

HELLO

I'm Nurul Liyana

PROJECT PROPOSAL

TikTok videos from top 5 content creators in 9 selected countries

CONTENTS

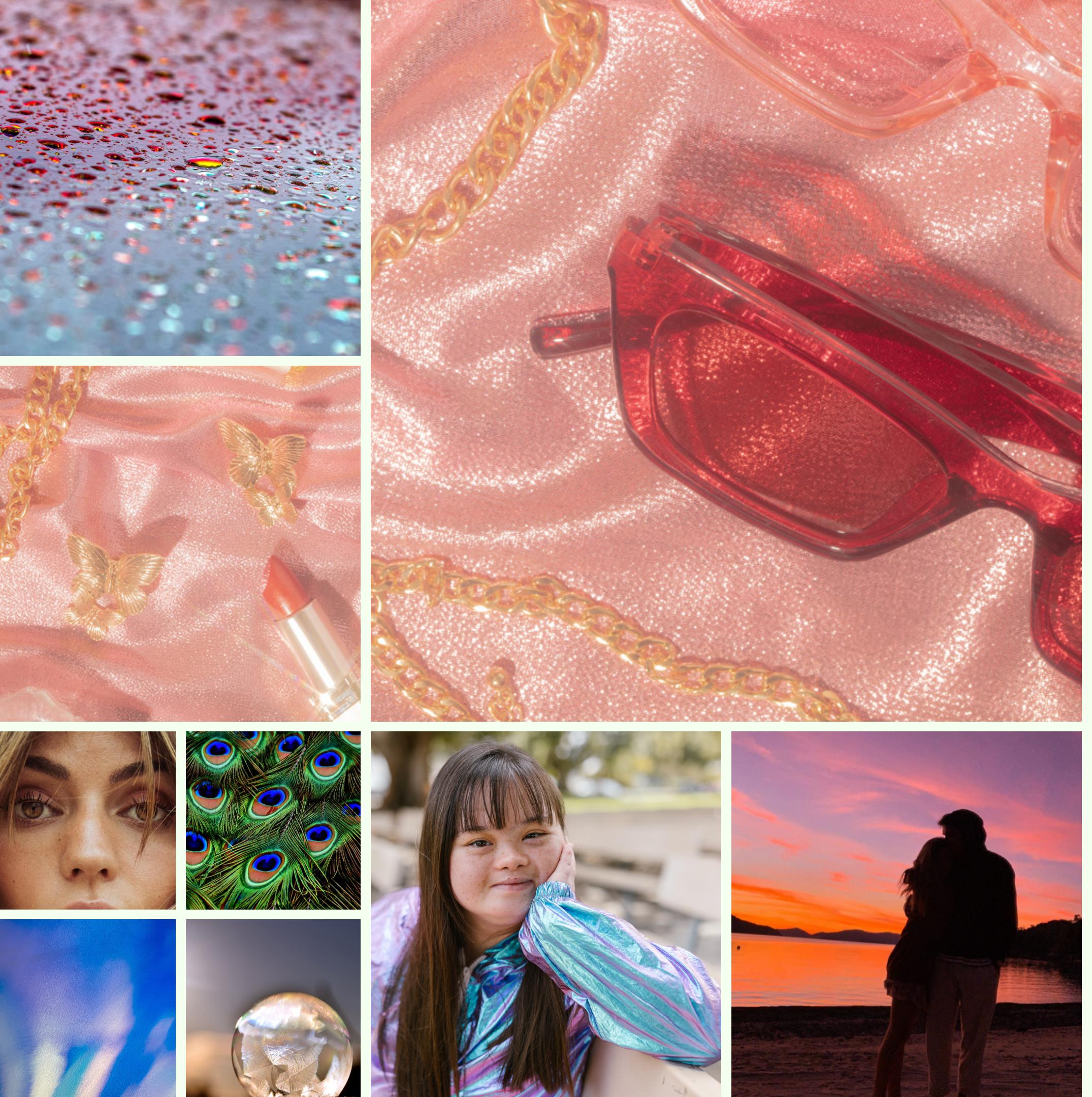
DATA SET

1. Chosen Dataset
2. Source of Data
3. Records found in Dataset
4. Dataset.shape()
5. Important columns
6. Sample of Data
7. Challenges I faced

GOALS

1. Questionism
2. Points

DATA SET



DATASET. CHOSEN DATASET

Most of my time is used to watching TikTok videos. They are educational, entertaining and escaping. Found this dataset as '**The top 5 creators of 9 countries**' in TikTok. Relates to my daily routine of watching them.



DATASET. SOURCE OF DATA

<https://www.kaggle.com/datasets/didymusne/tiktok-videos-from-top-5-creators>

The user's inspired by 'What sort of content in TikTok makes people happy?'



RECORDS FOUND IN DATASET

USER_NAME

User name of the user who posted the video

USER_ID

The ID of the user recorded

VIDEO_ID

The ID of the video posted

VIDEO_DESC

The description or caption of the video, written by the user

VIDEO_TIME

The time of posting of the video in UTC format

VIDEO_LENGTH

The length of the video in seconds

VIDEO_LINK

The URL link to the video

N_LIKES

The number of likes received by the video

N SHARES

The number of shares received by the video

N_COMMENTS

The number of comments received by the video

N_PLAYS

The number of plays recorded by the video

VIDEO_TIMESTAMP

The date of posting of the video, converted from UTC

COUNTRY

The country of the user who posted the video

YEAR

The year of posting the video

YEAR_2

The year of posting the video

VIDEO_ID_SHORTENED

The shortened ID of the video posted



DATASET. SHAPE()

43746

NUMBER OF ROWS

16

NUMBER OF
COLUMNS

2809

NUMBER OF
MISSING VALUES

DATA
TYPES

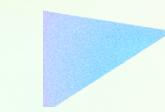
- OBJECT
- FLOAT64
- INT64

DATA SET. IMPORTANT COLUMNS

01

USER_NAME

Get to know the 5 popular accounts in the chosen country



02

VIDEO_DESC

Get to know what the caption from the user is telling about the video which leads to viewers want to like/share/comment/play



03

**N_LIKES/
N SHARES/
N COMMENTS/
N PLAYS**

Get to know the number of likes/shares/ comments/plays of the user to see the different ways of viewers reacting to the video



04

VIDEO_TIMESTAMP

Get to know when user uploaded the video to see if it affects the number of viewers watching

05

COUNTRY

Get to know where the top 5 users is from.



topfive

user_name	user_id	video_id	video_desc	video_time	video_length	video_link	n_likes	n_shares	n_comments	n_plays	video_timestamp	country	year	YEAR_2	video_id_shortened
zachking	6.86165E+16	6.99743E+18	#behindthescenes @shaymitchell	1629215827	13	https://www.tiktok.com/@shaymitchell/video/6.99743E+18	660000	824	1423	5600000	8/17/2021	US	2021	2021	20433
zachking	6.86165E+16	6.99554E+18	#ad You never know when you might wir	1628776286	16	https://www.tiktok.com/@shaymitchell/video/6.99554E+18	394300	951	1045	4900000	8/12/21	US	2021	2021	488520
zachking	6.86165E+16	6.99301E+18	The #lighthouse keeper,Ã¶s job never er	1628186532	55	https://www.tiktok.com/@shaymitchell/video/6.99301E+18	379700	980	1740	3200000	8/5/21	US	2021	2021	413990
zachking	6.86165E+16	6.98965E+18	The magical #travel agency @shaymitch	1627405829	28	https://www.tiktok.com/@shaymitchell/video/6.98965E+18	884000	2223	3441	5400000	7/27/2021	US	2021	2021	126841
zachking	6.86165E+16	6.98745E+18	#ad Even lunchtime is an adventure! Bui	1626892734	17	https://www.tiktok.com/@shaymitchell/video/6.98745E+18	622500	906	1422	5000000	7/21/2021	US	2021	2021	158925
zachking	6.86165E+16	6.98668E+18	Teaching @shaymitchell how to always v	1626712810	11	https://www.tiktok.com/@shaymitchell/video/6.98668E+18	882800	1363	1589	6800000	7/19/2021	US	2021	2021	378878
zachking	6.86165E+16	6.98443E+18	I love capturing the magic in-camera #be	1626188575	13	https://www.tiktok.com/@shaymitchell/video/6.98443E+18	2100000	1559	2312	47900000	7/13/2021	US	2021	2021	30798
zachking	6.86165E+16	6.98256E+18	This is always a strange room to #paint	1625753522	21	https://www.tiktok.com/@shaymitchell/video/6.98256E+18	1800000	3084	5831	22400000	7/8/21	US	2021	2021	542499
zachking	6.86165E+16	6.97737E+18	Playing a #videogame in real life	1624546239	17	https://www.tiktok.com/@shaymitchell/video/6.97737E+18	4700000	26300	20900	36600000	6/24/2021	US	2021	2021	353720
zachking	6.86165E+16	6.97516E+18	The Day to Night #chair	1624030231	18	https://www.tiktok.com/@shaymitchell/video/6.97516E+18	782300	1310	2578	6200000	6/18/2021	US	2021	2021	509649
zachking	6.86165E+16	6.97404E+18	Never walk on someone,Ã¶s #chalk #art	1623769668	14	https://www.tiktok.com/@shaymitchell/video/6.97404E+18	16300000	83500	57000	166700000	6/15/2021	US	2021	2021	231584
zachking	6.86165E+16	6.97217E+18	It,Ã¶s an amBUSH! Check out my home	1623335083	23	https://www.tiktok.com/@shaymitchell/video/6.97217E+18	2100000	3967	2393	31300000	6/10/21	US	2021	2021	93782
zachking	6.86165E+16	6.97181E+18	Sorry @marshmellomusic , I,Ã¶ll get you	1623251630	19	https://www.tiktok.com/@shaymitchell/video/6.97181E+18	1200000	3387	3432	8600000	6/9/21	US	2021	2021	996522
zachking	6.86165E+16	6.96399E+18	The weight is over...→appleüße,ÄçappleöÄappleü	1621429852	22	https://www.tiktok.com/@shaymitchell/video/6.96399E+18	3400000	9916	10600	65600000	5/19/2021	US	2021	2021	574921
zachking	6.86165E+16	6.96108E+18	#behindthescenes	1620753540	23	https://www.tiktok.com/@shaymitchell/video/6.96108E+18	1700000	2509	2141	21400000	5/11/21	US	2021	2021	458854
zachking	6.86165E+16	6.95994E+18	Museums in #2021	1620486260	23	https://www.tiktok.com/@shaymitchell/video/6.95994E+18	4600000	58900	13500	33200000	5/8/21	US	2021	2021	420718
zachking	6.86165E+16	6.95848E+18	Statue of no limitations	1620148510	12	https://www.tiktok.com/@shaymitchell/video/6.95848E+18	2100000	30700	8633	27700000	5/4/21	US	2021	2021	459388
zachking	6.86165E+16	6.95698E+18	the best hiding spot	1619797244	17	https://www.tiktok.com/@shaymitchell/video/6.95698E+18	2900000	4983	12000	29300000	4/30/2021	US	2021	2021	595253
zachking	6.86165E+16	6.95587E+18	#behindthescene	1619539068	27	https://www.tiktok.com/@shaymitchell/video/6.95587E+18	2300000	1440	3002	19200000	4/27/2021	US	2021	2021	616900
zachking	6.86165E+16	6.95288E+18	Turn your phone upside down when I tel	1618843014	17	https://www.tiktok.com/@shaymitchell/video/6.95288E+18	845400	2417	3497	7800000	4/19/2021	US	2021	2021	872662
zachking	6.86165E+16	6.95098E+18	LIKE to find your fortune	1618402075	36	https://www.tiktok.com/@shaymitchell/video/6.95098E+18	810400	1076	2739	6700000	4/14/2021	US	2021	2021	106557
zachking	6.86165E+16	6.94809E+18	How to climb a wall like spiderman #lifeh	1617727263	12	https://www.tiktok.com/@shaymitchell/video/6.94809E+18	850900	2200	2100	8000000	4/6/21	US	2021	2021	775815
zachking	6.86165E+16	6.94657E+18	Let me know if you want me to keep pos	1617374753	14	https://www.tiktok.com/@shaymitchell/video/6.94657E+18	7600000	9354	8881	102400000	4/2/21	US	2021	2021	526316
zachking	6.86165E+16	6.94547E+18	How I got my #YouTube Diamond Play B	1617117408	32	https://www.tiktok.com/@shaymitchell/video/6.94547E+18	706900	1138	1906	7300000	3/30/2021	US	2021	2021	613686
zachking	6.86165E+16	6.94397E+18	Life isn,Ã¶t always as it seems. #dinner :	1616768608	14	https://www.tiktok.com/@shaymitchell/video/6.94397E+18	1900000	4124	3542	18800000	3/26/2021	US	2021	2021	703920

✓ 1. 8 or more attributes for analysing the data. | ✓ 2. 10,000 (or more) records to classify in the data. | ✓ 3. Good documentation of the dataset attributes

SAMPLE OF DATA

DATA SET. CHALLENGES I FACED

DUPLICATE DATA

Having two columns with the same input and unsure on what both columns can be used together

MISSING INFORMATION

Certain information like number of followers is not in the dataset to look in detail the number of followers for the particular user

MISINTERPRET DATA

video_timestamp should have been named as date_of_post or video_date something that can relate to the entries, as for a viewer who don't know the legends of the data would be thinking timestamp is time with different formats and not includes date.

REDUNDANT DATA

Having a video_ID and another shortened video_ID, don't think there's a need to shortened the ID or have another set of ID. Its like creating 2 unique entries for a video.

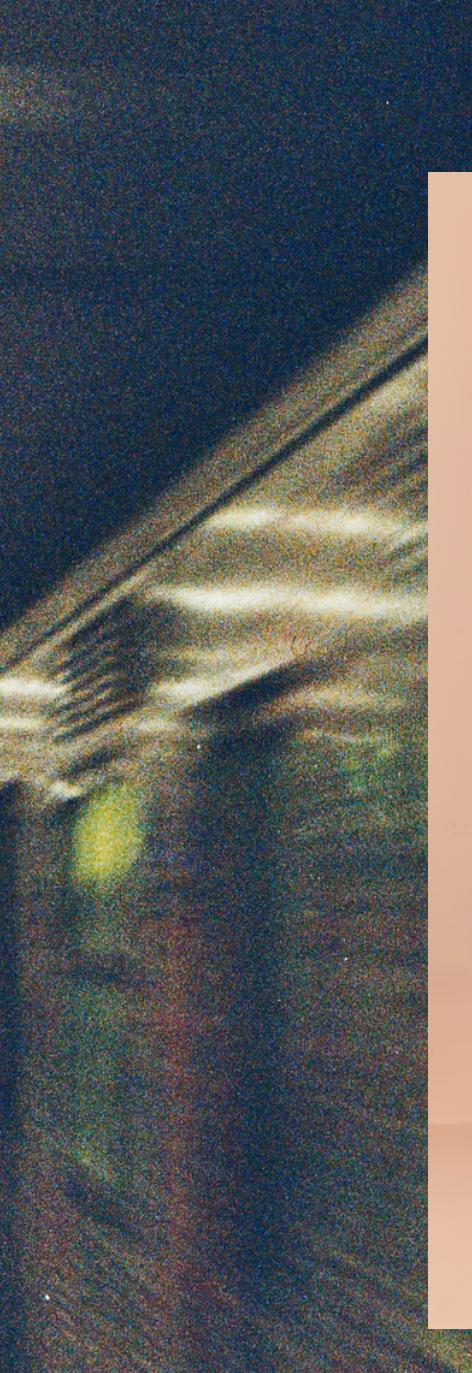
DIFFERENT DATATYPE

For the timestamp is an object type and not a datetime, will need to figure out how to make it as a datetime format

THE CONTENT ENTRIES

If there is a column with the type of content (horror/humour/education/lifestyle/etc) could help with more findings.

GOALS



GOALS. QUESTIONISM

1. Whose video have the highest number of likes and what is the video description?
2. Which country has the highest number of plays?
3. Is there a different amount of plays in 2015 and 2021?
4. what is video_time? Create a column 'time' to datetime the time of video taken.
5. Any matching number of number of plays, how many number of likes were there?
6. Most top content creators are Mens or Ladies?
7. What year has the least number of number of plays do they also have the least number of likes and number of shares?
8. From 2015 to 2021, what is the average number of number of likes in every country for all the top 5 content creator?



G O A L S . P O I N T S 1

Ensure that everyone who reads the dataset understand what they are reading, assuming has many possibilities which can be good or bad.

G O A L S . P O I N T S 2

With a dataset which has more than 10,000 video links, a person reading it can't open 10,000 links to watch what is in it, best to have a general explanation or type of video

G O A L S . P O I N T S 3

Creating a few columns such as video_type
to make new conclusions

THANK YOU

