```python
#@title install
!pip install lib
!pip install isodate
!pip install kneed
import pandas as pd
import numpy as np
from dateutil import parser
import isodate
# Data visualization libraries
import matplotlib.pyplot as plt
import matplotlib.ticker as ticker
import seaborn as sns
sns.set(style="darkgrid", color_codes=True)
# Google API
from googleapiclient.discovery import build
# NLP libraries
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
nltk.download('stopwords')
nltk.download('punkt')
from wordcloud import WordCloud
from yellowbrick.cluster import KElbowVisualizer
from sklearn.decomposition import PCA
from sklearn.cluster import KMeans
from kneed import KneeLocator
import plotly.graph_objects as go
from plotly.subplots import make_subplots
```

```
Looking in indexes: https://pypi.org/simple, https://us-
python.pkg.dev/colab-wheels/public/simple/
Collecting lib
  Downloading lib-4.0.0-py3-none-any.whl (4.0 kB)
Installing collected packages: lib
Successfully installed lib-4.0.0
Looking in indexes: https://pypi.org/simple, https://us-
python.pkg.dev/colab-wheels/public/simple/
Collecting isodate
  Downloading isodate-0.6.1-py2.py3-none-any.whl (41 kB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 41.7/41.7 KB 1.2 MB/s eta
0:00:00
ent already satisfied: six in /usr/local/lib/python3.8/dist-packages
(from isodate) (1.15.0)
Installing collected packages: isodate
Successfully installed isodate-0.6.1
Looking in indexes: https://pypi.org/simple, https://us-
python.pkg.dev/colab-wheels/public/simple/
Collecting kneed
  Downloading kneed-0.8.2-py3-none-any.whl (10 kB)
Requirement already satisfied: numpy>=1.14.2 in
```

```
/usr/local/lib/python3.8/dist-packages (from kneed) (1.21.6)
Requirement already satisfied: scipy>=1.0.0 in
/usr/local/lib/python3.8/dist-packages (from kneed) (1.7.3)
Installing collected packages: kneed
Successfully installed kneed-0.8.2
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
```

```
api_key = 'AIzaSyDhvPkboT1Se80yLyOhHz0f8dqKuRHojzQ'

channel_ids = ['UCKqH_9mk1waLgBiL2vT5b9g', # vannossgaming
               'UC7_YxT-KID8kRbqZo7MyscQ', # markiplier
               'UCYzPXprvl5Y-Sf0g4vX-m6g', # jacksepticeye
               'UCS5Oz6CHmeoF7vSad0qqXfw', # DanTDM
               'UCpGdL9Sn3Q5YWUH2DVUW1Ug', # PopularMMOs
               'UCjtLOfx1yt1NlnFIDyAX3Ug', # W2S
               'UCYVinkwSX7szARULgYpvhLw', # Ali-A
               'UCClNRixXlagwAd--5MwJKCw', # H2ODelirious
               'UC1ieoHqKW-yYgDhLHIcx28w', # Syndicate
              ]

youtube = build('youtube', 'v3', developerKey=api_key)

def get_channel_stats(youtube, channel_ids):
    """
    Get channel statistics: title, subscriber count, view count, video
count, upload playlist
    Params:

    youtube: the build object from googleapiclient.discovery
    channels_ids: list of channel IDs

    Returns:
    Dataframe containing the channel statistics for all channels in
the provided list: title, subscriber count, view count, video count,
upload playlist

    """
    all_data = []
    request = youtube.channels().list(
                part='snippet,contentDetails,statistics',
                id=','.join(channel_ids))
    response = request.execute()

    for i in range(len(response['items'])):
        data = dict(channelName = response['items'][i]['snippet']
['title'],
                    subscribers = response['items'][i]['statistics']
['subscriberCount'],
                    views = response['items'][i]['statistics']
['viewCount'],
                    totalVideos = response['items'][i]['statistics']
['videoCount'],
                    playlistId = response['items'][i]
['contentDetails']['relatedPlaylists']['uploads'])
        all_data.append(data)

    return pd.DataFrame(all_data)

def get_video_ids(youtube, playlist_id):
    """
    Get list of video IDs of all videos in the given playlist
    Params:

    youtube: the build object from googleapiclient.discovery
    playlist_id: playlist ID of the channel

    Returns:
    List of video IDs of all videos in the playlist

    """

    request = youtube.playlistItems().list(
                part='contentDetails',
                playlistId = playlist_id,
                maxResults = 50)
    response = request.execute()

    video_ids = []

    for i in range(len(response['items'])):
        video_ids.append(response['items'][i]['contentDetails']
['videoId'])

    next_page_token = response.get('nextPageToken')
    more_pages = True

    while more_pages:
        if next_page_token is None:
            more_pages = False
        else:
            request = youtube.playlistItems().list(
                        part='contentDetails',
                        playlistId = playlist_id,
                        maxResults = 50,
                        pageToken = next_page_token)
            response = request.execute()
            for i in range(len(response['items'])):
                video_ids.append(response['items'][i]
['contentDetails']['videoId'])

            next_page_token = response.get('nextPageToken')

    return video_ids

def get_video_details(youtube, video_ids):
    """
    Get video statistics of all videos with given IDs
    Params:

    youtube: the build object from googleapiclient.discovery
    video_ids: list of video IDs

    Returns:
    Dataframe with statistics of videos, i.e.:
        'channelTitle', 'title', 'description', 'tags', 'publishedAt'
        'viewCount', 'likeCount', 'favoriteCount', 'commentCount'
        'duration', 'definition', 'caption'
    """

    all_video_info = []

    for i in range(0, len(video_ids), 50):
        request = youtube.videos().list(
            part="snippet,contentDetails,statistics",
            id=','.join(video_ids[i:i+50])
        )
        response = request.execute()

        for video in response['items']:
            stats_to_keep = {'snippet': ['channelTitle', 'title',
'description', 'tags', 'publishedAt'],
                             'statistics': ['viewCount', 'likeCount',
'favouriteCount', 'commentCount'],
                             'contentDetails': ['duration',
'definition', 'caption']
                            }
            video_info = {}
            video_info['video_id'] = video['id']

            for k in stats_to_keep.keys():
                for v in stats_to_keep[k]:
                    try:
                        video_info[v] = video[k][v]
                    except:
                        video_info[v] = None
            all_video_info.append(video_info)
    return pd.DataFrame(all_video_info)

def get_comments_in_videos(youtube, video_ids):
    """
    Get top level comments as text from all videos with given IDs
(only the first 10 comments due to quote limit of Youtube API)
    Params:

    youtube: the build object from googleapiclient.discovery
    video_ids: list of video IDs

    Returns:
    Dataframe with video IDs and associated top level comment in text.

    """
    all_comments = []

    for video_id in video_ids:
        try:
            request = youtube.commentThreads().list(
                part="snippet,replies",
                videoId=video_id
            )
            response = request.execute()

            comments_in_video = [comment['snippet']['topLevelComment']
['snippet']['textOriginal'] for comment in response['items'][0:10]]
            comments_in_video_info = {'video_id': video_id,
'comments': comments_in_video}

            all_comments.append(comments_in_video_info)

        except:
            # When error occurs - most likely because comments are
disabled on a video
            print('Could not get comments for video ' + video_id)

    return pd.DataFrame(all_comments)

def plot_cloud(wordcloud,title='none'):
  plt.figure(figsize=(10, 8))
  plt.imshow(wordcloud)
  if (title != 'none'):
    plt.title(title)
  plt.axis("off");


# Create a dataframe with video statistics and comments from all
channels
```

```
channel_data = get_channel_stats(youtube, channel_ids)
video_df = pd.DataFrame()
comments_df = pd.DataFrame()

for c in channel_data['channelName'].unique():
    print("Getting video information from channel: " + c)
    playlist_id = channel_data.loc[channel_data['channelName']== c,
'playlistId'].iloc[0]
    video_ids = get_video_ids(youtube, playlist_id)

    # get video data
    video_data = get_video_details(youtube, video_ids)
    # get comment data
    '''comments_data = get_comments_in_videos(youtube, video_ids)'''

    # append video data together and comment data toghether
    video_df = video_df.append(video_data, ignore_index=True)
    '''comments_df = comments_df.append(comments_data,
ignore_index=True)'''
```

```
Getting video information from channel: VanossGaming
Getting video information from channel: Markiplier
Getting video information from channel: Ali-A
Getting video information from channel: H2ODelirious
Getting video information from channel: DanTDM
Getting video information from channel: jacksepticeye
Getting video information from channel: PopularMMOs
Getting video information from channel: Syndicate
Getting video information from channel: W2S
```

```
# Write video data to CSV file for future references
video_df.to_csv('video_data.csv')
```

```
'''video_df = pd.read_csv('')'''
```

```
{"type":"string"}
```

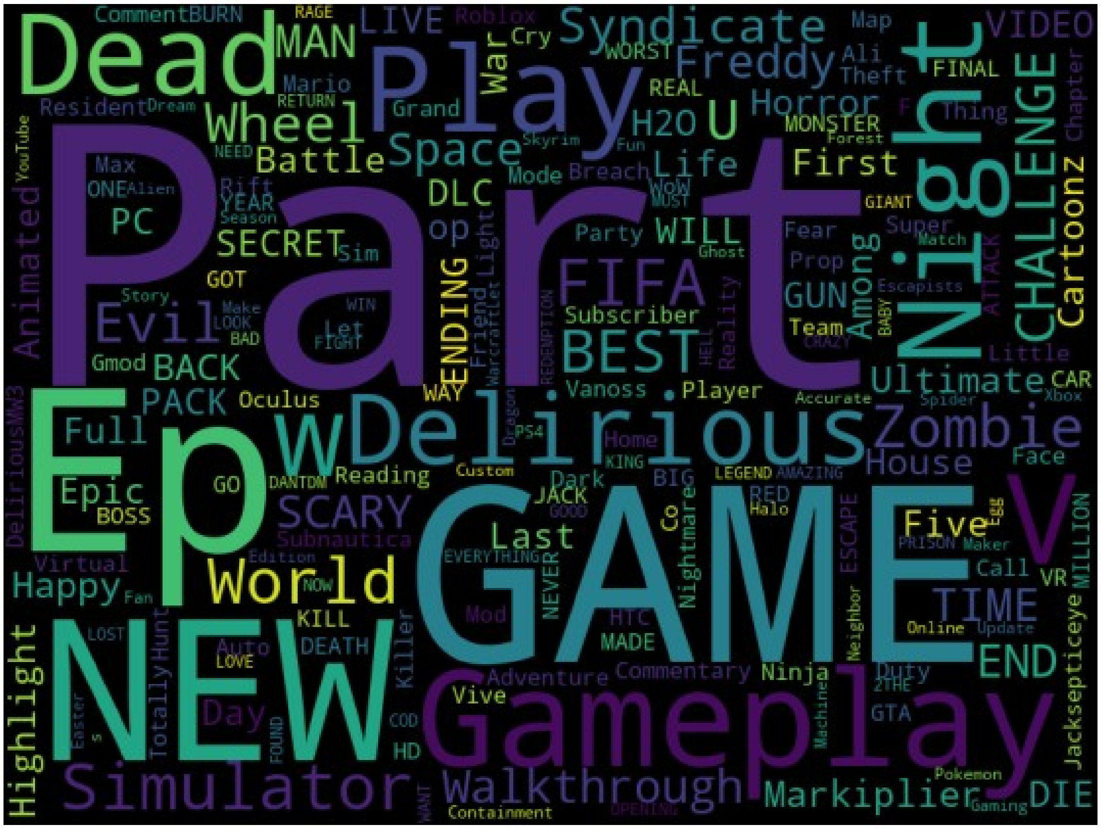
```
cols = ['viewCount', 'likeCount', 'commentCount']
video_df[cols] = video_df[cols].apply(pd.to_numeric, errors='coerce',
axis=1)
```

```
# Create publish day (in the week) column
video_df['publishedAt'] =  video_df['publishedAt'].apply(lambda x:
parser.parse(x))
video_df['pushblishDayName'] = video_df['publishedAt'].apply(lambda x:
x.strftime("%A"))
# convert duration to seconds
video_df['durationSecs'] = video_df['duration'].apply(lambda x:
isodate.parse_duration(x))
video_df['durationSecs'] =
video_df['durationSecs'].astype('timedelta64[s]')
```

```
# Add number of tags
video_df['tagsCount'] = video_df['tags'].apply(lambda x: 0 if x is
None else len(x))
# Comments and likes per 1000 view ratio
video_df['likeRatio'] = video_df['likeCount']/ video_df['viewCount'] *
1000
video_df['commentRatio'] = video_df['commentCount']/
video_df['viewCount'] * 1000
# Title character length
video_df['titleLength'] = video_df['title'].apply(lambda x: len(x))

# remove the outlier
indexDrop = video_df[ (video_df['video_id'] == 'ndsaoMFz9J4') ].index
video_df.drop(indexDrop , inplace=True)
video_df['viewCount'] = video_df['viewCount'].astype('float')
video_df['likeCount'] = video_df['likeCount'].astype('float')
video_df['commentCount'] = video_df['commentCount'].astype('float')
video_df['titleLength'] = video_df['titleLength'].astype('float')
video_df['durationSecs'] = video_df['durationSecs'].astype('float')
video_df = video_df.fillna('0')
indexDrop = video_df[ (video_df['viewCount'] == '0') |
(video_df['likeCount'] == '0') | (video_df['commentCount'] == '0') |
(video_df['titleLength'] == '0') | (video_df['durationSecs'] ==
'0') ].index
video_df.drop(indexDrop , inplace=True)
X = video_df.iloc[:, [2]].values
X = X.tolist()
X = [x[0] for x in X]

# TF-IDF Feature Generation
from sklearn.feature_extraction.text import TfidfVectorizer
from nltk.tokenize import RegexpTokenizer


# Initialize regex tokenizer
tokenizer = RegexpTokenizer(r'\w+')

# # Vectorize document using TF-IDF
tf_idf_vect = TfidfVectorizer(lowercase=True,
                        stop_words='english',
                        ngram_range = (1,1),
                        tokenizer = tokenizer.tokenize)

# Fit and Transfrom Text Data
X_train_counts = tf_idf_vect.fit_transform(X)

# Check Shape of Count Vector
X_train_counts.shape
```

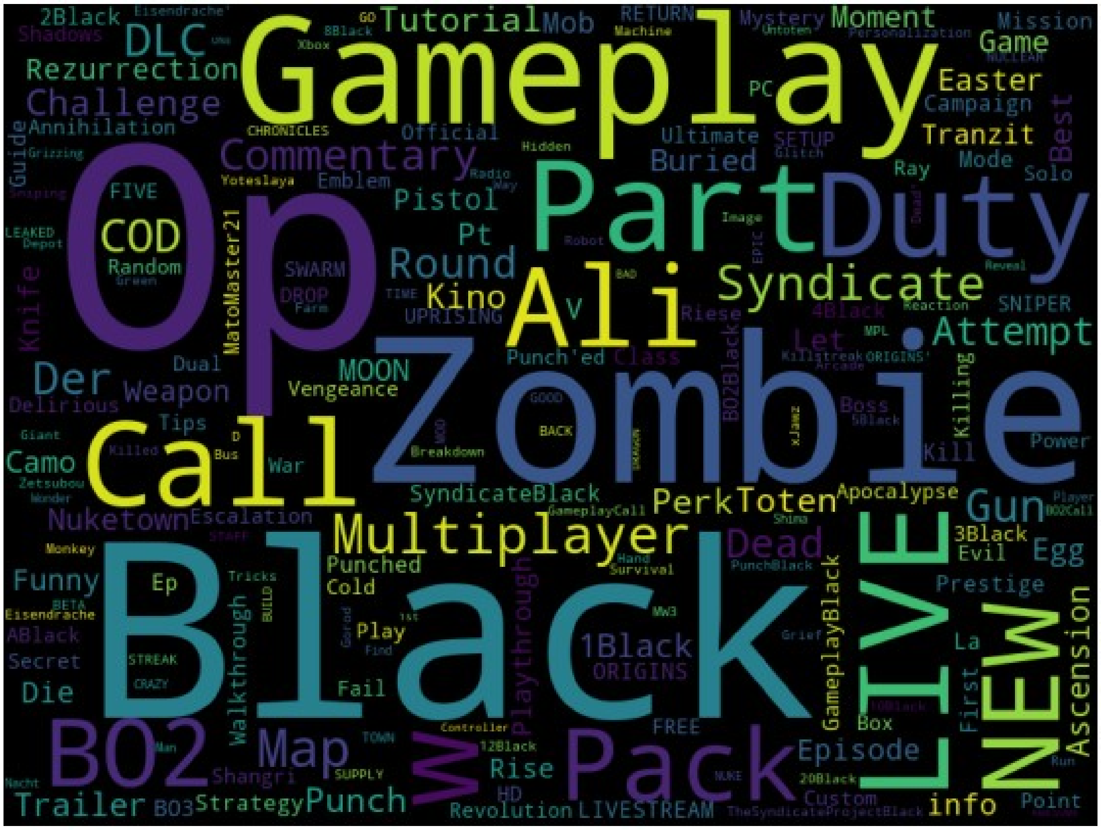
```
(31750, 13653)
```

```
# Import KMeans Model
from sklearn.cluster import KMeans

# Create Kmeans object and fit it to the training data
kmeans = KMeans(n_clusters=8).fit(X_train_counts)

# Get the labels using KMeans
pred_labels = kmeans.labels_

video_df['Cluster'] = kmeans.labels_

for i in range(8):
    new_df=video_df[video_df.Cluster==i]
    text="".join(new_df.title.tolist())
    wordcloud = WordCloud(width = 800, height = 600, random_state=1,
background_color='black',
                      colormap='viridis',
collocations=False).generate(text)
    plot_cloud(wordcloud,str(i+1))
```

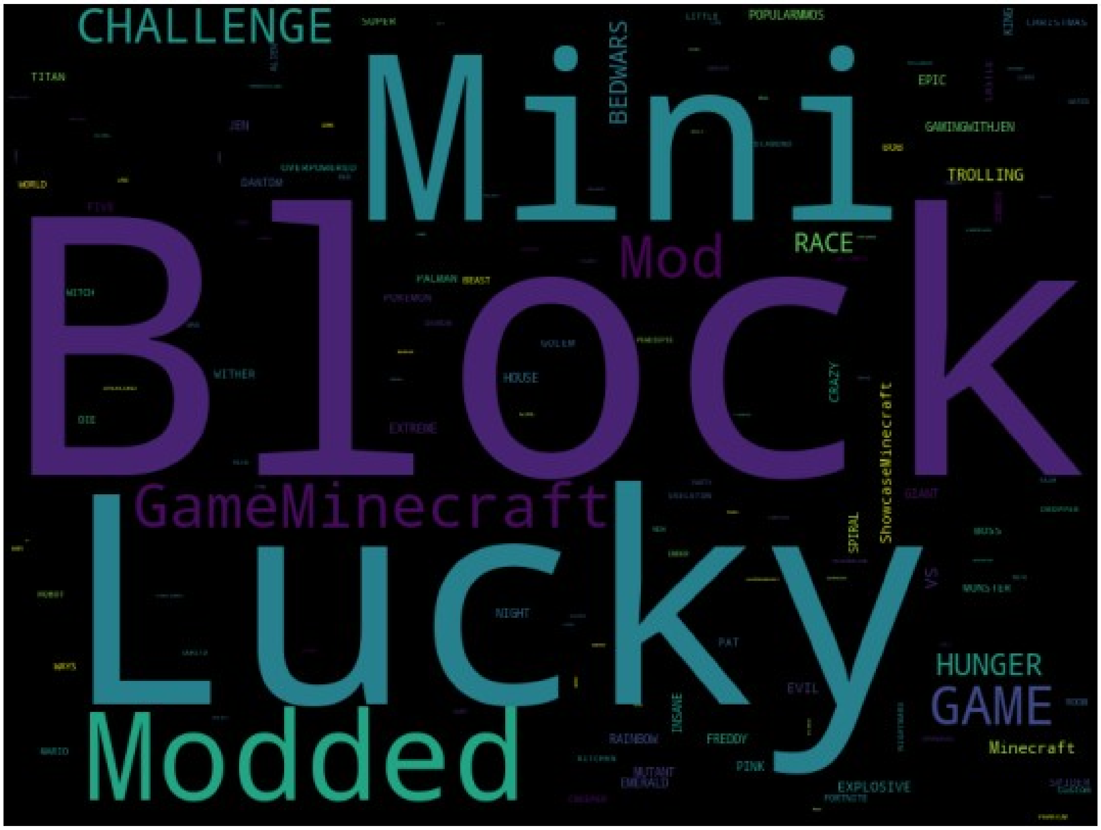


1

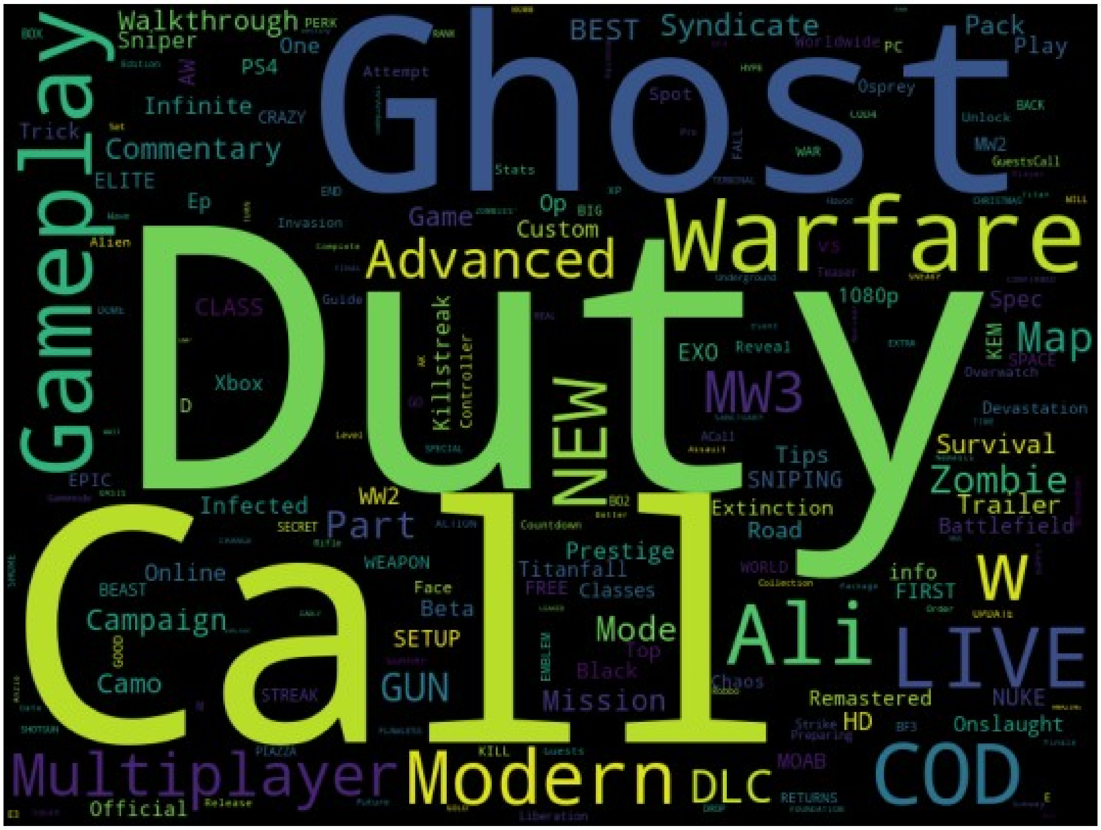
2

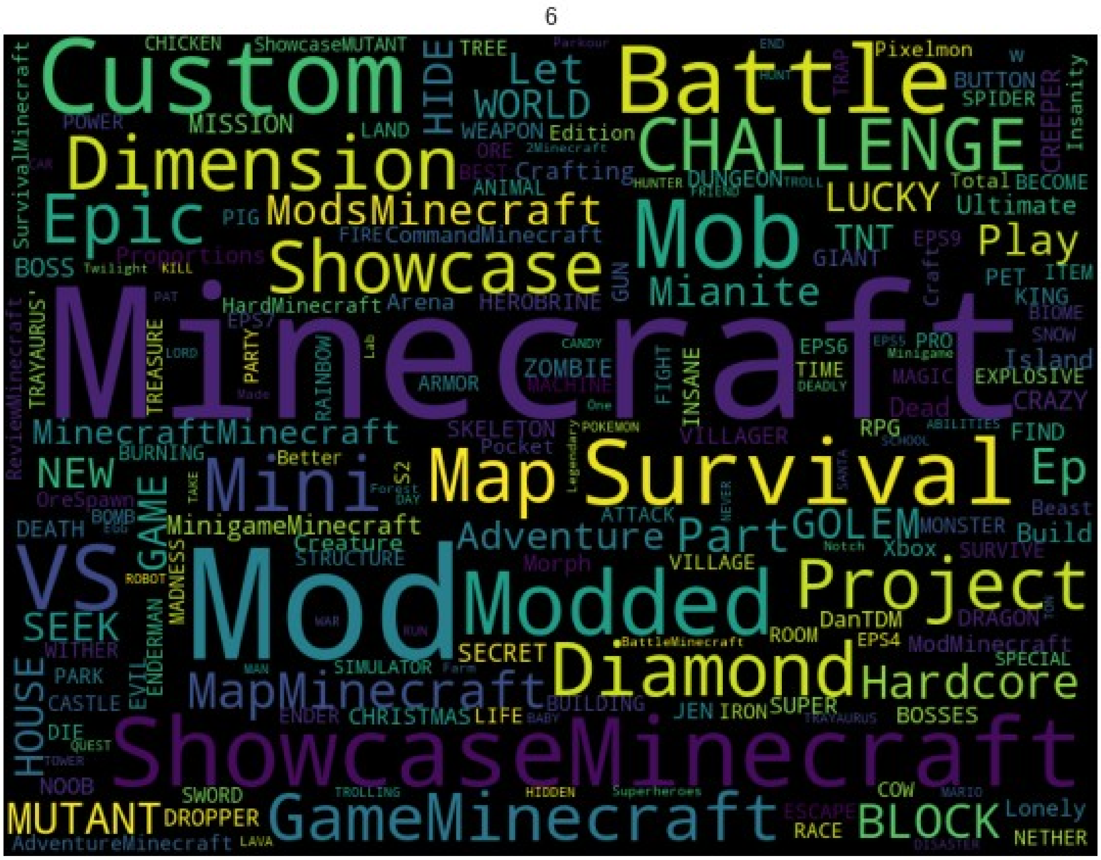

Gameplay
Op
Zombie
Black
Part
Duty
Call
Ali
LIVE
NEW
Pack
BO2
w
Multiplayer
Syndicate
COD
Commentary
Challenge
DLC
Rezurrection
Tutorial
Attempt
Round
Map
Der
Weapon
Camo
Nuketown
Funny
Knife
Kino
Pistol
Buried
Tranzit
Easter
Egg
Gun
PerkToten
Dead
Ascension
Episode
Punch
Trailer
Strategy
Walkthrough
Playthrough
1Black
Moment
Game
Mission
Mob
Campaign
Mystery
Best
Official
Annihilation
Emblem
Ultimate
Mode
Solo
Ray
Pt
Riese
Let
MOON
Vengeance
SyndicateBlack
Apocalypse
Escalation
Punched
Cold
Prestige
Evil
3Black
Kill
Killing
Boss
Power
Point
Custom
info
Revolution
LIVESTREAM
BO3
Shangri
Rise
HD
FREE
Box
First
La
GameplayBlack
ORIGINS
Fail
Play
Ep
Die
Secret
ABlack
Tips
Delirious
Dual
War
Random
FIVE
Guide
Shadows
2Black
MatoMaster21
UPRISING
SWARM
DROP
Class
Punch'ed
SNIPER
4Black
BO2Black
Reveal
Glitch
SETUP
PC
RETURN
Untoten
Personalization
NUCLEAR
CHRONICLES
Yoteslaya
Hidden
Radio
Image
Robot
EPIC
Dead
Killstreak
ORIGINS'
Arcade
xJawz
5Black
Breakdown
BACK
GOOD
DRAGON
Zetsubou
Wonder
Giant
Killed
Bus
D
Monkey
Tricks
Eisendrache
BETA
STREAK
CRAZY
Nacht
Man
TOWN
SUPPLY
Controller
12Black
NUKE
20Black
TheSyndicateProjectBlack
Run
Grief
Find
Gorod
1st
BUILD
MW3
Hand
Survival
Punch@Black
Shima
GameplayCall
Player
BO2Call
STAFF
8Black
Xbox
GO
Machine
Eisendrache'
Grizzing
Sniping
LEAKED
Depot
Green
Farm
TIME
BAD
MPL
Reaction
10Black

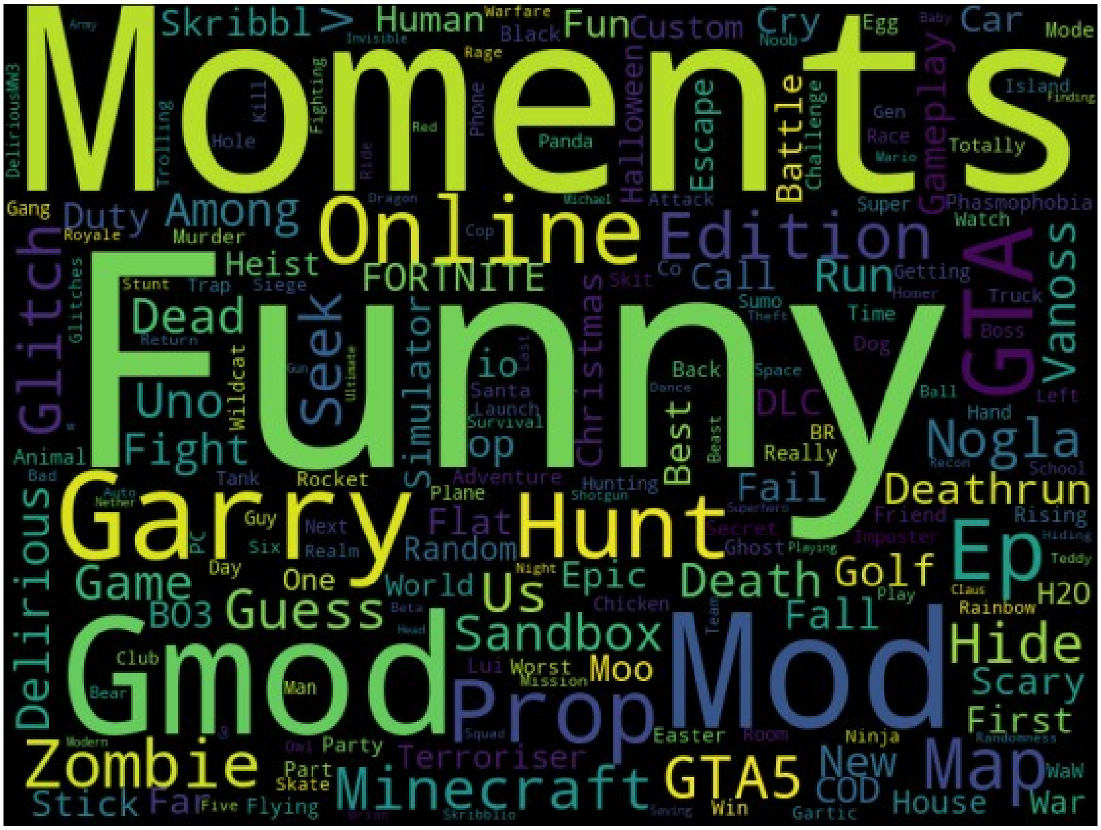
3


Block
Lucky
Mini
Modded
GameMinecraft
CHALLENGE
Mod
GAME
HUNGER
RACE
BEDWARS
TROLLING
GAMINGWITHJEN
POPULARMMOS
ShowcaseMinecraft
Minecraft
EXPLOSIVE
SUPER
TITAN
EPIC
JEN
DANTOM
OVERPOWERED
WORLD
WITCH
PALMAN
BEAST
POKEMON
GOLEM
WITHER
HOUSE
EXTREME
CRAZY
DIE
SPIRAL
GIANT
VS
MONSTER
NIGHT
BOSS
ROBOT
PAT
EVIL
WAYS
INSANE
RAINBOW
FREDDY
MUTANT
EMERALD
PINK
MARIO
KITCHEN
CREEPER
FORTNITE
KING
CHRISTMAS
LITTLE
FIVE

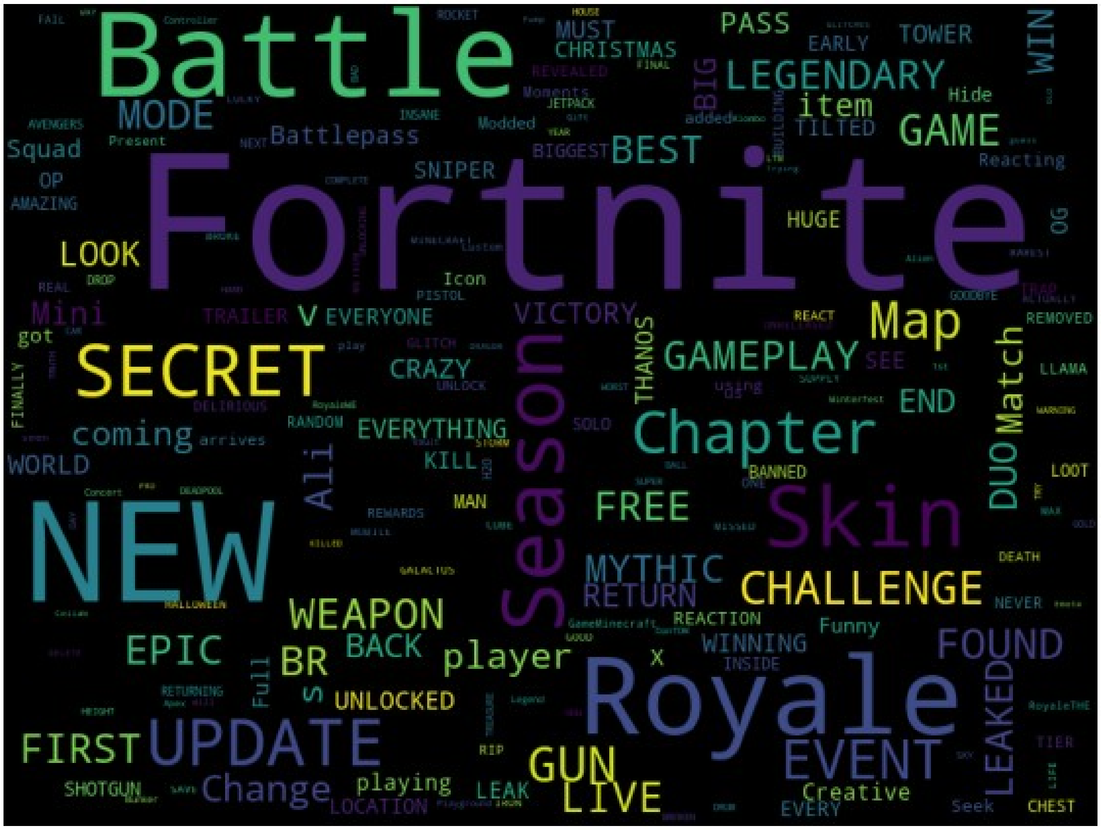
Episode
Minecraft
Syndicate
Play
Race
Project
Let
WoW
Epic
Prestige
Hunting
OpTic
World
Island
Proportions
Season
Warcraft
Batman
Arkham
Isle
Mod
Mianitian
S2
PrestigeThe
STORY
Walking
Modded
Last
Part
MODE
Dead
Survival
Parallel
Racing
Knight
Einschlag
Boss
BATTLE
Dungeon
Die
Fear
Hall
BIG
AMAZING
Magic
W
Happen
Power
BACK
Gameplay
DOOM
Hardcore
Among
Nether
Attack
Begin
WAY
Tower
Guild
S2Minecraft
Layers
Forest
SyndicateThe
Within
Series
Shift
BigTymer
Time
Truth
Fight
DRAGON
Castle
Must
PIRATE
3The
Mission
Vlogs
make
ESCAPE
Ultimate
1The
SECRET
Little
SCRAP
Wild
Work
Wolf
HARD
Bonus
GAME
GUARDIAN
LOVE
Quest
Got
Kill
Simdicate
vs
Thieves
MIANITE
Evil
Mad
GameplayEVIL
Man
Ultra
Adventure
Tales
Finding
Plan
Strange
Taking
Gun
ENDING
Breaking
Diamond
JOURNEY
Crafting
House
HUNTED
MEGA
Isolation
First
Borderlands
FTB
Legend
Virtual
DEATH
MOB
War
Friend
WalkthroughThe
Cry
Telltale
FINALE
INSANE
Dawn
Storm
Will
Unboxing
Gaming
DEADLY
Best
Base
Team
Wall
Notch
Live
Revealed
VIDEO
Machine
PrestigeEpisode
ENDER
Great
PAST
Sims
Sea
TDM
Lily
Prop
PVP
SAVE
Tekkit
GETS
Twilight
PrestigeMinecraft
EVERYWHERE
New
END
Pet
VanossHunter
Enemy
Hunt
DayZ
Killer
Dimension
PLACE
Map
Mall
EXPLORING
Life
ZTHE
Saving
U
STREAM
GO
Chase
Ep
Final
Max
Village
HUSBAND
CRAZY
GameplayMinecraft
Lost
Delirious
Outlast
WIFE
Zombie


Duty
Call
Ghost
Warfare
Gameplay
Advanced
Modern
Multiplayer
COD
LIVE
Ali
W
MW3
NEW
DLC
Map
Zombie
Walkthrough
Syndicate
BEST
Pack
Play
Sniper
One
PS4
Infinite
Commentary
Trick
ELITE
Ep
Game
Op
Custom
CLASS
Killstreak
Controller
1080p
Spec
EXO
Reveal
KEM
Xbox
D
Devastation
Survival
Tips
SNIPING
EPIC
Infected
Part
WW2
Extinction
Trailer
Battlefield
Road
Prestige
Titanfall
Online
BEAST
WEAPON
Face
Beta
FREE
Classes
info
FIRST
Campaign
Mode
SETUP
Top
Black
Camo
GUN
Mission
Chaos
Remastered
NUKE
HD
Onslaught
STREAK
MOAB
Official
RETURNS
Osprey
Unlock
MW2
Spot
Attempt
CRAZY
AW
PC
Worldwide
Invasion
Guide
Overwatch
SPACE
Countdown
Liberation
Release
KILL


6
Minecraft
Mod
ShowcaseMinecraft
Survival
Custom
Battle
Mob
CHALLENGE
Dimension
Showcase
Map
Modded
Project
Diamond
Mini
VS
Epic
GameMinecraft
MapMinecraft
ModsMinecraft
Hardcore
Adventure
Part
GOLEM
Ep
Mianite
LUCKY
Play
TNT
WORLD
Let
HIDE
SEEK
GAME
NEW
MUTANT
BLOCK
MinecraftMinecraft
MinigameMinecraft
HOUSE
Pixelmon
BUTTON
SPIDER
CREEPER
Insanity
Ultimate
Total
BECOME
ShowcaseMUTANT
CHICKEN
TREE
MISSION
LAND
WEAPON
Edition
Crafting
ORE
ANIMAL
DUNGEON
HUNTER
TROLL
FRIEND
CommandMinecraft
PIG
FIRE
Proportions
BOSS
Twilight
KILL
HardMinecraft
Arena
HEROBRINE
GUN
GIANT
EPS9
PET
ITEM
KING
BIOME
SNOW
Island
EPS6
EPS5
PRO
Minigame
TIME
DEADLY
EXPLOSIVE
MAGIC
CRAZY
Dead
RPG
FIND
ABILITIES
SCHOOL
ZOMBIE
ARMOR
MACHINE
FIGHT
INSANE
SKELETON
POKEMON
VILLAGER
Pocket
Better
S2
Forest
DAY
Legendary
NEVER
SANTA
Beast
Build
MONSTER
SURVIVE
Xbox
Notch
DRAGON
ModMinecraft
SPECIAL
EPS4
DanTDM
ROOM
BattleMinecraft
ATTACK
VILLAGE
Morph
Creature
STRUCTURE
SECRET
SIMULATOR
CHRISTMAS
LIFE
BUILDING
BABY
JEN
IRON
SUPER
BOSSES
ENDER
DEATH
BOMB
EGG
ROBOT
MADNESS
EVIL
ENDERMAN
WITHER
PARK
CASTLE
DIE
QUEST
TOWER
NOOB
SWORD
DROPPER
TROLLING
HIDDEN
Superheroes
COW
MARIO
Lonely
ESCAPE
RACE
DISASTER
NETHER
AdventureMinecraft
LAVA
OreSpawn
BURNING
TAKE
ReviewMinecraft
SurvivalMinecraft
TRAYAURUS'
TREASURE
LORD
PAT
EPS7
PARTY
RAINBOW
Lab
Made
CANDY
One
POWER
CAR
BEST
2Minecraft
Parkour
END
HUNT
TRAP
W
Craft
WAR
RUN
MAN
Farm
TRAYAURUS


Moments
Funny
Gmod
Mod
Garry
Prop
Hunt
Online
Edition
GTA
Ep
Map
Minecraft
GTA5
Zombie
Glitch
Seek
Delirious
Nogla
Vanoss
Deathrun
Hide
Sandbox
Guess
Death
Game
Among
Uno
Dead
Fight
Heist
FORTNITE
Simulator
Christmas
Halloween
Escape
Battle
Gameplay
Skribbl
Human
Fun
Custom
Cry
Car
Egg
Mode
Black
Warfare
Invisible
Noob
Rage
Baby
Army
DeliriousMW3
Trolling
Kill
Fighting
Phone
Red
Panda
Hole
Ride
Challenge
Gen
Race
Totally
Island
Finding
Mario
Dragon
Michael
Attack
Super
Phasmophobia
Watch
Gang
Royale
Duty
Murder
Cop
Call
Run
Getting
Stunt
Trap
Siege
Skit
Co
Truck
Glitches
Sumo
Theft
Time
Homer
Boss
Return
Wildcat
Gun
Ultimate
Lost
io
Dog
Back
Space
Santa
Launch
Survival
Dance
DLC
Ball
Left
Hand
Best
Beast
BR
Really
Animal
Bad
op
Tank
Rocket
Adventure
Hunting
Recon
School
Auto
Nether
Plane
Shotgun
Fail
Superhero
Friend
Rising
Flat
Secret
Imposter
Hiding
Guy
Next
PC
Random
Ghost
Playing
Six
Realm
Day
Night
Epic
Golf
Teddy
World
Us
One
BO3
Chicken
Play
Claus
Rainbow
H2O
Beta
Head
Team
Fall
Club
Bear
Man
Lui
Worst
Mission
Moo
Scary
First
Modern
Owl
Party
Squad
Easter
Room
Ninja
Randomness
Terroriser
New
Part
Skate
COD
WaW
Stick
Far
Five
Flying
Brian
Skribblio
Saving
Win
Gartic
House
War

8



```python
#number of instances in each clusters
video_df['Cluster'].value_counts()
```

```
0    17821
5     6124
6     1782
1     1753
7     1396
4     1124
3     1042
2      708
Name: Cluster, dtype: int64
```

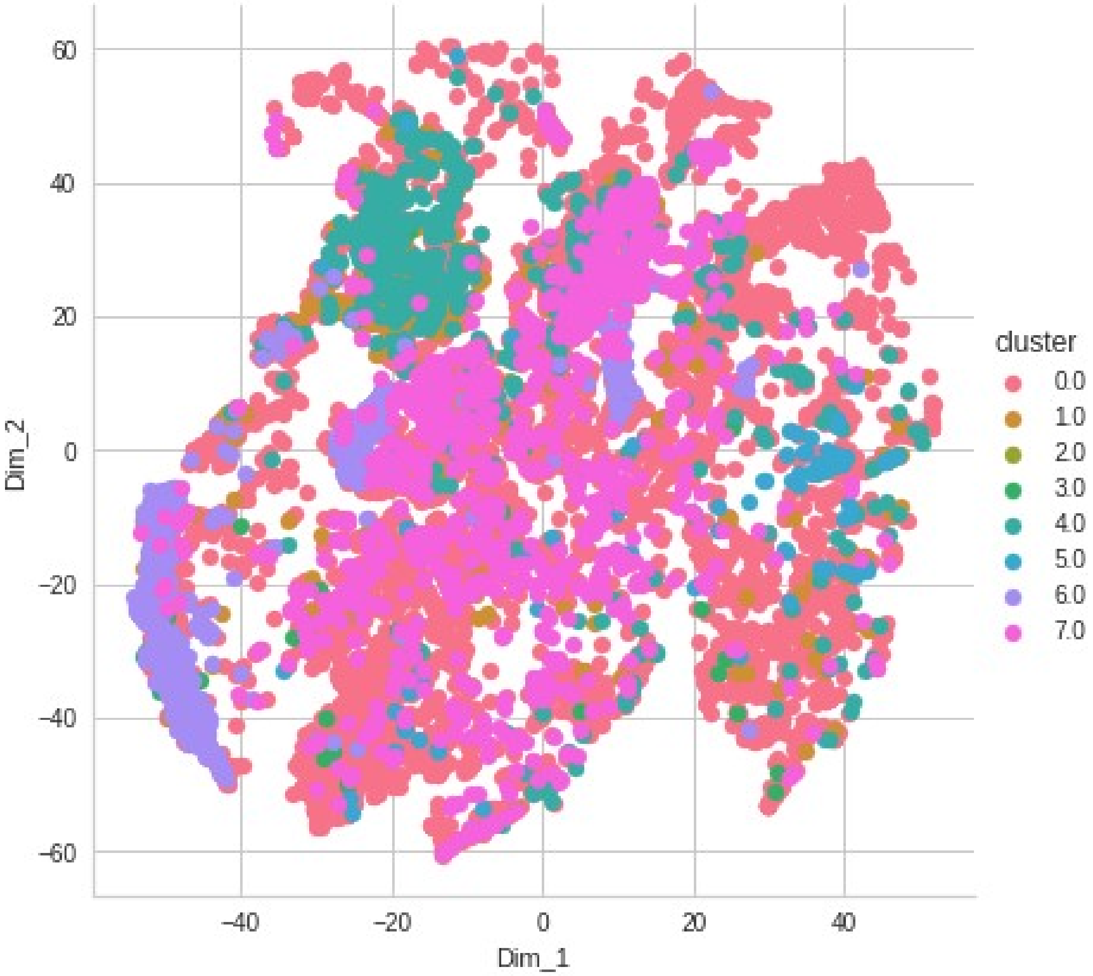
```python
# assign cluster
cluster_1 = video_df[video_df.Cluster == 0]
cluster_2 = video_df[video_df.Cluster == 1]
cluster_3 = video_df[video_df.Cluster == 2]
cluster_4 = video_df[video_df.Cluster == 3]
cluster_5 = video_df[video_df.Cluster == 4]
cluster_6 = video_df[video_df.Cluster == 5]
cluster_7 = video_df[video_df.Cluster == 6]
cluster_8 = video_df[video_df.Cluster == 7]

video_df.head()
```

```
      video_id  channelTitle  \
0  SrDiTlWK88w  VanossGaming   
1  7boaXv3qNKI  VanossGaming   
2  -3yRHG-8T90  VanossGaming   
3  ANYiX53ZnKY  VanossGaming   
4  HY1R4kgSr8g  VanossGaming   

                                               title  \
0     Gmod Guess Who - Invasion at the COD Terminal!   
1  Mario Party Superstars - Driving Terroriser In...   
2  Fortnite Funny Moments - Super Anime Moves and...   
3    Minecraft Funny Moments - Lui Meets the Warden!   
4    Escape the Backrooms - A Terrifying New Update!   

                                         description  \
0  Outro Song: https://www.youtube.com/watch?v=b5...   
1  Outro Song: https://www.youtube.com/watch?v=b5...   
2  Thank you to Playstation for sponsoring this v...   
3  Outro Song: https://www.youtube.com/watch?v=b5...   
4  Outro Song: https://www.youtube.com/watch?v=8_...   

                                                tags  \
0  [Funny Moments, Montage video games, gaming, V...   
1  [Funny Moments, Montage video games, gaming, V...   
2  [Funny Moments, Montage video games, gaming, V...   
3  [Funny Moments, Montage video games, gaming, V...   
4  [Funny Moments, Montage video games, gaming, V...   

                publishedAt  viewCount likeCount favouriteCount 
commentCount  \
0 2023-02-08 00:47:37+00:00    89016.0    9739.0             0 
516.0   
1 2023-02-06 00:25:07+00:00   823635.0   37543.0             0 
1031.0   
2 2023-02-04 01:20:16+00:00   860764.0   44331.0             0 
2096.0   
3 2023-02-02 04:36:57+00:00  1272354.0   58497.0             0 
1469.0   
4 2023-01-31 02:58:33+00:00  1383753.0   59327.0             0 
1410.0   

   duration definition caption pushblishDayName  durationSecs 
tagsCount  \
0  PT16M56S         hd   false        Wednesday        1016.0 
16   
1  PT30M36S         hd   false           Monday        1836.0 
16   
2  PT17M25S         hd   false         Saturday        1045.0 
16   
3  PT23M59S         hd   false         Thursday        1439.0 
```

```
16
4  PT24M33S          hd   false          Tuesday        1473.0
16

    likeRatio commentRatio  titleLength  Cluster
0  109.407298     5.796711         46.0        0
1   45.582084     1.251768         51.0        0
2   51.501922     2.435046         63.0        6
3   45.975413     1.154553         47.0        6
4   42.873981     1.018968         47.0        0
```

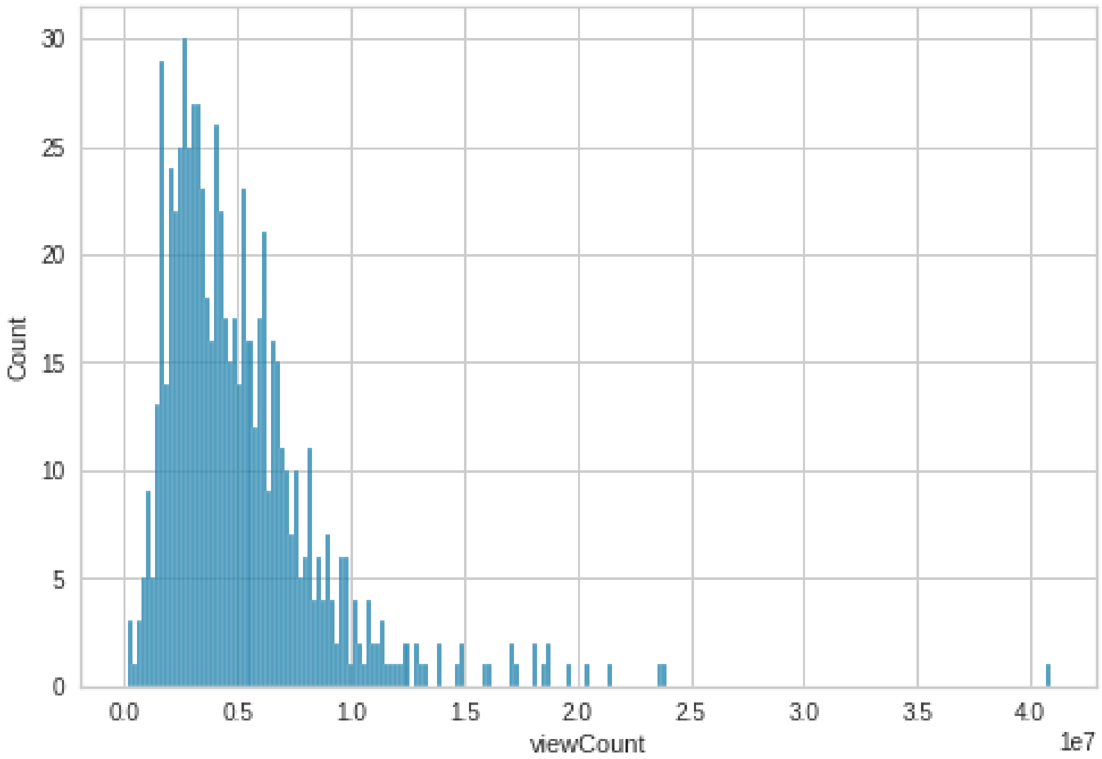
```python
Y = video_df.iloc[:, [6,7,9,14]].values
from sklearn.manifold import TSNE
from sklearn.preprocessing import StandardScaler
standardized_data = StandardScaler().fit_transform(Y)
model = TSNE(n_components = 2, random_state = 0,perplexity=100)
tsne_data = model.fit_transform(standardized_data)

l = video_df['Cluster']
tsne_data1 = np.vstack((tsne_data.T, l)).T
tsne_data2 = tsne_data1[0:10000]
tsne_df = pd.DataFrame(data = tsne_data2,
     columns =("Dim_1", "Dim_2", "cluster"))

# Plotting the result of tsne
sns.FacetGrid(tsne_df, hue ="cluster", size = 6).map(
       plt.scatter, 'Dim_1', 'Dim_2').add_legend()

plt.show()
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/axisgrid.py:337:
UserWarning: The `size` parameter has been renamed to `height`; please
update your code.
  warnings.warn(msg, UserWarning)
```



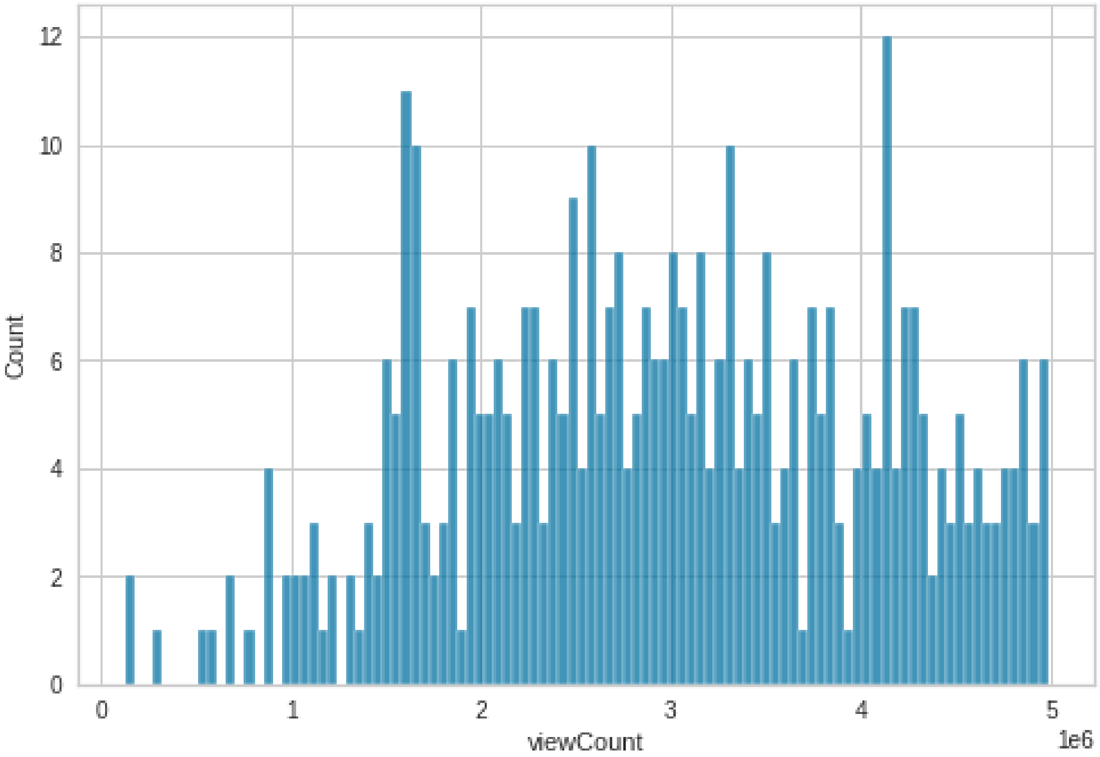
```
#Cluster 3
sns.histplot(data=cluster_3, x="viewCount", bins=200)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7abbb2b9d0>
```



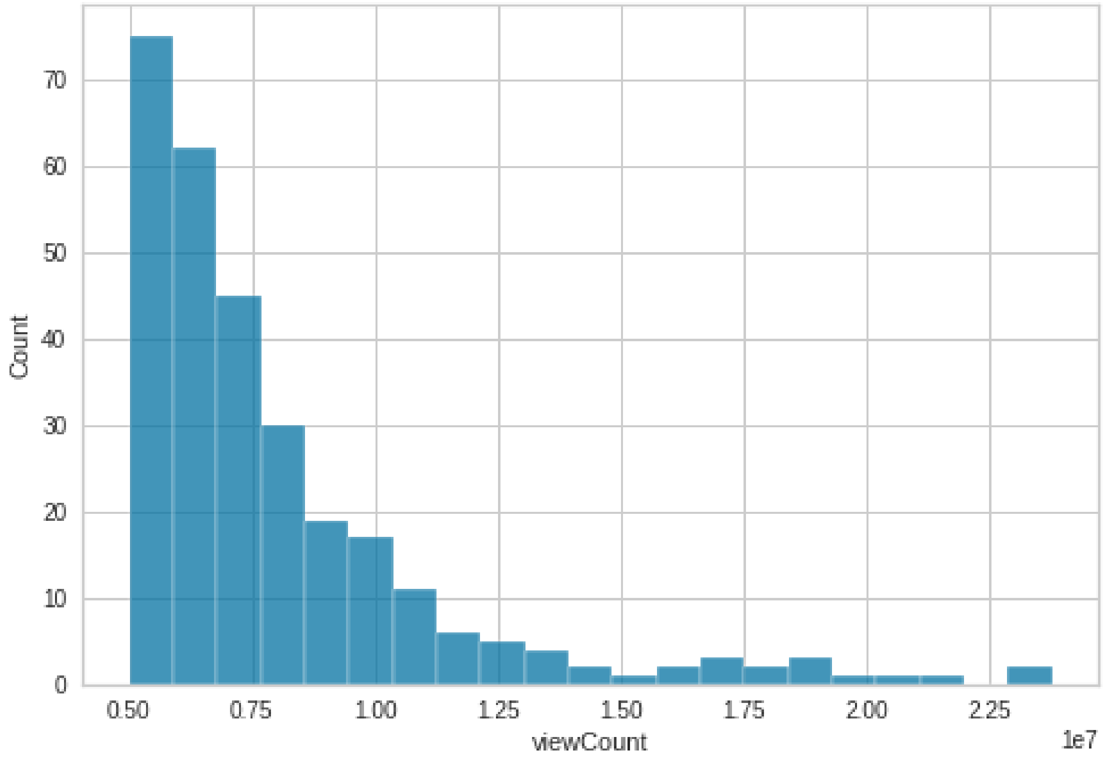
```
sns.histplot(data=cluster_3[cluster_3.viewCount <= 5000000],
x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ac1c93fd0>
```



```
sns.histplot(data=cluster_3[(cluster_3.viewCount > 5000000) &
(cluster_3.viewCount <= 25000000)], x="viewCount")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7abb572220>
```



```
cluster_3['viewCount'].mean()
```

```
5107430.745762712
```

```
cluster_3['channelTitle'].value_counts()
```

```
PopularMMOs     698
DanTDM            5
H2ODelirious      4
Ali-A             1
Name: channelTitle, dtype: int64
```

```
cluster_3['title'].value_counts()
```

```
Minecraft: FROZEN HUNGER GAMES - Lucky Block Mod - Modded Mini-Game
2
Minecraft: EXTREME DARK SIDE LUCKY BLOCK RACE - Lucky Block Mod -
Modded Mini-Game                2
AM I LUCKY?!
1
Minecraft: EVIL JEN'S EMERALD LUCKY BLOCK RACE - Lucky Block Mod -
Modded Mini-Game               1
Minecraft: CREEPER SPIDER MUTATION CHALLENGE GAMES - Lucky Block Mod -
Modded Mini-Game           1

..
Minecraft: GAMINGWITHJEN VS GAMINGWITHJEN CHALLENGE GAMES - Lucky
Block Mod - Modded Mini-Game    1
```

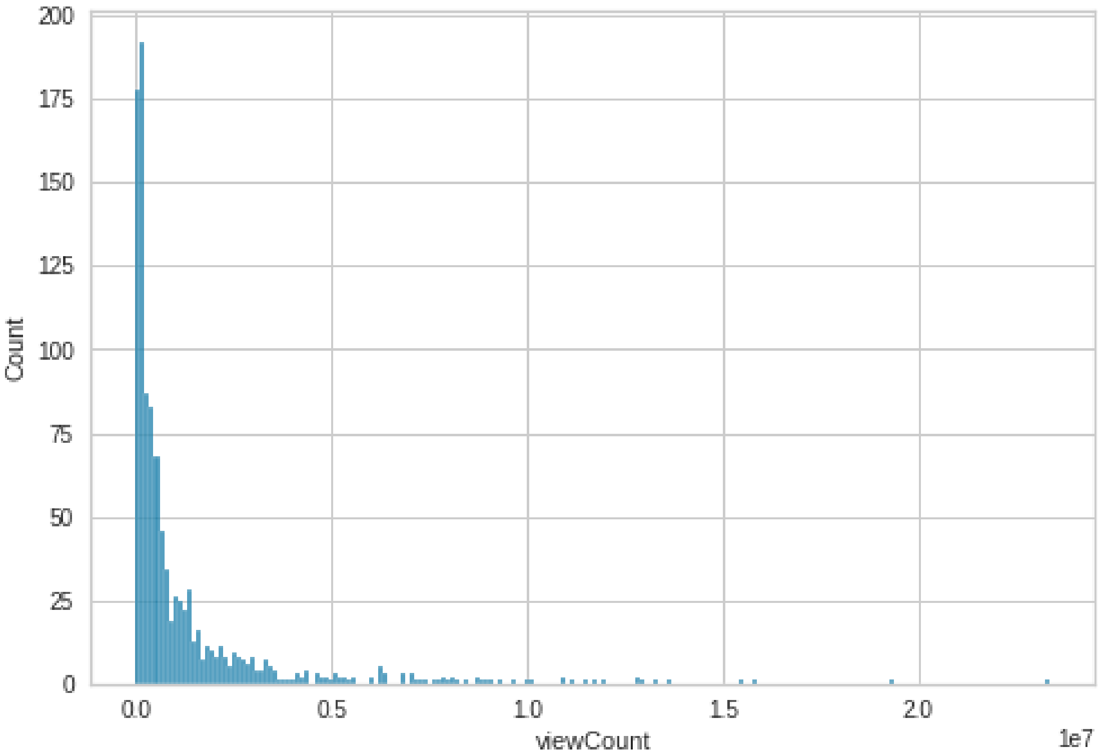
```
Minecraft: SCARY CLOWN CHALLENGE GAMES - Lucky Block Mod - Modded
Mini-Game                               1
Minecraft: POPULARMMOS VS POPULARMMOS CHALLENGE GAMES - Lucky Block
Mod - Modded Mini-Game           1
Minecraft: CRABZILLA CHALLENGE GAMES - Lucky Block Mod - Modded Mini-
Game                           1
Minecraft: MUTANT ZOMBIE CHALLENGE GAMES - RUINS MOD - Modded Mini-
Game                             1
Name: title, Length: 706, dtype: int64
```

This cluster contains videos about Minecraft mainly from PopularMMos. Most of videos have less than 5000000 views and the mean view count is 5107430.

#Cluster 4

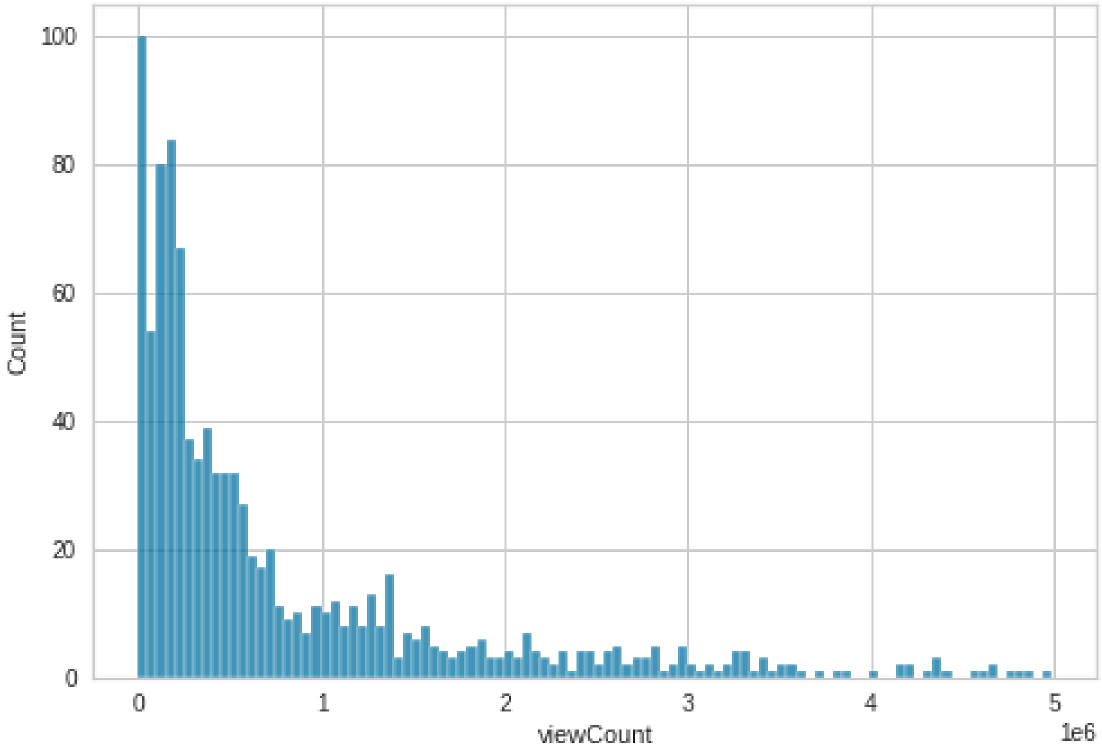
```
sns.histplot(data=cluster_4, x="viewCount", bins=200)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7abb4489d0>
```



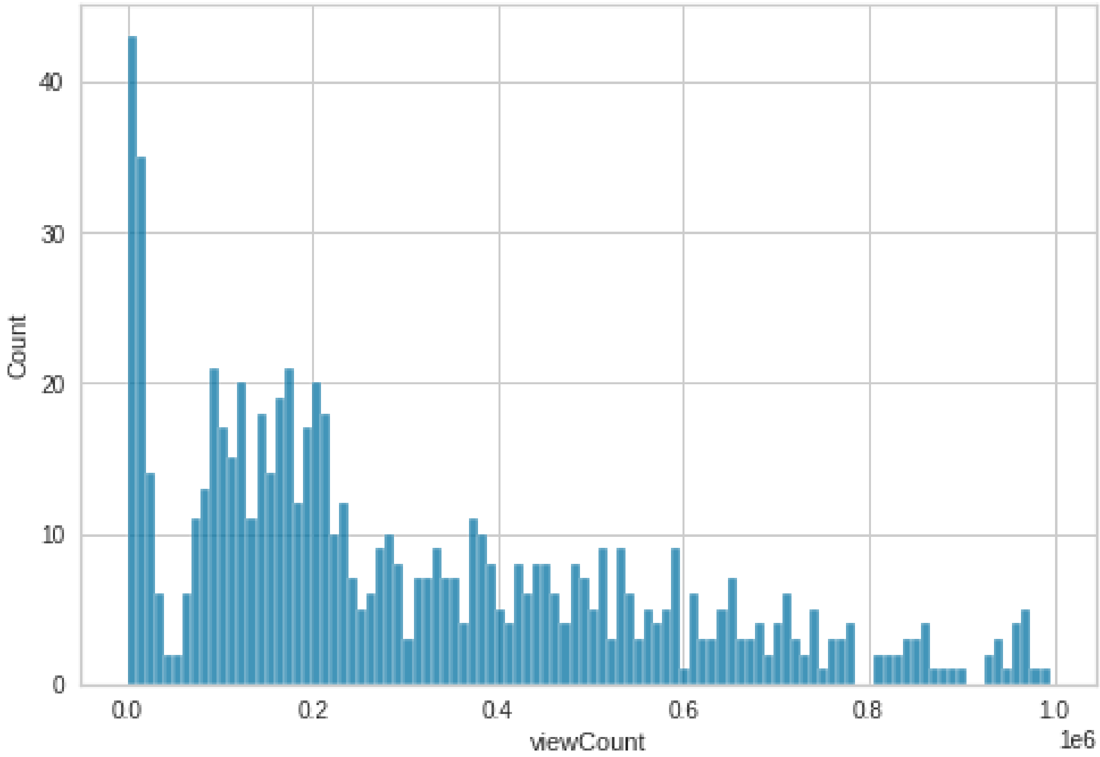
```
sns.histplot(data=cluster_4[cluster_4.viewCount <= 5000000],
x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7abb1c08b0>
```



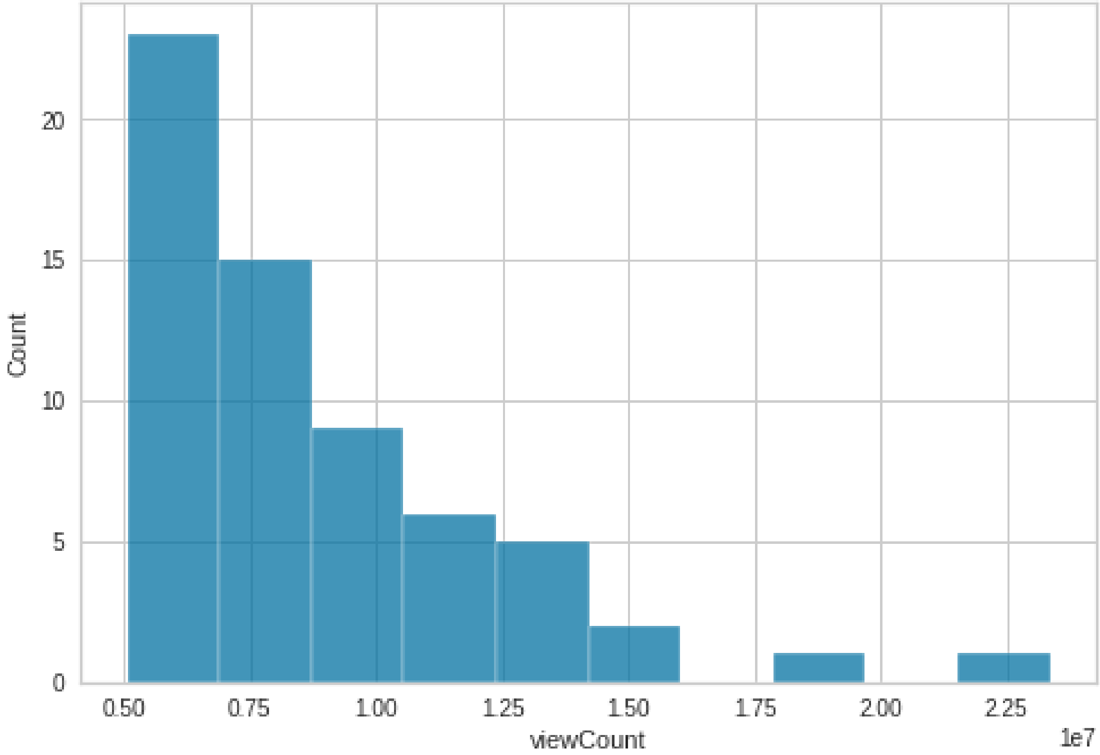
```
sns.histplot(data=cluster_4[cluster_4.viewCount <= 999999],
x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7abb1655b0>
```



```
sns.histplot(data=cluster_4[(cluster_4.viewCount > 5000000) ],
x="viewCount")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7abaf62b50>
```



```
cluster_4['viewCount'].mean()
```

```
1255906.4596928982
```

```
cluster_4['title'].value_counts()
```

```
Emotional Damage - Fears To Fathom (Episode 1 & 2 STREAM!)
2
Power Wash Simulator - The Vanoss Podcast, Episode 2!
1
Building My REAL House! - The Minecraft Project Episode #398
1
Minecraft: Racing OpTic - "Parallel Island" - Episode 8
1
Minecraft: Racing OpTic - "Parallel Island" - Episode 7
1

..
IMPOSSIBLE TO ESCAPE PARKOUR!! - STORY MODE SEASON 2 - [EPISODE 2] [3]
1
CHOOSE THE RIGHT MINECART OR DIE!!! - STORY MODE SEASON 2 - [EPISODE
2] [2]    1
A GIANT EATS US!!! - STORY MODE SEASON 2 - [EPISODE 2] [1]
1
Minecraft: OMG I DIED! - STORY MODE [Episode 8] [4]
1
Pubic Moutain! - Minecraft - Lets Play - Episode 1 -
```

```
TheSyndicateProject       1
Name: title, Length: 1041, dtype: int64
```

```
cluster_4['channelTitle'].value_counts()
```

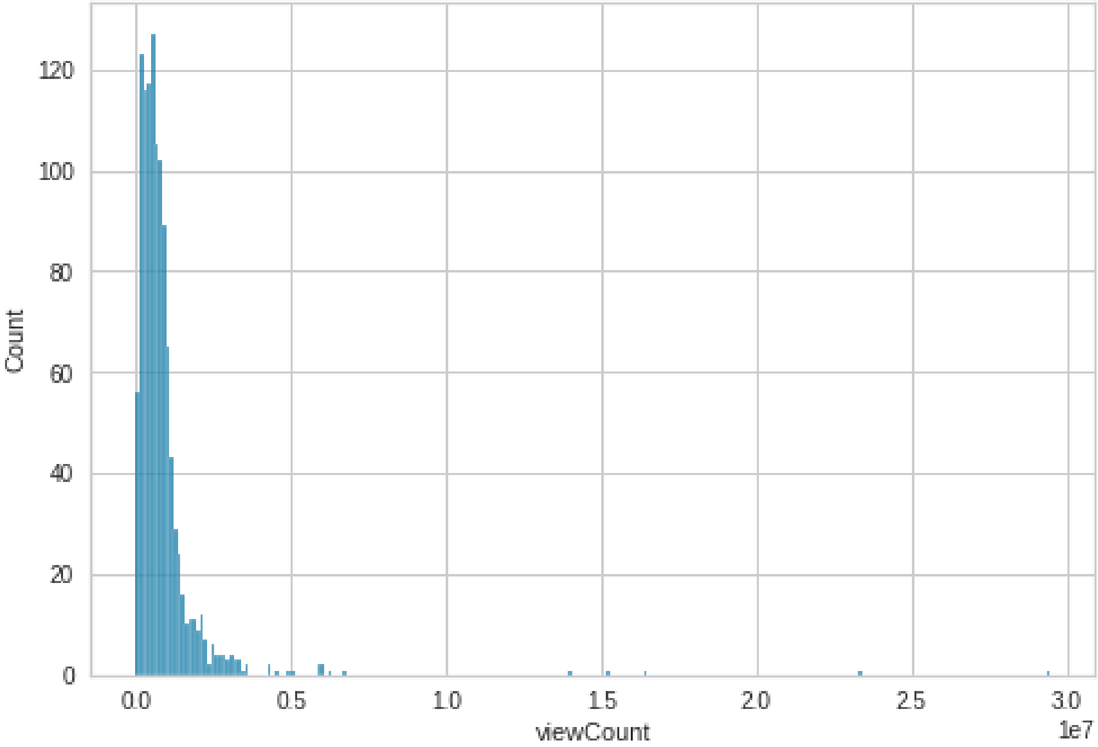
```
Syndicate        453
PopularMMOs      249
H2ODelirious     133
jacksepticeye     94
DanTDM            76
Markiplier        23
VanossGaming      13
Ali-A              1
Name: channelTitle, dtype: int64
```

This cluster has videos mostly from VanossGaming and popularMMOs. Videos are mostly about Minecraft and episode mode of Minecraft. The mean value of view is 1255906.

#Cluster 5

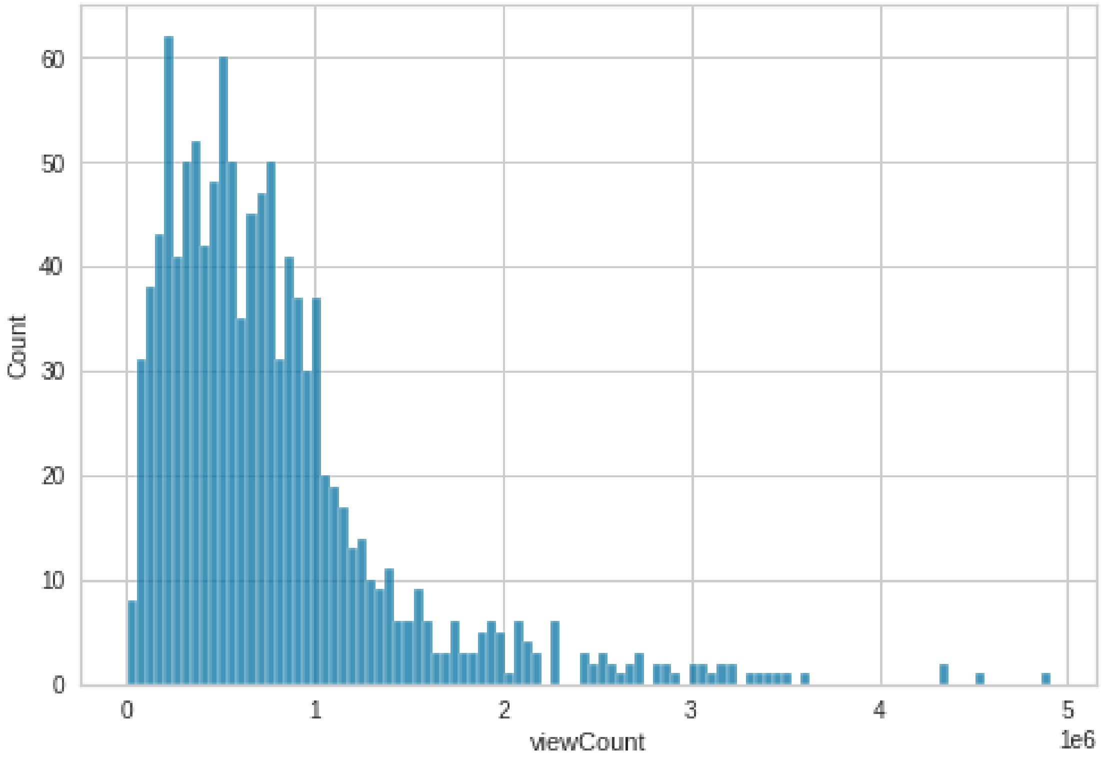
```
sns.histplot(data=cluster_5, x="viewCount")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7abbb44c40>
```



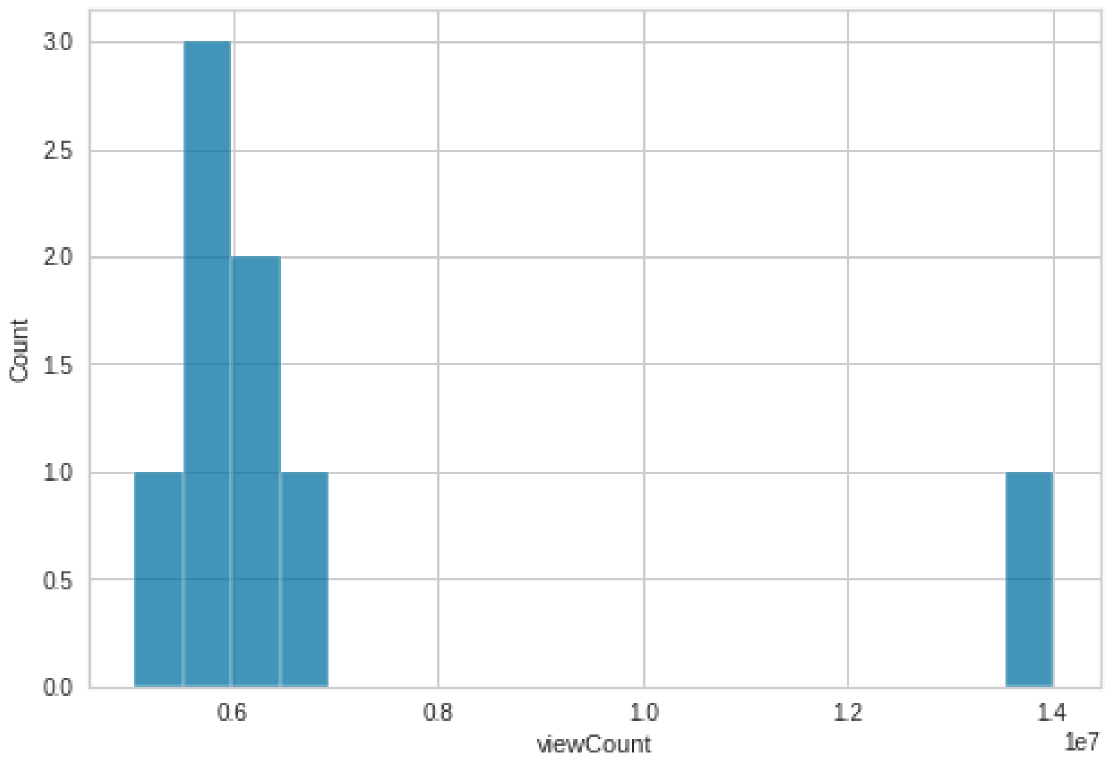
```
sns.histplot(data=cluster_5[cluster_5.viewCount <= 5000000],
x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7aba951e80>
```



```
sns.histplot(data=cluster_5[(cluster_5.viewCount > 5000000) &
(cluster_5.viewCount <= 15000000)], x="viewCount")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7aba80d250>
```



```
cluster_5['viewCount'].mean()
```

```
894246.6005338079
```

```
cluster_5['channelTitle'].value_counts()
```

```
Ali-A           990
Syndicate       124
H2ODelirious      6
VanossGaming      3
Markiplier        1
Name: channelTitle, dtype: int64
```

```
cluster_5['title'].value_counts()
```

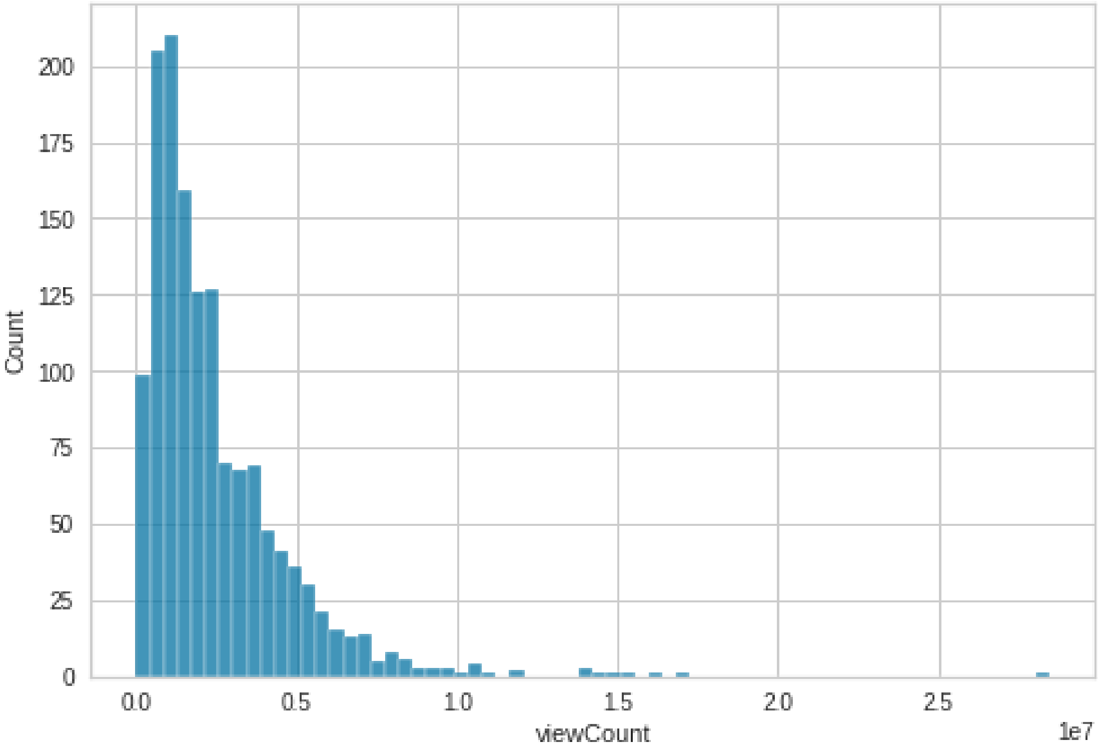
```
Call Of Duty Zombies: Custom Map: Anzio - Live Commentary w/ Syndicate
& Guests                             4
Call Of Duty Zombies: Custom Map: Misery - Live Commentary w/
Syndicate & Guests                            3
Call Of Duty Zombies: Custom Map: Gould - Live Commentary w/
Syndicate, Yoteslaya & MrWonanother    2
Call Of Duty Zombies: Custom Map: Cavern - Live Commentary w/
Syndicate & Guests                            2
NEW GUNS, KILLSTREAKS + MAX RANK! - Call of Duty: WW2 Beta Gameplay
*LIVE* w/ Ali-A!                    2

..
"BEHEAD HIM!" - Assassin's Creed: Unity GAMEPLAY! w/ Ali-A! - (E3 2014
Multiplayer)                      1
"OVR 2 QWIK!" - COD GHOSTS LIVE w/ Ali-A #33 - (Call of Duty Ghost
Gameplay)                                1
"EPIC COMEBACK?" - Gun Game LIVE w/ Ali-A #9! - (Call of Duty: Ghost)
1
"MULTIPLAYER GUNS!" - Call of Duty: Advanced Warfare gameplay! - (COD
AW 2014 HD)                          1
Call of Duty Ghosts - SNIPER Changes, NEW Spawns, Infected Guns & MORE
NEWS! (COD Ghost)                 1
Name: title, Length: 1116, dtype: int64
```

This cluster contains videos about black ops (a fps game), COD, zombies mode in black ops from 2 main channels: Ali-A and Syndicate. Most of videos have less than 1000000 views and the mean view count is 894247.

#Cluster 8

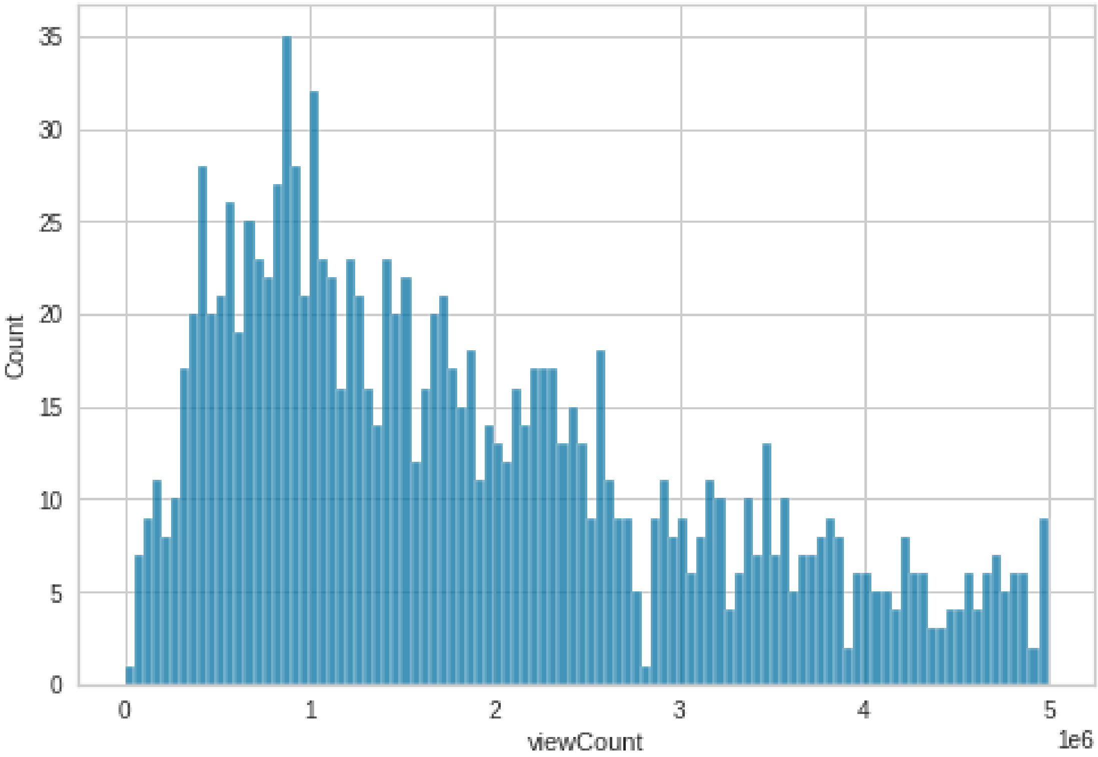
```
sns.histplot(data=cluster_8, x="viewCount")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7aba7f2370>
```



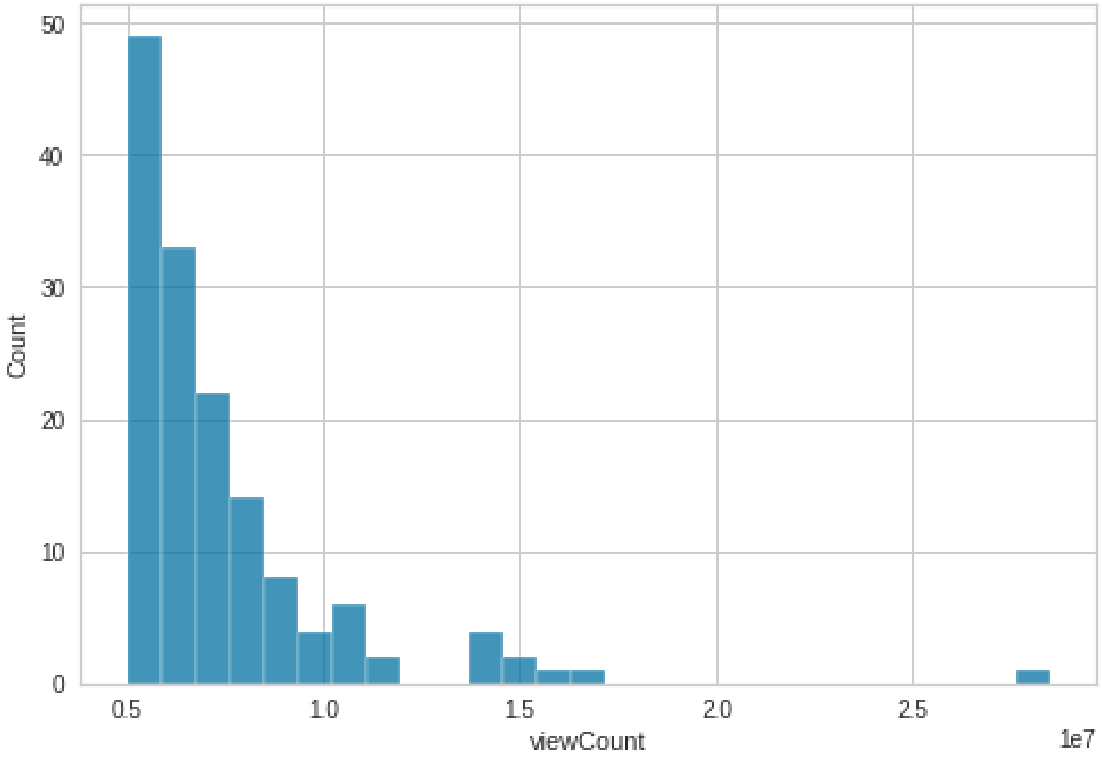
```
sns.histplot(data=cluster_8[cluster_8.viewCount <= 5000000],
x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7aba69e9d0>
```



```
sns.histplot(data=cluster_8[(cluster_8.viewCount > 5000000)],
x="viewCount")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7aba4d0bb0>
```



```
cluster_8['viewCount'].mean()
```

```
2460825.2528653294
```

```
cluster_8['channelTitle'].value_counts()
```

```
Ali-A            1117
H2ODelirious      113
DanTDM             92
PopularMMOs        31
jacksepticeye      16
Syndicate          14
Markiplier          5
W2S                 5
VanossGaming        3
Name: channelTitle, dtype: int64
```

```
cluster_8['title'].value_counts()
```

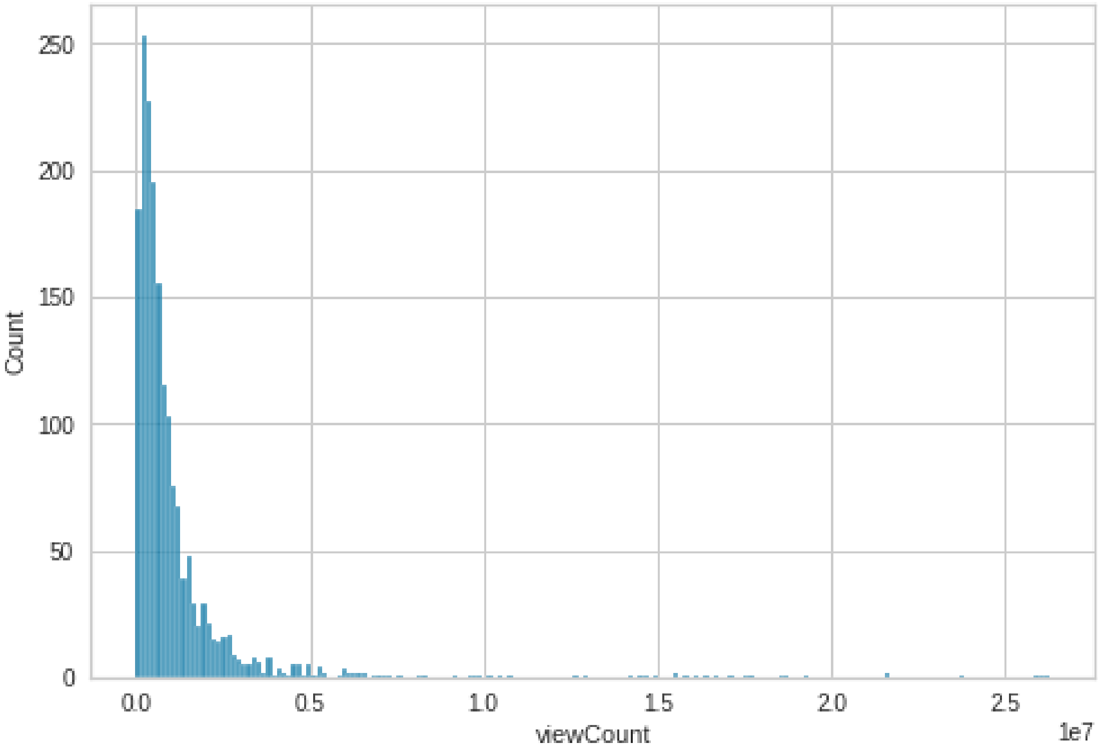
```
*NEW* HIDDEN SECRETS found in Fortnite! (MUST SEE)
2
*NEW* 100 Player Hide & Seek in Fortnite! (Chapter 2 Season 2)
2
CHRISTMAS DAY in Fortnite: Battle Royale!
2
The NEW Fortnite item...
2
NEW *SECRET* BLOCKBUSTER SKIN GAMEPLAY - Fortnite: Battle Royale! (The
Visitor)    1

..
*NEW* HIDDEN SPACESHIP has a BIG SECRET in Fortnite!
1
*NEW* FLOPPERS are INSANE in Fortnite! (FISHING UPDATE LEAK)
1
*NEW* SECRET UPDATE in Fortnite! (MAP CHANGES + MORE)
1
*NEW* ATLANTIS UPDATE in Fortnite! (HUGE SECRETS FOUND)
1
1 FORTNITE WIN = DUOS VS NINJA
1
Name: title, Length: 1392, dtype: int64
```

This cluster contains videos about Fortnite (a fps game) from Ali-A channel. Most of videos have less than 5000000 views and the mean view count is 2460825.

#Cluster 2

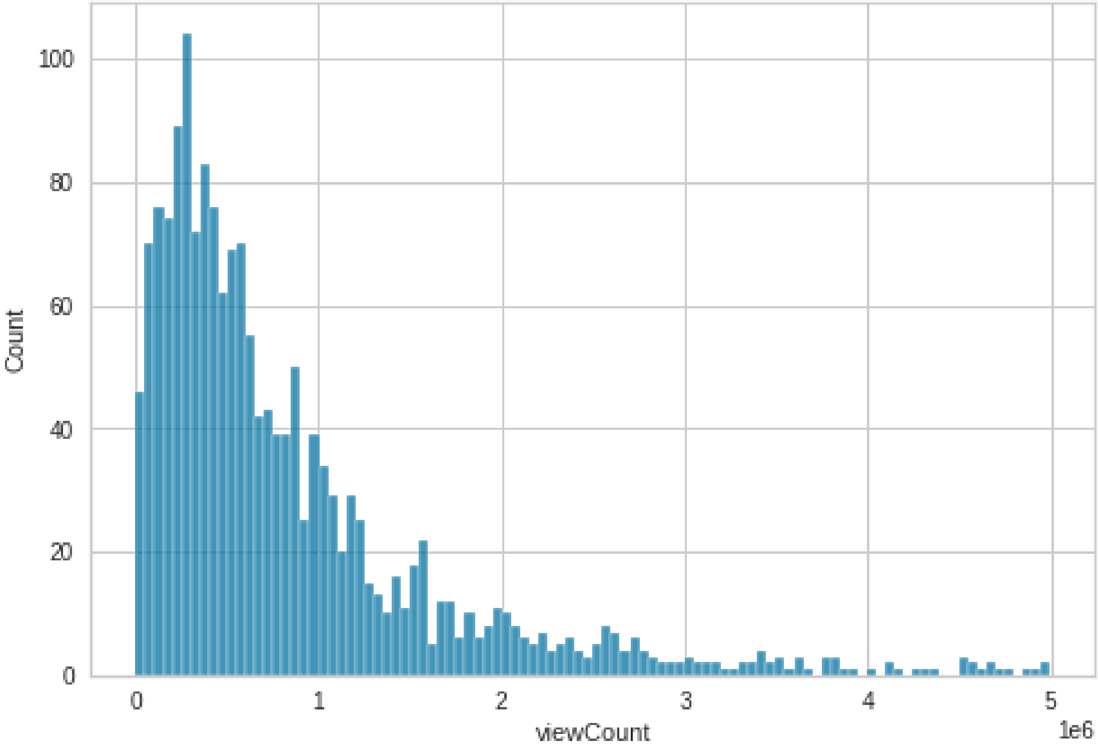
```
sns.histplot(data=cluster_2, x="viewCount")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7aba46fd00>
```



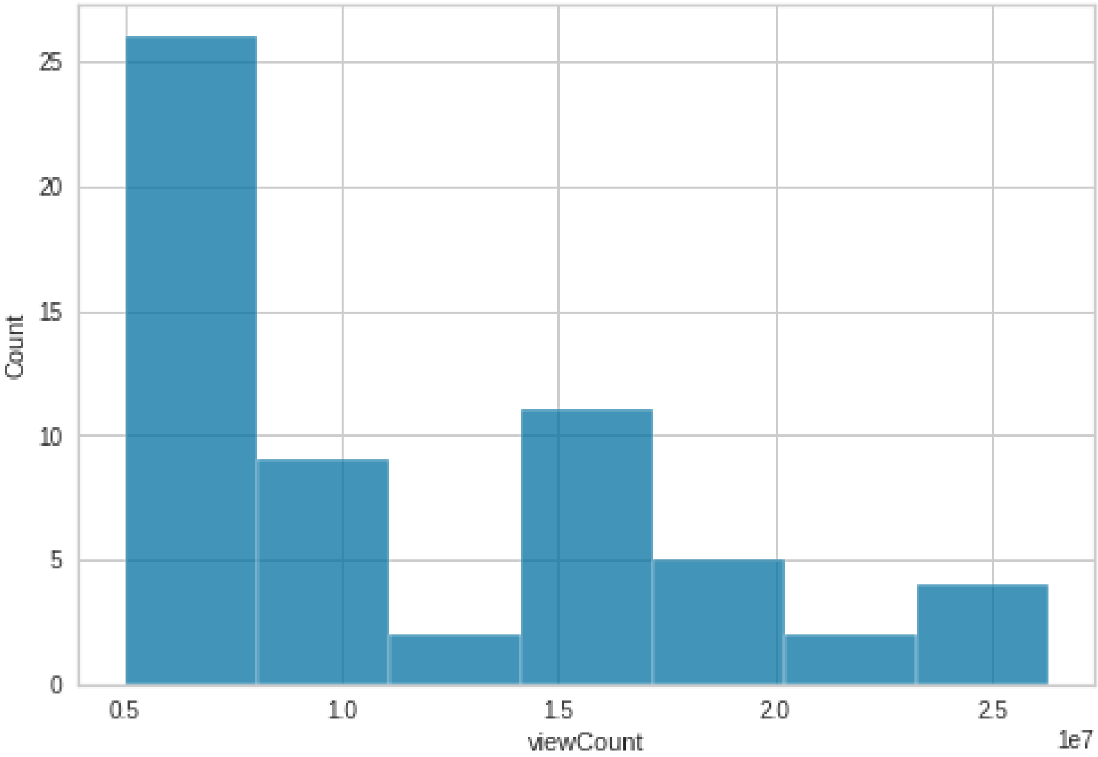
```
sns.histplot(data=cluster_2[cluster_2.viewCount <= 5000000],
x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7aba321910>
```



```
sns.histplot(data=cluster_2[(cluster_2.viewCount > 5000000)],
x="viewCount")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7aba0e5790>
```



```
cluster_2['viewCount'].mean()
```

```
1204758.660011409
```

```
cluster_2['channelTitle'].value_counts()
```

```
Ali-A           870
Syndicate       785
H2ODelirious     49
VanossGaming     48
W2S               1
Name: channelTitle, dtype: int64
```

```
cluster_2['title'].value_counts()
```

```
Black Ops Moon - Quad Commentary Syndicate, WoodysGamerTag, TheAv8tqr
& Quantumzz                      4
Lets Play, Call Of Duty Black Ops, Episode 3 - TheSyndicateProject
2
Black Ops 3 Zombies - Nogla Loses His Place in Heaven (Minecraft Mod)
1
Black Ops 2 Zombies 'Mob Of The Dead' *Afterlife Aggravation! Gameplay
Live w/Syndicate (Part 6)    1
Black Ops 2 Zombies 'Mob Of The Dead' Finale! Gameplay Live
w/Syndicate (Part 7)                    1

..
"ZEN MODE!" - BO2 LIVE w/ Ali-A #13 - Black Ops 2 Multiplayer Gameplay
```

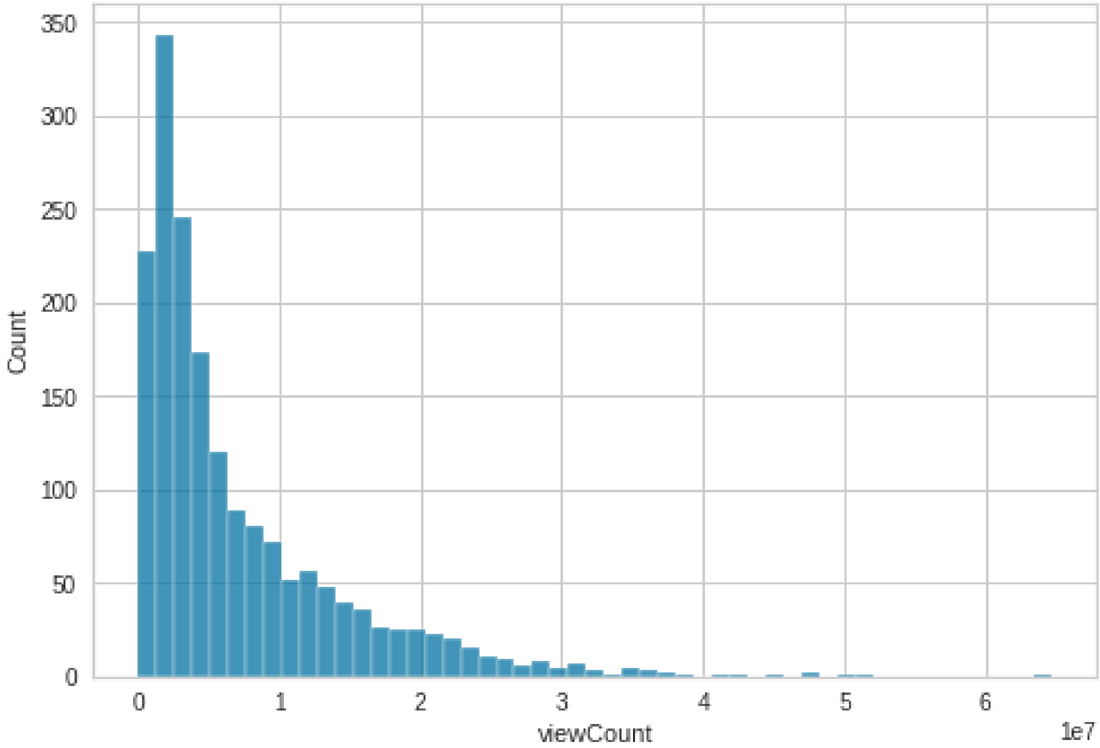
```
1
"CONTINUE...?!" - Die Rise Zombies w/ Ali-A #4 - (Black Ops 2 Zombies
Gameplay)                         1
"☢ NUCLEAR BABY!" - BO2 LIVE w/ Ali-A #14 - Black Ops 2 Multiplayer
Gameplay                            1
"EXCEPTIONAL VIEWERS!" - Die Rise Zombies w/ Ali-A #5 - (Black Ops 2
Zombies Gameplay)                  1
FIFA 16 - BLACK OPS 3 DISCARD CHALLENGE
1
Name: title, Length: 1749, dtype: int64
```

This cluster contains videos about Black ops, zombies mode in Black ops, Call of duty from Ali-A and Syndicate channel. Most of videos have less than 2000000 views and the mean view count is 1204758.

#Cluster 7

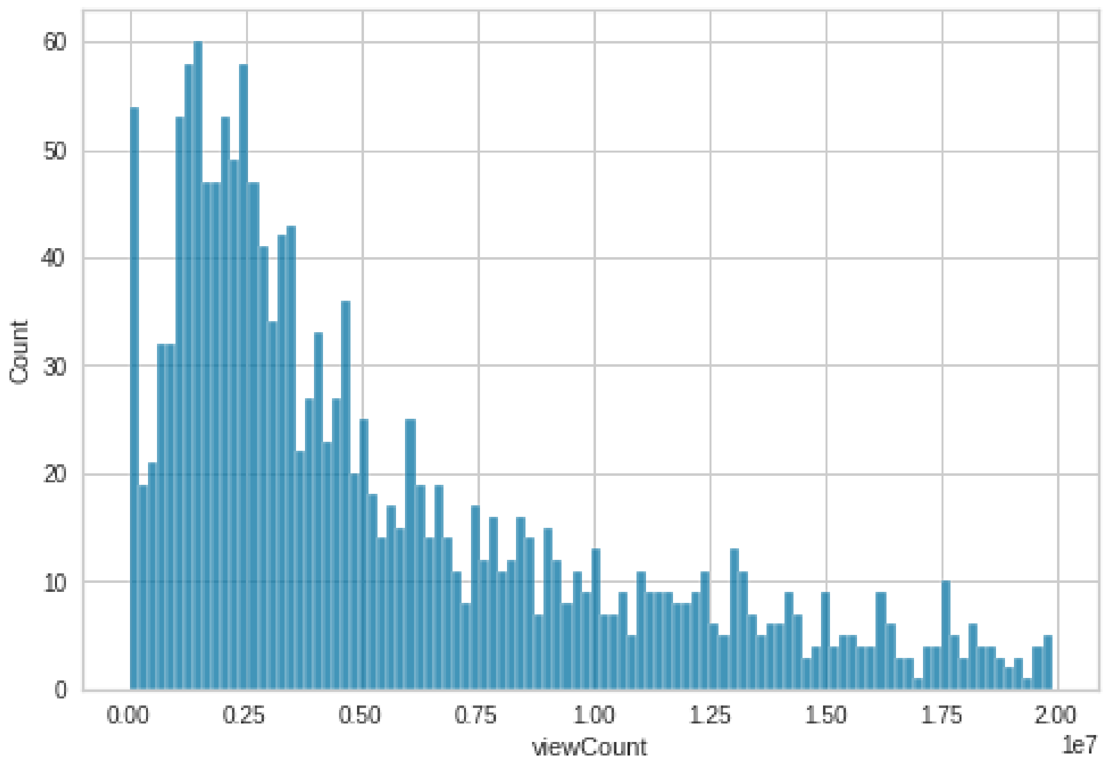
```
sns.histplot(data=cluster_7, x="viewCount")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7aba042a90>
```



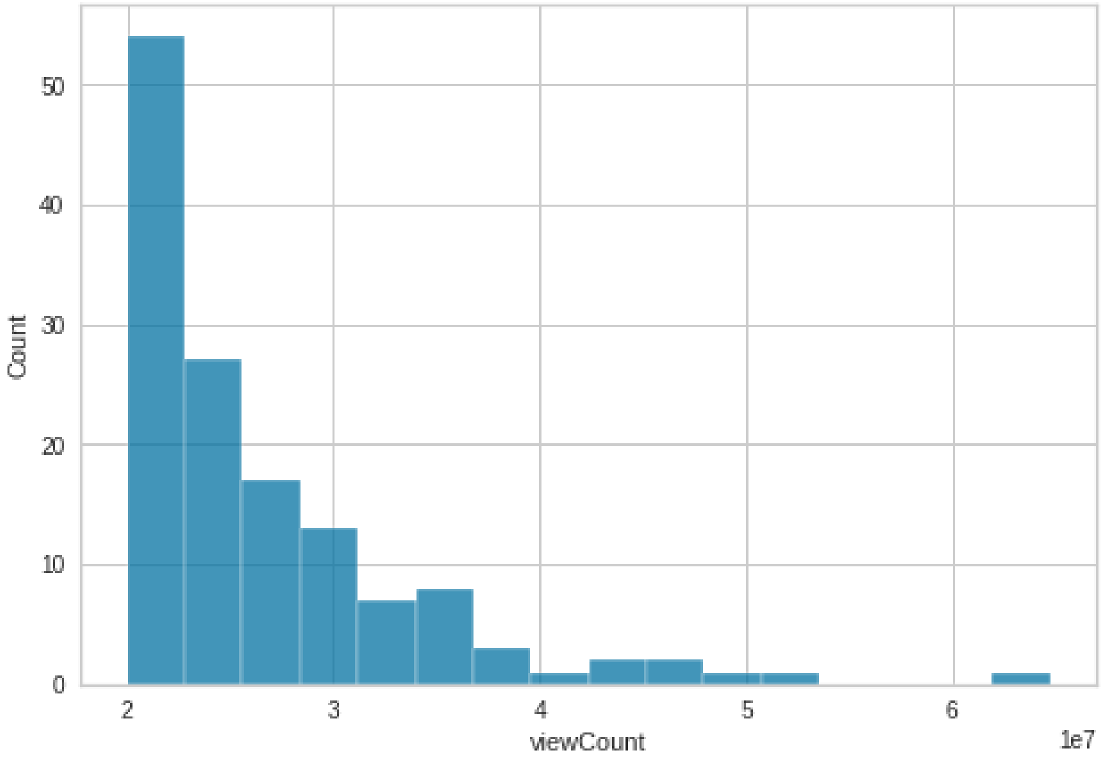
```
sns.histplot(data=cluster_7[cluster_7.viewCount <= 20000000],
x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7abaf03a90>
```



```
sns.histplot(data=cluster_7[(cluster_7.viewCount > 20000000)],
x="viewCount")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab97f2a90>
```



```
cluster_7['viewCount'].mean()
```

```
7184094.312008979
```

```
cluster_7['channelTitle'].value_counts()
```

```
VanossGaming     1272
H2ODelirious      377
jacksepticeye      48
Markiplier         37
DanTDM             15
Syndicate          13
W2S                 8
PopularMMOs         7
Ali-A               5
Name: channelTitle, dtype: int64
```

```
cluster_7['title'].value_counts()
```

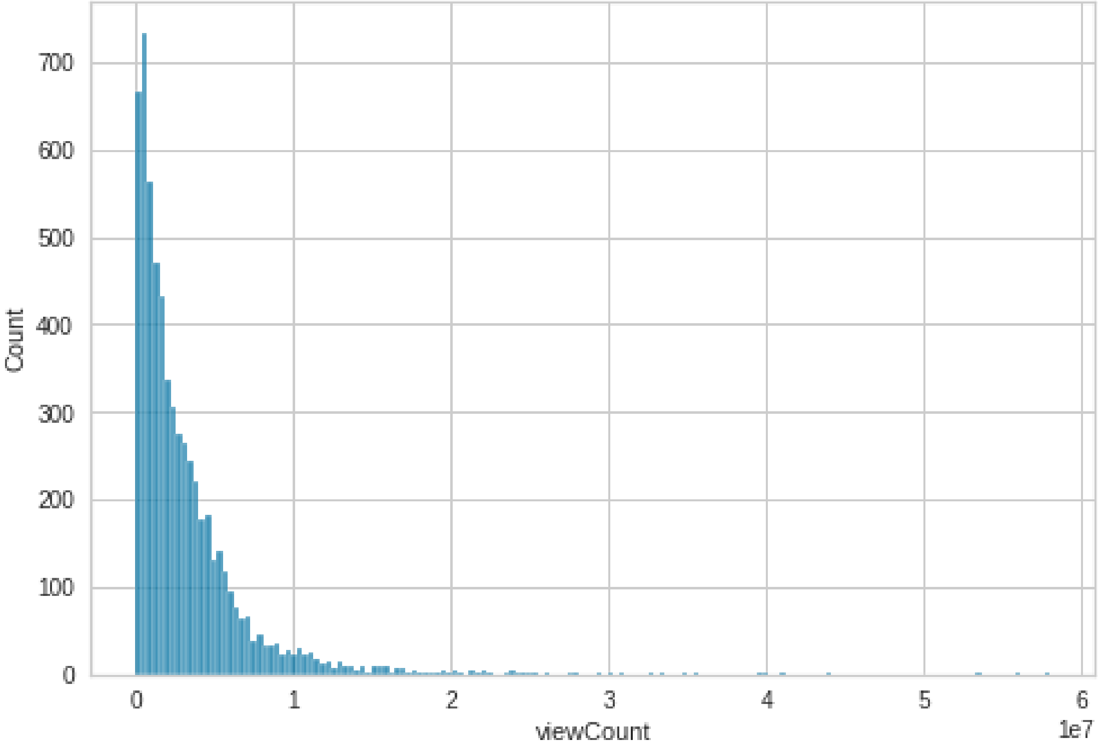
```
Fortnite Funny Moments - Super Anime Moves and Sky Kidnappings!
1
Gmod Sandbox Funny Moments - Fish Tank, Wii Sports, Trippy Maps, Crazy
Bombs! (Garry's Mod)         1
Gmod Minecraft!: Tutorials, Pictionary, Ender Dragon (Garry's Mod
Sandbox Funny Moments & Skits)    1
Gmod Scary Maps - Pull the Schnitzel! (Garry's Mod Funny Moments)
1
Gmod Prop Hunt Funny Moments - Whack a Baby, Kitchen Role Play,
Where's Grandpa?! (Garry's Mod)     1

..
The Experiment: Escape Room Funny Moments - Finding Bodies and
BLUBBAGS!                           1
Fortnite Creative Mode - Raging in a COD Map! (Funny Moments and
Fails)                             1
Uno Funny Moments - Al Duty, National Disaster!
1
GTA5 Online Funny Moments - Lui Fanclub and Demolition Derby!
1
THE BEST MOMENTS OF PINK SLIPS!!
1
Name: title, Length: 1782, dtype: int64
```

This cluster contains videos about Fortnite, Gmod, Gta,... mostly from VanossGaming and H2ODelirious channels. Most of videos have less than 20000000 views and the mean view count is 7184094.

#Cluster 6

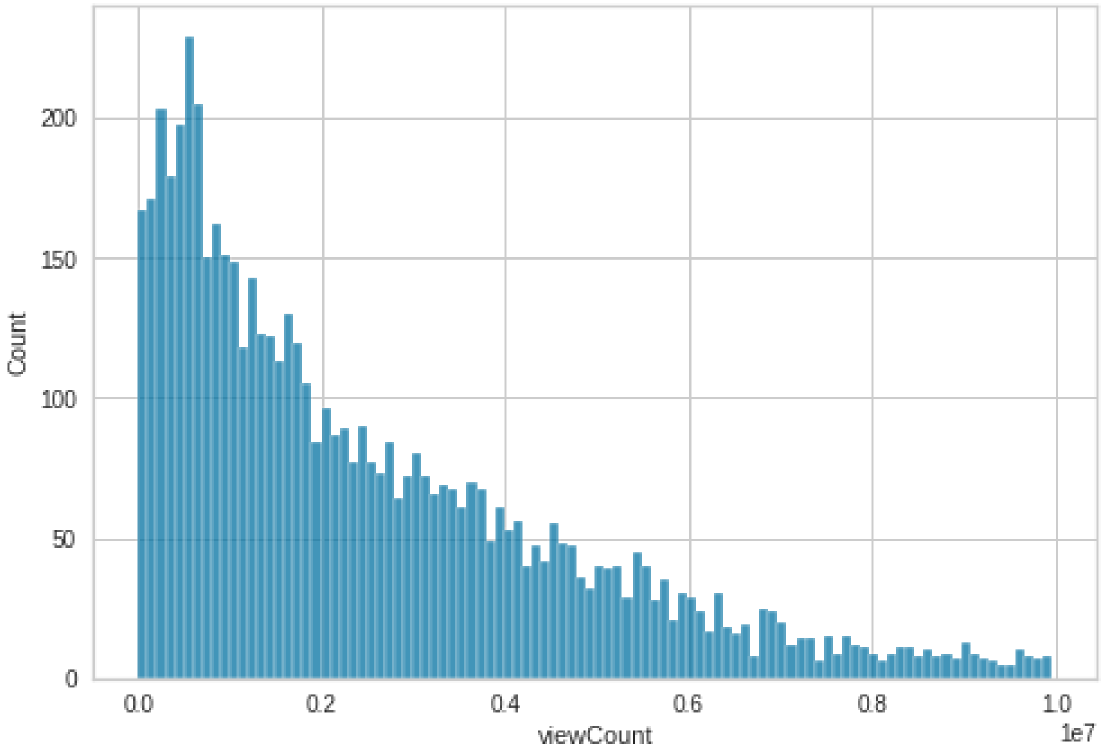
```
sns.histplot(data=cluster_6, x="viewCount")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab9c30df0>
```



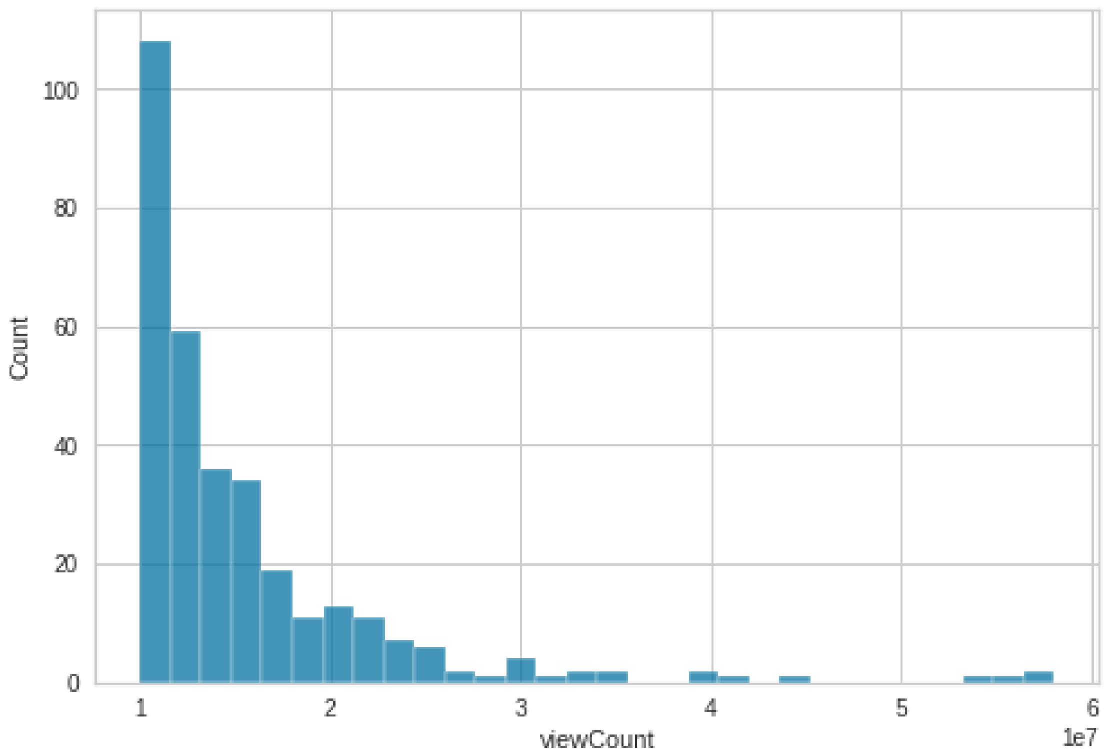
```
sns.histplot(data=cluster_6[cluster_6.viewCount <= 10000000],
x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab9a6e970>
```



```
sns.histplot(data=cluster_6[(cluster_6.viewCount > 10000000)],
x="viewCount")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab98c2940>
```



```
cluster_6['viewCount'].mean()
```

```
3214618.21097322
```

```
cluster_6['channelTitle'].value_counts()
```

```
PopularMMOs      3153
DanTDM           1898
Syndicate         808
Markiplier         95
H2ODelirious       83
jacksepticeye      62
VanossGaming       20
Ali-A               5
Name: channelTitle, dtype: int64
```

```
cluster_6['title'].value_counts()
```

```
The time I pranked Vanoss on Minecraft (Who Remembers?)
2
Minecraft Fan Event Follow Up - Thanks for Coming!
2
Minecraft Prop Hunt - Vanoss is Sofa King!
1
MUTANT YETI VS QUARTZ GOLEM, MUTANT CREEPER, & MUTANT OBSIDIAN GOLEM -
Minecraft Mob Battles - Mods    1
Minecraft: GRAVITY (MOB TRAPS, HOME PROTECTION, & MINI PLANETS!) Mod
Showcase                          1
```

```
..
Minecraft | MUTANT SNOW GOLEM MOD! | They throw presents :3 [1.4.7]
1
Minecraft does TWILIGHT in 30 seconds!
1
Minecraft | OCARINA MOD! | Play the Ocarina of Time from Zelda!
[1.4.7]                                1
Minecraft | PET MOBS MOD! | Befriend Endermen, Zombies, Magma Cubes &
More! [1.4.7]                    1
Lets Play MineCraft !? Episodes - TheSyndicateProject
1
Name: title, Length: 6122, dtype: int64
```

This cluster contains videos about Minecraft mostly from PopularMMos, Syndicate and DanTDM channels. Most of videos have less than 4000000 views and the mean view count is 3214618.

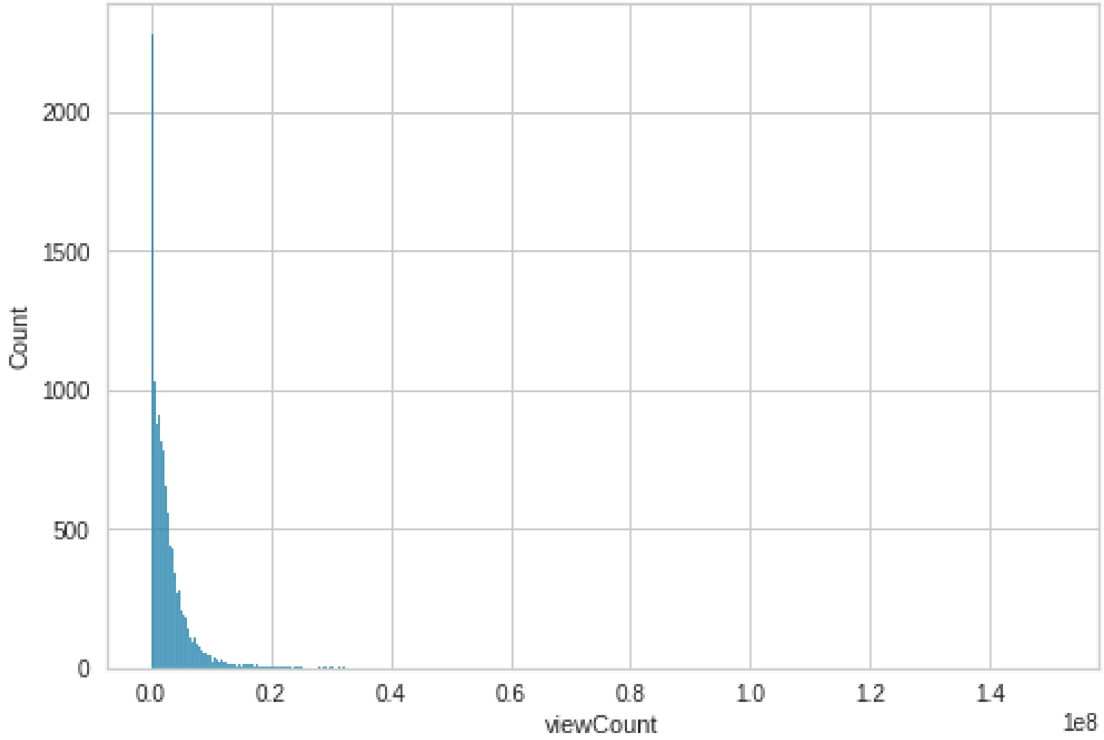
```
X = cluster_1.iloc[:, [2]].values
X = X.tolist()
X = [x[0] for x in X]
# Initialize regex tokenizer
tokenizer = RegexpTokenizer(r'\w+')

# # Vectorize document using TF-IDF
tf_idf_vect = TfidfVectorizer(lowercase=True,
                        stop_words='english',
                        ngram_range = (1,1),
                        tokenizer = tokenizer.tokenize)

# Fit and Transfrom Text Data
X_train_counts = tf_idf_vect.fit_transform(X)
# Import KMeans Model
from sklearn.cluster import KMeans

# Create Kmeans object and fit it to the training data
kmeans = KMeans(n_clusters=4).fit(X_train_counts)

# Get the labels using KMeans
pred_labels = kmeans.labels_
cluster_1['Cluster_1_extend'] = kmeans.labels_
```

```
<ipython-input-134-143d9b8045f3>:23: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  cluster_1['Cluster_1_extend'] = kmeans.labels_
```

#Cluster 1

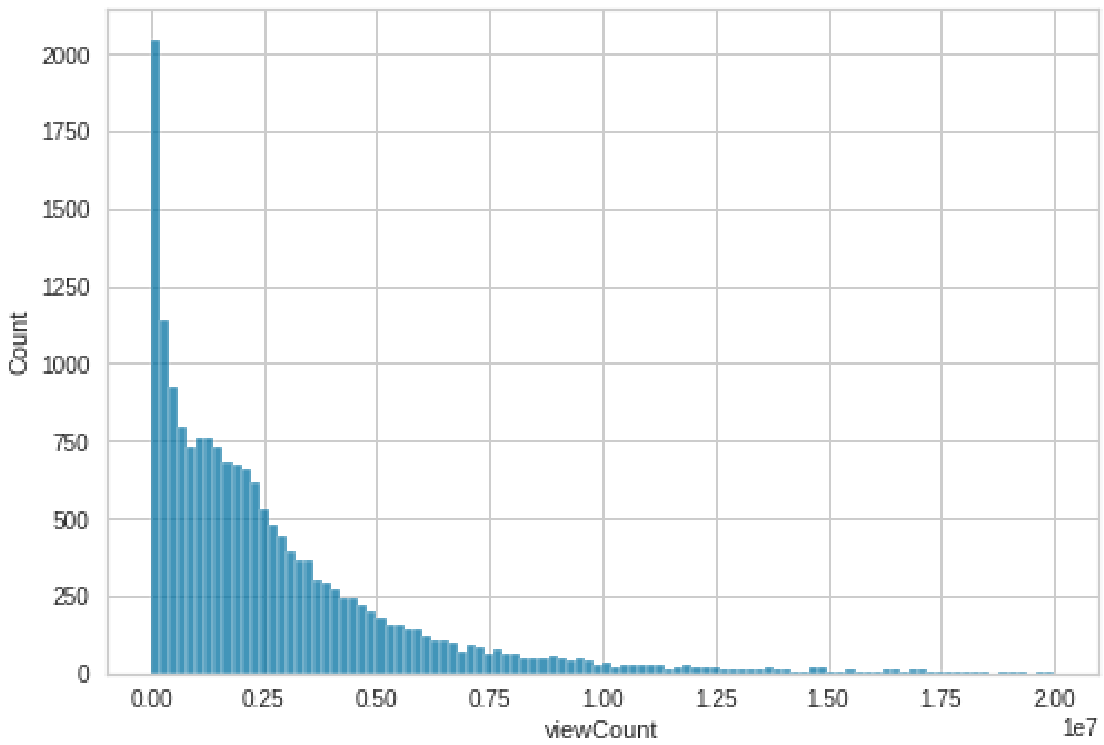
```
sns.histplot(data=cluster_1, x="viewCount")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab9768070>
```



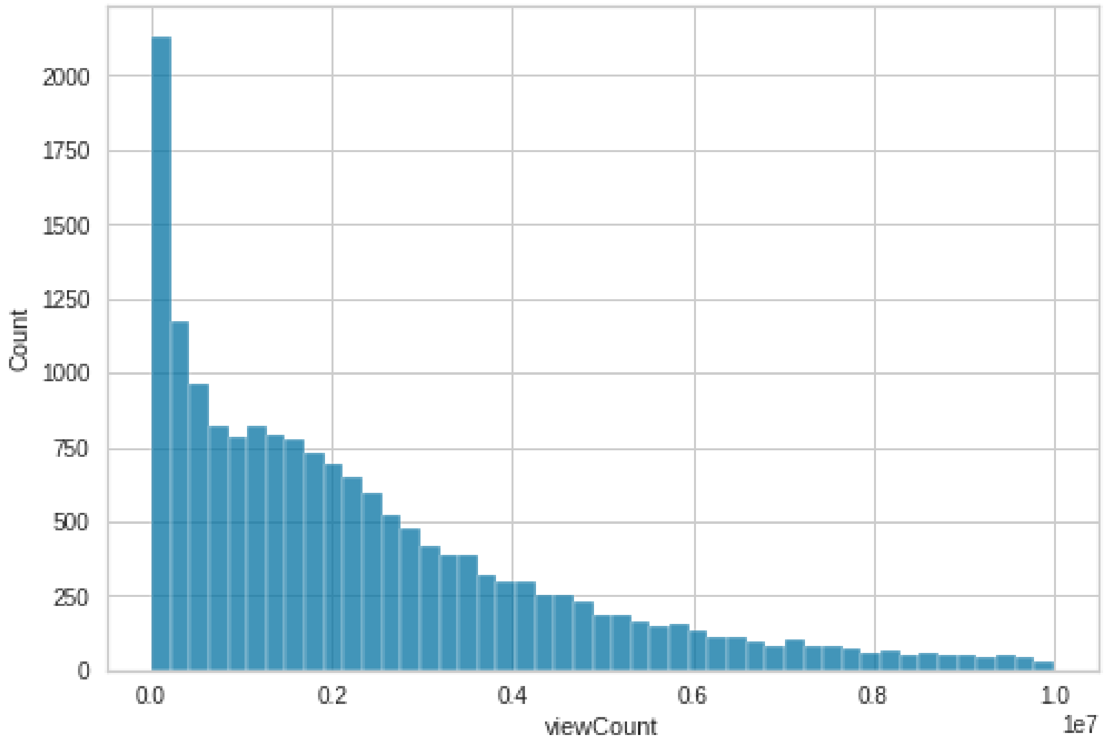
```
sns.histplot(data=cluster_1[cluster_1.viewCount <= 20000000],
x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab9615e20>
```



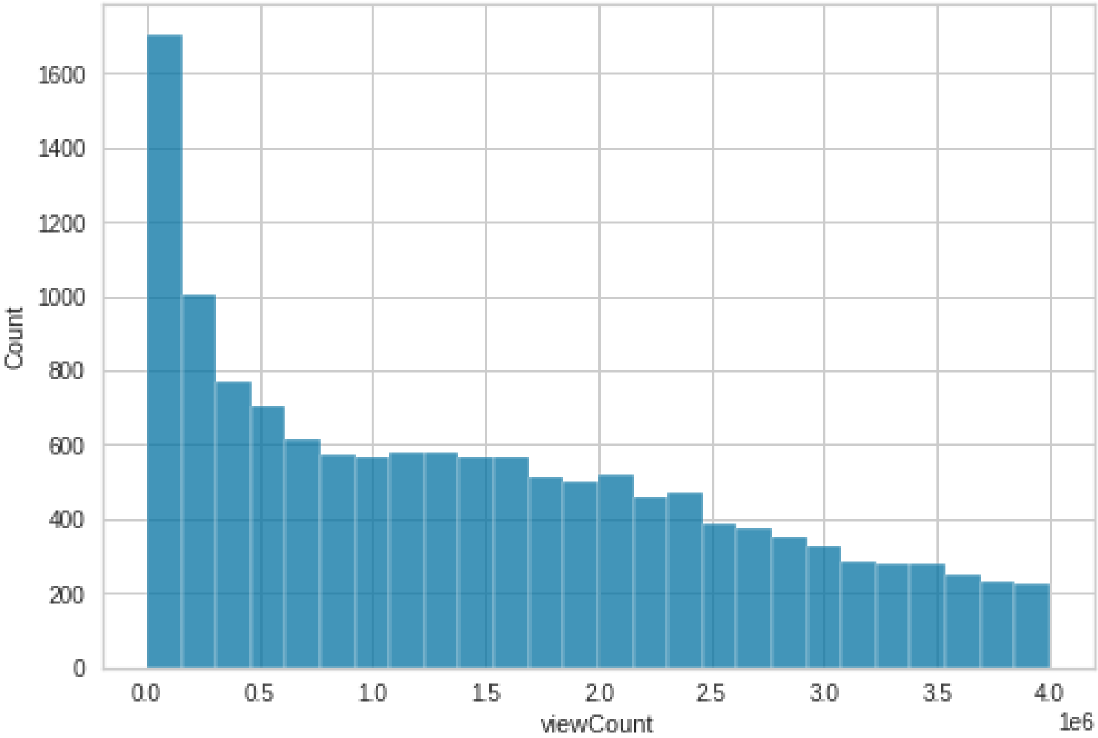
```
sns.histplot(data=cluster_1[(cluster_1.viewCount < 10000000)],
x="viewCount")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab8e21820>
```



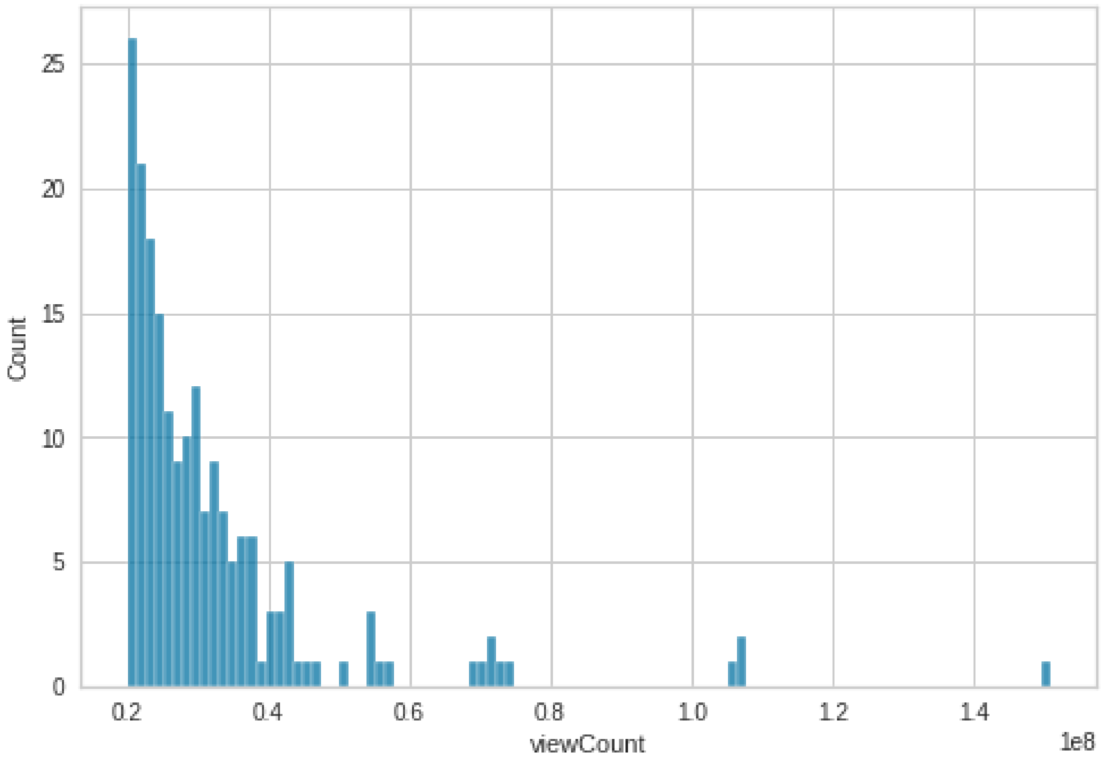
```
sns.histplot(data=cluster_1[(cluster_1.viewCount < 4000000)],
x="viewCount")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab8d77130>
```



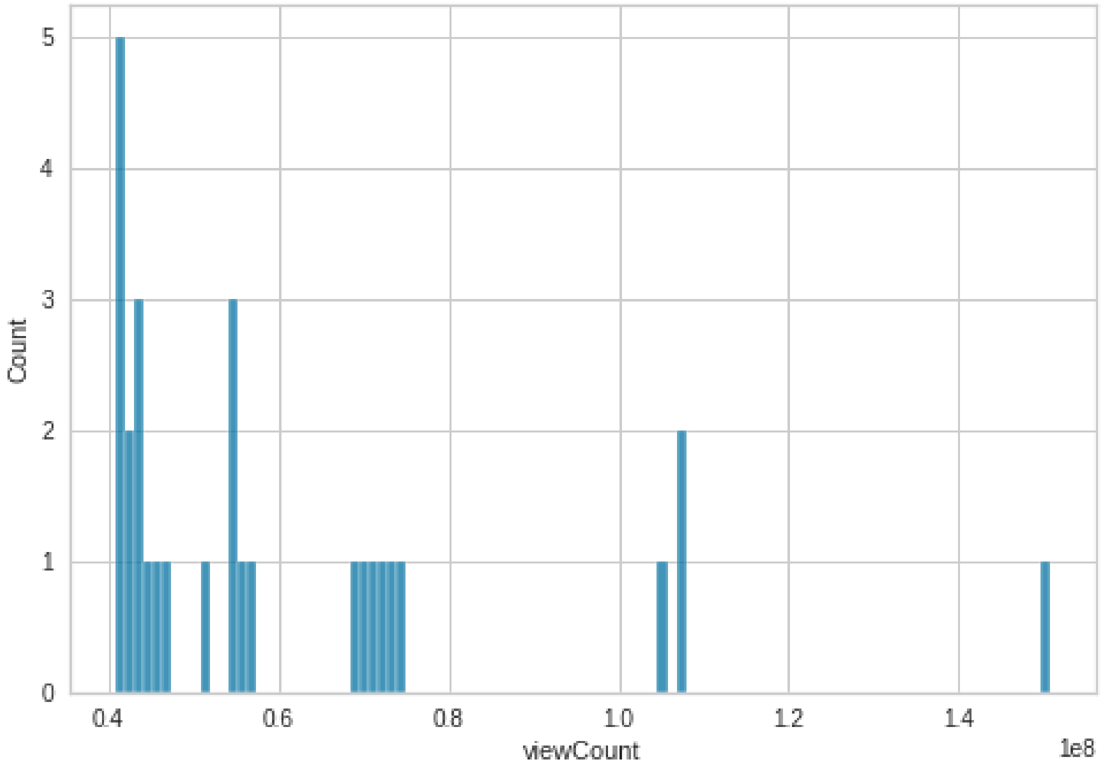
```
sns.histplot(data=cluster_1[cluster_1.viewCount > 20000000],
x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab8c970d0>
```



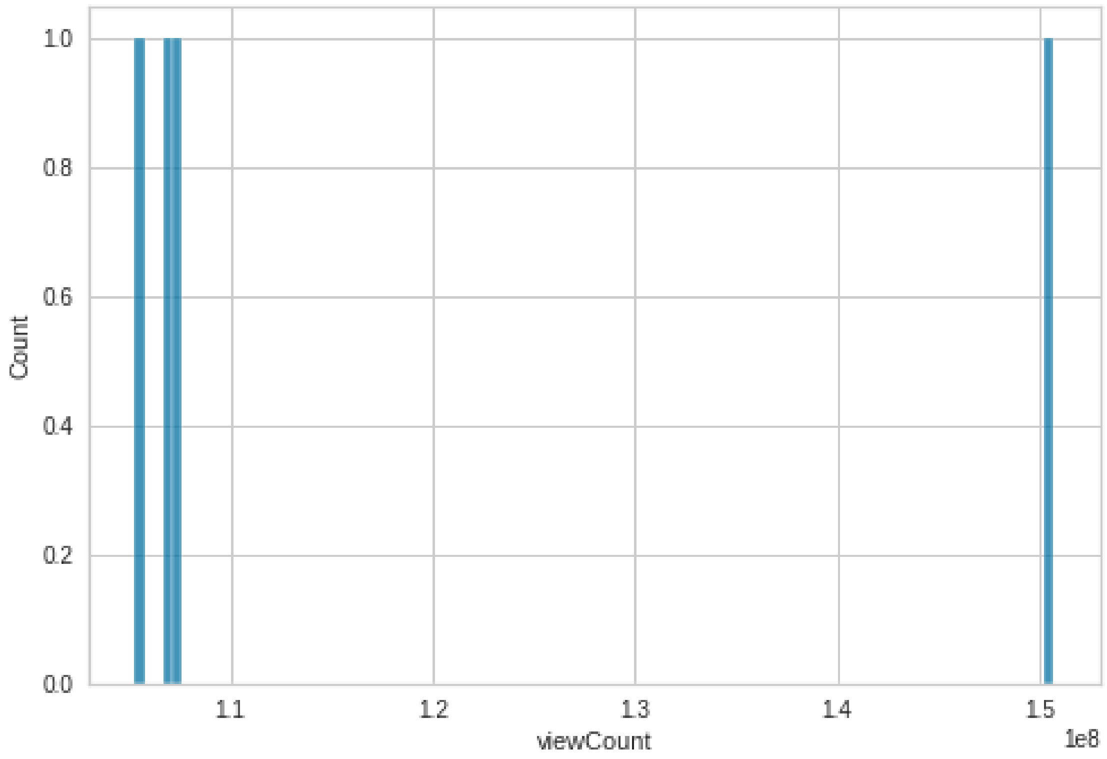
```
sns.histplot(data=cluster_1[cluster_1.viewCount > 40000000],
x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab8a1b400>
```



```
sns.histplot(data=cluster_1[cluster_1.viewCount  > 100000000],
x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab9cb95b0>
```



```
cluster_1['viewCount'].mean()
```

```
3074408.9591493183
```

```
cluster_1['channelTitle'].value_counts()
```

```
Markiplier       5153
jacksepticeye    4830
H2ODelirious     2566
DanTDM           1548
Syndicate        1314
Ali-A            1034
W2S               628
PopularMMOs       542
VanossGaming      206
Name: channelTitle, dtype: int64
```

```
cluster_1['title'].value_counts()
```

```
Thank You                                        5
Funniest TikToks I Could Find                    4
NO WAY...                                        2
Syndicate Visits Universal Studios!              2
The Impossible Quiz Book: Chapter 1              2
                                                ..
WHERE HAS THIS GUN GONE?!?                       1
REAL LIFE NUKETOWN + NEW ZOMBIES!!!              1
ALI - WTF ARE YOU DOING?                         1
```

```
THE NEXT BIG THING IN GAMING! (Ali-A HTC VIVE)    1
Fifa 12 | Trade To Transfer Ep 1 | Podolski       1
Name: title, Length: 17795, dtype: int64
```

This cluster is too large to have a reasonable analysis, we need to split it into smaller clusters to analyse.

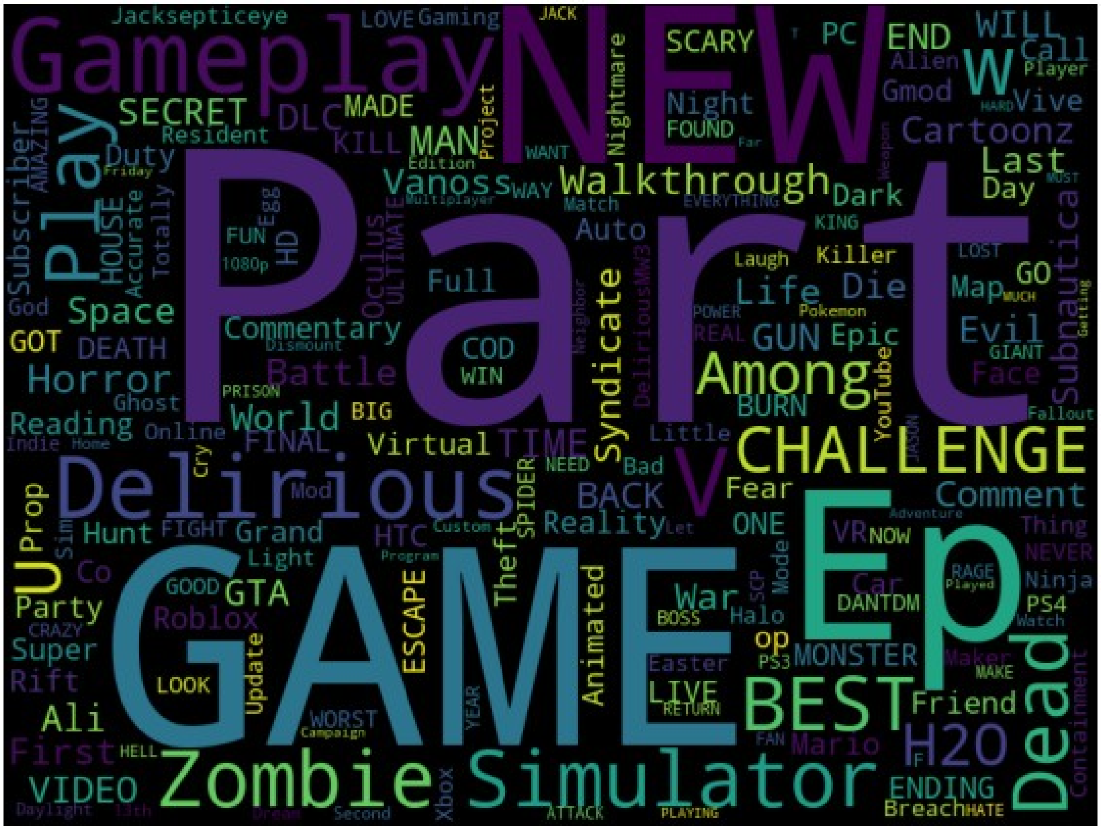
```
cluster_1.to_csv('cluster_1.csv')
cluster_2.to_csv('cluster_2.csv')
cluster_3.to_csv('cluster_3.csv')
cluster_4.to_csv('cluster_4.csv')
cluster_5.to_csv('cluster_5.csv')
cluster_6.to_csv('cluster_6.csv')
cluster_7.to_csv('cluster_7.csv')
cluster_8.to_csv('cluster_8.csv')
```

Cluster 1 - extend

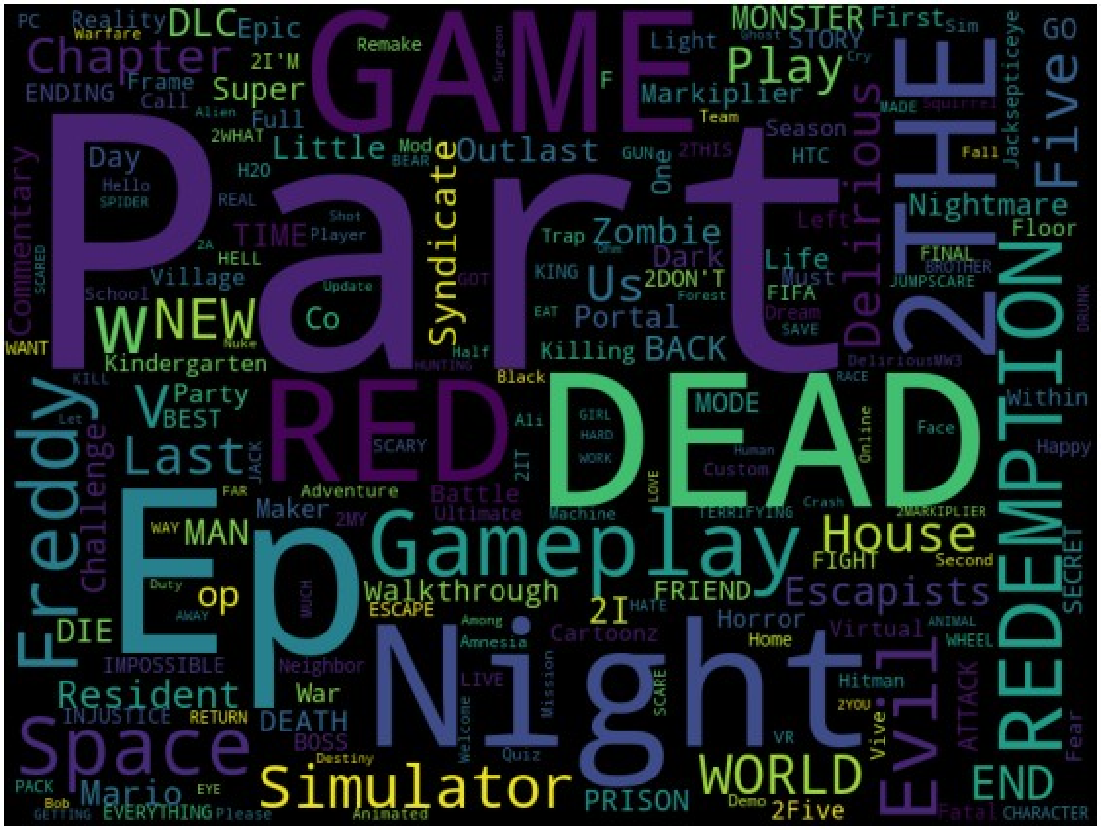
```
cluster_1
```

```
for i in range(4):
    new_df=cluster_1[cluster_1.Cluster_1_extend==i]
    text="".join(new_df.title.tolist())
    wordcloud = WordCloud(width = 800, height = 600, random_state=1, 
background_color='black', 
                          colormap='viridis', 
collocations=False).generate(text)
    plot_cloud(wordcloud,str(i+1))
```

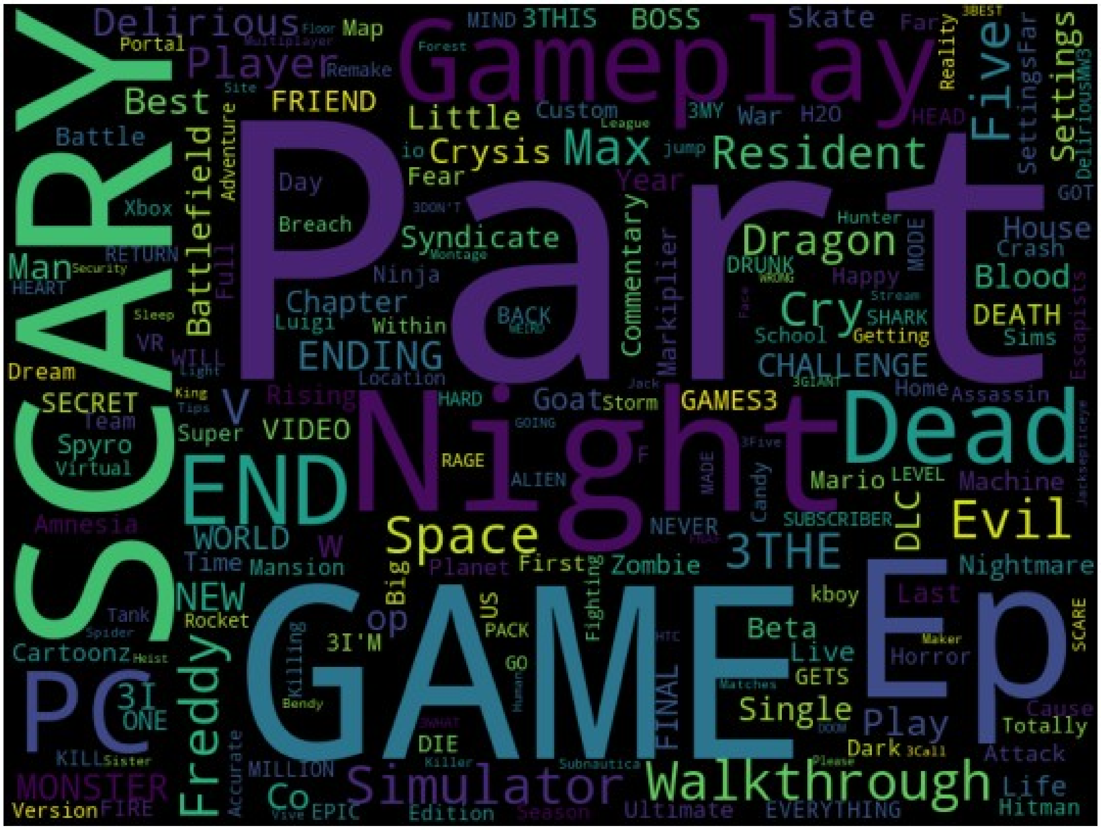
1

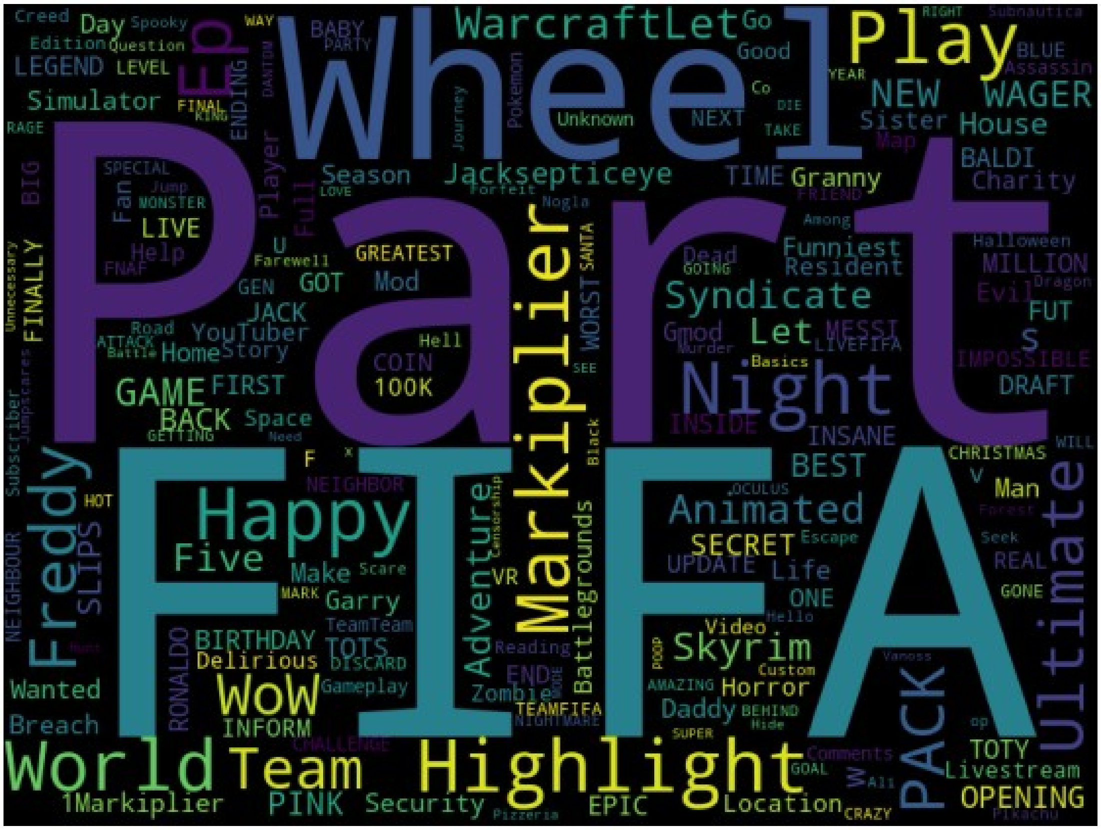

Part
NEW
GAME
Ep
Gameplay
Delirious
Zombie
Simulator
CHALLENGE
BEST
Dead
Among
Play
Walkthrough
H2O
Jacksepticeye
LOVE
Gaming
JACK
SCARY
PC
END
WILL
W
Call
Alien
Player
Gmod
Vive
SECRET
DLC
MADE
MAN
Night
Cartoonz
Nightmare
FOUND
Resident
KILL
Project
WANT
Duty
Subscriber
AMAZING
Edition
Vanoss
WAY
Last
Day
Dark
Subnautica
HOUSE
Accurate
Totally
Egg
FUN
HD
1080p
Oculus
ULTIMATE
Multiplayer
Match
EVERYTHING
KING
Auto
Killer
Laugh
LOST
Full
Life
Die
Map
GO
Space
Commentary
Syndicate
Epic
GUN
Evil
Pokemon
God
GOT
DEATH
Dismount
COD
Horror
Battle
Face
WIN
BURN
PRISON
Ghost
Reading
Online
BIG
World
Virtual
TIME
YouTube
Fallout
Indie
Home
FINAL
Little
Cry
Prop
Mod
SPIDER
NEED
Bad
BACK
Fear
Comment
Hunt
FIGHT
Grand
HTC
Custom
Reality
ONE
Adventure
Thing
Sim
Light
Theft
VR
NOW
NEVER
Co
GOOD
GTA
War
SCP
Mode
Car
RAGE
Played
Ninja
Party
ESCAPE
Animated
DANTDM
PS4
Roblox
Halo
BOSS
Watch
CRAZY
Super
Update
op
MONSTER
Easter
PS3
Maker
MAKE
Rift
LOOK
LIVE
RETURN
Friend
Ali
WORST
YEAR
Containment
First
HELL
Campaign
Mario
FAN
VIDEO
ENDING
Xbox
Daylight
13th
Dream
Second
ATTACK
PLAYING
Breach
HATE

2


Part
GAME
2THE
DEAD
RED
Ep
Night
REDEMPTION
Gameplay
Freddy
Space
Evil
Simulator
Delirious
Chapter
Play
Five
WNEW
Last
WORLD
House
Escapists
Syndicate
Zombie
Us
Outlast
Resident
Mario
END
Challenge
Commentary
Walkthrough
MONSTER
First
STORY
DLC
Epic
Reality
Warfare
PC
Remake
Light
Sim
GO
Surgeon
Ghost
Cry
Jacksepticeye
Super
Markiplier
F
Frame
Call
ENDING
2I'M
Full
Alien
Team
MADE
Squirrel
Little
Mod
Season
2WHAT
Day
BEAR
GUN
2THIS
HTC
Fall
H2O
One
Nightmare
Hello
SPIDER
REAL
Floor
TIME
Shot
Player
Trap
Ohm
Left
Dark
Life
2A
HELL
FINAL
BROTHER
Village
KING
2DON'T
Must
JUMPSCARE
SCARED
School
Update
GOT
Forest
FIFA
Co
Portal
Dream
DRUNK
EAT
SAVE
WANT
Nuke
Kindergarten
Half
HUNTING
Killing
BACK
DelirousMW3
KILL
Black
RACE
V
Party
BEST
MODE
Within
Let
Ali
GIRL
HARD
Online
Human
Face
JACK
SCARY
WORK
Custom
Happy
Adventure
2IT
LOVE
FAR
Battle
Maker
2MY
Ultimate
Machine
TERRIFYING
Crash
2MARKIPLIER
WAY
MAN
FIGHT
Second
SECRET
op
Duty
MUCH
2I
FRIEND
ESCAPE
HATE
AWAY
Among
Horror
ANIMAL
DIE
Cartoonz
Virtual
Amnesia
Home
WHEEL
IMPOSSIBLE
Neighbor
LIVE
Mission
SCARE
ATTACK
Hitman
War
Welcome
INJUSTICE
RETURN
DEATH
2YOU
Fear
BOSS
VR
Vive
Destiny
Quiz
PACK
EYE
Bob
Demo
2Five
GETTING
EVERYTHING
Please
Animated
Fatal
CHARACTER
PRISON

3


Part
GAME
Night
SCARY
Ep
Gameplay
Dead
END
PC
Walkthrough
Simulator
Space
Evil
Freddy
3THE
Resident
Dragon
Cry
Max
Crysis
Five
Delirious
Player
Best
FRIEND
Little
Syndicate
Battlefield
Chapter
ENDING
CHALLENGE
Blood
House
DEATH
Settings
SettingsFar
Skate
BOSS
3THIS
MIND
Map
Portal
Battle
Man
Xbox
Breach
Day
Fear
Year
Commentary
Markiplier
Ninja
Luigi
Within
BACK
DRUNK
Happy
Crash
Hunter
MODE
SHARK
School
Getting
Sims
Escapists
Dream
SECRET
Team
Spyro
Virtual
Amnesia
WORLD
Time
Mansion
NEW
Rising
VIDEO
Super
Goat
Storm
GAMES3
Location
HARD
RAGE
ALIEN
Mario
Machine
LEVEL
Candy
DLC
SUBSCRIBER
NEVER
Zombie
First
Planet
Nightmare
Last
kboy
Beta
Live
Horror
Single
Play
GETS
Dark
Attack
Life
Hitman
Totally
Cause
SCARE
Cartoonz
Tank
Rocket
3I'M
ONE
3I
MONSTER
KILL
Sister
FIRE
Version
Accurate
MILLION
DIE
Killing
Bendy
EPIC
Edition
Season
Ultimate
EVERYTHING
FINAL
Fighting
PACK
GO
US
Big
op
Co
V
W
Subnautica
Killer
HTC
Matches
Human
Maker
3Call
Please
DOOM
Adventure
Full
RETURN
HEART
Security
Sleep
VR
WILL
King
Tips
Light
Floor
Multiplayer
Remake
Forest
Site
io
Custom
League
jump
3MY
War
H2O
HEAD
Far
Reality
3BEST
DeliriousMW3
GOT
Montage
WRONG
Stream
Face
3DON'T
Jack
3GIANT
Home
Assassin
GOING
MADE
F
3Five
Jacksepticeye
Vive
Heist
Spider
3WHAT
WEIRD



```
#number of instances in each clusters
cluster_1['Cluster_1_extend'].value_counts()
```

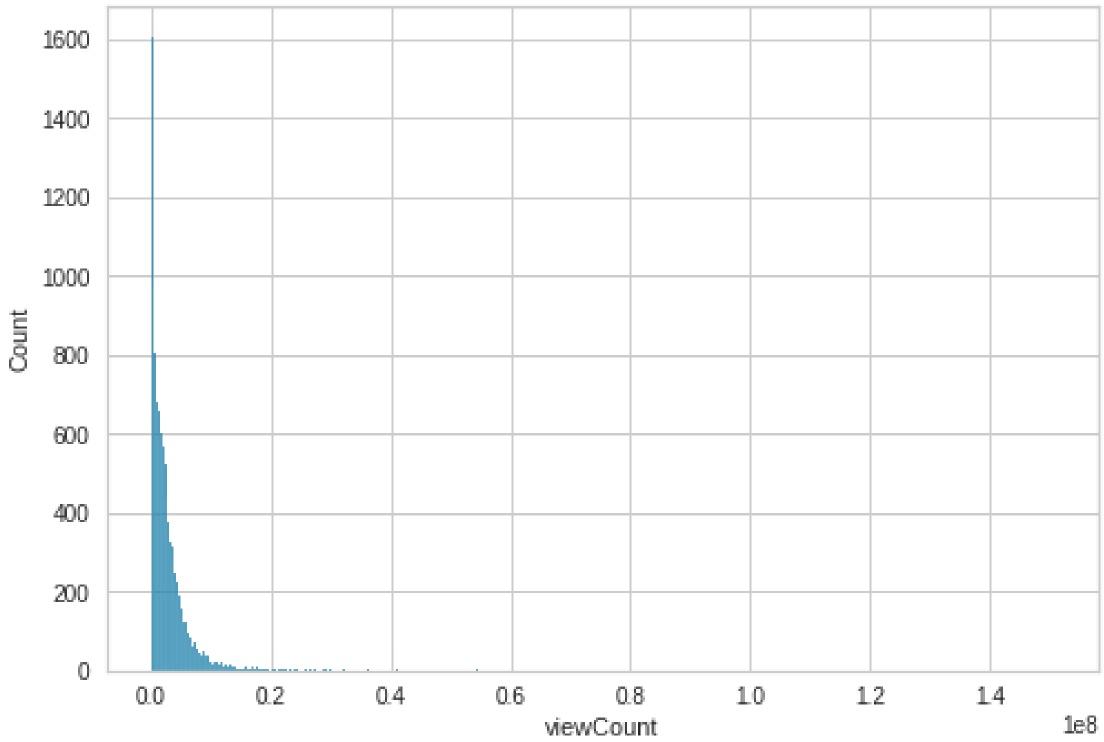
```
0    12252
3     2470
1     1774
2     1325
Name: Cluster_1_extend, dtype: int64
```

```
# assign cluster
cluster_1_extend = cluster_1[cluster_1.Cluster_1_extend == 0]
cluster_2_extend = cluster_1[cluster_1.Cluster_1_extend == 1]
cluster_3_extend = cluster_1[cluster_1.Cluster_1_extend == 2]
cluster_4_extend = cluster_1[cluster_1.Cluster_1_extend == 3]
```

##Cluster 1

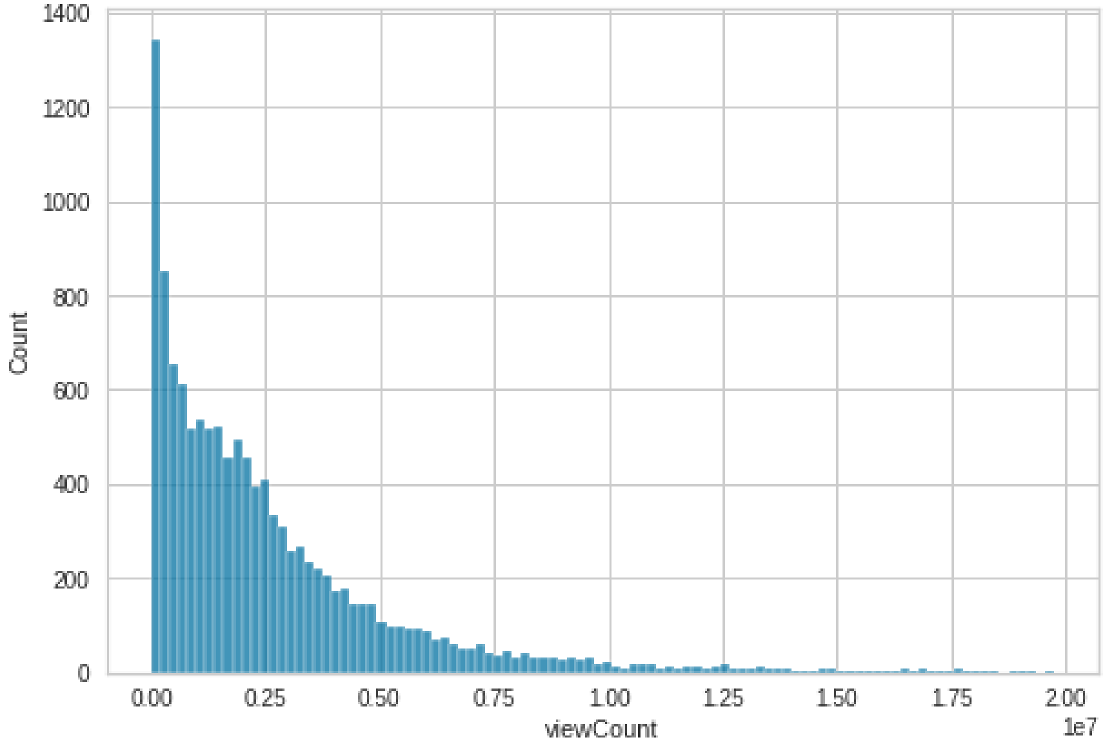
```
sns.histplot(data=cluster_1_extend, x="viewCount")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab860ac70>
```



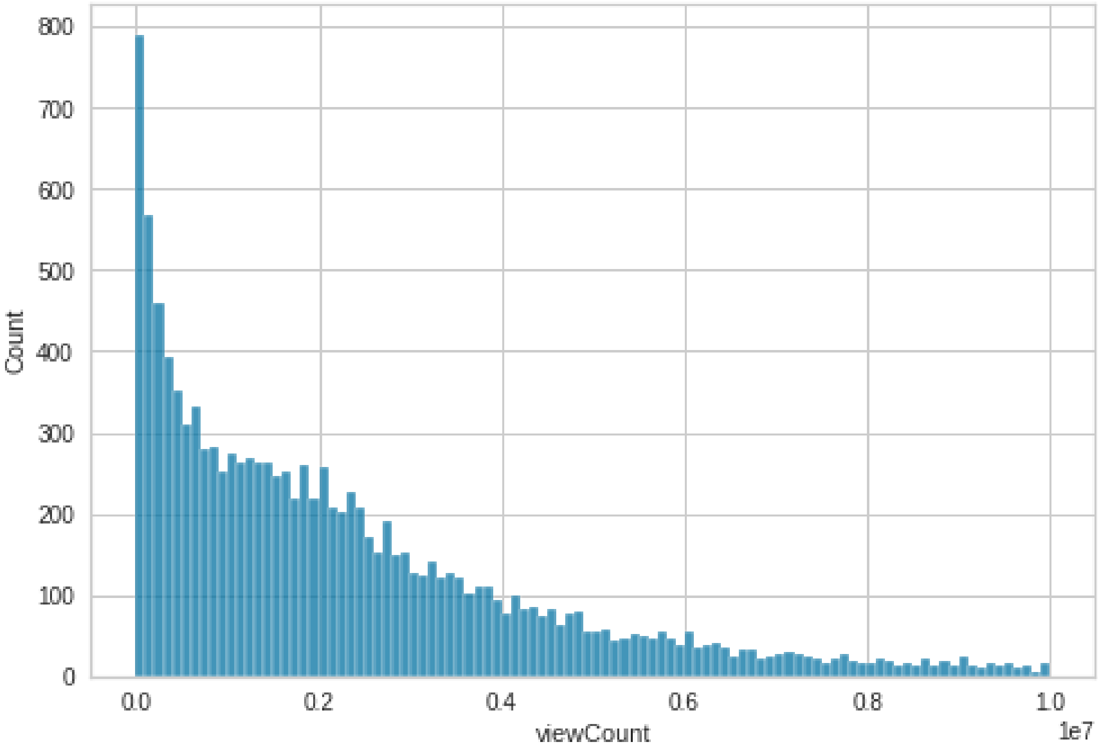
```
sns.histplot(data=cluster_1_extend[cluster_1_extend.viewCount <=
20000000], x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab7b40430>
```



```
sns.histplot(data=cluster_1_extend[cluster_1_extend.viewCount <=
10000000], x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab78e95e0>
```



```
cluster_1_extend['viewCount'].mean()
```

```
2905253.1233268036
```

```
cluster_1_extend['channelTitle'].value_counts()
```

```
jacksepticeye    3255
Markiplier       3238
H2ODelirious     2005
DanTDM           1142
Syndicate        1039
Ali-A             886
PopularMMOs       360
W2S               176
VanossGaming      151
Name: channelTitle, dtype: int64
```

```
cluster_1_extend['title'].value_counts()
```

```
Thank You                                                     5
Funniest TikToks I Could Find                                 4
Donating To Smaller Streamers                                 2
it is my birthday                                             2
```

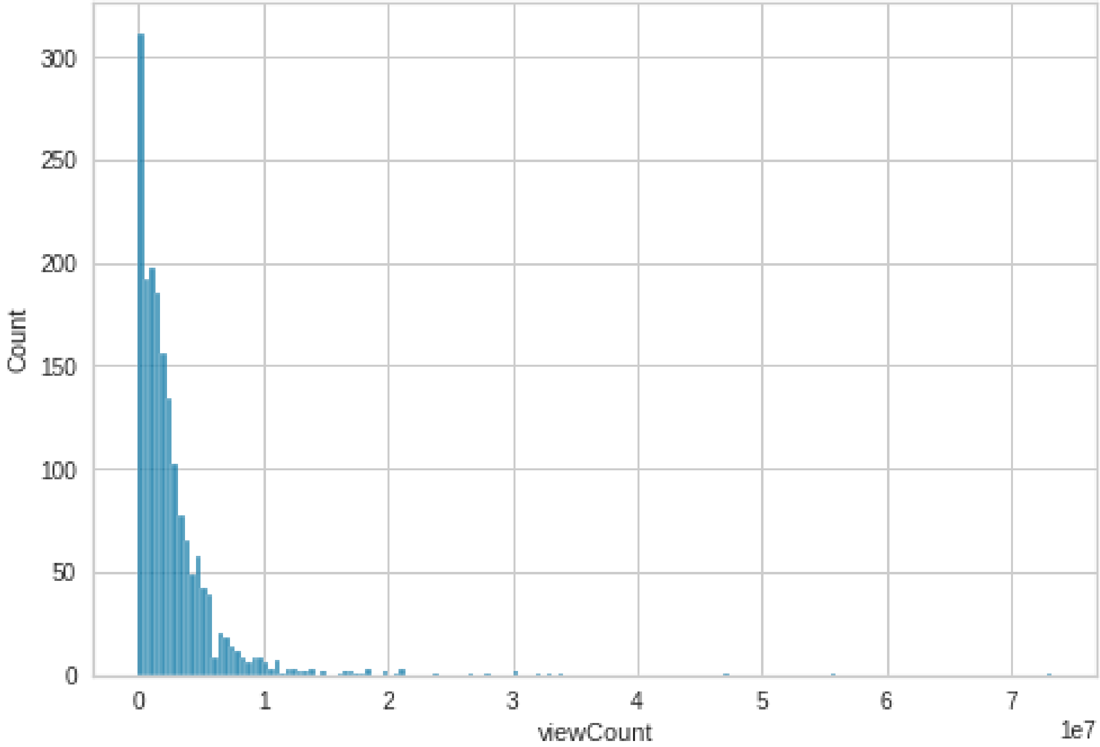
```
Prison Simulator                                            2
                                                           ..
I SEE YOUR COMMENTS! :) ❤                                  1
"STOP - POLICE!" (Battlefield Hardline Gameplay w/ Ali-A)   1
GAMING = A CHORE...? (Ali-A Q&A)                            1
FRHANK SHOPS FOR EXPRESSIONS... (w/ Ali-A)                  1
Fifa 12 | Trade To Transfer Ep 1 | Podolski                 1
Name: title, Length: 12230, dtype: int64
```

This cluster contains not only videos games but types of videos like lifestyle, Q&A,... Most videos are below 10000000 and the mean view count is 2905253.

##Cluster 2

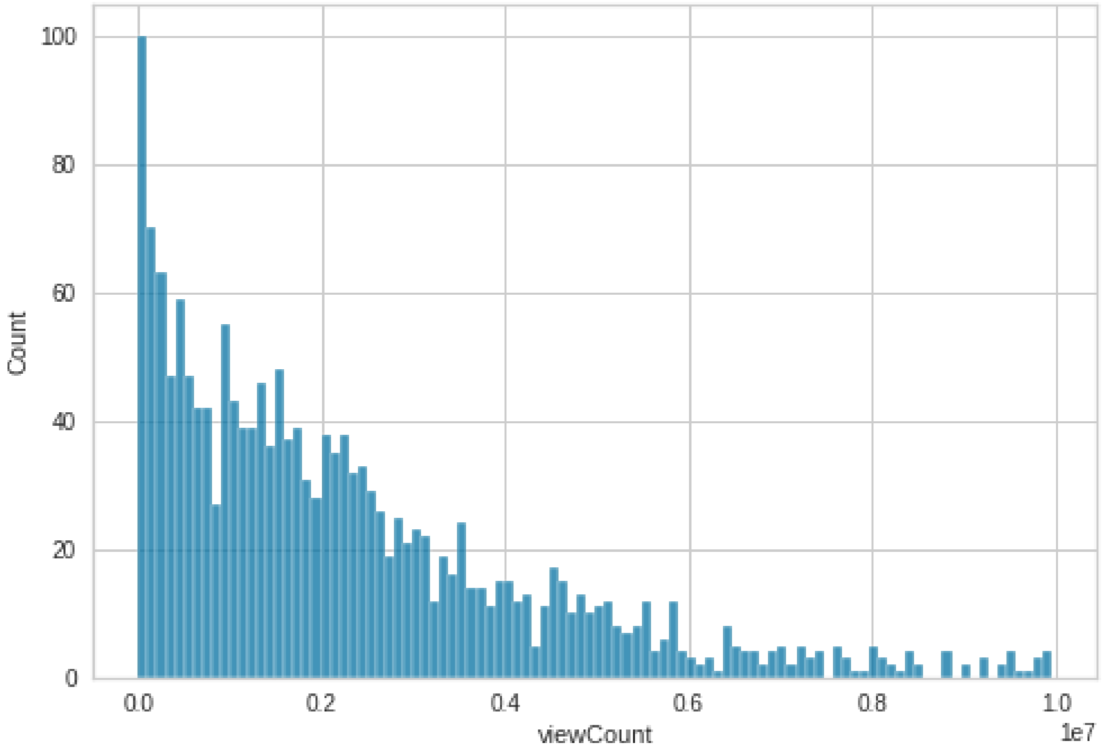
```
sns.histplot(data=cluster_2_extend, x="viewCount")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab775eeb0>
```



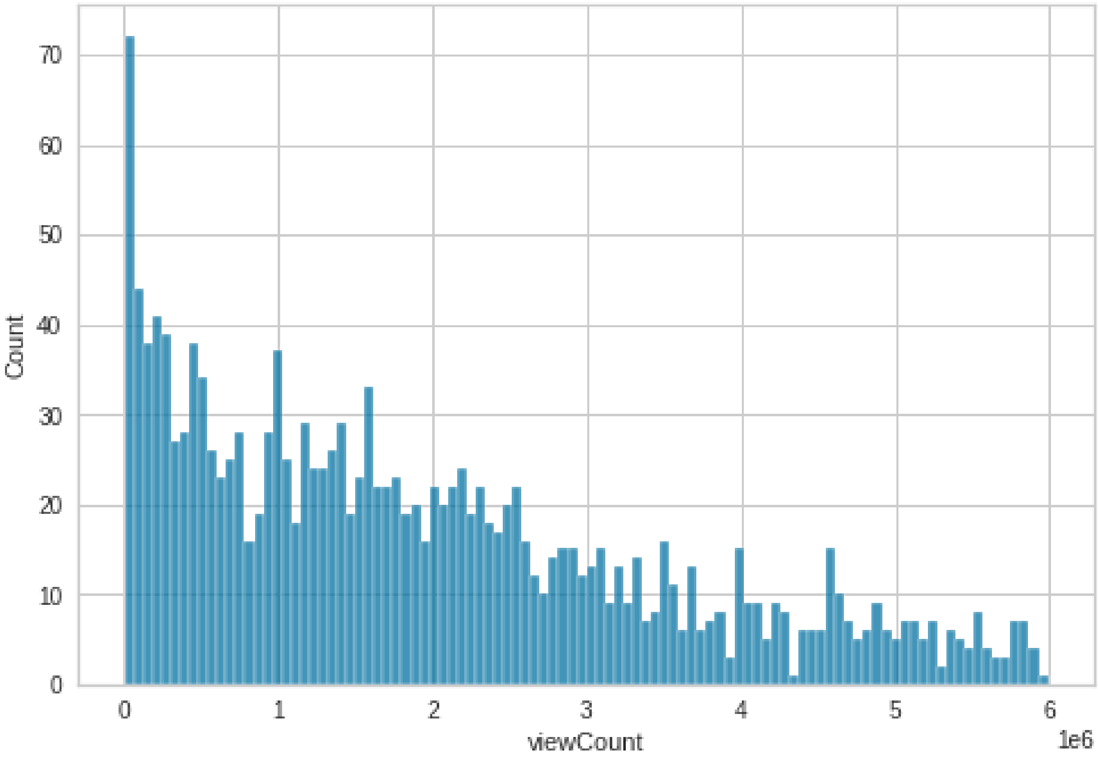
```
sns.histplot(data=cluster_2_extend[cluster_2_extend.viewCount <=
10000000], x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab74019a0>
```



```
sns.histplot(data=cluster_2_extend[cluster_2_extend.viewCount <=
6000000], x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab710f400>
```



```
cluster_2_extend['viewCount'].mean()
```

```
2826628.4272829765
```

```
cluster_2_extend['channelTitle'].value_counts()
```

```
jacksepticeye    646
Markiplier       542
H2ODelirious     281
DanTDM           125
Syndicate         80
PopularMMOs       36
Ali-A             34
W2S               24
VanossGaming       6
Name: channelTitle, dtype: int64
```

```
cluster_2_extend['title'].value_counts()
```

```
Killing Floor 2 - Helping Santalirious Save Christmas!
1
FATHER SON BOAT TRIP | Human Fall Flat w/Robin #2
1
KISSING SOME EGGS | Everybodys Golf #2
1
GETTING FRUSTRATED | Cuphead - Part 2
1
THE HERO WE NEED | South Park: The Fractured But Whole - Part 2
```

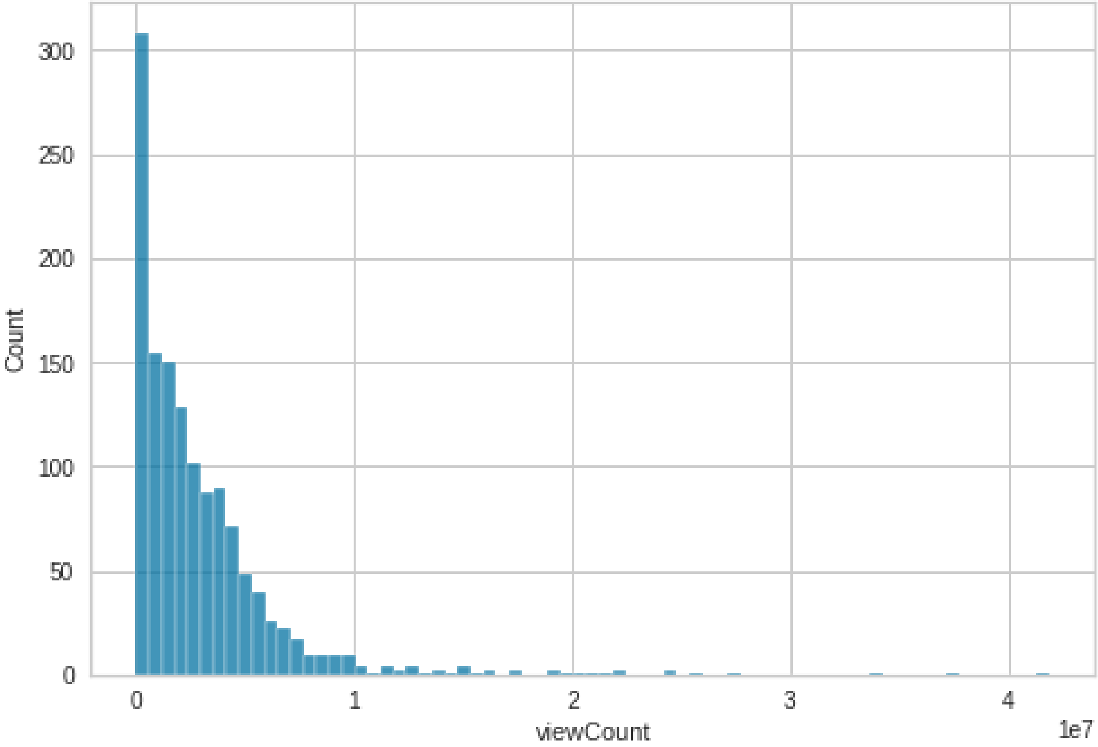
```
1

..
EVIL RABBITS EVERYWHERE - Happy Game - Ep. 2
1
H2ODelirious Plays Dying Light 2 👀
1
MOMMY LONG LEGS IS NIGHTMARE FUEL! | *NEW* Poppy Playtime CHAPTER 2
1
Saving Buzz From Mommy D Luffy! (Poppy Playtime Chapter 2 Modded)
1
Fifa 12 | Trade To Transfer Ep 2 | Lavezzi
1
Name: title, Length: 1774, dtype: int64
```

The cluster has videos of various games: freedy, redemption, simulator from numerous channels: jacksepticeye, Markiplier,... Most videos are below 4000000 views. Mean view count is 2826628 views.

##Cluster 3

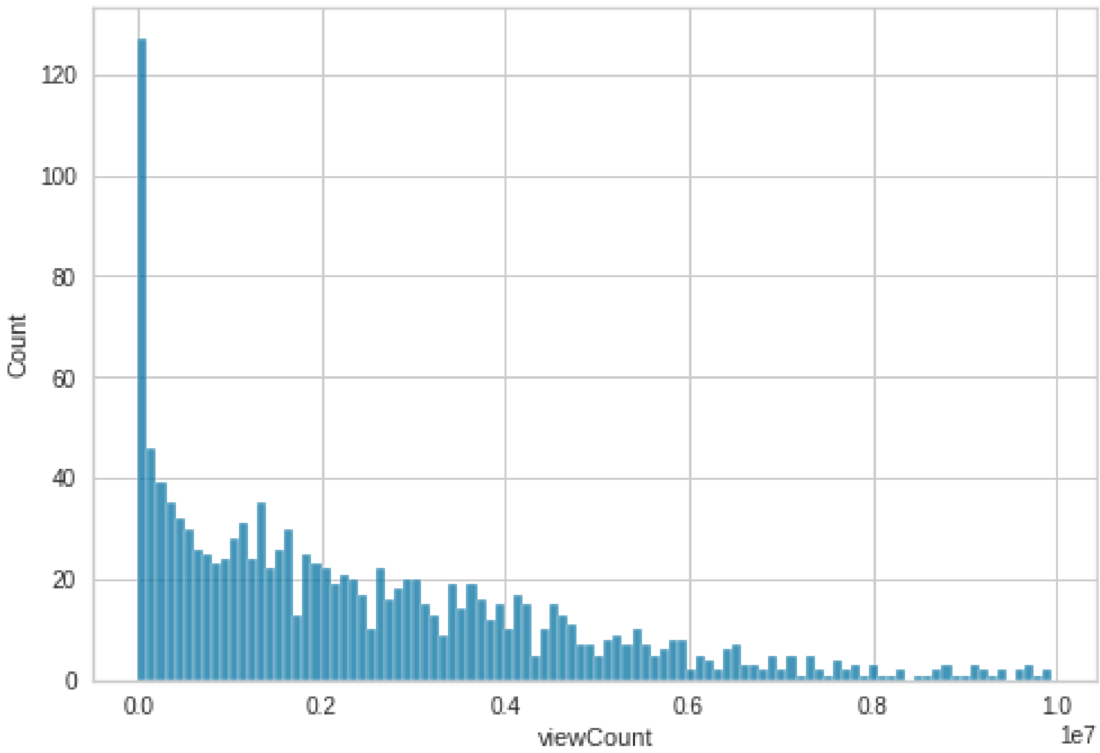
```
sns.histplot(data=cluster_3_extend, x="viewCount")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab6fed8e0>
```



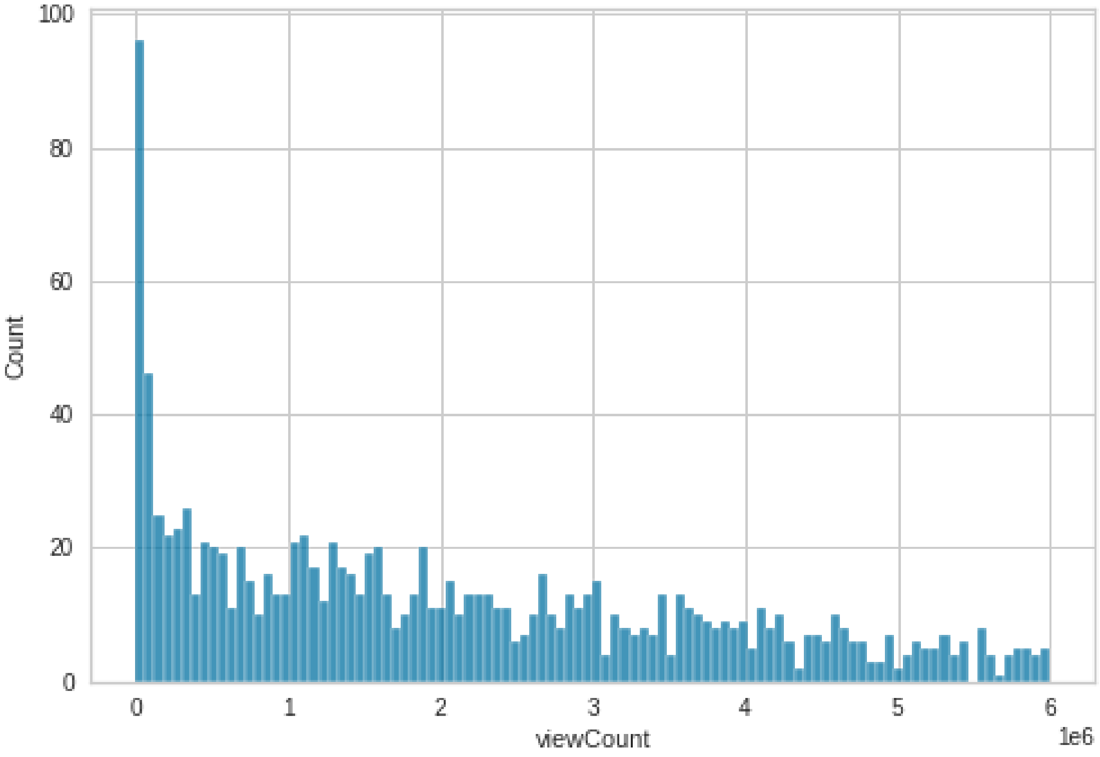
```
sns.histplot(data=cluster_3_extend[cluster_3_extend.viewCount <=
10000000], x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab6d4f490>
```



```
sns.histplot(data=cluster_3_extend[cluster_3_extend.viewCount <=
6000000], x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab6c2f0d0>
```



```
cluster_3_extend['viewCount'].mean()
```

```
2933744.795471698
```

```
cluster_3_extend['channelTitle'].value_counts()
```

```
jacksepticeye    513
Markiplier       463
H2ODelirious     151
DanTDM            80
Syndicate         59
Ali-A             23
PopularMMOs       17
W2S               15
VanossGaming       4
Name: channelTitle, dtype: int64
```

```
cluster_3_extend['title'].value_counts()
```

```
Goat Simulator 3 - Corpse Launching, The Floor is Lava and Demo Derby!
1
JUST TRY AND HIT ME | Mr. Shifty #3
1
50,000 CHICKENS VS. NUCLEAR BOMB | Ultimate Epic Battle Simulator #3
1
I'M THE JUDGE | Peace Death #3
1
BUTCHERED AND BEATEN!! | Outlast 2 - Part 3
```

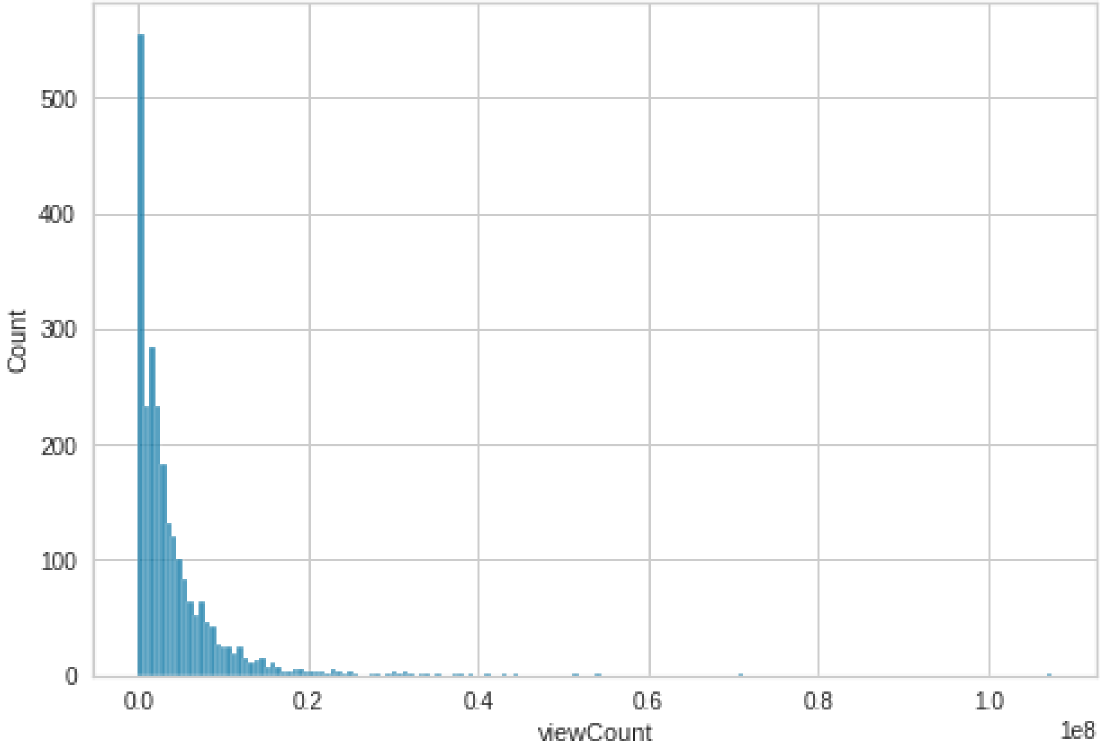
```
1

..
Poisonous | Part 3 | MAKING BACON
1
Disponentia | Part 3 | DEADLY DOGGIES!
1
SCP Containment Breach | Part 3 | BEYOND TERROR
1
Through the Portal | Part 3 | NOSTALGIA
1
Fifa 12 | Trade To Team Ep 3 | Finishing Touches
1
Name: title, Length: 1325, dtype: int64
```

The cluster has videos mostly about horror games from numerous channels but mainly: jacksepticeye, Markiplier,... Most videos are below 4000000 views. Mean view count is 2933745 views.

##Cluster 4

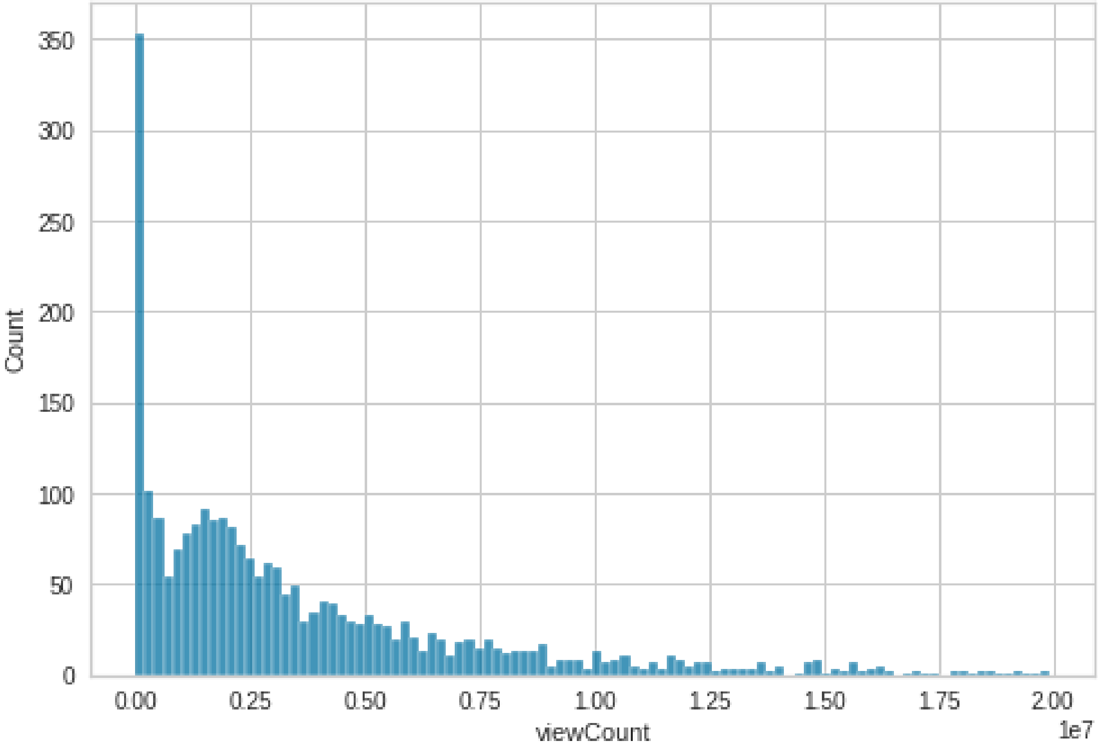
```
sns.histplot(data=cluster_4_extend, x="viewCount")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab613d850>
```



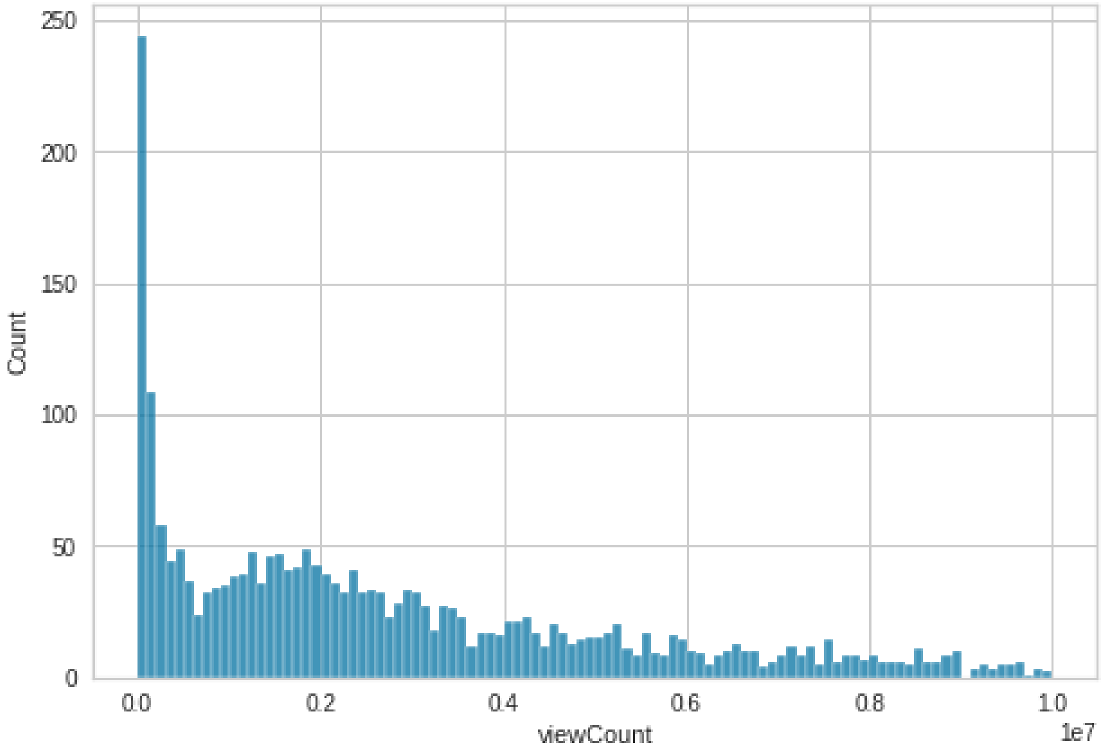
```
sns.histplot(data=cluster_4_extend[cluster_4_extend.viewCount <=
20000000], x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab6379100>
```



```
sns.histplot(data=cluster_4_extend[cluster_4_extend.viewCount <=
10000000], x="viewCount", bins=100)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7ab6319cd0>
```



```
cluster_4_extend['viewCount'].mean()
```

```
4166894.7813765183
```

```
cluster_4_extend['channelTitle'].value_counts()
```

```
Markiplier       910
jacksepticeye    416
W2S              413
DanTDM           201
Syndicate        136
H2ODelirious     129
PopularMMOs      129
Ali-A             91
VanossGaming      45
Name: channelTitle, dtype: int64
```

```
cluster_4_extend['title'].value_counts()
```

```
It's GONE...                                                    2
PINK SLIPS!! - WIN A COPY OF FIFA 14 - Fifa 13 Ultimate Team    2
Happy's Humble Burger Barn                                      2
IT'S BACK!                                                      2
YOU DON'T NEED LEGS! | Happy Wheels - Part 69                   1
                                                               ..
Markiplier Mail #6                                              1
Surgeon Simulator 2013 | Part 4 | MARKIPLIER LOSES HIS MIND!!   1
SCP Containment Breach | Part 30 | NEW SCP'S + ZOMBIES!!        1
```

```
Happy Wheels Highlights #15                                       1
Fifa 12 | Farewell Fifa Ep 1 | Starting Out                       1
Name: title, Length: 2466, dtype: int64
```

The cluster has videos mostly about fifa, happy wheels games from 3 main channels : jacksepticeye, Markiplier, W2S. Most videos are below 6000000 views. Mean view count is 4166895 views.

```
cluster_1_extend.to_csv('cluster_1_extend.csv')
cluster_2_extend.to_csv('cluster_2_extend.csv')
cluster_3_extend.to_csv('cluster_3_extend.csv')
cluster_4_extend.to_csv('cluster_4_extend.csv')
```