Data Wrangling Final Project

– Most Profitable Live Shows

Content

- Overview and business questions
- Data sources and data sets
- Data wrangling process
- Techniques, methods and challenges
- Analysis and output

Overview

Lethal India company

Sponsor Youtube creators to make more profits and plan a nationwide tour for our producers to give their work the widest possible exposure.

Business Analysis

Analyze different types of videos between 3 creators in 5 cities to get the final outcome.

Business Questions

- 1. Where is the most profitable place to hold a live show?
- 2. Which Youtube creator should we invite to the city?
- 3. What should the Youtube creator perform on stage?
- 4. How can we analyze this case?
- 5. How can we benefit from the whole process?

Survey

Data Sources

YouTube Analytics

• Social Blade

Survey : Primary Data Source

We posted a survey on Creators Social Media to access Direct data from their followers in order to have Raw DataSets.

We received a total of 106 Responses from the Survey posted on their Social Media

YouTube Analytics : Secondary Data Source

We got the access to creators Personal Youtube channel and took a dive in the analytics section provided to us by YouTube Studio

We were able to get the desired data sets with attributes such as Likes, Views, Shares, impression etc

We collected a total of 9 data sets and then combined it into 3 enriched data sets

Final dataset about video information from Youtube Analytics

Creator	Con-tent Video title	Video pub-lish time	Sub-scribers	Shares	Likes	Views	Watch time (hours)	Im-pres-sions
Vasu Kainth	ZFU7hqNNN0U THE THARA BHAI JOGINDER SONG (Full Version) - From @Mythpat 's Video!	Nov 21, 2021	26883	8137	69805	1262968	18411.9189	7022308
Vasu Kainth	4L8gj2r6Mvs EMIWAY BANTAI - MACHAYENGE 3 (100% NOT CLICKBAIT) Ft. Raftaar Vasu Kainth	Mar 29, 2020	19880	27268	94527	970080	21530.1253	8583771
Vasu Kainth	j3P-ffCVeLg PUBG BAN - Tribute Song Vasu Kainth	Sep 3, 2020	18946	25087	132658	1198456	21795.3652	8350659
Vasu Kainth	iKkgMANlw9Y YALGAAR - CARRYMINATI (Refix Version) Vasu Kainth	Jun 8, 2020	18506	15087	101488	1691267	34655.788	11555867
Vasu Kainth	skdMv_kB47A TIKTOK DISS × 100 Ft. Carryminati (Clickbait) Vasu Kainth	May 26, 2020	8648	5670	45915	483800	12891.59	4376929
Vasu Kainth	wAjzgheRvvk CARTOON MASHUP (Nostalgic) Vasu Kainth	Jun 13, 2020	7093	6674	32621	381777	8588.9997	2450665
Vasu Kainth	bHIUyVIAaQI HOW TO MAKE A TONY KAKKAR SONG	Feb 19, 2021	6625	3067	37716	614326	13854.2479	4120425
Vasu Kainth	1RNIM_83SJQ AYE SUN (Rap Song) Vasu Kainth	Jun 17, 2020	5472	5000	24404	277446	6335.0544	2187002
Vasu Kainth	h_tpl9TtsCl I REMADE "BELLA CIAO" FROM MONEY HEIST @Netflix India Vasu Kainth	Dec 3, 2021	5076	1381	24078	404624	5322.9785	2654925
Vasu Kainth	fLRcqmV2rpo Skechers - Hindi Version Vasu Kainth	May 14, 2020	4818	6991	33316	353016	6802.2668	2965057
Vasu Kainth	aBEXX5BQuXk EMIWAY BANTAI - MACHAYENGE 4 (100% NOT CLICKBAIT) Ft. Raftaar Vasu Kainth	Aug 27, 2020	4289	2444	28251	254531	5730.3393	2091386
	YU_0b0IAVx4 KYA HOGAYA BHAI (Rap Song) Vasu Kainth (PROD - ISWARE)	Jun 23, 2020	3051	6520	18777	209339	5916.506	1655362
Vasu Kainth	CINVH_Lr6dg IT'S HIS CHOICE (Anti - Girlfriend Song) VASU KAINTH	Jun 3, 2020	2840	2282	14698	137830	2859.4897	1211004
Vasu Kainth	EP9mEqP4q3k BROWN MUNDE X DESPACITO - AP DHILLON LUIS FONSI VASU KAINTH	Apr 1, 2021	2726	4643	21944	225105	4089.8029	1393525
Vasu Kainth	YFpmYnJhZhk THARA BHAI JOGINDER DISSTRACK!!! My reply to "BHAWANDAR" Vasu Kainth	Nov 23, 2021	2710	912	20550	263329	3893.4108	1508580
Vasu Kainth	hkR1b7uyZMk THARA BHAI JOGINDER "BHUCHAL" DELETED?? My Reply To "BHUCHAL" Vasu Kainth	Dec 9, 2021	2520	541	18937	238296	3316.7274	1530861
Vasu Kainth	h036X9KqXTQ CRADLES X ZARA ZARA X XXXTentacion (Mashup) Vasu Kainth	May 21, 2020	2365	3233	16794	163345	3563.7079	1560327
Vasu Kainth	RjX5g_frxas Kesariya - Brahmastra (Cover) Vasu Kainth	Jul 19, 2022	2316	2658	16611	193269	4679.7302	1517525
Vasu Kainth	MOICMSENAWI ON MY WAY INDIAN MASHUP KHAIRIYAT X LEHENGA X FILHAAL X MANY MORE VASU KAINTH	Apr 23, 2020	2295	1866	14650	143873	3066.9106	1504299
Vasu Kainth	qCBRGtEvigY BOHAT HUA MUJRA - TIKTOK BAN (Rap Song) Vasu Kainth	Jul 1, 2020	1973	3219	21293	169484	3938.4713	1369657
Vasu Kainth	YjCQpgP3AA4 Remade Carryminati Songs Vasu Kainth	Oct 9, 2020	1719	1213	24655	147016	4839.8363	1443943
Vasu Kainth	6NLdpo5zt6U Vasu Kainth - Sarphira (Lyric Video)	Feb 1, 2020	1688	2114	8709	102823	1246.5905	1137189

Final dataset about geography information

Creators	Cit-ies	City name	Views	Watch time (hours)	Av-er-age view dur-a-tion
Vasu Kainth	0x390d047309fff32f:0xfc5606ed1b5d46c3	Delhi	1210315	26027.4592	0:01:17
Vasu Kainth	0x3bc2bf2e67461101:0x828d43bf9d9ee3	Pune	767082	16802.2001	0:01:18
Vasu Kainth	0x3be7c6306644edc1:0x5da4ed8f8d648c	Mumbai	760515	16176.1498	0:01:16
Vasu Kainth	0x395e848aba5bd449:0x4fcedd11614f65	Ahmedabad	654316	13771.8295	0:01:15
Vasu Kainth	0x399bfd991f32b16b:0x93ccba8909978b	Lucknow	645027	13770.5364	0:01:16
Vasu Kainth	0x3962fcad1b410ddb:0x96ec4da356240f	Indore	620759	12974.275	0:01:15
Vasu Kainth	0x39f29937c52d4f05:0x831a0e05f607b27	Patna	580971	12409.1108	0:01:16
Vasu Kainth	0x39f882db4908f667:0x43e330e68f6c2cb	Kolkata	567260	12457.9635	0:01:19
Vasu Kainth	0x396c4adf4c57e281:0xce1c63a0cf22e09	Jaipur	494838	10688.6145	0:01:17
Vasu Kainth	0x39740d857c2f41d9:0x784aef38a9523b	Agra	450669	9494.9273	0:01:15
Vasu Kainth	0x375a5a287f9133ff:0x2bbd1332436bde3	Guwahati	369246	7636.2425	0:01:14
Vasu Kainth	0x3a1909d2d5170aa5:0xfc580e2b68b33fa	Bhubaneswar	347644	7204.7703	0:01:14
Vasu Kainth	0x39e44114f5441dcd:0xdeb5c4702063ed	Siliguri	335395	7143.7381	0:01:16
Vasu Kainth	0x391a837462345a7d:0x681102348ec60	Ludhiana	306302	6346.8538	0:01:14
Vasu Kainth	0x3bae1670c9b44e6d:0xf8dfc3e8517e4fe	Bengaluru	288487	6110.7945	0:01:16
Vasu Kainth	0x3755b8b087026b81:0x8fa563bbdd5904	Dhaka	256188	4700.42	0:01:06
Vasu Kainth	0x3bd4c0a5a31faf13:0x19b37d06d0bb3e	Nagpur	250272	4912.2776	0:01:10
Vasu Kainth	0x3bcb99daeaebd2c7:0xae93b78392bafb	Hyderabad	239964	4859.0821	0:01:12
Vasu Kainth	0x3eb33e06651d4bbf:0x9cf92f44555a0c2	Karachi	155777	3089.0985	0:01:11
Vasu Kainth	0x390cfd5b347eb62d:0x52c2b7494e204d	New Delhi	151670	3335.3923	0:01:19
Vasu Kainth	0x39eb198a307baabf:0xb5137c1bf18db1	Kathmandu	145705	2772.9522	0:01:08
Vasu Kainth	0x397c428f8fd68fbd:0x2155716d572d4f8	Bhopal	131759	2836.8418	0:01:17
Vasu Kainth	0x390578e3e35d6e67:0x1f7e7ff6ff9f54b7	Shimla	124053	2613.3411	0:01:15
Vasu Kainth	0x3be04e59411d1563:0xfe4558290938b0	Surat	110701	2276.2519	0:01:14
Vasu Kainth	0x3749265bf16390db:0x19499aae263c54	Imphal	96199	1921.5824	0:01:11

Final dataset about top videos after data enrichment

Creator	Con-tent Video title Video pu	b- Shares	Likes	Views	Watch time	Sub-scriber I	m-pres-sions	Display		Thumbnail	Time	Guest	Sentiment link
Vanshaj Singh	SvnLWeko(How Met Aug 12,	20 7287	446774	5209521	59918.31	8459	744554		Funny	Both	60s	Yes	0.56 https://www.youtube.com/shorts/SvnLWeko008
Vanshaj Singh	-wvR75gCv Have You E Jul 27, 20	02 3987	245016	5298230	61229.84	27747	420532		Funny	Both	60s	Yes	0.8 https://www.youtube.com/shorts/QX5eyCHe8tg
Vanshaj Singh	Ivdl6gaW2 That Rich K Aug 4, 2	02 2525	255802	3391283	37439.62	8600	473755		Funny	Both	60s	No	0.7 https://www.youtube.com/shorts/Ivdl6gaW2rA
Vanshaj Singh	-E9pTGe6Y TRIGGERED Dec 14, 2	20: 899	46522	426618	21502.75	4756	3782059		Roast	Both	6m10s	Yes	0.48 https://www.youtube.com/watch?v=-E9pTGe6YNU
Vanshaj Singh	X5Uc4gLxL Technical S Sep 29, 2	20: 3635	30748	369285	10171.09	2724	3157001		Music	Both	2m25s	No	0.56 https://www.youtube.com/watch?v=X5Uc4gLxLIM
Vanshaj Singh	tyVkUc9-bJ PAPA KI PA Jun 4, 20	20 4025	28575	362704	9127.004	2204	2649714		Music	Both	3m11s	Yes	0.67 https://www.youtube.com/watch?v=tyVkUc9-bJo
Vanshaj Singh	VytRwqrRS BEST FEMA Apr 1, 20	02: 1203	36219	300465	25269.73	6618	3260733		Roast	Both	10m55s	Yes	0.52 https://www.youtube.com/watch?v=VytRwqrRS9E
Vanshaj Singh	dDLnGYnof CAPTAIN IN Nov 25,	20 516	14560	160744	3632.683	572	1936205		Music	Figure	3m43	No	0.51 https://www.youtube.com/watch?v=BzJSucqM9 4
Vanshaj Singh	LIT7co-tHA ONBOARD Feb 7, 20	02: 508	18423	132958	9771.571	528	1544812		Roast	Both	8m11s	No	-0.34 https://www.youtube.com/watch?v=LJT7co-tHA4
Vasu Kainth	iKkgMANIv YALGAAR - Jun 8, 20	18506	15087	101488	1691267	34655.79	11555867		Music	Both	2m41s	No	0.85 https://www.youtube.com/watch?v=iKkgMANlw9Y
Vasu Kainth	j3P-lfCVeLg PUBG BAN Sep 3, 20	18946	25087	132658	1198456	21795.37	8350659		Roast	Both	3m20s	No	0.67 https://www.youtube.com/watch?v=j3P-lfCVeLg
Vasu Kainth	ZFU7hqNN THE THARA Nov 21,	20 26883	8137	69805	1262968	18411.92	7022308		Funny	Both	2m10s	Yes	0.74 https://www.youtube.com/watch?v=ZFU7hqNNN0U
Vasu Kainth	4L8gj2r6M EMIWAY B, Mar 29,	20 19880	27268	94527	970080	21530.13	8583771		Funny	Figure	1m58s	Yes	0.82 https://www.youtube.com/watch?v=4L8gi2r6Mvs
Vasu Kainth	bHiUyVlAa HOW TO N Feb 19, 2	20: 6625	3067	37716	614326	13854.25	4120425		Roast	Both	6m16s	No	0.73 https://www.youtube.com/watch?v=bHiUyVIAaQI
Vasu Kainth	skdMv_kB ² TIKTOK DIS May 26,	20 8648	5670	45915	483800	12891.59	4376929		Funny	Both	3m27s	Yes	0.77 https://www.youtube.com/watch?v=skdMv kB47A
Vasu Kainth	h_tpl9TtsC REMADE Dec 3, 20	02: 5076	1381	24078	404624	5322.979	2654925		Music	Both	2m16s	No	0.90 https://www.youtube.com/watch?v=h tpl9TtsCl
Vasu Kainth	wAjzgheRv CARTOON I Jun 13, 2	202 7093	6674	32621	381777	8589	2450665		Roast	Both	3m01s	No	0.55 https://www.youtube.com/watch?v=wAjzgheRvvk
Vasu Kainth	fLRcqmV2r Skechers - May 14,	20 4818	6991	33316	353016	6802.267	2965057		Music	Both	2m34s	No	0.84 https://www.youtube.com/watch?v=fLRcqmV2rpo
Adit Minocha	g15vz15P9 Girls Answ Mar 24,	20 8857	121607	2905883	278334.6	29887	51686916		Funny	Both	13m57s	Yes	0.61 https://www.youtube.com/watch?v=g15vz15P9WQ
Adit Minocha	yCUeRVw4 Boys Answ Apr 14, 2	20: 9567	70601	1204919	118539.4	10120	32975613		Funny	Both	13m22s	Yes	0.67 https://www.youtube.com/watch?v=CAs hBuUZ g
Adit Minocha	cE6W7q-SC JUSTIN BIEF Nov 25,	20 8778	85841	1011637	77571.04	20934	15374283		Roast	Both	11m25s	No	-0.68 https://www.youtube.com/watch?v=cE6W7q-SGaw&
Adit Minocha	sYLmGIExu NEWTELLA May 23,	20 18571	67789	866718	83439.63	14836	11489861		Roast	Both	14m41s	No	-0.29 https://www.youtube.com/watch?v=sYLmGIExuIM
Adit Minocha	YTQdaNKP MOST YOU Sep 17, 2	20: 5502	64699	818657	73186.91	7580	12830689		Roast	Both	11m21s	No	0.49 https://www.youtube.com/watch?v=YTQdaNKPYmo
Adit Minocha	kfmBt14ul- My Indian Apr 6, 20	5000	51055	616631	56893.2	5485	12582533		Funny	Both	12m07s	Yes	0.55 https://www.youtube.com/watch?v=kfmBt14uHPU
Adit Minocha	fk32X22K4 Shayad x N May 5, 2	02 630	13464	104480	1901.651	445	1030494		Music	Both	2m35s	No	0.82 https://www.youtube.com/watch?v=fk32X22K410
Adit Minocha	VmwAMiA 1 GUY 1 Feb 14, 2	20: 1005	15704	94262	2868.247	763	1068335		Music	Both	4m12s	No	0.78 https://www.youtube.com/watch?v=VmwAMiAJ8YQ
Adit Minocha	5ZG0X9-m Yummy - Jt Jan 22, 2	202 231	7677	55162	701.6572	110	664367		Music	Both	1m20s	No	0.65 https://www.youtube.com/watch?v=5ZG0X9-mK2o

Data Wrangling

Generated a dataset for video information

The final scores have been scaled to be interpretable

Well, we tried to make it look good

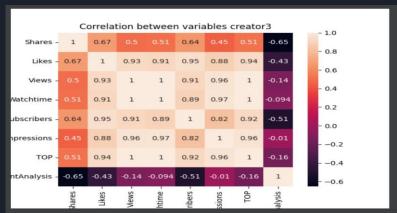


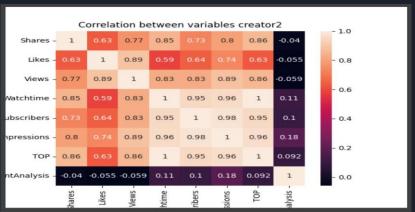
Adjusted the units of time and date, removal of redundant data

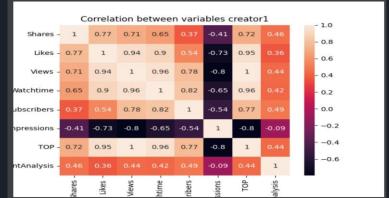
Additional data has been generated for generation if the score

Data Pre-Processing

Redundant columns have been removed using the regression results.







Data formatting

- Structured Data
 we obtained parts of the survey in which no description or writing is required, which are
 presented in a pre-defined data model or format.
- Unstructured data
 - Such as short-answer survey questions about "Who do you want him to invite some guests into the type of his videos you like", "Do you want him to invite some guests into the type of his videos you like?"
 - ❖ To ensure that the data is consistent and relevant, all required fields must be made mandatory, and we can replace missing values and irrelevant data with the values entered by the majority of survey respondents.
 - ❖ In addition, we analyzed the sentiment of the top 20 comments on each video using the top 20 comments. So while collecting the comments, it included all the subjective and objective emotions of the users, along with emoticons, which also come in unstructured data. To resolve this, we used IBM Watson to do the sentiment analysis, with the score varying from -1 to +1, where 0 is a neutral score.

Data Transformation

The duration of each video has been given in minutes, it has been transformed into seconds.

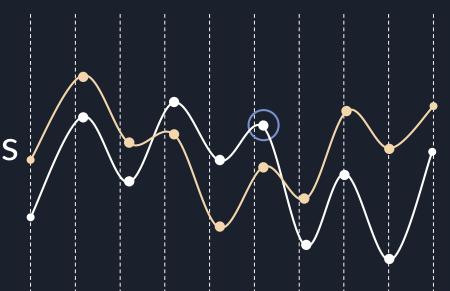
Data Profiling

The score obtained for each cell in the final dataset using the formula mentioned in the previous section was a huge number. In order to make it interpretable and useful, it has been scaled. The logarithmic function was applied to the scores and the resulting number was multiplied by 100

Data Enrichment

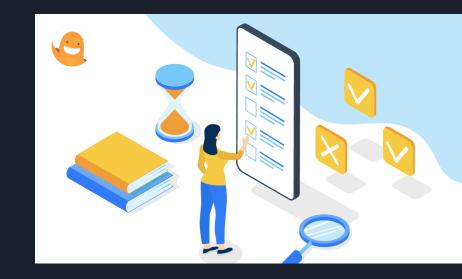
- After getting the dataset about all videos information and filtering the **top videos** for each creator, it is necessary to incorporate additional data, such as a sentiment analysis score, thumbnails, figures, guests, time, and display, which improved the quality of the data set by making it easier to analyze and by adding more valuable output to the final analysis.
- After receiving the data in the form of survey responses, the data was in raw form, which can be a
 quality issue. To improve the data, we added more tables based on the number of responses to each
 creator based on the type of content each creator produces, as well as the percentage for each
 category, making the data more relevant and easier to analyze.
- In the end, we have utilized the data sets that we have obtained from a variety of sources in order to get the final score table data set based on each city and content type.

Methods and Techniques /



Method: Survey for primary data

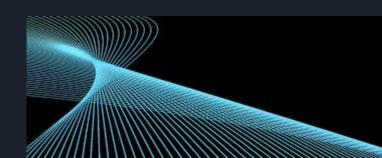
Tools & Technologies: Google Survey forms



Methods

- Linear Regression
 Linear regression has helped in estimating the significance of each
 attribute
- Logistic Regression
 Logistic regression has helped in calculating mean the mean values of watch time for each category.

Tools & Technologies: Python



Method: Sentimental Analysis

Tool: IBM Watson Engine

We have analysed the comments on the top videos of each creator to understand how the videos are being received.



Method: Score Generation

Final Formula:

((0.625*(likes)) + (0.25*(Sentimental Analysis score) + (12.5*Duration))

* (Subscriber Score) * (Category Score) * (Geographical Score)

Tools & Technologies: Python, Excel

Challenges: Human Beings Having to Work

- Human Judgement involved in the mathematical formula
- Manually categorising the videos
- Sentimental analysis had to be manually done for each comment
- Reliability of the survey results

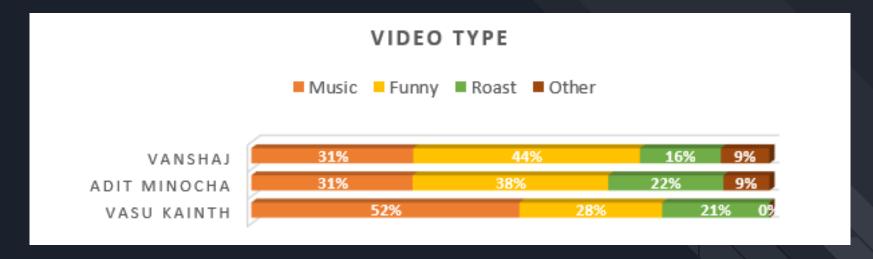
Output and Analysis

Scores

	Creators											
Cities	V	anshaj Sing	h	١	/asu Kainth		Adit Minocha					
	Funny	Roast	Music	Music	Funny	Roast	Music	Roast	Funny			
Delhi	1832.825	1729.697	1639.443	2029.632	2086.652	2053.477	1391.571	1767.232	1833.374			
Mumbai	1770.402	1667.273	1577.019	1983.168	2040.188	2007.013	1391.224	1766.884	1833.026			
Hyderabad	1644.741	1541.612	1451.358	1867.817	1924.838	1891.663	1269.547	1645.207	1711.349			
Bengaluru	1673.97	1570.841	1480.587	1886.233	1943.254	1910.079	1321.3	1696.96	1763.102			
Ahmedabad	1723.059	1619.93	1529.676	1968.127	2025.148	1991.973	1266.407	1642.067	1708.209			

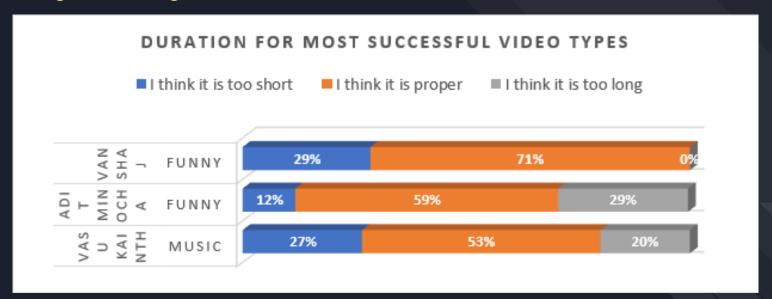
- Scores about estimating each creator's performance in each city
- The higher, the better

Survey Analysis for Video Types



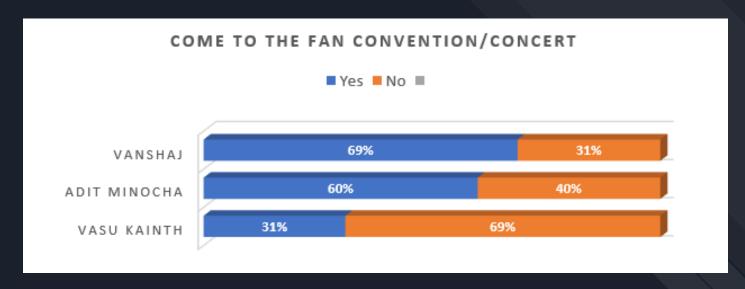
- Figure out the most successful video type for each creator from viewer's opinion
- Then dig out the exact pattern for these successful videos viewers like

Survey Analysis for Duration



- For the most successful video type in viewers' opinion, people choose it is long, short or proper
- For Adit, Funny videos are over 10 mins, 59% people think it is proper
- For Vanshaj, Funny videos are around 60s, 71% people think it is proper
- Duration is not a main factor in successful videos, they need keep their unique style

Survey Analysis for Fan Convention/ Concert



- For Vasu, 69% people show they don't want to come his concerts

 For Adit and Vanshaj, over 60% people show they want to come their concerts
- Consider people's opinions, we will measure the score again

Suggestion from Scores

		Creators										
Cities	V	anshaj Sing	h	\	Vasu Kainth Adit Minocha							
	Funny	Roast	Music	Music	Funny	Roast	Music	Roast	Funny			
Delhi	1832.825	1729.697	1639.443	2029.632	2086.652	2053.477	1391.571	1767.232	1833.374			
Mumbai	1770.402	1667.273	1577.019	1983.168	2040.188	2007.013	1391.224	1766.884	1833.026			
Hyderabad	1644.741	1541.612	1451.358	1867.817	1924.838	1891.663	1269.547	1645.207	1711.349			
Bengaluru	1673.97	1570.841	1480.587	1886.233	1943.254	1910.079	1321.3	1696.96	1763.102			
Ahmedabad	1723.059	1619.93	1529.676	1968.127	2025.148	1991.973	1266.407	1642.067	1708.209			

- For Delhi, we suggest Adit and Vanshaj perform funny shows together.
- For Mumbai, we suggest Adit and Vanshaj perform funny shows together.
- For Hyderabad, we suggest Adit perform Funny and Roast style shows on the stage.
- For Bengaluru, we suggest Adit perform Funny and Roast style shows on the stage.
- For Ahmedabad, we suggest Adit and Vanshaj perform funny shows together.

Thank You!