

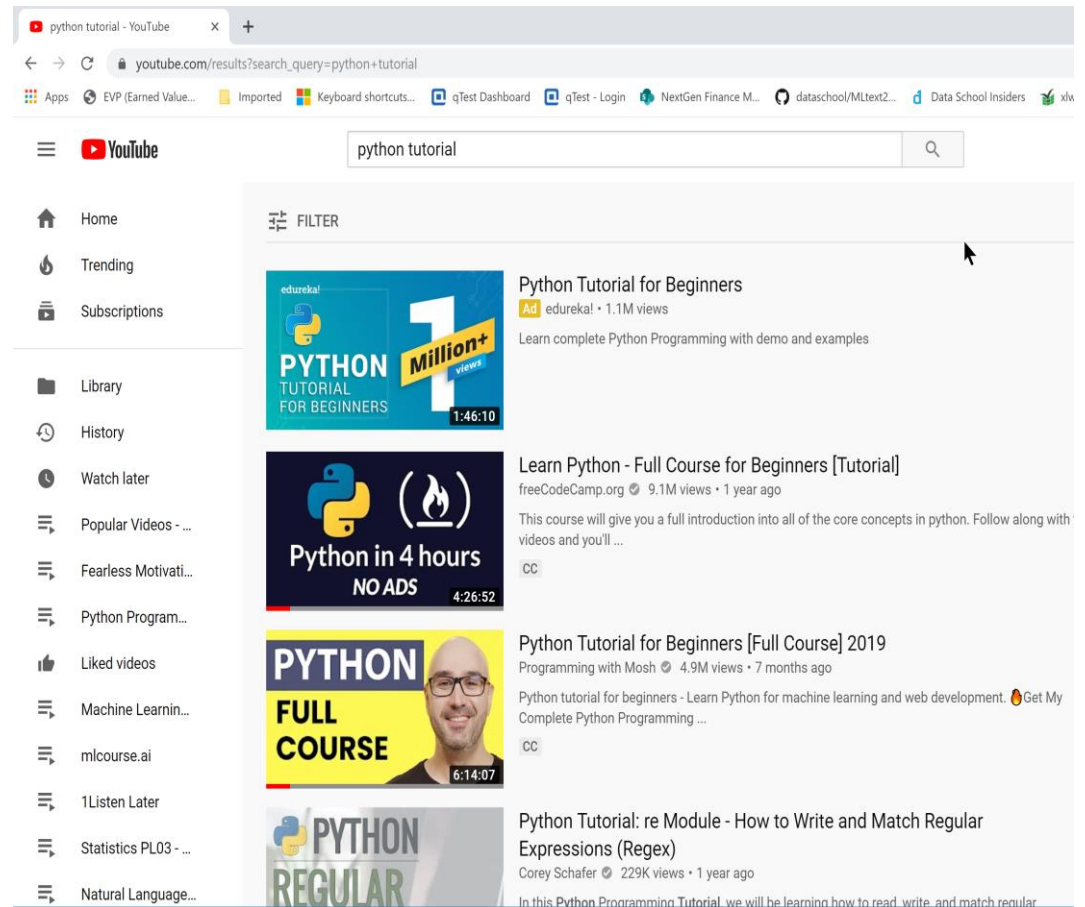
SPRINGBOARD - CAPSTONE 2 - FINDING HIDDEN HIGH QUALITY YOUTUBE CONTENT

Tom Preston

OCTOBER 16, 2019

AGENDA

- Problem Statement
- The Data
- Exploration
- Naïve Bayes
- Fuzzy Matching
- Results
- Insights



PROBLEM STATEMENT

- Millions of people search Youtube (YT) videos daily to learn new things
- YT uses a proprietary algorithm to present the “top videos” for a search request.
- Often, there are high quality, less viewed videos that show up below the “top 10” search results – a user has to manually review the results to find the “really good ones”
- Can we use machine learning to find the high quality but less viewed content ?
- Can we create a different way to score and rank videos ?

YT QUERY APPROACH

- Perform 8 separate YT queries around python topics
- Gather the metrics via the YT API for the top 25 videos for each query (8 queries * 25 results = 200 videos)
 - Metrics include number of views, likes, dislikes and comments
 - Calculate the age of each video in the top 25 results
- Gather the top 100 comments for each of the 200 video results
- Perform machine learning (Naïve Bayes) to classify the “positive comments” for each video and use this as a new video scoring approach
- Determine a way to have the high-quality (positive comments) but less seen videos higher in the query results

Python Queries	
Python tutorial	Python dictionaries
Python reading CSV Files	Python sort functions
Python pandas DataFrames	Python for loops
Python lists	Python tuples

THE DATA

YT API – 3 QUERIES

- (1) Video Search
 - Perform a YT search via the YT API
 - YT API Returns to 25 results for each query (video ID, title, and description (first 175 characters only))
 - Results are in a nested dictionary
- (2) Video Statistics
 - Requires Video ID (the 25 videos IDs for each query sent as 1 group query for the video statistics)
 - Video statistics include video views, likes, dislikes, and comment counts
 - Results are in a nested dictionary
 - Video search results and video statistics results combined into 1 excel file for each query
- (3) Video Comments
 - Requires Video ID (sent one at a time given)
 - Top 100 comments (chosen using YT “relevance” setting) taken for each video and full comment text is returned
 - Most of the “top 10” videos have more than 100 comments per video
 - Several of the “top 11 – 25” (lesser viewed videos) do not have 100 comments
- Custom calculation – video age (in weeks) calculated using Datetime method

DATA - YT VIDEO SUMMARY

- Python Query 1 – “python tutorial” Top 25 results shown below.
- Separate files created for 8 separate YT queries around python

Top video
seen 9M
times

Query Num	Query	Search Rank	Channel	Video ID	Video Title	Video Description	Video Age	view count	like count	dislike count	comment count
P1	python tutorial	1	freeCodeCamp.org	rfscVS0vtbw	Learn Python - Full (This course will give you everything you need to know to get started with Python)		64	9008855	211081	2703	11961
P1	python tutorial	2	Programming with Mosh	uQrJ0TkZlc	Python Tutorial for Beginners		32	4810630	177478	1251	11936
P1	python tutorial	3	CS Dojo	Z1Yd7upQsXY	Python Tutorial for Beginners		92	3587298	70809	1130	4649
P1	python tutorial	4	Programming with Mosh	f79MRyMsjrQ	Python Tutorial for Beginners		49	359370	6532	184	601
P1	python tutorial	5	CS Dojo	kLZuut1fyZQ	What Can You Do with Python		67	1535629	44954	681	1505
P1	python tutorial	6	Derek Banas	H1elmMBnyKA	Python Tutorial 201: Get my Ultimate Python		6	39582	1415	14	391
P1	python tutorial	7	edureka!	vaysJAMDaZw	Python Tutorial For Beginners		30	521642	9283	175	302
P1	python tutorial	8	Derek Banas	N4mEzFDjqtA	Python Programming: Get my Ultimate Python		255	5158257	69531	1661	6481
P1	python tutorial	9	TechLead	5mj_Qftw2_0	How to Learn Python: Ex-Google Tech Lead		58	281454	12430	324	908
P1	python tutorial	10	Corey Schafer	ZDa-Z5JzLYM	Python OOP Tutorial: In this Python Object		171	1432106	36996	242	1891
P1	python tutorial	11	freeCodeCamp.org	8DvywoWv6fI	Python for Everybody: This Python 3 tutorial		22	604481	22924	192	864
P1	python tutorial	12	Socratica	apACNr7DC_s	Python Classes and Objects		120	435342	11972	282	639
P1	python tutorial	13	freeCodeCamp.org	C6jjg9Zan7w	Python Game Tutorial: A Pong clone game		41	109128	2163	33	492
P1	python tutorial	14	ProgrammingKnowledge	bZ6NL59FMoc	Full Python Program: Python is one of the		23	184368	2799	61	106
P1	python tutorial	15	Corey Schafer	90s0o3wzS_I	Python Tutorial for Beginners		124	379673	6420	106	402
P1	python tutorial	16	IntelliPaat	5GYeia8IRbg	Python Tutorial PyIntelliPaat Python		19	136765	3897	84	250
P1	python tutorial	17	CS Dojo	NSbOtYzIQiO	How To Use Functions		90	760748	10462	163	1381
P1	python tutorial	18	IntelliPaat	pJ3IPRqID2M	Python Tutorial for Beginners		7	293894	11636	239	469
P1	python tutorial	19	Corey Schafer	W8KRzm-HUcc	Python Tutorial for Beginners		124	352165	8524	46	487
P1	python tutorial	20	Multimedia Channel	3cZsjOclmoM	Zero to Hero with Python		101	168003	1946	83	129
P1	python tutorial	21	Academind	kDdTgxv2Vv0	Python Tutorial for Beginners		30	87192	2228	41	193
P1	python tutorial	22	freeCodeCamp.org	CD4qAhfFuLo	Snake Game Python		48	155749	1801	66	188
P1	python tutorial	23	kjdElectronics	cpPGObKHYKc	Python Beginner Tutorial		298	2741687	11750	765	1191
P1	python tutorial	24	Durga Software Solutions	v_S64kldryc	Learn Python - Full Course		29	510953	12052	388	799
P1	python tutorial	25	kjdElectronics	uPwztoPBVWI	Python Beginner Tutorial		123	73602	711	7	92

Significant
drop off in
views below
top 10
search
results

Applying
machine
learning on
comments
to find
“positive”
ones

DATA - YT VIDEO SUMMARY CONT.

- Python Query 1 – “python tutorial” Top 100 comments for video results #1
- Using Naïve Bayes to identify “positive” vs “other” comments

Query Num	Query	Channel	Video Num	Video ID	Video Title	Video Descript	Comment	Comment ID	Replies	Likes
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course Hey everyone! Thanks for watching my course Follow me on twitter at https://twitter.com/	UgxKkKnul	276	4940	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course [Comment Deleted] **You'll never how I got these likes*	UgzhpJCKH	38	2484	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course 2400 dislikes are people that offer paid courses for programming.	Ugyz2bKLv	14	741	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course One minute in : Assigning a = 54 hours in : *Hacking the Pentagon*	Ugwayb-l	3	302	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course print('friends')console: Error. You have no friends	UgywKRt	4	222	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course 0:00 introduction1:45 installing python and pycharm6:40 setup and hello world10:23 Dr	UgzulBKGr	199	7129	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course I retired this last year after having been a professional programmer for 39 years. Started	UgwYxzUg	10	513	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course Funny that it counts characters in the string including the "" but it starts with G = 0 hmm	UgxuHnZzl	0	1	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course print("hello world")	UgwIEPc	2	3	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course /oo ? Agency = [AISA,NASA]For i in Agency:Print(i)	UgxZjlpkv	0	0	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course Haha, i have a python exam tomorrow, im only 45 mins into this video and ive already let	Ugw2wn1t	116	2333	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course This dude is brilliant at explaining things for a beginner, I've learned more from this 4 hou	UgwiL_hLv	13	170	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course Im 50 and dont know anything about programing , but this is what I need to learn to built	Ugxv7Y07Z	7	83	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course friends = ["no one", "me", "myself"] :ccc	UgysCVQyl	0	6	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course Don't mind me, just setting my timestamp: 8:30	Ugx6HymZ	0	1	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course How many people in JULY /AUGUST/SEPTEMBER 2019?	UgwA5S8V	65	1619	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course Colleges after this course "I declare BANKRUPTCY!!"	Ugz-c3z67	0	10	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course python: *clean and concise*java: User uSeR = nEw User();	Ugy3kmBC	3	60	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course Thank you so much for creating this course, I'm teaching myself python as I'll be able to	Ugy56n2cc	0	0	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course Instructions unclear ended up hacking nasa headquarters naw jk	Ugz6B8Gfj	4	85	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course Who is learning Python in March 2019???	Ugwy1P4B	246	3138	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course This course is extremely well done and performed... showing to beginners that Python is	UgwT9Tub	3	50	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course elf not(male) and is_wearingmansize13nikes: print("Beat that")	UgzkYBdKf	0	1	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course 4 hours set it to 2x speed only 2 hours. Understandable but very usefull	Ugy3VeDlc	4	44	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course My god i love how fast fire this is. Took a course in college and it was educational, but we	UgzWii_fnl	0	19	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course 4 and a half hours long, yet no ads, Greatest man in the world!	Ugya39Tpf	11	2421	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course 1:09:22 all these names are from the Office 😊 Nice to see a fan of the show	UgxyTA5Af	0	1	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course I've spent well over an hour going through the first part of the tutorial and I have to say t	Ugwq3wQj	0	0	
P1	python tutorial	freeCodeCamp	1	rfscVS0vrtbw	Learn Python - Full Course	This course /oo? 4h is too long split this in episodes. youtube recommend 10mn	UgzluMBpl	0	0	

“Error – you have no friends” an example of an “other” type comment

“Thank you so much.....” example of a positive comment

DATA - YT VIDEO SUMMARY STATISTICS

- Python summary statistics for all 8 python queries
- Surprisingly small % of likes, dislikes and comments (compared to views)
- 1,120 of the 8,492 comments manually tagged as “positive” or “other” for Naïve Bayes classification

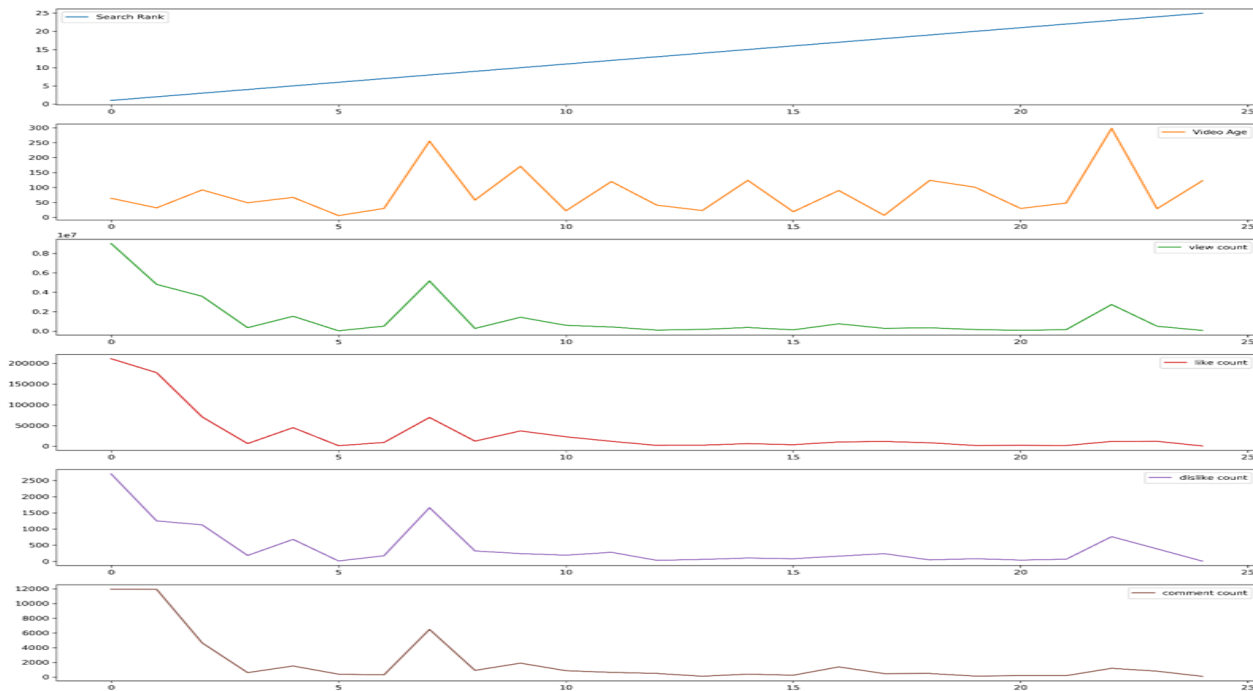
Python Query	Total Views for top 25 results	Total Likes for top 25 results	Total Dislikes for top 25 results	Total Comments	Total Extracted Comments	Hand Tagged Comments
1) Python tutorial	33,728,573	751,794 (2.2%)	10,921 (1.5%)	48,307	2,366 (5%)	595
2) Python reading CSV files	2,231,893	27,462 (1.2%)	616 (2.2%)	2,263	808 (36%)	264
3) Python pandas DataFrames	1,214,931	18,761 (1.5%)	378 (2.0%)	1,912	721 (38%)	111
4) Python lists	2,956,419	50,627 (1.7%)	616 (1.2%)	3,869	1,282 (33%)	50
5) Python dictionaries	1,650,572	28,577 (1.7%)	472 (1.7%)	2,170	1,016 (47%)	25
6) Python sort functions	419,081	7,567 (1.8%)	185 (2.4%)	678	432 (64%)	25
7) Python for loops	3,195,745	36,107 (1.1%)	836 (2.3%)	4,358	1,184 (27%)	25
8) Python tuples	1,443,193	26,187 (1.8%)	329 (1.3%)	1,801	683 (38%)	25
Totals:	46,840,407	947,082 (1.6%)	14,353 (1.8%)	65,358 (0.1%)	8,492 (13%)	1,120 (13%)

YT VIDEO STATISTICS EDA

EDA FOR “TOP 25 RESULTS” FOR EACH OF THE 8PYTHON QUERIES

- Likes, Dislikes and Comments per video follow the view pattern
- Most videos in top 25 are less than 2 years old
- After the top 10 videos, the views, likes, dislikes and comments drop off significantly

P1 python tutorial



MACHINE LEARNING

NAÏVE BAYES TO CLASSIFY EDA FOR PRICE SWINGS AND TREND MOVES

- YT provides video search results based off a proprietary algorithm using multiple features
- Naïve Bayes used to classify “positive” versus “other” comments to develop a new ranking indicator – “Positive Comments”
- 8,492 comments extracted (top 100 “relevant” comments from each of the 200 videos)
- 1,120 of the 8,492 comments hand tagged as “positive” or “other”

Hand Tagged Comments Distribution

Comment Type	# / %	Observations
Positive	635 (57%)	Very similar types of comments – “great job, great video, I learned a lot”, etc – generally very concise and very positive.
Other	485 (43%)	Broad range of comments often more editorial than praising the video. Also, if a video is showing how to code something in python and it’s confusing, there will be many questions in the video comments.

MACHINE LEARNING CONT.

NAÏVE BAYES CLASSIFICATION RESULTS

- Overall accuracy score of 78% (trained on 840 comments, tested on 280 comments)
- Top 20 positive comment tokens clarify the type of positive comments given (awesome, excellent, great, wonderful, useful, etc.)

Token	Positive Count	Other Count	Ratio
thank	157.0	11.0	14.272727
much	78.0	6.0	13.000000
awesome	36.0	3.0	12.000000
excellent	23.0	2.0	11.500000
very	80.0	7.0	11.428571
best	55.0	5.0	11.000000
amazing	30.0	3.0	10.000000
great	78.0	8.0	9.750000
videos	60.0	7.0	8.571429
watched	17.0	2.0	8.500000
simple	15.0	2.0	7.500000
teacher	22.0	3.0	7.333333
far	14.0	2.0	7.000000
useful	13.0	2.0	6.500000
clear	19.0	3.0	6.333333
helped	12.0	2.0	6.000000
really	41.0	7.0	5.857143
2017	11.0	2.0	5.500000
thanks	117.0	23.0	5.086957
wonderful	10.0	2.0	5.000000

MACHINE LEARNING OPTIMIZATION

NAÏVE BAYES OPTIMIZATION

- Ran multiple tests with different CountVectorizer parameters
- Improved accuracy from 78.2% to 82.8% (4.6% increase)
- Final parameters of Ngram_range = (1,2) and Min_df = 3 used in predicting comment types for remaining 7,372 comments

CountVectorizer Parameter Testing

Parameter	Description	Prediction Impact
Default	Default Settings	78.2% (baseline)
Lowercase = false	Do not make all tokens lowercase	74.2% (-4%)
Ngram_range = (1,2)	Use 1 and 2 word ngrams	77.8% (-0.4%)
Max_df = 0,5	Ignore terms that appear in 50% 2.8% of the documents	78.2% (no change)
Min_df = 2	Term must appear in at least 2 documents	81.0% (2.8% increase)
Combining 2 settings: Ngram_range = (1,2) Min_df = 3	Use 1 and 2 word ngrams Term must appear in at least 2 documents	82.8% (4.6% increase)

MACHINE LEARNING PREDICTION

NAÏVE BAYES PREDICTION

- Ran predictions on 7,372 comments with optimized CountVectorizer parameters
- Bayes Prediction probabilities “>= .80” classified as a positive comment
- 7,372 classified comments distribution of 56% “positive” / 44% “other” closely match hand tagging results

Bayes Prediction Probabilities

Bayes Probability	Comment Count	Comment Disposition
(0.9, 1.0]	3851	Positive
(0.8, 0.9]	305	Positive
(0.7, 0.8]	166	Other
(0.6, 0.7]	129	Other
(0.5, 0.6]	293	Other
(0.4, 0.5]	126	Other
(0.3, 0.4]	136	Other
(0.2, 0.3]	188	Other
(0.1, 0.2]	252	Other
(0.0, 0.1]	1926	Other
	7,372 (Total Comments)	

Comment Manual Tagging vs. Bayes Prediction Probabilities

Comment Type	Manually Flagged	Bayes Prediction (>= .80 = Positive Comment)	Totals
Positive	635 (57%)	4156 (56%)	4,791
Other	485 (43%)	3216 (44%)	3,701
Total:	1,120	7,372	8,492

POSITIVE COMMENTS RESULTS

- Updated the Video IDs with the 8,492 “positive comments
- 200 overall video IDs but 18 videos appear in multiple query results
- Top six “positive comment” videos make up 1,065 of the 8,492 positive comments
- Even distribution (92 comments down to 0 comments for the remaining 176 videos

Video Positive Comment Count (descending order)

Query N	Query_x	Search Rank	Channel	Video ID	Video Title	Video Age	view count	like count	dislike count	comment count	Very_Pos
P1	python tutorial	19	Corey Schafer	W8KRzm-HUci	Python Tutorial for Beginners 4: Lists, Tuples, and Sets	124	352165	8524	46	487	256
P4	python lists	2	Corey Schafer	W8KRzm-HUci	Python Tutorial for Beginners 4: Lists, Tuples, and Sets	124	352184	8525	46	487	256
P8	python tuples	2	Corey Schafer	W8KRzm-HUci	Python Tutorial for Beginners 4: Lists, Tuples, and Sets	124	352186	8525	46	487	256
P4	python lists	10	Joe James	R-HLU9FI5ug	Python: Data Structures - Lists, Tuples, Sets & Dictionaries	227	174020	3830	60	179	228
P5	python dictionary	14	Joe James	R-HLU9FI5ug	Python: Data Structures - Lists, Tuples, Sets & Dictionaries	227	174020	3830	60	179	228
P8	python tuples	6	Joe James	R-HLU9FI5ug	Python: Data Structures - Lists, Tuples, Sets & Dictionaries	227	174020	3830	60	179	228
P4	python lists	16	Socratica	AhSvKGTh28Q	List Comprehension Python Tutorial Learn Python	126	156160	5055	50	249	156
P7	python for Loops	19	Socratica	AhSvKGTh28Q	List Comprehension Python Tutorial Learn Python	126	156160	5055	50	249	156
P4	python lists	18	Corey Schafer	3dt4OGnU5sM	Python Tutorial: Comprehensions - How they work and why	214	182583	3960	30	288	154
P5	python dictionary	12	Corey Schafer	3dt4OGnU5sM	Python Tutorial: Comprehensions - How they work and why	214	182583	3960	30	288	154
P2	python reading CSV	8	Keith Galli	vmEHClJofslg	Complete Python Pandas Data Science Tutorial! (Reading CSV/JSON files)	49	155424	4032	41	266	153
P3	python pandas data	4	Keith Galli	vmEHClJofslg	Complete Python Pandas Data Science Tutorial! (Reading CSV/JSON files)	49	155424	4032	41	266	153
P4	python lists	11	Corey Schafer	6iF8Xb7Z3wQ	Python Tutorial for Beginners 7: Loops and Iterations	124	199765	4460	47	186	118
P7	python for Loops	3	Corey Schafer	6iF8Xb7Z3wQ	Python Tutorial for Beginners 7: Loops and Iterations	124	199766	4460	47	186	118
P1	python tutorial	10	Corey Schafer	ZDa-Z5JzLYM	Python OOP Tutorial 1: Classes and Instances	171	1432106	36996	242	1891	92
P3	python pandas data	16	Data School	xvpNA7bC8cs	How do I select multiple rows and columns from a pandas DataFrame?	173	116593	2487	21	452	88
P3	python pandas data	10	Data School	2AFGPdNn4FN	How do I filter rows of a pandas DataFrame by column values?	179	120842	2068	27	482	82
P1	python tutorial	15	Corey Schafer	90s0o3wzS_I	Python Tutorial for Beginners 8: Functions	124	379673	6420	106	402	77
P8	python tuples	1	Socratica	NI26dghs2Rk	Python Tuples Python Tutorial Learn Python	166	167433	3619	69	203	77
P5	python dictionary	1	Corey Schafer	daefalGnkw0	Python Tutorial for Beginners 5: Dictionaries - Working with	124	330708	7244	48	360	76
P1	python tutorial	5	CS Dojo	kLZuut1fyZQ	What Can You Do with Python? - The 3 Main Applications	67	1535629	44954	681	1505	73
P1	python tutorial	6	Derek Banas	H1elmMBnykF	Python Tutorial 2019	6	39582	1415	14	391	72
P2	python reading CSV	23	Corey Schafer	Uh2ebFW8OYI	Python Tutorial: File Objects - Reading and Writing Text Files	179	519861	6244	97	518	69
P2	python reading CSV	3	Socratica	Xi52xbpHRU	CSV Files in Python Python Tutorial Learn Python	140	259268	4758	126	298	68
P4	python lists	1	CS Dojo	tw7ror9x32s	Introduction To Lists In Python (Python Tutorial #4)	90	430234	6101	66	713	66
P1	python tutorial	24	Durga Software Solutions	v_S64klndryc	Learn Python - Full Fundamental Course for Beginners	29	510953	12052	388	799	65
P1	python tutorial	3	CS Dojo	Z1Yd7upQsXY	Python Tutorial for Absolute Beginners #1 - What Are Variables?	92	3587298	70809	1130	4649	64
P3	python pandas data	2	codebasics	F6kmpjWWEd	Python Pandas Tutorial 2: DataFrame Basics	140	231584	2993	65	206	64
P1	python tutorial	12	Socratica	apACNr7DC_s	Python Classes and Objects Python Tutorial Learn Python	120	435342	11972	282	639	63

FUZZY MATCHING

- Additional ranking metric developed by “fuzzy matching” each python query to video title & description
- FuzzyWuzzy Method uses Levenshtein Distances to develop 4 types of text matching ratios
- 8 ratios calculated for each video (4 ratios each for the title and description)

FuzzyWuzzy Method Matching Ration Types

Ratio Type / Query Notation	Description	Comment
Ratio / (Title_R1, Descrip_R1)	Compares full query to full text	Needs exact query match for high ratio. Very sensitive to minor differences in word order and text length
Partial ratio / (Title_R2, Descrip_R2)	Compares partial query to full text	Provides higher score for partial text matches.
Token_set ratio / (Title_R3, Descrip_R3)	Breaks query and text into tokens and matches the sorted intersection	Provides most flexibility for sorted intersections of tokens.
Token_sort ratio / (Title_R4, Descrip_R4)	Breaks query and text into tokens, sorts token and text and then calculates the match	More flexible but longer strings provide poor matching against shorter queries

FUZZY MATCHING RESULTS

- 8 metrics generated for comparing the query with the title and description.
- Full and Partial string matched (_R1, _R2) had consistently lower rankings
- Token Set and Sort (_R3, _R4) performed better but Token Sort penalized the results for longer comparison strings (this impacted the longer video description)
- Token Set (_R3 – in yellow) was chosen as the most consistent metric to use for Fuzzy Matching

Fuzzy Matching Ratio Scores

Query Num	Query	Search Rank	Video ID	Video Title	Video Description	Title_R1	Title_R2	Title_R3	Title_R4	Descrip_R1	Descrip_R2	Descrip_R3	Descrip_R4
P1	python tutorial	1	rfsCVS0vbt	Learn Python - Full Course f	This course will give you a full intrc	39	60	100	48	10	67	57	11
P1	python tutorial	2	_uQrJ0TKZl	Python Tutorial for Beginne	Python tutorial for beginners - Lea	41	87	100	49	17	93	100	19
P1	python tutorial	3	Z1Yd7upQs	Python Tutorial for Absolut	Learn Python programming with tl	33	87	100	41	16	93	100	17
P1	python tutorial	4	f79MRyMs	Python Tutorial for Program	Finally a Python tutorial for progra	38	87	100	45	17	93	100	19
P1	python tutorial	5	kLZuut1fYz	What Can You Do with Pyth	What is Python used for? What ca	23	40	57	27	15	93	100	18
P1	python tutorial	6	H1elmMBR	Python Tutorial 2019	Get my Ultimate Python Tutorial S	74	87	100	86	16	87	100	20
P1	python tutorial	7	vaysJAMDc	Python Tutorial For Beginne	Edureka Python Training (Flat 20%	24	87	100	29	18	87	100	19
P1	python tutorial	8	N4mEzFDj	Python Programming	Get my Ultimate Python Tutorial S	55	60	61	36	15	87	100	18
P1	python tutorial	9	5mJ_Qftw	How to Learn Python Tutori	Ex-Google Tech Lead teaches you l	29	87	100	35	16	73	100	18
P1	python tutorial	10	ZDa-Z5JzLY	Python OOP Tutorial 1: Clas	In this Python Object-Oriented Tut	44	67	100	52	15	60	100	17
P1	python tutorial	11	8DvywoW\	Python for Everybody - Full	This Python 3 tutorial course aims	28	53	57	26	16	87	100	17
P1	python tutorial	12	apACNr7Dl	Python Classes and Objects	Classes are a fundamental part of	30	87	100	37	15	47	57	14
P1	python tutorial	13	C6jJg9Zan7	Python Game Tutorial: Pong	A Pong clone game tutorial for beg	63	60	100	75	14	60	100	21
P1	python tutorial	14	bZ6NL59Fh	Full Python Programming C	Python is one of top 5 programmi	28	87	100	32	14	47	57	13
P1	python tutorial	15	9Oso03wz!	Python Tutorial for Beginne	In this Python Beginner Tutorial, w	46	87	100	54	15	60	100	18
P1	python tutorial	16	5GYeia8IRt	Python Tutorial Python Tu	Intellipaat Python course: https://i	34	87	100	42	16	93	100	17
P1	python tutorial	17	NSbOtYziQ	How To Use Functions In Py	This entire series in a playlist: https	42	87	100	48	14	40	15	13
P1	python tutorial	18	pJ3IPRqID2	Python Tutorial for Beginne	Intellipaat Python for Data Science	45	87	100	54	18	93	100	19
P1	python tutorial	19	W8KRzm-H	Python Tutorial for Beginne	In this Python Beginner Tutorial, w	37	87	100	44	14	60	100	17
P1	python tutorial	20	3cZsjOclm	Zero to Hero with Python Pi	Are you brand new to coding? Wa	24	93	100	27	11	40	16	15
P1	python tutorial	21	kDdTgxv2v	Python Tutorial for Beginne	Learn Python from scratch by buil	29	87	100	34	18	93	100	19
P1	python tutorial	22	CD4qAhffu	Snake Game Python Tutorial	Learn to code a snake game using	63	87	100	73	14	47	57	15
P1	python tutorial	23	cpPG0bKH'	Python Beginner Tutorial 1	(This Python Programming Tutorial	39	60	100	47	16	60	100	18
P1	python tutorial	24	v_S64klDry	Learn Python - Full Fundam	This course will provides you a full	26	87	100	30	14	53	57	12
P1	python tutorial	25	uPwzt0PB\	Python Beginner Tutorial 3	This tutorial covers declaring funct	32	60	100	38	12	67	70	17
P2	python reading C	1	q5uM4VKy	Python Tutorial: CSV Modul	In this Python Programming Tutori	37	54	80	50	19	54	80	26
P2	python reading C	2	efSjcrp87O	Reading CSV Files In Python	Reading CSV files is a common tas	59	62	100	94	18	67	100	29
P2	python reading C	3	Xi52tx6phf	CSV Files in Python Pyth	Working with CSV files is simple in	26	46	80	45	16	62	80	27
P2	python reading C	4	k5zkC7H9A	How to read CSV file in Pyth	Python Certification Training: http	31	67	59	44	21	54	59	28

NEW VIDEO RANKING INDICATOR - “NEW SCORE”

- New ranking metric developed by combining the Fuzzy Matching ratio and the video positive comment counts
- Goal is to have the high quality but lower viewed videos appear higher in the search results
- Example 1 below – Python query “python tutorial” video previously number 19 of 25 results is moved to position 1 based on new ranking
- Example 2 below – video previously number 1 in the same query is moved to number 19 based on lower R3 scores and Positive Comment scores
- New Score = Average (Title_R3 Score + Descrip_R3 Score) + Very_Pos Count

New ranking indicator ranking example 1

Title R3 Score	Descrip R3 Score	Very Pos Count	New Score
100	100	156	$(100+100)/2 = 100$ $+ 156 = 256$

New ranking indicator ranking example 2

Title R3 Score	Descrip R3 Score	Very Pos Count	New Score
100	57	46	$(100+57)/2 = 78.5$ $+ 46 = 124.5$

“NEW SCORE” VIDEO RANKING RESULTS

- “New Score” ranking metric moved up high quality, lower viewed content up in the search results.
- Total positive comments has a big impact on the new score therefore it behooves video developers to develop high quality content and get users to leave positive comments developed by combining the Fuzzy Matching ratio and the video positive comment counts
- All 8 python queries has similar reranking results

“New Score” ranking indicator – Python Query 1 reranking example

Query Num	Query	Search R	Channel	Video ID	Video Title	Video Description	Descrip	Descrip_R3	Title_R3	Very_Pos	New Score	New_Order
P1	python tutorial	19	Corey Schafer	W8KRzm-HUcc	Python Tutorial for Beginners 4: Lists, Tuples, and Sets	In this Python Beginner Tutorial we will cover Lists, Tuples, and Sets.	60	100	100	256	356	1
P1	python tutorial	10	Corey Schafer	ZDa-Z5JzLYM	Python OOP Tutorial 1: Classes and Instances	In this Python Object-Oriented Tutorial we will cover Classes and Instances.	60	100	100	92	192	2
P1	python tutorial	15	Corey Schafer	90s0o3wzS_0	Python Tutorial for Beginners 8: Functions	In this Python Beginner Tutorial we will cover Functions.	60	100	100	77	177	3
P1	python tutorial	6	Derek Banas	H1elmMBnykA	Python Tutorial 2019	Get my Ultimate Python Tutorial	87	100	100	72	172	4
P1	python tutorial	3	CS Dojo	Z1Yd7upQsXY	Python Tutorial for Absolute Beginners #1 - What is Python programming?	Learn Python programming from scratch.	93	100	100	64	164	5
P1	python tutorial	7	edureka!	vaysJAMDaZw	Python Tutorial For Beginners Python Full Course 2019 Python tutorial for beginners	Edureka Python Training (Full Course)	87	100	100	55	155	6
P1	python tutorial	2	Programming with Mosh	uQrj0TKZlc	Python Tutorial for Beginners [Full Course]	Python tutorial for beginners	93	100	100	52	152	7
P1	python tutorial	21	Academind	kDdTgxv2Vv0	Python Tutorial for Beginners - Crash Course	Learn Python from scratch	93	100	100	52	152	8
P1	python tutorial	5	CS Dojo	kLZuut1fYzQ	What Can You Do with Python? - The 3 Main Applications	What is Python used for? What can you do with Python?	93	100	57	73	151.5	9
P1	python tutorial	9	TechLead	5mJ_Qftw2_0	How to Learn Python Tutorial - Easy & Simple	Ex-Google Tech Lead teaches you how to learn Python.	73	100	100	49	149	10
P1	python tutorial	4	Programming with Mosh	f79MRyMsrQ	Python Tutorial for Programmers - Python Crash Course	Finally a Python tutorial for programmers	93	100	100	46	146	11
P1	python tutorial	24	Durga Software Solutions	v_S64klidrc	Learn Python - Full Fundamental Course for Beginners	This course will provide you with a comprehensive understanding of Python.	53	57	100	65	143.5	12
P1	python tutorial	12	Socratica	apACNr7DC_s	Python Classes and Objects Python Tutorial	Classes are a fundamental part of Python.	47	57	100	63	141.5	13
P1	python tutorial	11	freeCodeCamp.org	8DvywoWv6fI	Python for Everybody - Full Course with Dr. Charles Severack	This Python 3 tutorial covers everything you need to know to get started with Python.	87	100	57	60	138.5	14
P1	python tutorial	16	IntelliPaat	5GYeia8IRBg	Python Tutorial Python Tutorial for Beginners	IntelliPaat Python course: Python for Data Science	93	100	100	37	137	15
P1	python tutorial	8	Derek Banas	N4mEzFDjqtA	Python Programming	Get my Ultimate Python Tutorial	87	100	61	49	129.5	16
P1	python tutorial	13	freeCodeCamp.org	C6jlg9Zan7w	Python Game Tutorial: Pong	A Pong clone game tutorial	60	100	100	29	129	17
P1	python tutorial	23	kjElectronics	cpG0bKHVKc	Python Beginner Tutorial 1 (For Absolute Beginners)	This Python Programming Tutorial is for absolute beginners.	60	100	100	27	127	18
P1	python tutorial	1	freeCodeCamp.org	rfscVS0vtbw	Learn Python - Full Course for Beginners	This course will give you a full understanding of Python.	67	57	100	46	124.5	19
P1	python tutorial	18	IntelliPaat	pJ3IPRqiD2M	Python Tutorial for Beginners IntelliPaat	IntelliPaat Python for Data Science	93	100	100	21	121	20
P1	python tutorial	14	ProgrammingKnowledge	bZ6NL59FMoc	Full Python Programming Course Python Tutorial	Python is one of the top 5 programming languages.	47	57	100	37	115.5	21
P1	python tutorial	25	kjElectronics	uPwztoPBVWI	Python Beginner Tutorial 3 - Function Return Value	This tutorial covers declaring and using functions.	67	70	100	26	111	22
P1	python tutorial	17	CS Dojo	NSbOtYzIQIO	How To Use Functions In Python (Python Tutorial)	This entire series in a playlist.	40	15	100	48	105.5	23
P1	python tutorial	22	freeCodeCamp.org	CD4qAhfFuLo	Snake Game Python Tutorial	Learn to code a snake game	47	57	100	12	90.5	24
P1	python tutorial	20	Multimedia Channel	3cZsJOclmOM	Zero to Hero with Python Programming - 2017	Are you brand new to coding?	40	16	100	30	88	25

RESULTS / FINAL INSIGHTS

- Results:
 - Using Naïve Bayes to tag video comments as “positive” and “other” was a useful exercise to understand the broad range of video comments and the fact that many comments are just general comments / statements
 - Using Naïve Bayes and Fuzzy Matching did provide a way to build an indicator for reranking videos – how well these videos are liked is still user dependent
 - Extracting and summarizing the video data provided addition insights (top video channels, top video getting the most likes)
- Python Queries Final Insights:
 - YT proprietary algorithm based on “relevance” does a good job providing query results but it’s difficult for lesser viewed videos to be noticed
 - The view statistics (likes, dislikes, comments) are useful but it’s a very small percentage(<2%) of the total views
 - The YT API is helpful but it’s only a subset of the data YT uses for it’s own proprietary algorithm