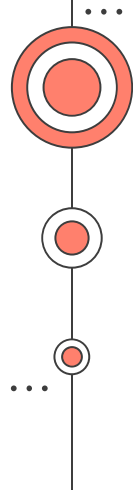


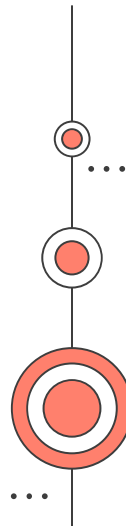
A Study of the Effect of **YouTube** Tech Channels to the Number Units Sold of Newly Released Devices





01

Introduction



Datasets

BeautifulSoup

	Manufacturer	Model	Form factor	Smartphone?	Year	Graph (million units)	x
0	Nokia	1100	Bar	No	2003	250[3][1][4]	250
1	Nokia	1110	Bar	No	2005	250[3][1][5]	250
114	Nokia	N97	Tilt slider	Yes	2009	2[4]	2

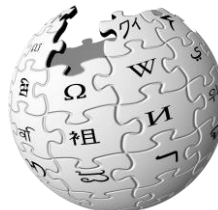
115 rows × 7 columns

	company_name	device	year	units
0	APPLE	IPHONE 6	2014	222000000
1	APPLE	IPHONE 6 PLUS	2014	222000000
103	NOKIA	N97	2009	2000000

104 rows × 4 columns

	company_name	device
0	APPLE	iPhone 13 Pro Max
1	APPLE	iPhone 13 Pro
4211	XIAOMI	Mi Note Plus

4212 rows × 2 columns



WIKIPEDIA
Die freie Enzyklopädie

GSMarena

Datasets YouTube

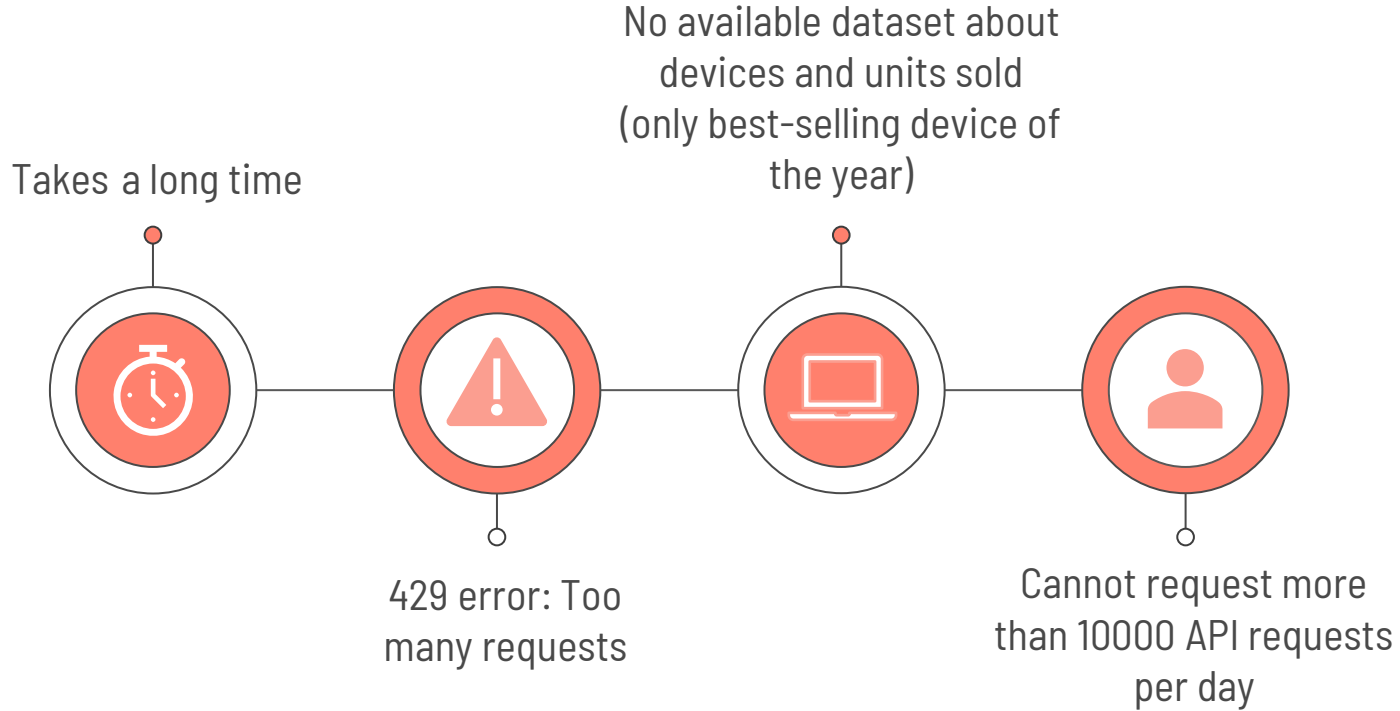
	channel_name	titles	tags	likes_number	views_count	dislikes_number	commens_count	video_duration
0	MKBHD	Surface Laptop Studio Impressions: Windows 11 ...	SURFACE LAPTOP	80208.0	2184860.0	1472.0	5720.0	687
1	MKBHD	Surface Laptop Studio Impressions: Windows 11 ...	SURFACE STUDIO	80208.0	2184860.0	1472.0	5720.0	687
127119	UNBOX_THERAPY	Lacie 2big Quadra Unboxing & Overview in HD!	7200RPM	170.0	10440.0	5.0	17.0	283
127120	UNBOX_THERAPY	Lacie 2big Quadra Unboxing & Overview in HD!	32MB	170.0	10440.0	5.0	17.0	283
127121	UNBOX_THERAPY	Lacie 2big Quadra Unboxing & Overview in HD!	CACHE	170.0	10440.0	5.0	17.0	283

127122 rows × 8 columns

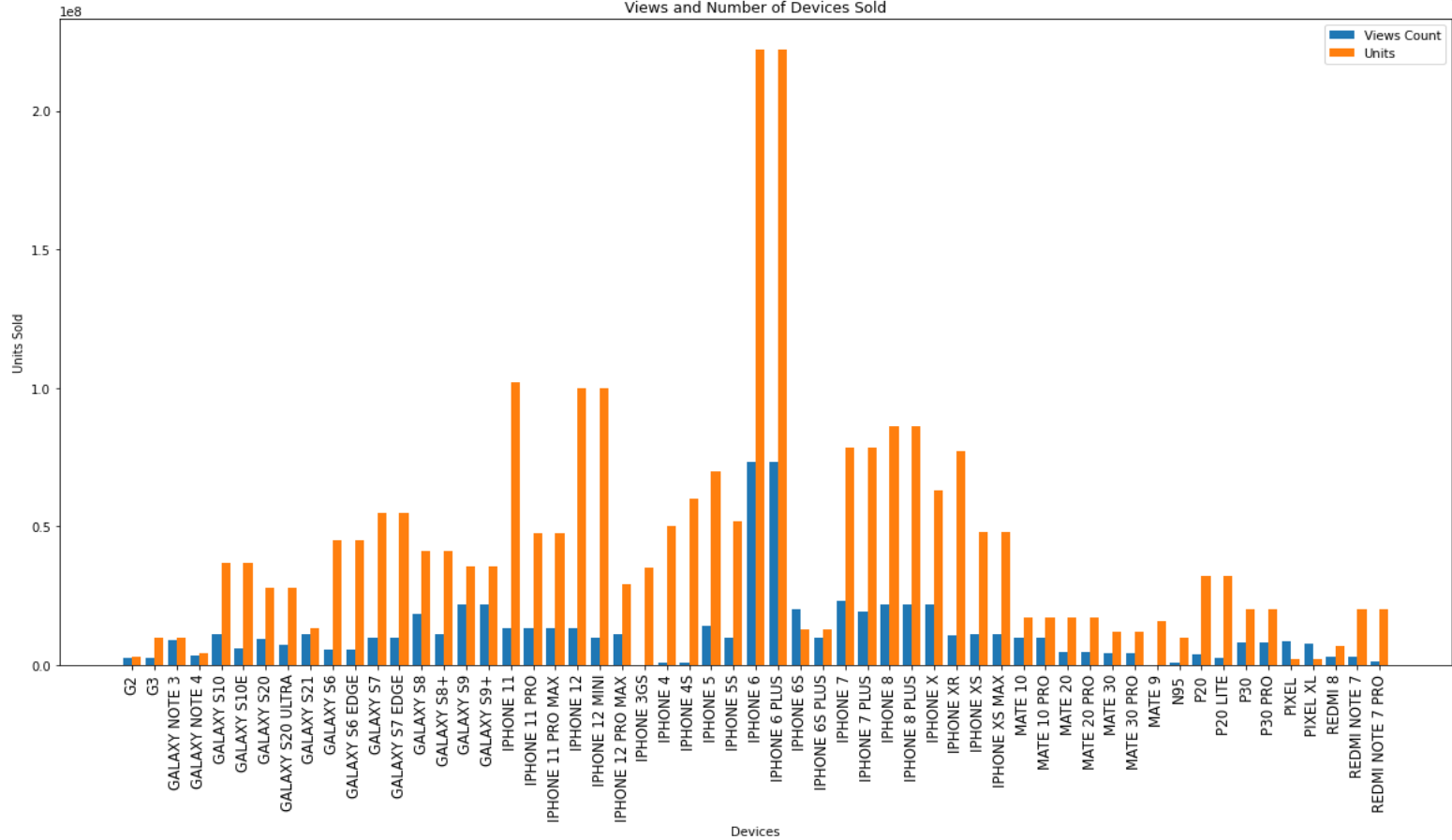
Final Datasets

	company_name_x	device	units	channel_name	likes_number	views_count	dislikes_number	commens_count	video_duration
0	APPLE	IPHONE 12 PRO MAX	2.900000e+07	MKBHD	289750.0	6380863.0	6569.0	19865.0	1505
1	APPLE	IPHONE 12 PRO MAX	2.900000e+07	MKBHD	170900.0	5512865.0	3293.0	12763.0	609
2	APPLE	IPHONE 12 PRO MAX	2.900000e+07	MKBHD	230897.0	6736834.0	6498.0	27074.0	1105
4859	XIAOMI	REDMI NOTE 7 PRO	2.000000e+07	UNBOX_THERAPY	36921.0	1248998.0	1057.0	5495.0	393
4860	XIAOMI	REDMI NOTE 7	2.000000e+07	MRWHOSETHEBOSS	68267.0	3046416.0	2970.0	2037.0	436
4861	XIAOMI	REDMI NOTE 7	2.000000e+07	UNBOX_THERAPY	36921.0	1248998.0	1057.0	5495.0	393

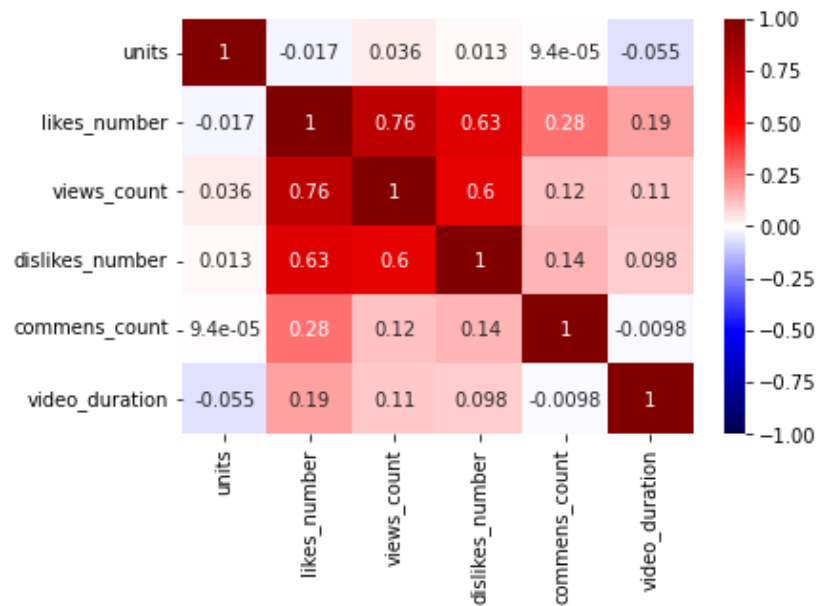
Problems Faced During Web Scrapping



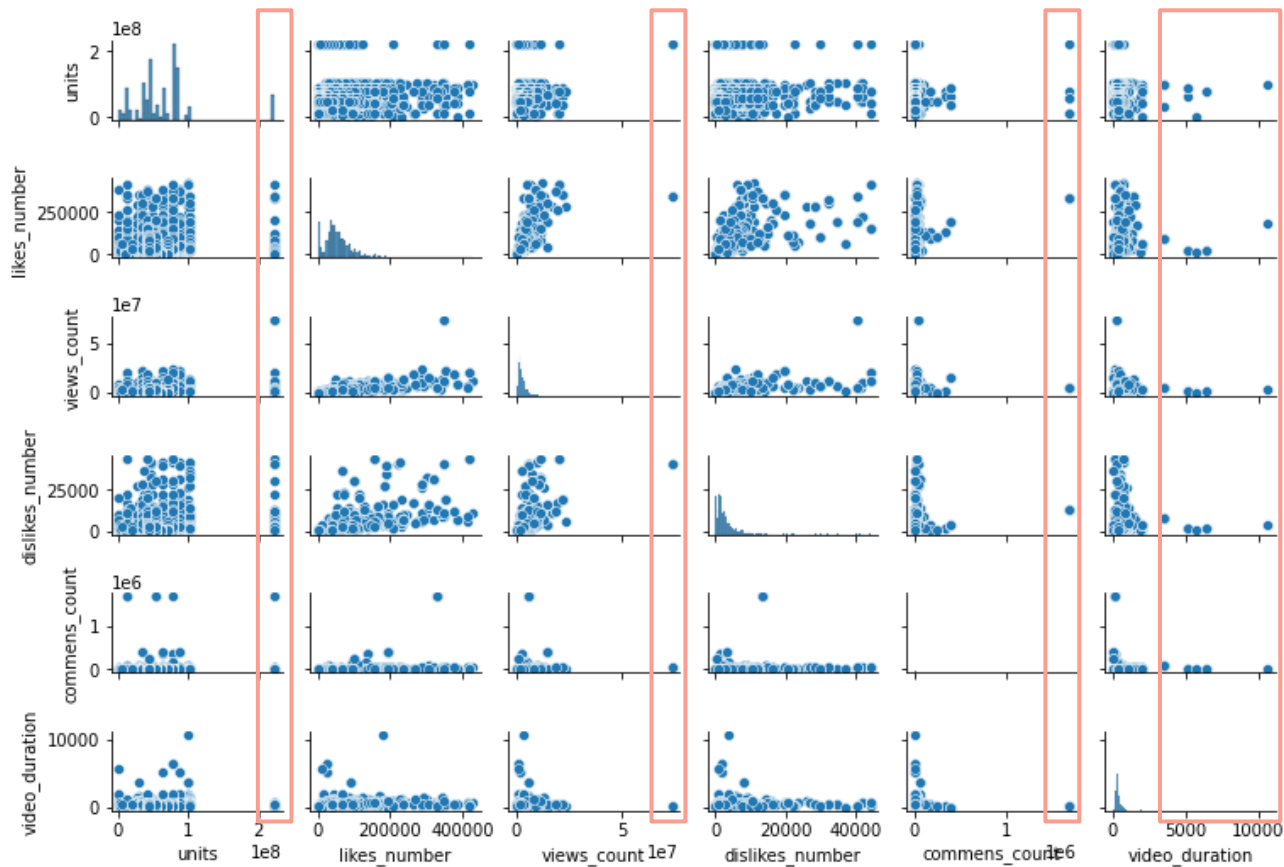
EDA



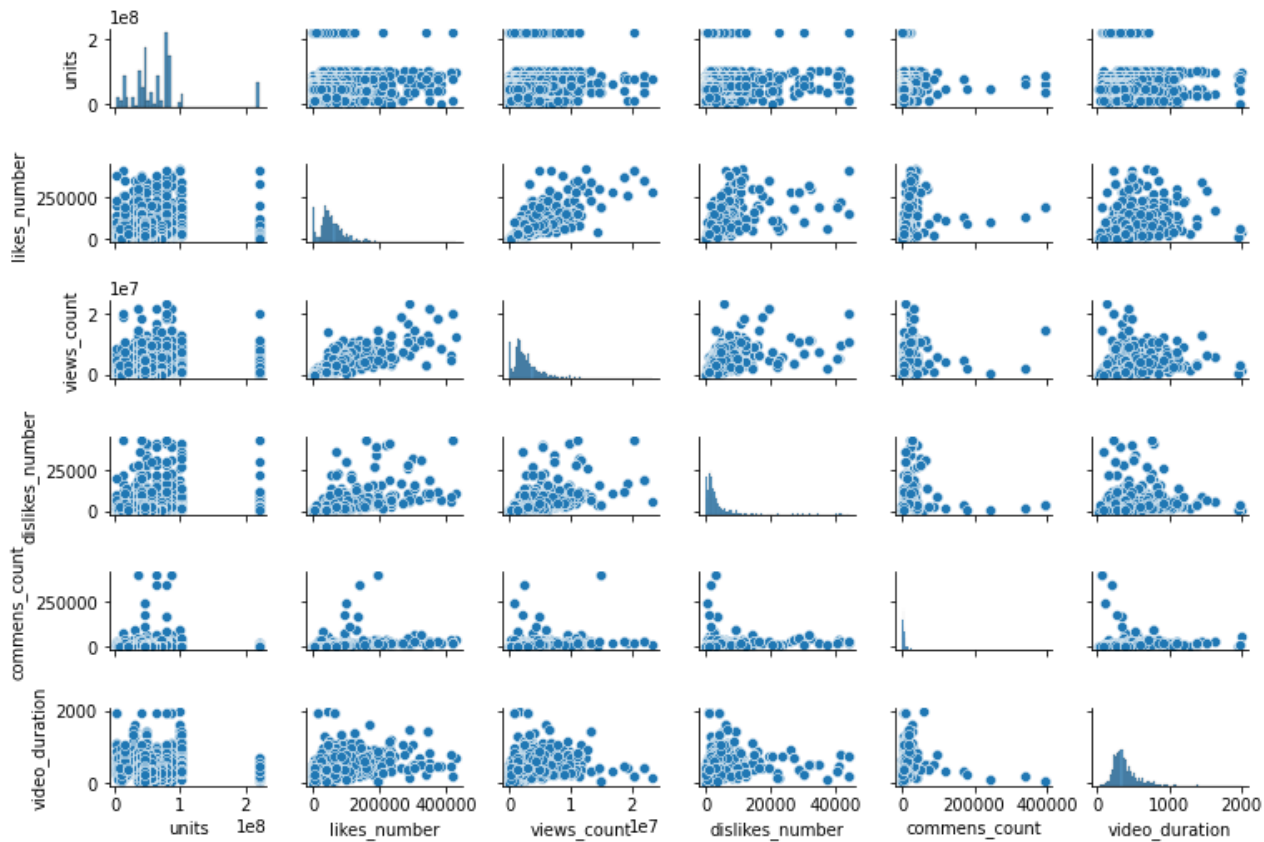
EDA



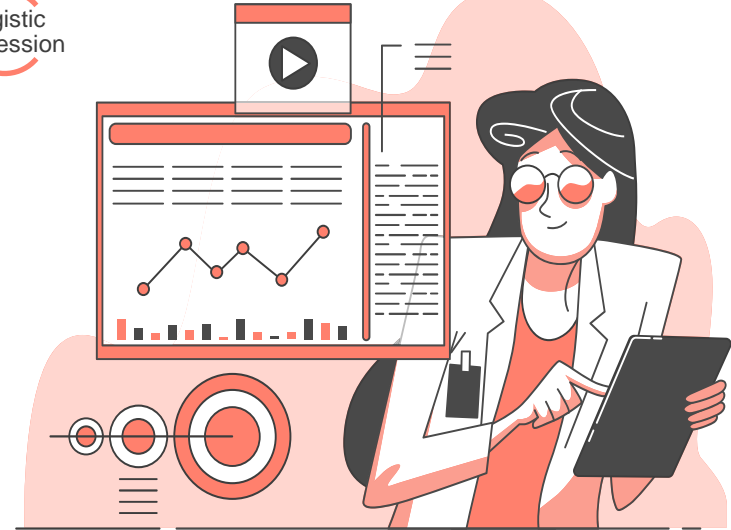
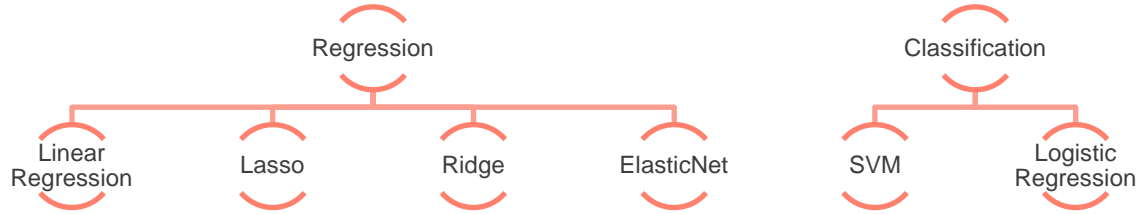
EDA



EDA



Results + Experiments



Regression Results

	Mean Squared Error	Mean Absolute Error	r2	adj r2
LR_scores	2.222646e+12	5.738715e+04	0.998973	0.998667
Lasso_scores	5.178937e+12	9.331484e+04	1.000000	1.000000
Ridge_scores	3.920545e+12	7.924187e+04	1.000000	1.000000
ElasticNet_scores	4.007793e+12	1.411187e+05	0.999981	0.999981
LR_scores_no_dummies	2.525750e+15	3.245676e+07	0.002687	-0.013831
Lasso_scores_no_dummies	2.251499e+15	3.172793e+07	0.010961	0.009352
Ridge_scores_no_dummies	2.251499e+15	3.172793e+07	0.010961	0.009352
ElasticNet_scores_no_dummies	2.251499e+15	3.172793e+07	0.010961	0.009352
LR_scores_polynomial	1.924481e+17	1.944450e+07	-61.839057	-80.493904
Lasso_scores_polynomial	4.306357e+11	1.508230e+05	0.999996	0.999996
Ridge_scores_polynomial	2.643683e+14	2.380044e+06	0.999009	0.998986
ElasticNet_scores_polynomial	5.242582e+14	1.071803e+07	0.910808	0.908722
LR_scores_GBU	3.120487e+14	1.505590e+07	-6.264618	4.600054
Lasso_scores_GBU	5.880708e+14	2.135536e+07	0.193170	-0.008537
Ridge_scores_GBU	5.880744e+14	2.135542e+07	0.193170	-0.008537
ElasticNet_scores_GBU	5.881176e+14	2.135613e+07	0.193170	-0.008537



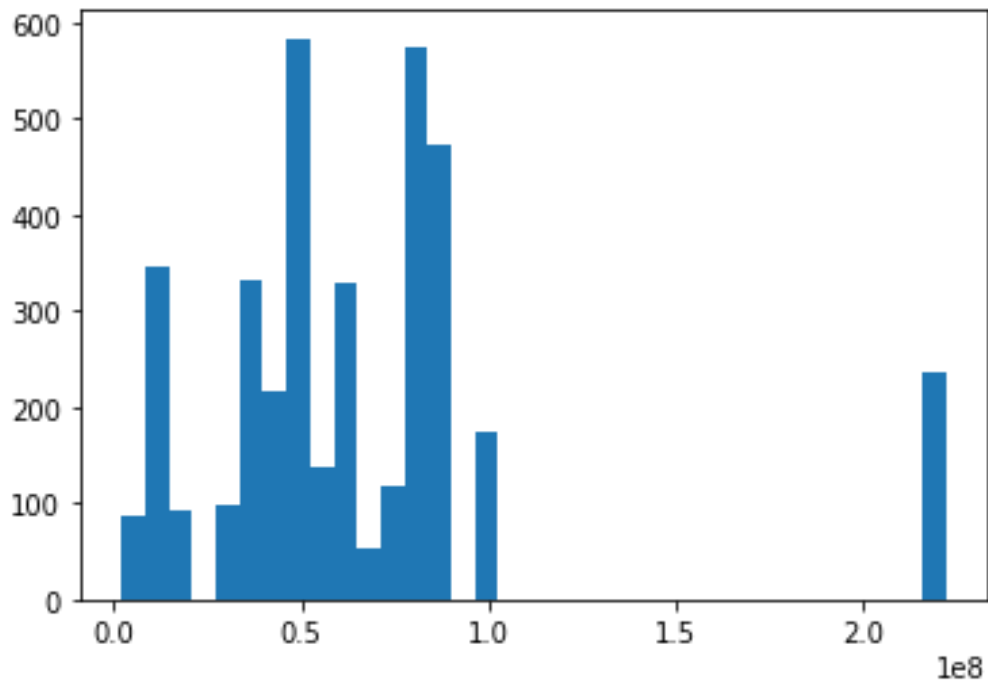
Datasets



	company_name_x	device	units	channel_name	likes_number	views_count	dislikes_number	commens_count	video_duration
0	APPLE	IPHONE 12 PRO MAX	2.900000e+07	MKBHD	289750.0	6380863.0	6569.0	19865.0	1505
1	APPLE	IPHONE 12 PRO MAX	2.900000e+07	MKBHD	170900.0	5512865.0	3293.0	12763.0	609
2	APPLE	IPHONE 12 PRO MAX	2.900000e+07	MKBHD	230897.0	6736834.0	6498.0	27074.0	1105
4859	XIAOMI	REDMI NOTE 7 PRO	2.000000e+07	UNBOX_THERAPY	36921.0	1248998.0	1057.0	5495.0	393
4860	XIAOMI	REDMI NOTE 7	2.000000e+07	MRWHOSETHEBOSS	68267.0	3046416.0	2970.0	2037.0	436
4861	XIAOMI	REDMI NOTE 7	2.000000e+07	UNBOX_THERAPY	36921.0	1248998.0	1057.0	5495.0	393

3849 rows × 9 columns

Classification



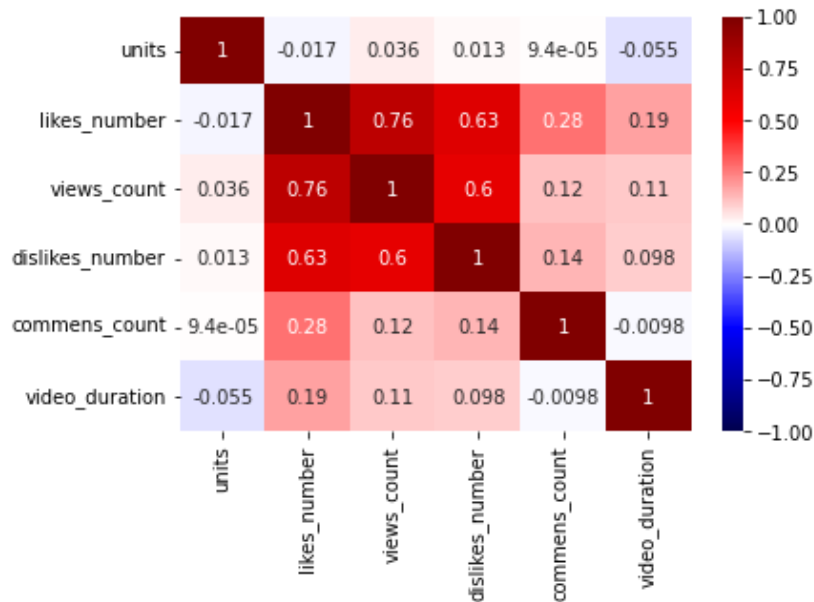
...

Classification

	r2	adj r2	accuracy_score
log_reg_scores	0.465450	0.442214	0.465450
svm_scores	0.495437	0.484298	0.495437

EXTRA Experiments

- Taking only likes, video_duration, dislikes, commens_count with View_count as our target.



EXTRA Experiments

- Taking only likes, video_duration, dislikes, commens_count with View_count as our target.

	Mean Squared Error	Mean Absolute Error	r2	adj r2
LR_scores_EXTRA	4.040743e+12	822268.709997	0.630839	0.625964
Lasso_scores_EXTRA	1.896448e+12	763973.217957	0.599129	0.598608
Ridge_scores_EXTRA	1.896448e+12	763973.219117	0.599129	0.598608
ElasticNet_scores_EXTRA	1.896448e+12	763973.054112	0.599129	0.598608

- Splitting the data into 3 Dataframes based on channel.
- Splitting the Dataframe based on the brand (company name).



Thanks!

CAUTION: PLEASE DON'T TRY THIS DATASET
AT HOME!