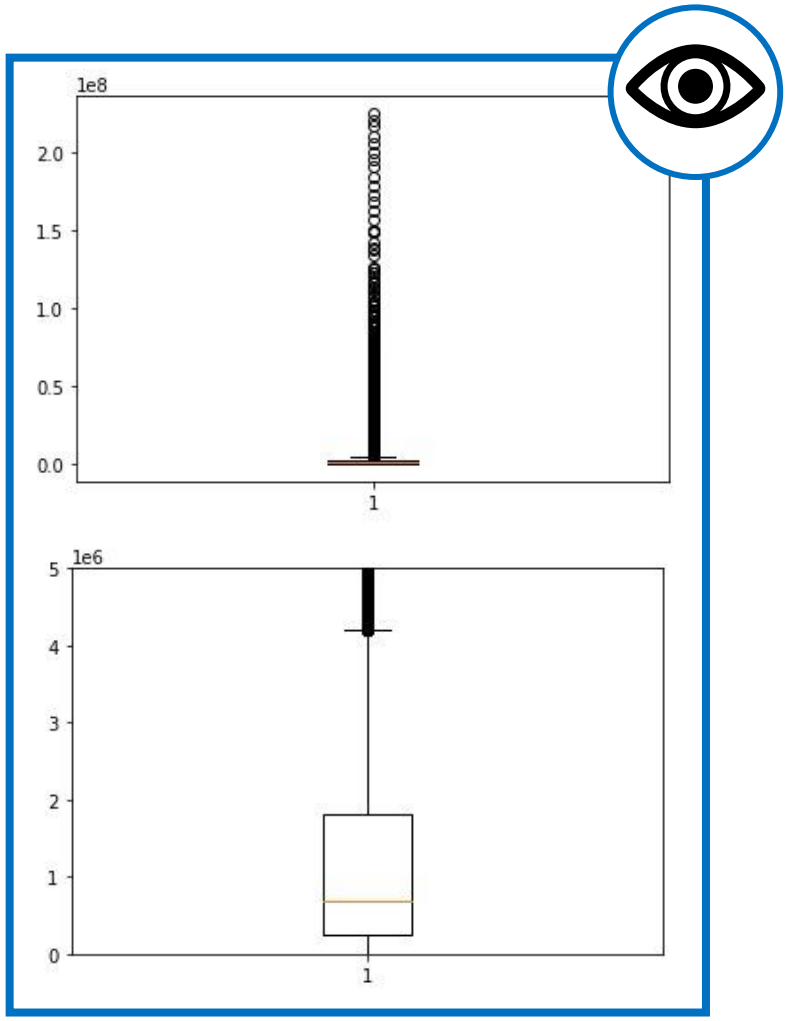
PEMROGRAMAN: HISTOGRAM

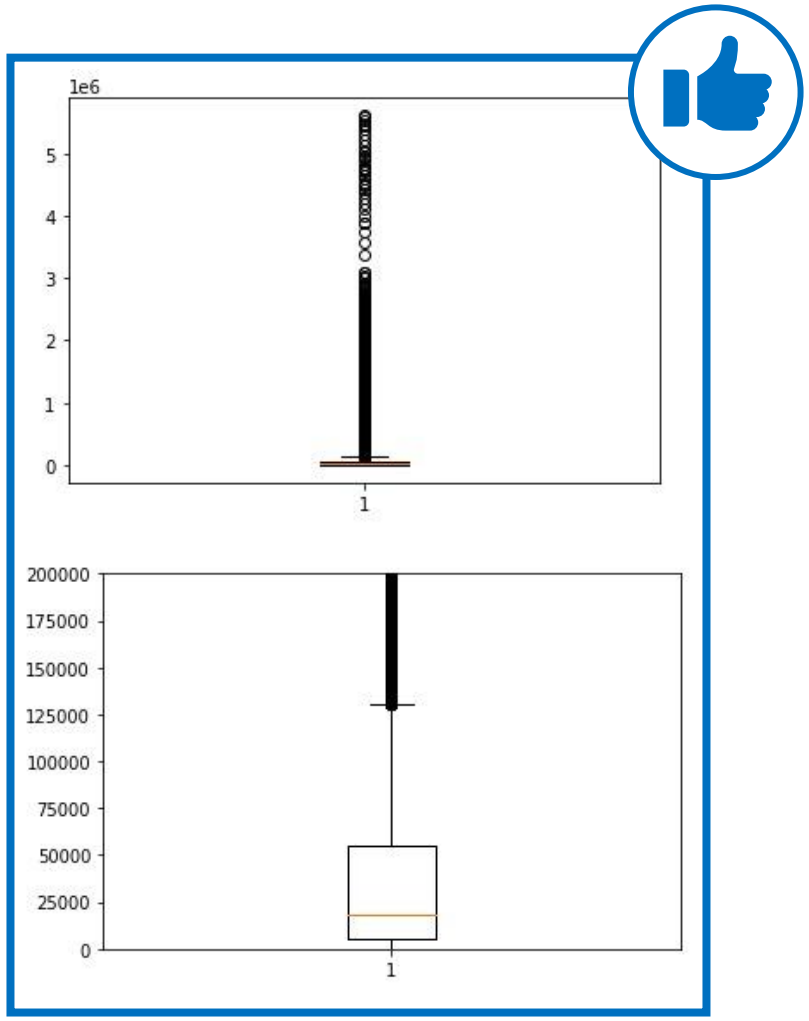


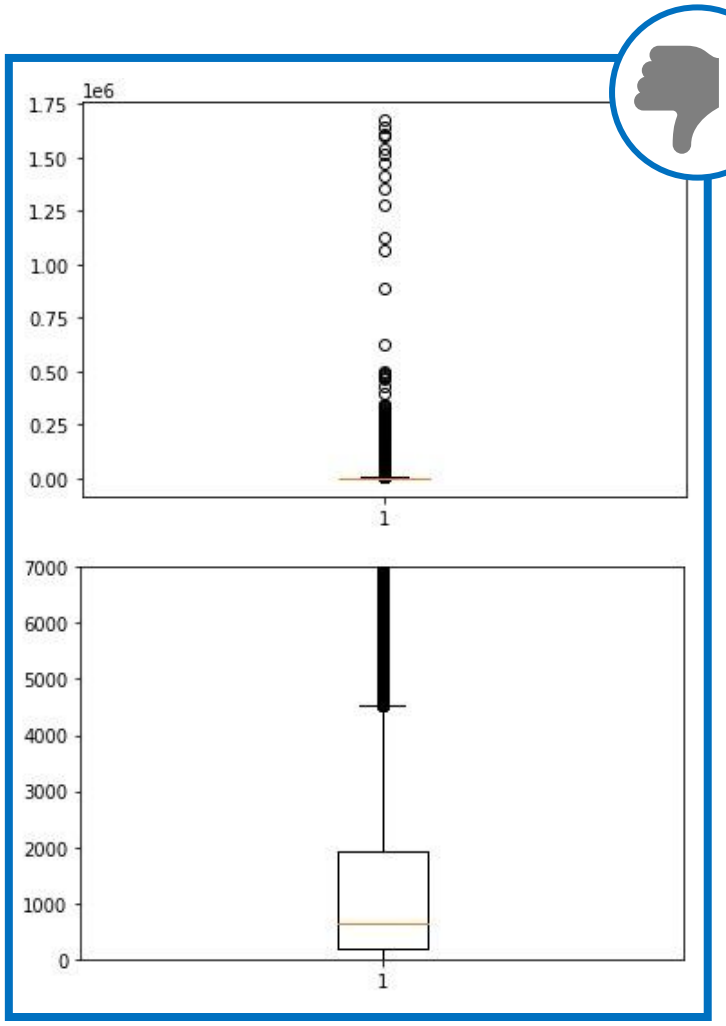




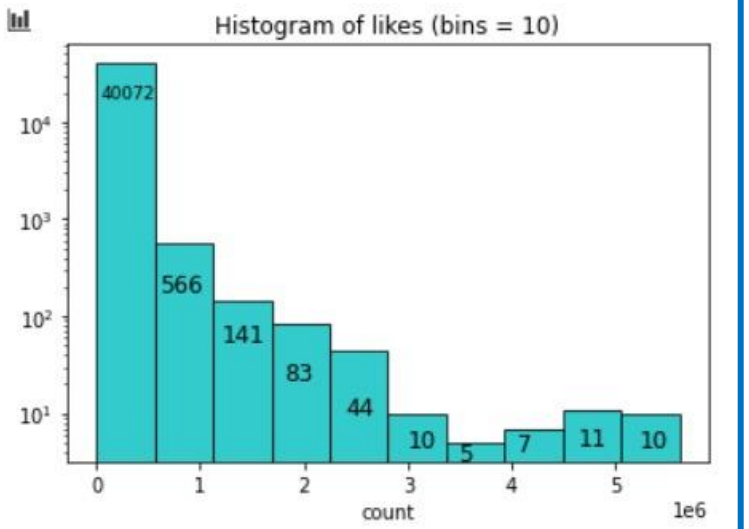
PEMROGRAMAN: BOXPLOT





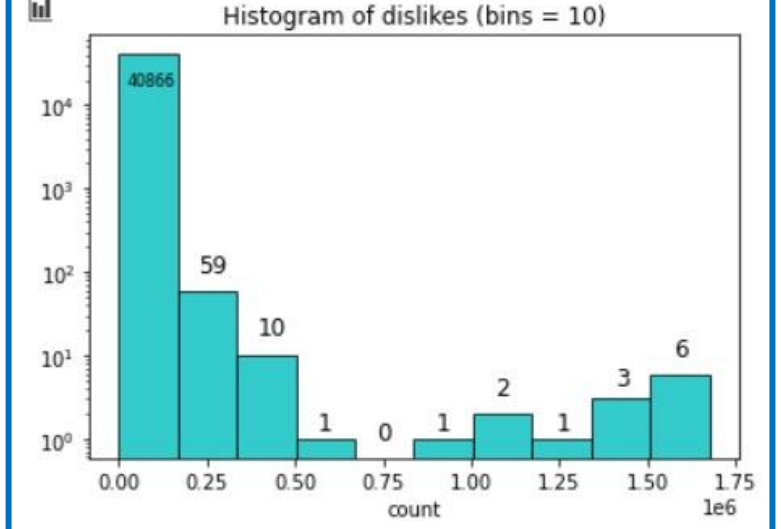


LAPORAN: FAKTA MENARIK




```
(-5613.828, 561382.7]      40072
(561382.7, 1122765.4]        566
(1122765.4, 1684148.1]       141
(1684148.1, 2245530.8]        83
(2245530.8, 2806913.5]        44
(2806913.5, 3368296.2]        10
(3368296.2, 3929678.9]         5
(3929678.9, 4491061.6]         7
(4491061.6, 5052444.3]        11
(5052444.3, 5613827.0]        10
Name: likes, dtype: int64
```

±1000 video




```
(-1674.421, 167442.0]      40866
(167442.0, 334884.0]          59
(334884.0, 502326.0]          10
(502326.0, 669768.0]           1
(669768.0, 837210.0]           0
(837210.0, 1004652.0]          1
(1004652.0, 1172094.0]         2
(1172094.0, 1339536.0]         1
(1339536.0, 1506978.0]         3
(1506978.0, 1674420.0]         6
Name: dislikes, dtype: int64
```

±100 video

Video Youtube cenderung memperoleh likes daripada dislikes

menunjukkan

Algoritma Filter Bubble: hanya menayangkan video yang disukai penonton

LAPORAN: TERPENCILNYA VIDEO DENGAN VIEWS > 100 JT

```
more_than_100M_views = len(X[X["views"] >= 100_000_000])
more_than_100M_views
```

```
40
```

```
all_videos = len(X)
all_videos
```

```
40949
```

Dari 40.949 video, hanya 40 yang memiliki views > 100 juta, atau dapat dikatakan

hanya 0.098%

```
percentage_100M_views = 100 * more_than_100M_views/all_videos
print("Persentase video dengan views lebih dari 100 JT = %.3f" %(percentage_100M_views))
```

```
Persentase video dengan views lebih dari 100 JT = 0.098
```

LAPORAN: LIKES MINIMUM UNTUK VIDEO TERFAVORIT

Video terfavorit: 0.1% video dengan likes terbanyak

```
def get_minimum_point(df, key, proportion):
    x = df[key].sort_values(ascending=False)
    return x.iloc[int(len(x) * proportion)-1]
```

```
get_minimum_point(X, 'likes', 0.1/100)
```

```
2906264
```

Minimum likes:

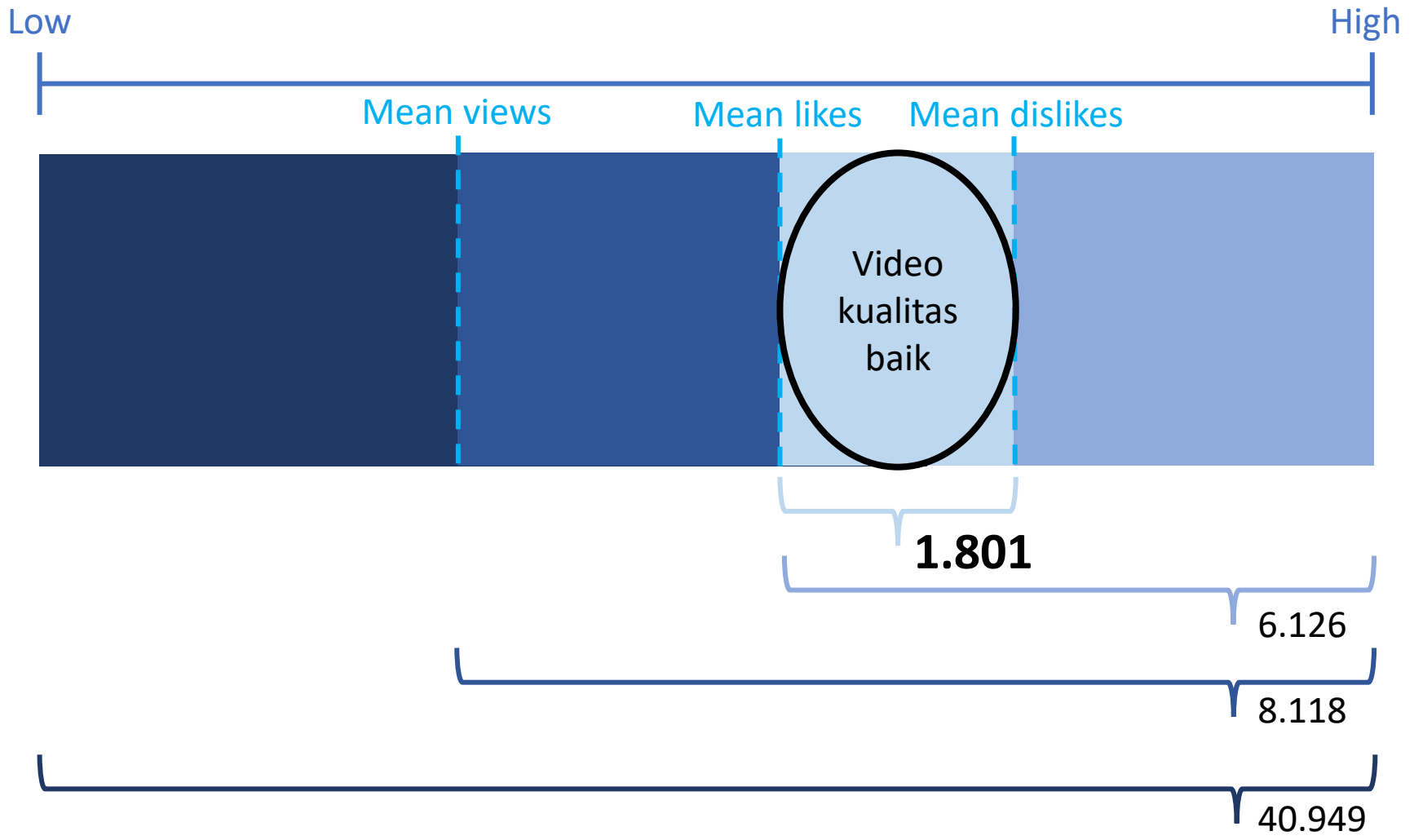
2.906.264

LAPORAN: IDENTIFIKASI VIDEO BERKUALITAS BAIK

Filter views di atas rata-rata

Filter likes di atas rata-rata

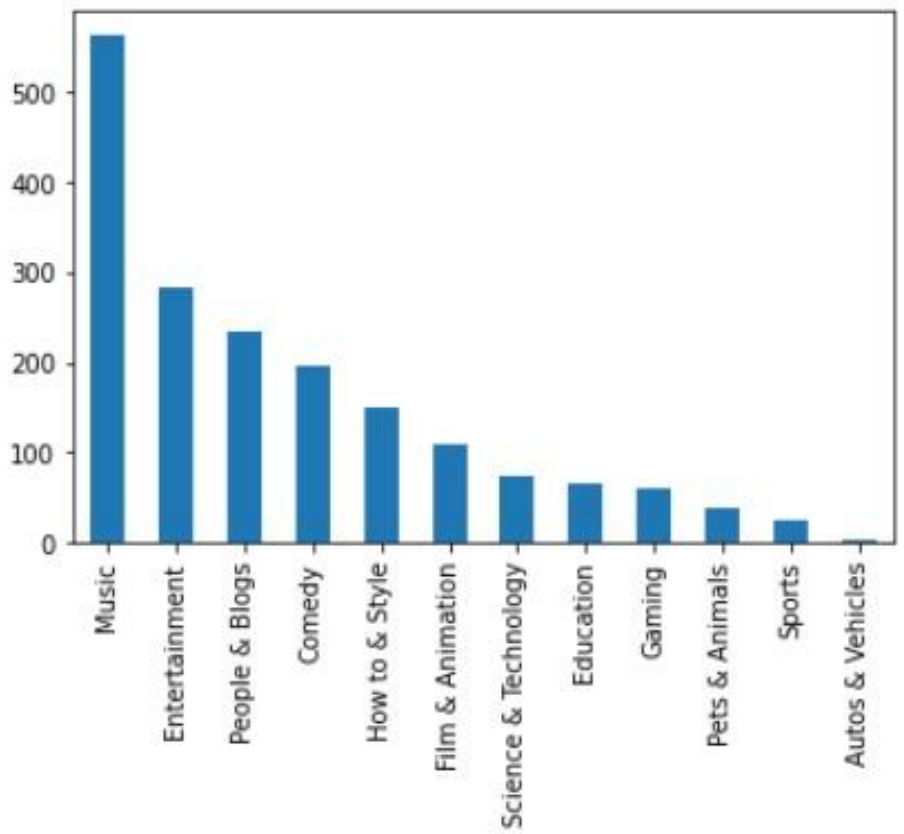
Filter dislikes di bawah rata-rata



LAPORAN: IDENTIFIKASI VIDEO BERKUALITAS BAIK

ID	Category name
1	Film & Animation
2	Autos & Vehicles
10	Music
15	Pets & Animals
17	Sports
19	Travel & Events
20	Gaming
22	People & Blogs
23	Comedy
24	Entertainment
25	News & Politics
26	Howto & Style
27	Education
28	Science & Technology
29	Nonprofits & Activism

Source: https://techpostplus.com/youtube-video-categories-list-faqs-and-solutions/



Video berkualitas baik didominasi oleh video kategori musik

KESIMPULAN DAN SARAN

KESIMPULAN

- Video Youtube **cenderung mendapat likes dibandingkan dislikes** yang menunjukkan adanya algoritma **Filter Bubble**.
- Hanya ada **0.098% video viral** dengan views > 100 juta.
- Untuk bisa dikategorikan **terfavorit**, video harus memiliki **minimal 2.906.264 likes**.
- Jika disaring dengan rata-rata views, likes, dan dislikes, jumlah video yang dapat dikatakan **berkualitas baik ada 1.801** yang **didominasi dengan video musik**.

SARAN

- Untuk **memperjelas visualisasi** data yang terlalu luas, dapat dilakukan **zoom in**.
- Salah satu cara untuk menentukan **kualitas suatu video** adalah dengan mengkategorikannya **berdasarkan rata-rata views, likes, dan dislikes**.


Thank
You