

DATA MINING PRINCIPLES
MSCA 31008

Final Project Submission

Youtube Analysis

Austin Sky-Kappler
Carrie Wang

Agenda

- 1) Executive Summary
- 2) Data
 - Explain
 - Analyze
 - Visualize
- 1) Data Prep
 - Cleaning / Transformation
 - Resulting Data Variables
- 1) Modeling
 - Model Fitting
 - Model Performance
 - Model Predictions
 - Feature Relevance
- 1) Conclusion
 - Findings
 - Next Steps

Executive summary



Executive Summary

Problem Statement: As Youtube continues to grow in popularity, content creation on the platform has become mainstream and can be a viable source of income for the creators that operate on it. One issue that these channels may run into is trying to determine what subjects to create content on and then how to structure the video and structure description in a manner that optimizes engagement for the channel.

Project Overview: By connecting to the Youtube API and generating a sample dataset of videos and channels, the team mined and analyzed features to create a model that will predict a video's weighted engagement score ($\text{Likes} + 2 \times \text{Comments}$).

Data Overview - Video - 15,000 Videos Analyzed

<u>Feature</u>	<u>Description</u>	<u>Data Type</u>
video_id	Unique identifier for each video	String
video_title	Title for the video	String
video_description_short	Short description of the video that displays when it is pulled up	String
video_description_full	Displays when user clicks show more	String
video_thumbnail	Image that displays for the video	Object (jpg)
video_upload_date	Date of video upload	
video_view_count	Count of views by users	Int64
video_like_count	Count of likes by users	Int64
video_dislike_count	Count of dislikes by users	Int64
video_comment_count	Count of comments left by users	int64

<u>Feature</u>	<u>Description</u>	<u>Data Type</u>
video_tags	Meta tags to support SEO	Object
video_categoryID	ID determining video category	Int64
video_definition	Video quality – SD vs HD	String
video_duration	Length of video	Object
video_caption	Indicates whether captions are available for the video.	Boolean
video_licensedContent	Indicates whether the video represents licensed content, which means that the content was uploaded to a channel linked to a YouTube content partner and then claimed by that partner.	Boolean

Extracted from Youtube.com via Google Dev Platform API
Connection from 12/1/21 - 12/7/21



Video Data Overview

#datamining #predictivemodeling #predictiveanalytics
What is Data Mining and Why is it Important?
30,378 views · Mar 12, 2021

Eye on Tech
26.2K subscribers

KDD in data mining assists data prep for ML: <https://searchenterpriseai.techtarget.com/>
What is data mining: <https://searchsqlserver.techtarget.com/>

SHOW MORE

25 Comments SORT BY


Add a public comment...

Search filters: All From your search Data mining Mac

- What is a Data Warehouse? 365 Data Science 122K views · 1 year ago
- Data Mining Fundamentals Dave Sullivan 33K views · 4 years ago
- What is Data Mining? Cubeware GmbH 2.5K views · 2 months ago
- What REALLY is Data Science? Told by a Data Scientist Joma Tech 2.5M views · 3 years ago
- Mix - Eye on Tech YouTube
- DATA MINING IN HEALTHCARE Herbert Chase 2.8K views · 2 years ago
- Becoming Alpha Males Thanks To Gary Vee Coffeezilla 197K views · 3 days ago
- Top 10 Craziest Rogue Traders Patrick Boyle 191K views · 1 year ago
- Evergrande Makes Bond Payment! Patrick Boyle 111K views · 1 month ago
- What is a Data Warehouse - Explained with real life examples IT & Funds 94K views · 1 year ago
- Data Mining: How You're Revealing More Than You Think SciShow

Video Data Overview

Title



What is Data Mining and Why is it Important?

30,378 views · Mar 12, 2021

Eye on Tech
26.2K subscribers

KDD in data mining assists data prep for ML: <https://searchenterpriseai.techtarget.com/>
What is data mining: <https://searchsqlserver.techtarget.com/>

SHOW MORE

25 Comments SORT BY

Add a public comment...

432 DISLIKE SHARE SAVE ...

SUBSCRIBE

From your search Data mining Mac

- What Is a Data Warehouse? 365 Data Science 122K views · 1 year ago 3:32
- Data Mining Fundamentals Dave Sullivan 33K views · 4 years ago 7:48
- What is Data Mining Cuberware GmbH 2.5K views · 2 months ago 6:05
- What REALLY is Data Science? Told by a Data Scientist Joma Tech 2.5M views · 3 years ago 11:09
- Data Mining Mix - Eye on Tech YouTube 36:28
- DATA MINING IN HEALTHCARE Herbert Chase 2.8K views · 2 years ago
- Becoming Alpha Males Thanks To Gary Vee Coffeezilla 197K views · 3 days ago New
- Top 10 Craziest Rogue Traders Patrick Boyle 191K views · 1 year ago 33:20
- Evergrande Makes Bond Payment! Patrick Boyle 111K views · 1 month ago 9:49
- What is a Data Warehouse - Explained with real life example IT & Funde 34K views · 1 year ago 9:37
- Data Mining: How You're Revealing More Than You Think SciShow



Video Data Overview

Views

#datamining #predictivemodeling #predictiveanalytics
What Is Data Mining and Why is it Important?
30,378 views • Mar 12, 2021

Eye on Tech
26.2K subscribers

KDD in data mining assists data prep for ML: <https://searchenterpriseai.techtarget.com/>
What is data mining: <https://searchsqlserver.techtarget.com/>

SHOW MORE

25 Comments SORT BY

Add a public comment...

Annotations
Playback speed Normal
Subtitles/CC (2) Off
Quality Auto 720p60

432 DISLIKE SHARE SAVE ...

SUBSCRIBE

What Is a Data Warehouse? 365 Data Science 122K views • 1 year ago 3:32

Data Mining Fundamentals Dave Sullivan 33K views • 4 years ago 7:48

What is Data Mining Cuberware GmbH 2.5K views • 2 months ago 6:05

DATA SCIENCE Told by a Data Scientist Joma Tech 2.5M views • 3 years ago 11:09

Data Mining Mix - Eye on Tech YouTube 36:28

DATA MINING IN HEALTHCARE Herbert Chase 2.8K views • 2 years ago

Becoming Alpha Males Thanks To Gary Vee Coffeezilla 197K views • 3 days ago New

TOP 10 CRAZIEST ROGUE TRADERS Top 10 Craziest Rogue Traders Patrick Boyle 191K views • 1 year ago 33:20

Evergrande Makes Bond Payment! Patrick Boyle 111K views • 1 month ago 9:49

What is a Data Warehouse - Explained with real life example! IT & Funde 34K views • 1 year ago 9:37

Data Mining: How You're Revealing More Than You Think SciShow



Video Data Overview

Upload Date

What is Data Mining? Why is it important?

30,378 views • Mar 12, 2021

Eye on Tech
26.2K subscribers

KDD in data mining assists data prep for ML: <https://searchenterpriseai.techtarget.com/>
What is data mining: <https://searchsqlserver.techtarget.com/>

25 Comments SORT BY

Add a public comment...

Recommended videos:

- What Is a Data Warehouse? 122K views • 1 year ago
- Data Mining Fundamentals 33K views • 4 years ago
- What is Data Mining 2.5K views • 2 months ago
- What REALLY is Data Science? Told by a Data Scientist 2.5M views • 3 years ago
- Mix - Eye on Tech YouTube
- DATA MINING IN HEALTHCARE 2.8K views • 2 years ago
- Becoming Alpha Males Thanks To Gary Vee 197K views • 3 days ago
- Top 10 Craziest Rogue Traders 191K views • 1 year ago
- Evergrande Makes Bond Payment! 111K views • 1 month ago
- What is a Data Warehouse - Explained with real life exampl... 34K views • 1 year ago
- Data Mining: How You're Revealing More Than You Think SciShow



Video Data Overview

The screenshot shows a YouTube video player for the video "What is Data Mining and Why is it Important?" by the channel "Eye on Tech". The video player shows a silhouette of a person mining with a pickaxe. The channel name "Eye on Tech" is highlighted with a red box and a red arrow pointing to it from the text "Channel Name" on the left. The video has 30,378 views and was posted on Mar 12, 2021. The description includes links to KDD and data mining resources. The right sidebar shows a list of recommended videos related to data mining and data science.

Channel Name

Channel Name



Video Data Overview

Subscriber Count

The image shows a YouTube video player interface. The video is titled "What is Data Mining and Why is it Important?" by Joma Tech. The video player shows a silhouette of a person mining with a pickaxe. The video has 30,378 views and was uploaded on Mar 12, 2021. The description includes links to KDD and search engines. The subscriber count is 26.2K. The right sidebar shows a list of related videos.

What is Data Mining and Why is it Important?

30,378 views · Mar 12, 2021

432 DISLIKE SHARE SAVE ...

Joma Tech

26.2K subscribers

KDD in data mining assists data prep for ML: <https://searchenterpriseai.techtarget.com/>
What is data mining: <https://searchsqlserver.techtarget.com/>

SHOW MORE

25 Comments SORT BY

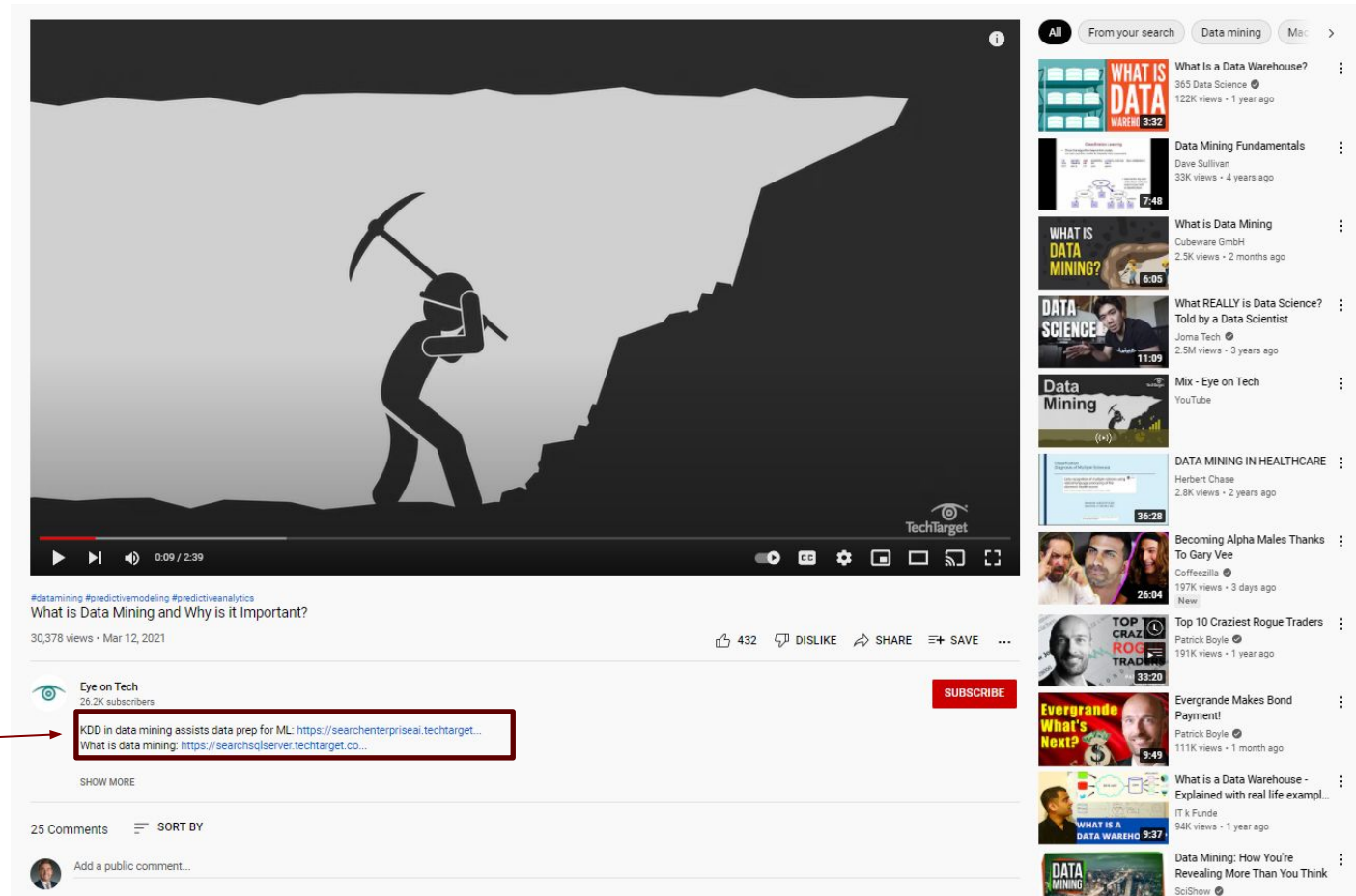
Add a public comment...

Related videos:

- What Is a Data Warehouse? 365 Data Science 122K views · 1 year ago
- Data Mining Fundamentals Dave Sullivan 33K views · 4 years ago
- What is Data Mining Cuberware GmbH 2.5K views · 2 months ago
- What REALLY is Data Science? Told by a Data Scientist Joma Tech 2.5M views · 3 years ago
- Mix - Eye on Tech YouTube
- DATA MINING IN HEALTHCARE Herbert Chase 2.8K views · 2 years ago
- Becoming Alpha Males Thanks To Gary Vee Coffeezilla 197K views · 3 days ago
- Top 10 Craziest Rogue Traders Patrick Boyle 191K views · 1 year ago
- Evergrande Makes Bond Payment! Patrick Boyle 111K views · 1 month ago
- What is a Data Warehouse - Explained with real life example IT & Funde 34K views · 1 year ago
- Data Mining: How You're Revealing More Than You Think SciShow



Video Data Overview



#datamining #predictivemodeling #predictiveanalytics
What is Data Mining and Why is it Important?
30,378 views · Mar 12, 2021

Eye on Tech
26.2K subscribers

KDD in data mining assists data prep for ML: <https://searchenterpriseai.techtarget.com/...>
What is data mining: <https://searchsqlserver.techtarget.com/...>

25 Comments SORT BY

Add a public comment...

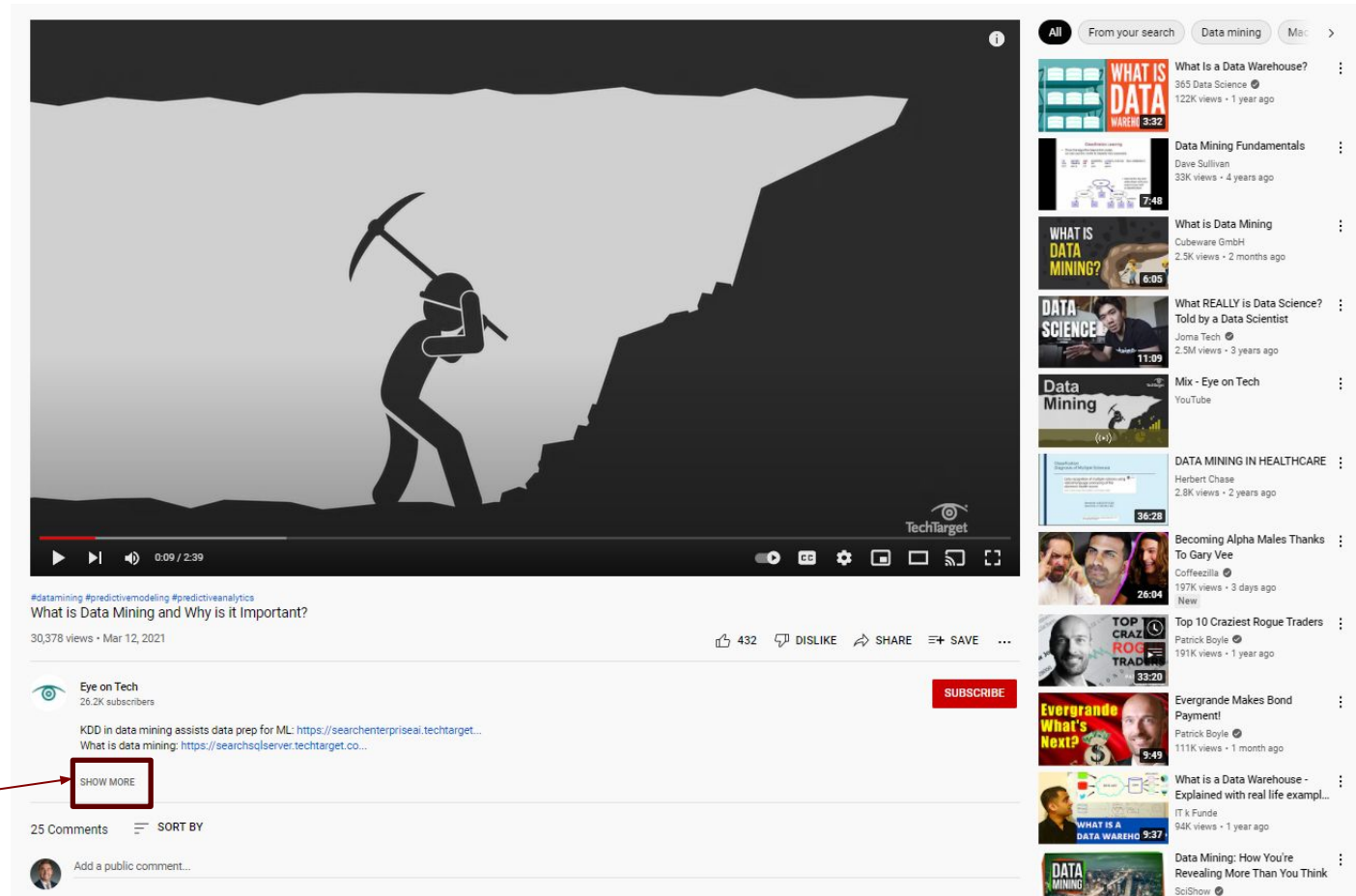
Recommended videos:

- What Is a Data Warehouse? 365 Data Science 122K views · 1 year ago
- Data Mining Fundamentals Dave Sullivan 33K views · 4 years ago
- What is Data Mining Cuberware GmbH 2.5K views · 2 months ago
- What REALLY is Data Science? Told by a Data Scientist Joma Tech 2.5M views · 3 years ago
- Mix - Eye on Tech YouTube
- DATA MINING IN HEALTHCARE Herbert Chase 2.8K views · 2 years ago
- Becoming Alpha Males Thanks To Gary Vee Coffeezilla 197K views · 3 days ago
- Top 10 Craziest Rogue Traders Patrick Boyle 191K views · 1 year ago
- Evergrande Makes Bond Payment! Patrick Boyle 111K views · 1 month ago
- What is a Data Warehouse - Explained with real life example IT & Funke 34K views · 1 year ago
- Data Mining: How You're Revealing More Than You Think SciShow

Short Description



Video Data Overview

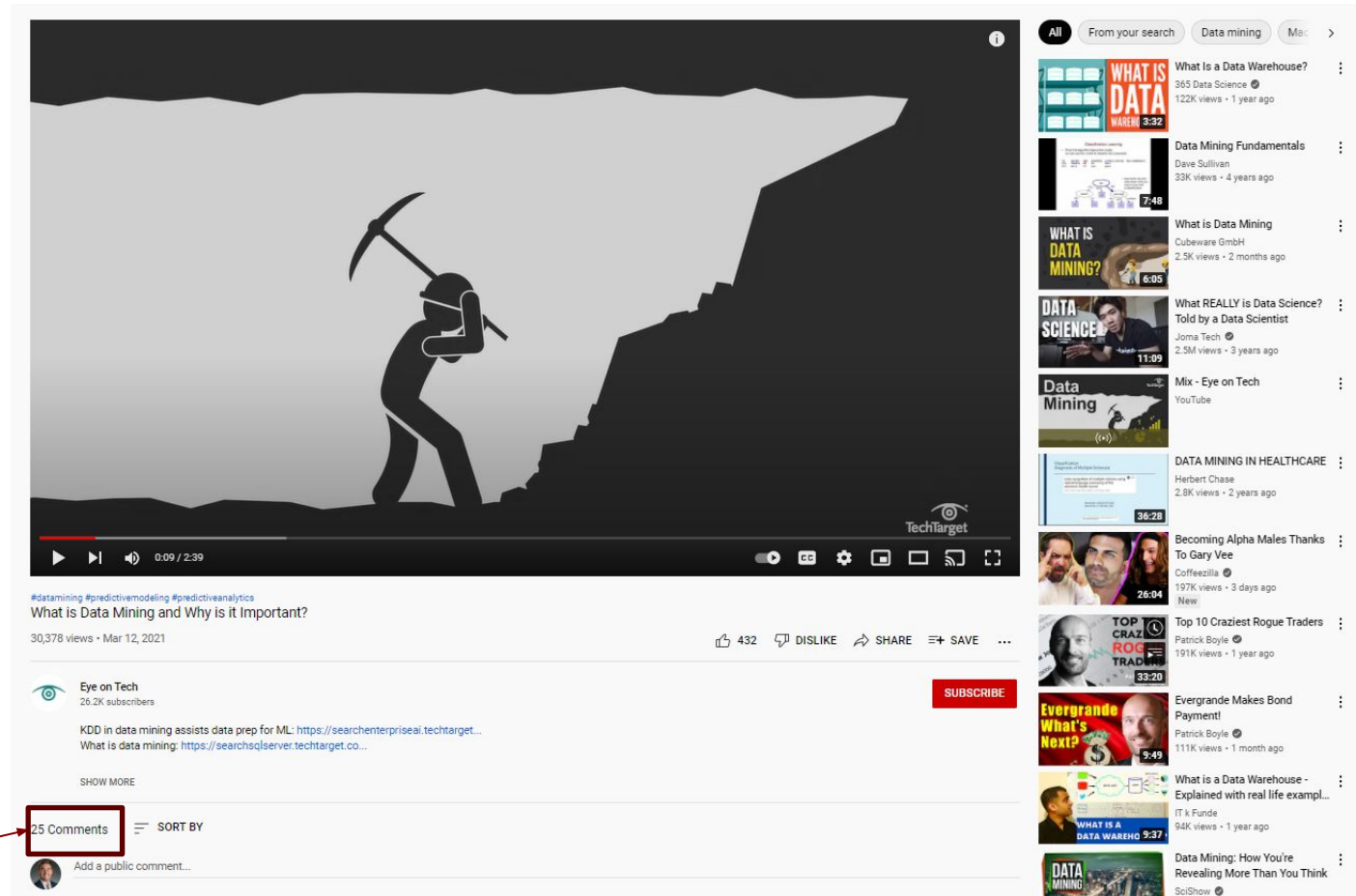


The image shows a YouTube video player interface. The main video is titled "What is Data Mining and Why is it Important?" by the channel "Eye on Tech". The video player shows a silhouette of a person mining with a pickaxe. Below the video player, the description is visible, starting with "#datamining #predictivemodeling #predictiveanalytics". A red box highlights the "SHOW MORE" button in the description. To the right of the video player, there is a sidebar with several recommended videos, including "What is a Data Warehouse?", "Data Mining Fundamentals", "What is Data Mining", "What REALLY is Data Science?", "Data Mining", "DATA MINING IN HEALTHCARE", "Becoming Alpha Males Thanks To Gary Vee", "Top 10 Craziest Rogue Traders", "Evergrande Makes Bond Payment!", "What is a Data Warehouse - Explained with real life exampl...", and "Data Mining: How You're Revealing More Than You Think".

Long Description



Video Data Overview

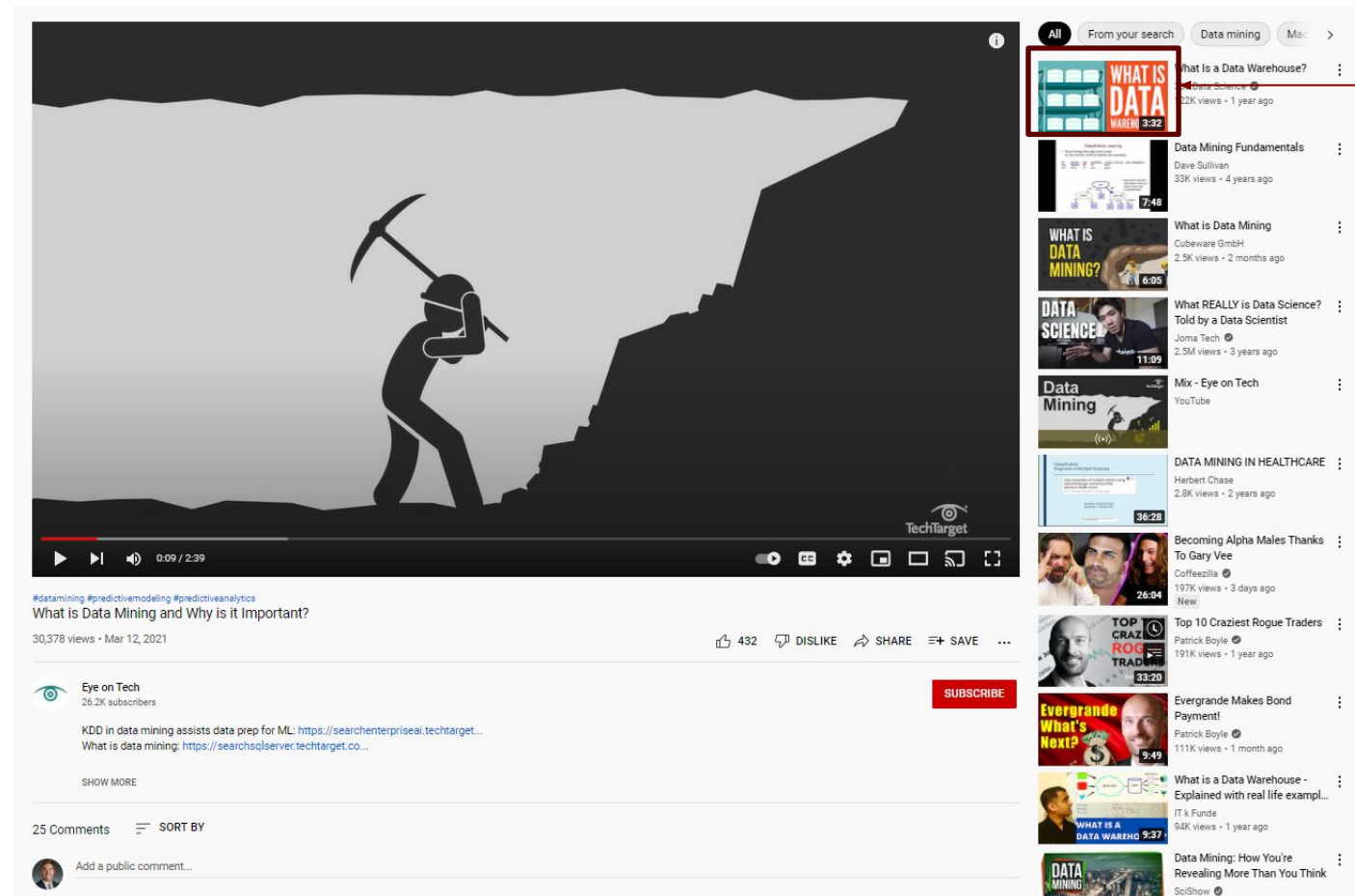


The image shows a YouTube video player interface. The main video is titled "What is Data Mining and Why is it Important?" by the channel "Eye on Tech". The video player shows a silhouette of a person mining with a pickaxe. Below the video player, the description reads: "KDD in data mining assists data prep for ML: <https://searchenterpriseai.techtarget.com/>. What is data mining: <https://searchsqlserver.techtarget.com/>". To the right of the video player, there is a list of recommended videos, including "What is a Data Warehouse?", "Data Mining Fundamentals", "What is Data Mining", "What REALLY is Data Science?", "Data Mining", "DATA MINING IN HEALTHCARE", "Becoming Alpha Males Thanks To Gary Vee", "Top 10 Craziest Rogue Traders", "Evergrande Makes Bond Payment!", "What is a Data Warehouse - Explained with real life exampl...", and "Data Mining: How You're Revealing More Than You Think".

Comments

25 Comments

Video Data Overview



Thumbnail



Video Data Overview

The screenshot shows a YouTube video player interface. The video is titled "What is Data Mining and Why is it Important?" by the channel "Eye on Tech". The video player shows a silhouette of a person mining. The right sidebar displays a list of related videos, including "What Is a Data Warehouse?", "Data Mining Fundamentals", "What is Data Mining", "What REALLY is Data Science?", "Mix - Eye on Tech", "DATA MINING IN HEALTHCARE", "Becoming Alpha Males Thanks To Gary Vee", "Top 10 Craziest Rogue Traders", "Evergrande Makes Bond Payment!", "What is a Data Warehouse - Explained with real life exampl...", and "Data Mining: How You're Revealing More Than You Think". A red box highlights the "Like" and "Dislike" buttons, with a red arrow pointing to the text "Likes / Dislikes".

Likes / Dislikes

Video Data Overview

The screenshot shows a YouTube video player interface. The video is titled "What is Data Mining and Why is it Important?". The video player shows a progress bar at 0:09 / 2:39. The video player is embedded in a page with a sidebar on the right showing a list of related videos. The video player is also showing the video title, description, and a list of comments.

Annotations

Playback speed Normal >

Subtitles (CC) Off >

Quality Auto 720p60 >

What is Data Mining and Why is it Important?

30,378 views · Mar 12, 2021

Eye on Tech
26.2K subscribers

KDD in data mining assists data prep for ML: <https://searchenterpriseai.techtarget.com/>
What is data mining: <https://searchsqlserver.techtarget.com/>

SHOW MORE

25 Comments SORT BY

Add a public comment...

Related videos:

- What Is a Data Warehouse? 365 Data Science 122K views · 1 year ago
- Data Mining Fundamentals Dave Sullivan 33K views · 4 years ago
- What is Data Mining Cuberware GmbH 2.5K views · 2 months ago
- What REALLY is Data Science? Told by a Data Scientist Joma Tech 2.5M views · 3 years ago
- Mix - Eye on Tech YouTube
- DATA MINING IN HEALTHCARE Herbert Chase 2.8K views · 2 years ago
- Becoming Alpha Males Thanks To Gary Vee Coffeezilla 197K views · 3 days ago
- Top 10 Craziest Rogue Traders Patrick Boyle 191K views · 1 year ago
- Evergrande Makes Bond Payment! Patrick Boyle 111K views · 1 month ago
- What is a Data Warehouse - Explained with real life example IT & Funke 34K views · 1 year ago
- Data Mining: How You're Revealing More Than You Think SciShow

Quality



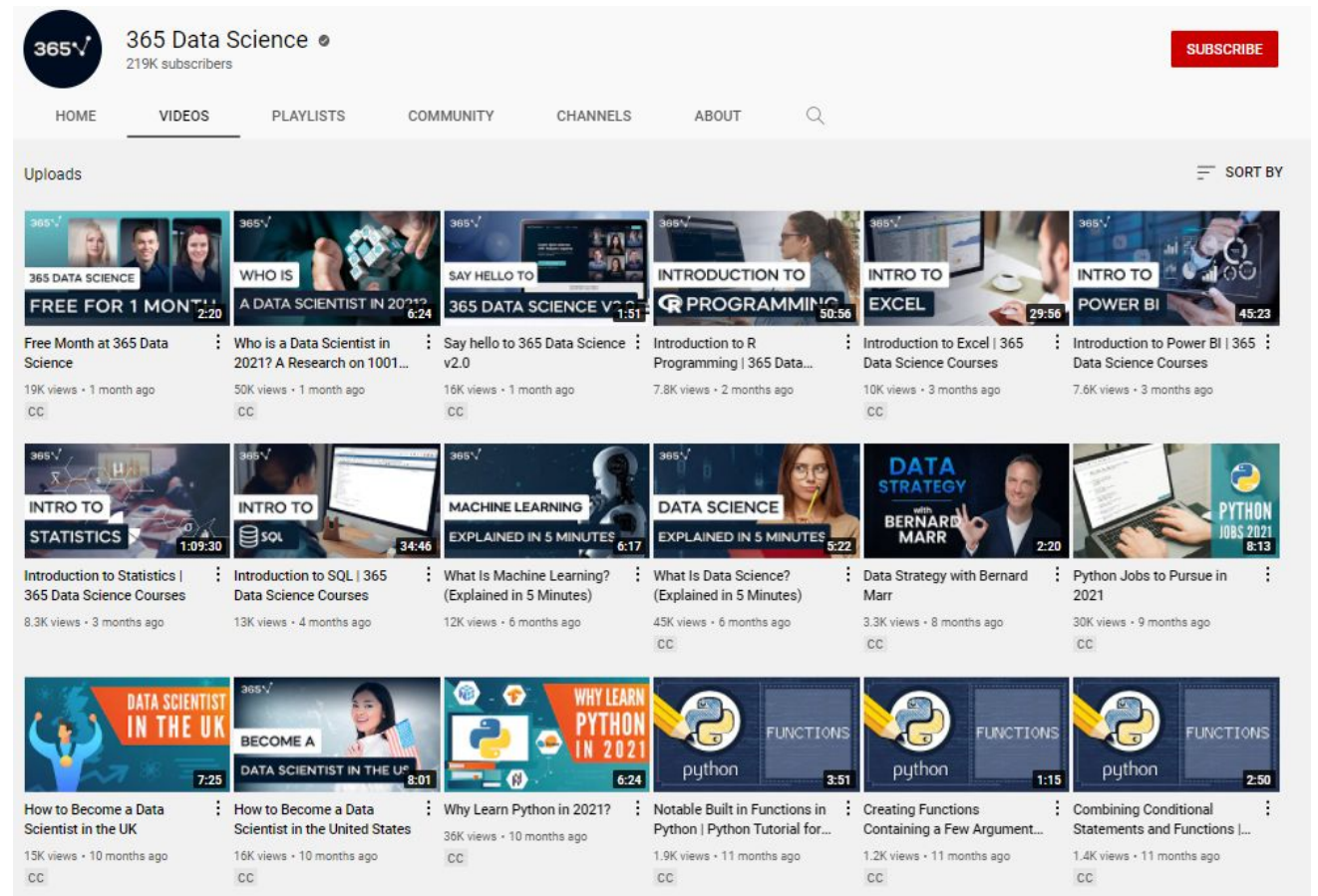
Data Overview - Channel - 200 Channels Analyzed

<u>Feature</u>	<u>Description</u>	<u>Data Type</u>
channel_title	Title for Channel	String
channel_description	Description of channel / content	String
channel_customUrl	Creator defined URL vs Youtube generated ID	String
channel_publishedAt	Date Channel first published	Datetime
channel_country	Country of origin for channel	String
channel_subscriberCount	Count of users that elected to receive notifications for the channel	Int64
channel_hiddenSubscriberCount	T/F creator elected to suppress count of subscribers	
channel_viewCount	Total views that all videos on the channel have received	Int64
channel_topicCategories	A list of Wikipedia URLs that describe the channel's content.	Int64
channel_privacyStatus	T/F is the video available for public access	

<u>Feature</u>	<u>Description</u>	<u>Data Type</u>
channel_longUploadsStatus	If the channel is eligible to post videos that are > 15 minutes	String
channel_madeForKids	whether the channel is designated as child-directed	Boolean
channel_keywords	Meta tags to support SEO	Int64
channel_videoCount	Total count of videos that have been uploaded by the channel	Int64

Extracted from Youtube.com via Google Dev Platform API
Connection from 12/1/21 - 12/7/21

Channel Data Overview



The screenshot displays the YouTube channel page for "365 Data Science", which has 219K subscribers. The channel's navigation bar includes links for HOME, VIDEOS, PLAYLISTS, COMMUNITY, CHANNELS, and ABOUT, along with a search icon. A red "SUBSCRIBE" button is located in the top right corner. The "Uploads" section is active, showing a grid of 18 video thumbnails arranged in three rows of six. Each video card includes a thumbnail, a title, a duration, and view statistics. The videos cover a range of topics including introductory content for various data science tools (R, Excel, Power BI, SQL, Python), machine learning, statistics, and career guidance. The channel's logo, a blue circle with "365" and a checkmark, is positioned in the top left of the page header.

Video Title	Duration	Views	Time Ago	License
Free Month at 365 Data Science	2:20	19K	1 month ago	CC
WHO IS A DATA SCIENTIST IN 2021? A Research on 1001...	6:24	50K	1 month ago	CC
SAY HELLO TO 365 DATA SCIENCE V2.0	1:51	16K	1 month ago	CC
INTRODUCTION TO R PROGRAMMING 365 Data...	50:56	7.8K	2 months ago	CC
INTRO TO EXCEL	29:56	10K	3 months ago	CC
INTRO TO POWER BI	45:23	7.6K	3 months ago	CC
INTRODUCTION TO STATISTICS 365 Data Science Courses	1:09:30	8.3K	3 months ago	CC
INTRODUCTION TO SQL 365 Data Science Courses	34:46	13K	4 months ago	CC
What Is Machine Learning? (Explained in 5 Minutes)	6:17	12K	6 months ago	CC
What Is Data Science? (Explained in 5 Minutes)	5:22	45K	6 months ago	CC
Data Strategy with Bernard Marr	2:20	3.3K	8 months ago	CC
Python Jobs to Pursue in 2021	8:13	30K	9 months ago	CC
How to Become a Data Scientist in the UK	7:25	15K	10 months ago	CC
How to Become a Data Scientist in the United States	8:01	16K	10 months ago	CC
Why Learn Python in 2021?	6:24	36K	10 months ago	CC
Notable Built in Functions in Python Python Tutorial for...	3:51	1.9K	11 months ago	CC
Creating Functions Containing a Few Argument...	1:15	1.2K	11 months ago	CC
Combining Conditional Statements and Functions ...	2:50	1.4K	11 months ago	CC

Channel Data Overview

youtube.com/c/365DataScience/about

365 data science

365 DataScience
BREAK INTO DATA SCIENCE

365 Data Science
219K subscribers

SUBSCRIBE

HOME VIDEOS PLAYLISTS COMMUNITY CHANNELS ABOUT

Description

At 365 Data Science we make #DataScience accessible to all by providing the highest quality online data science education. Our courses are designed to build up your skill set from beginner to job-ready, regardless of your current level, background education, and experience.

Our data science program helps students and young professionals embark on a career as a data scientist. Demand for qualified data scientists is at an all-time high and traditional degrees are simply too expensive, time-consuming or rigid to meet the majority of needs. Our mission is to close that gap.

We provide high quality, affordable training to open new opportunities for students and professionals all over the world. We constantly improve our program offering and the customer experience.
We've Helped More than 1 Million Student differentiate themselves by providing them with:

- 150+ Hours of video
- 1,980+ Lessons
- 35+ Courses
- 1.5M+ Students

Ready to Start Your Career?

Details

For business inquiries: [VIEW EMAIL ADDRESS](#)

Location: Bulgaria

Links

[Become a Data Science Expert](#) [Facebook](#)
[Instagram](#) [LinkedIn](#)

Stats

Joined Aug 7, 2017

8,606,133 views



Channel Data Overview

Custom URL

The screenshot shows a web browser window with the address bar displaying `youtube.com/c/365DataScience/about`. A red box highlights the `365DataScience` portion of the URL, and a red arrow points from the text "Custom URL" to this box. The browser's search bar contains the text "365 data science".

The YouTube channel page for "365 Data Science" is displayed. The header features a banner with the text "BREAK INTO DATA SCIENCE" and a group of six people. Below the banner, the channel name "365 Data Science" is shown with a verified badge and "219K subscribers". A red "SUBSCRIBE" button is visible in the top right corner of the channel header.

The navigation menu includes links for HOME, VIDEOS, PLAYLISTS, COMMUNITY, CHANNELS, and ABOUT (which is currently selected). The "ABOUT" section is divided into two columns:

- Description:**

At 365 Data Science we make #DataScience accessible to all by providing the highest quality online data science education. Our courses are designed to build up your skill set from beginner to job-ready, regardless of your current level, background education, and experience.

Our data science program helps students and young professionals embark on a career as a data scientist. Demand for qualified data scientists is at an all-time high and traditional degrees are simply too expensive, time-consuming or rigid to meet the majority of needs. Our mission is to close that gap.

We provide high quality, affordable training to open new opportunities for students and professionals all over the world. We constantly improve our program offering and the customer experience. We've Helped More than 1 Million Student differentiate themselves by providing them with:

 - 150+ Hours of video
 - 1,980+ Lessons
 - 35+ Courses
 - 1.5M+ Students

Ready to Start Your Career?
- Stats:**

Joined Aug 7, 2017

8,606,133 views

Below the description, there is a "Details" section with a "VIEW EMAIL ADDRESS" button and a "Location" field showing "Bulgaria". At the bottom, a "Links" section provides links to "Become a Data Science Expert", "Facebook", "Instagram", and "LinkedIn".

Channel Data Overview

youtube.com/c/365DataScience/about

365 data science

365 Data Science

BREAK INTO DATA SCIENCE

365 Data Science

365 Data Science

SUBSCRIBE

HOME VIDEOS PLAYLISTS COMMUNITY CHANNELS ABOUT

Description

At 365 Data Science we make #DataScience accessible to all by providing the highest quality online data science education. Our courses are designed to build up your skill set from beginner to job-ready, regardless of your current level, background education, and experience.

Our data science program helps students and young professionals embark on a career as a data scientist. Demand for qualified data scientists is at an all-time high and traditional degrees are simply too expensive, time-consuming or rigid to meet the majority of needs. Our mission is to close that gap.

We provide high quality, affordable training to open new opportunities for students and professionals all over the world. We constantly improve our program offering and the customer experience. We've Helped More than 1 Million Student differentiate themselves by providing them with:

- 150+ Hours of video
- 1,980+ Lessons
- 35+ Courses
- 1.5M+ Students

Ready to Start Your Career?

Details

For business inquiries: [VIEW EMAIL ADDRESS](#)

Location: Bulgaria

Links

[Become a Data Science Expert](#) [Facebook](#)

[Instagram](#) [LinkedIn](#)

Stats

Joined Aug 7, 2017

8,606,133 views

Title

Channel Data Overview

youtube.com/c/365DataScience/about

365 data science

365 Data Science
BREAK INTO DATA SCIENCE

365 Data Science
219K subscribers

SUBSCRIBE

HOME VIDEOS PLAYLISTS COMMUNITY CHANNELS ABOUT

Subscriber Count

Description

At 365 Data Science we make #DataScience accessible to all by providing the highest quality online data science education. Our courses are designed to build up your skill set from beginner to job-ready, regardless of your current level, background education, and experience.

Our data science program helps students and young professionals embark on a career as a data scientist. Demand for qualified data scientists is at an all-time high and traditional degrees are simply too expensive, time-consuming or rigid to meet the majority of needs. Our mission is to close that gap.

We provide high quality, affordable training to open new opportunities for students and professionals all over the world. We constantly improve our program offering and the customer experience. We've Helped More than 1 Million Student differentiate themselves by providing them with:

- 150+ Hours of video
- 1,980+ Lessons
- 35+ Courses
- 1.5M+ Students

Ready to Start Your Career?

Details

For business inquiries: [VIEW EMAIL ADDRESS](#)

Location: Bulgaria

Links

[Become a Data Science Expert](#) [Facebook](#)
[Instagram](#) [LinkedIn](#)

Stats

Joined Aug 7, 2017

8,606,133 views



Channel Data Overview

The screenshot shows the YouTube channel page for '365 Data Science'. The header features a banner with the text 'BREAK INTO DATA SCIENCE' and a group of people. Below the banner, the channel name '365 Data Science' is displayed with a verified badge and '219K subscribers'. A red 'SUBSCRIBE' button is on the right. The navigation bar includes links for HOME, VIDEOS, PLAYLISTS, COMMUNITY, CHANNELS, and ABOUT. The 'ABOUT' tab is selected. The 'Description' section is highlighted with a red box and contains the following text:

Description

At 365 Data Science we make #DataScience accessible to all by providing the highest quality online data science education. Our courses are designed to build up your skill set from beginner to job-ready, regardless of your current level, background education, and experience.

Our data science program helps students and young professionals embark on a career as a data scientist. Demand for qualified data scientists is at an all-time high and traditional degrees are simply too expensive, time-consuming or rigid to meet the majority of needs. Our mission is to close that gap.

We provide high quality, affordable training to open new opportunities for students and professionals all over the world. We constantly improve our program offering and the customer experience. We've Helped More than 1 Million Student differentiate themselves by providing them with:

- 150+ Hours of video
- 1,980+ Lessons
- 35+ Courses
- 1.5M+ Students

Ready to Start Your Career?

Details

For business inquiries: [VIEW EMAIL ADDRESS](#)

Location: Bulgaria

Links

[Become a Data Science Expert](#) [Facebook](#)

[Instagram](#) [LinkedIn](#)

Stats

Joined Aug 7, 2017

8,606,133 views

Description

Channel Data Overview

youtube.com/c/365DataScience/about

365 data science

365 DataScience
BREAK INTO DATA SCIENCE

365 Data Science
219K subscribers

SUBSCRIBE

HOME VIDEOS PLAYLISTS COMMUNITY CHANNELS ABOUT

Description

At 365 Data Science we make #DataScience accessible to all by providing the highest quality online data science education. Our courses are designed to build up your skill set from beginner to job-ready, regardless of your current level, background education, and experience.

Our data science program helps students and young professionals embark on a career as a data scientist. Demand for qualified data scientists is at an all-time high and traditional degrees are simply too expensive, time-consuming or rigid to meet the majority of needs. Our mission is to close that gap.

We provide high quality, affordable training to open new opportunities for students and professionals all over the world. We constantly improve our program offering and the customer experience. We've Helped More than 1 Million Student differentiate themselves by providing them with:

- 150+ Hours of video
- 1,980+ Lessons
- 35+ Courses
- 1.5M+ Students

Ready to Start Your Career?

Stats

Joined Aug 7, 2017

8,606,133 views

Details

For business inquiries: [VIEW EMAIL ADDRESS](#)

Location: Bulgaria

Links

[Become a Data Science Expert](#) [Facebook](#)

[Instagram](#) [LinkedIn](#)

Country

Channel Data Overview

The screenshot shows the YouTube channel page for '365 Data Science'. The channel has 219K subscribers and 8,606,133 views. The channel was joined on August 7, 2017. The description states that the channel provides high-quality online data science education. The channel is located in Bulgaria. The channel is published.

youtube.com/c/365DataScience/about

365 data science

365 Data Science
BREAK INTO DATA SCIENCE

365 Data Science
219K subscribers

HOME VIDEOS PLAYLISTS COMMUNITY CHANNELS ABOUT

Description

At 365 Data Science we make #DataScience accessible to all by providing the highest quality online data science education. Our courses are designed to build up your skill set from beginner to job-ready, regardless of your current level, background education, and experience.

Our data science program helps students and young professionals embark on a career as a data scientist. Demand for qualified data scientists is at an all-time high and traditional degrees are simply too expensive, time-consuming or rigid to meet the majority of needs. Our mission is to close that gap.

We provide high quality, affordable training to open new opportunities for students and professionals all over the world. We constantly improve our program offering and the customer experience. We've Helped More than 1 Million Student differentiate themselves by providing them with:

- 150+ Hours of video
- 1,980+ Lessons
- 35+ Courses
- 1.5M+ Students

Ready to Start Your Career?

Details

For business inquiries: [VIEW EMAIL ADDRESS](#)

Location: Bulgaria

Links

[Become a Data Science Expert](#) [Facebook](#)
[Instagram](#) [LinkedIn](#)

Stats

Joined Aug 7, 2017

8,606,133 views

Channel Published

Channel Data Overview

The screenshot shows the YouTube channel page for '365 Data Science'. The channel has 219K subscribers and 8,606,133 views. The page includes a banner with the text 'BREAK INTO DATA SCIENCE' and a group of people. The 'About' tab is selected, showing a description of the channel's mission and a list of statistics: 150+ Hours of video, 1,980+ Lessons, 35+ Courses, and 1.5M+ Students. A red box highlights the view count, and a red arrow points to it from the word 'Views'.

youtube.com/c/365DataScience/about

365 data science

365 Data Science
BREAK INTO DATA SCIENCE

365 Data Science
219K subscribers

SUBSCRIBE

HOME VIDEOS PLAYLISTS COMMUNITY CHANNELS ABOUT

Description

At 365 Data Science we make #DataScience accessible to all by providing the highest quality online data science education. Our courses are designed to build up your skill set from beginner to job-ready, regardless of your current level, background education, and experience.

Our data science program helps students and young professionals embark on a career as a data scientist. Demand for qualified data scientists is at an all-time high and traditional degrees are simply too expensive, time-consuming or rigid to meet the majority of needs. Our mission is to close that gap.

We provide high quality, affordable training to open new opportunities for students and professionals all over the world. We constantly improve our program offering and the customer experience. We've Helped More than 1 Million Student differentiate themselves by providing them with:

- 150+ Hours of video
- 1,980+ Lessons
- 35+ Courses
- 1.5M+ Students

Ready to Start Your Career?

Details

For business inquiries: VIEW EMAIL ADDRESS

Location: Bulgaria

Links

Become a Data Science Expert Facebook Instagram LinkedIn

Stats

Joined Aug 7, 2017

8,606,133 views

Views

Meta Data

Video Page Tags

```
"ue","kevlar_home_skeleton":true,"kevlar_home_skeleton_hide_later":true,"kevlar_include_query_in_search_endpoint":true,"kevlar_injector":true,"kevlar_inlined_html_templates_polymer_flags":true,"kevlar_js_fixes":true,"kevlar_k  
ttp..."><meta name="keywords" content="joma, vlog, data science, data scientist, facebook, google, netflix, amazon, data analyst, business intelligence, AI, deep learning, big data"><link rel="shortlinkUrl" href="https://you
```

Channel Keyword Tags

```
a\u003dtrue\u0026preskip_button_style_ads_backend\u003dcountdown_next_to_thumbnail\u0026qoe_nwl_downloads\u003dtrue\u0026qoe_send_and_write\u003dtrue\u0026record_app_crashed_web\u003dtrue\u0026render_unicode_emojis_as_images\u003dtr  
l_f9CZUw","keywords":["data analyst","data analyst jobs","data analyst" analytics "data analyst salary" "how to become a data analyst" "data analyst career" sql "analytics career" "data scientist" "how to become a d
```



Feature Heatmap



Data Prep



Data Exploration

RangeIndex: 9849 entries, 0 to 9848

Data columns (total 30 columns):

#	Column	Non-Null Count	Dtype
0	video_id	9849 non-null	object
1	video_title	9848 non-null	object
2	video_description_short	9404 non-null	object
3	video_description_full	9406 non-null	object
4	video_thumbnail	9849 non-null	object
5	upload_date	9849 non-null	object
6	view_count	9849 non-null	int64
7	like_count	9849 non-null	int64
8	dislike_count	9849 non-null	int64
9	comment_count	9823 non-null	float64
10	video_tags	8160 non-null	object
11	video_categoryID	9849 non-null	int64
12	video_definition	9849 non-null	object
13	video_duration	9849 non-null	object
14	video_caption	9849 non-null	bool
15	video_licensedContent	9849 non-null	bool

16	channel_title	9849 non-null	object
17	channel_description	8678 non-null	object
18	channel_customUrl	8502 non-null	object
19	channel_publishedAt	9849 non-null	object
20	channel_country	9406 non-null	object
21	channel_subscriberCount	9849 non-null	int64
22	channel_hiddenSubscriberCount	9849 non-null	bool
23	channel_viewCount	9849 non-null	int64
24	channel_topicCategories	9849 non-null	object
25	channel_privacyStatus	9849 non-null	object
26	channel_longUploadsStatus	9849 non-null	object
27	channel_madeForKids	7731 non-null	float64
28	channel_keywords	7574 non-null	object
29	channel_videoCount	9849 non-null	int64

dtypes: bool(3), float64(2), int64(7), object(18)

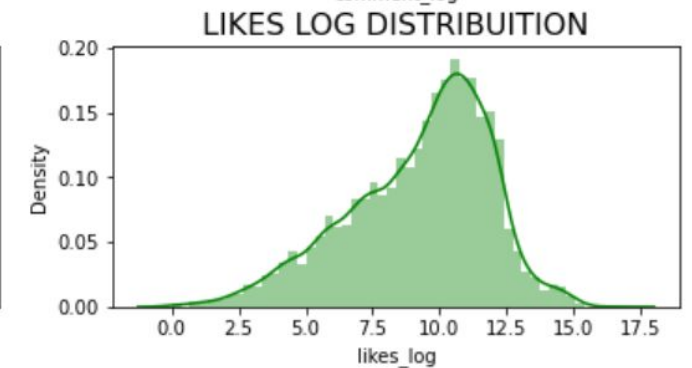
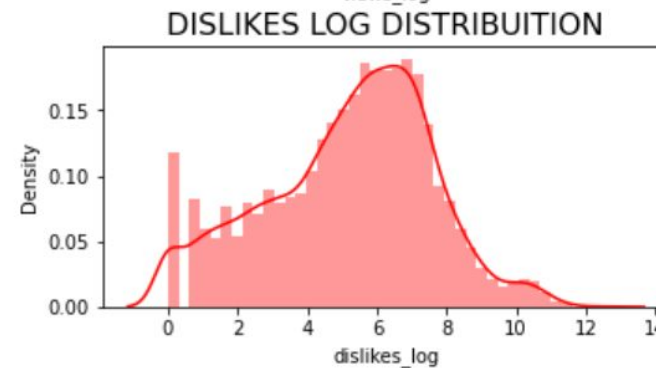
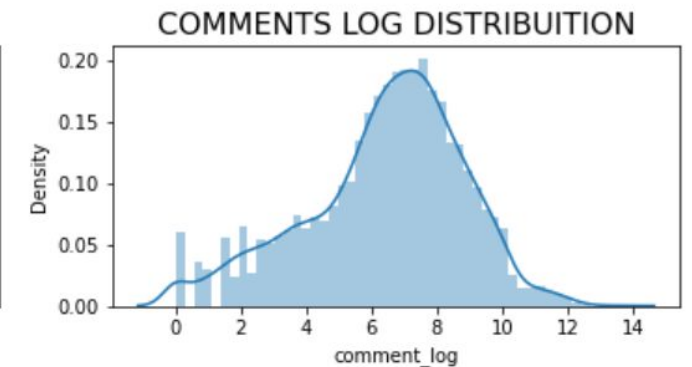
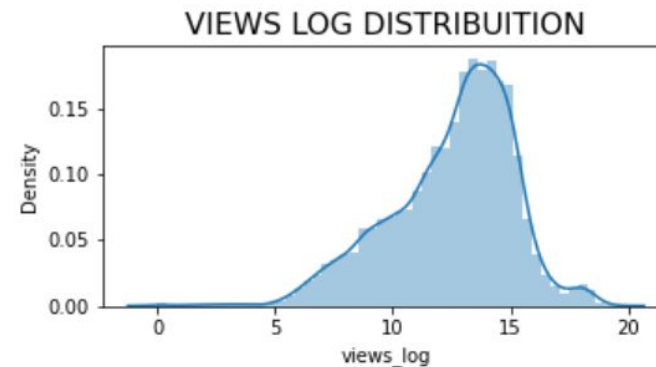
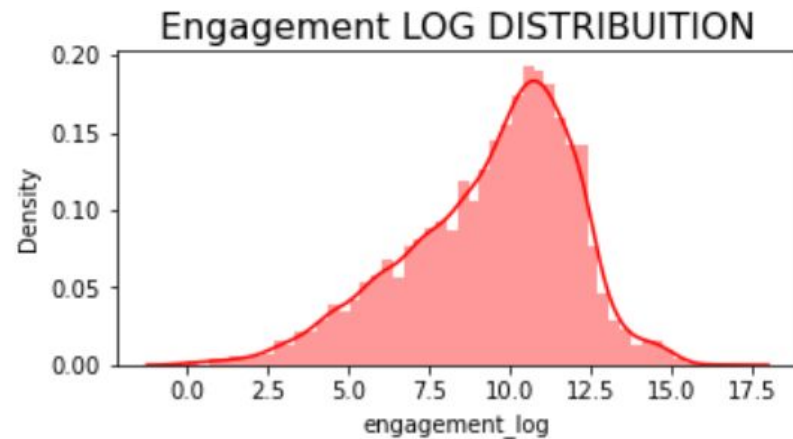
Removed unnecessary columns



Data Exploration

Target: Engagement!

Engagement = 2 * comment + likes
(we weight comments higher)



normal distribution?

Data Exploration

Categorical & Dummy Variables

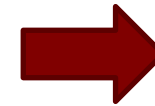
- Categories?
- Is HD?
- Is in US?
- Is for kids?
- Dummy variables transformed from other numerical variables, such as:
- Is video length between 20 and 60 minutes?
- Is the video new or old?

Numerical Variable

- Video duration: how many minutes
- Video upload dates
- Subscribers

Text Analysis

- Video title
- Video description
- Video tags
- Channel description
- Channel keywords



“Engagement”

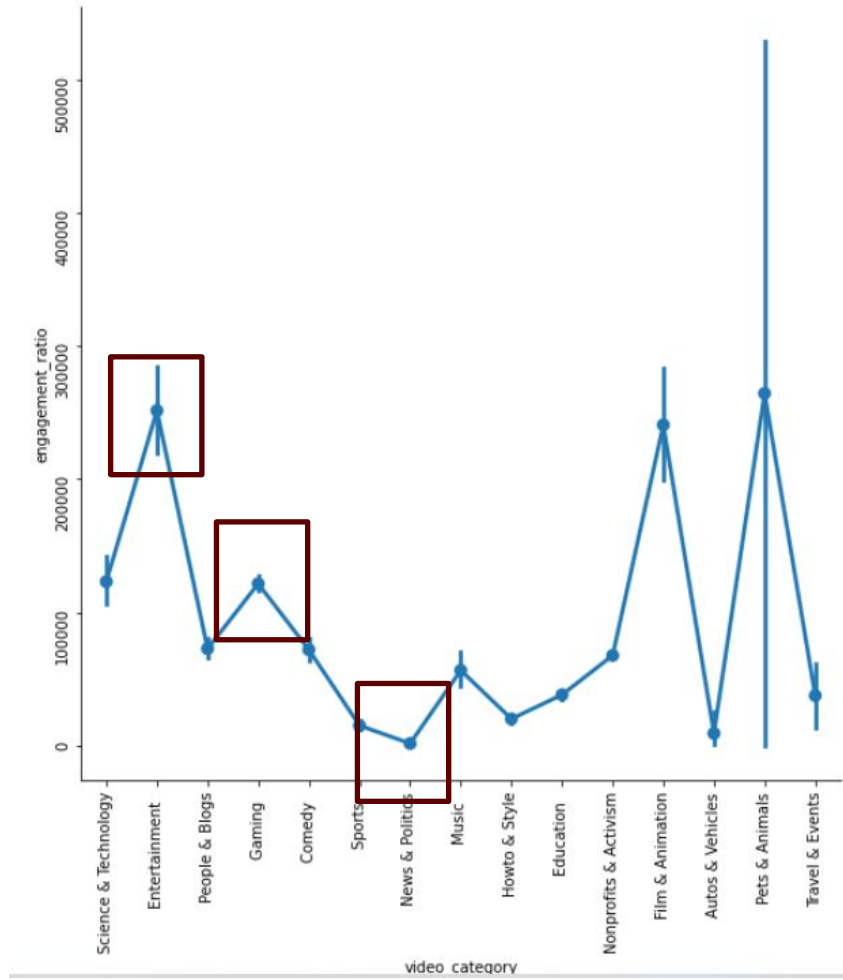
How to increase engagement from audience?



Data Exploration

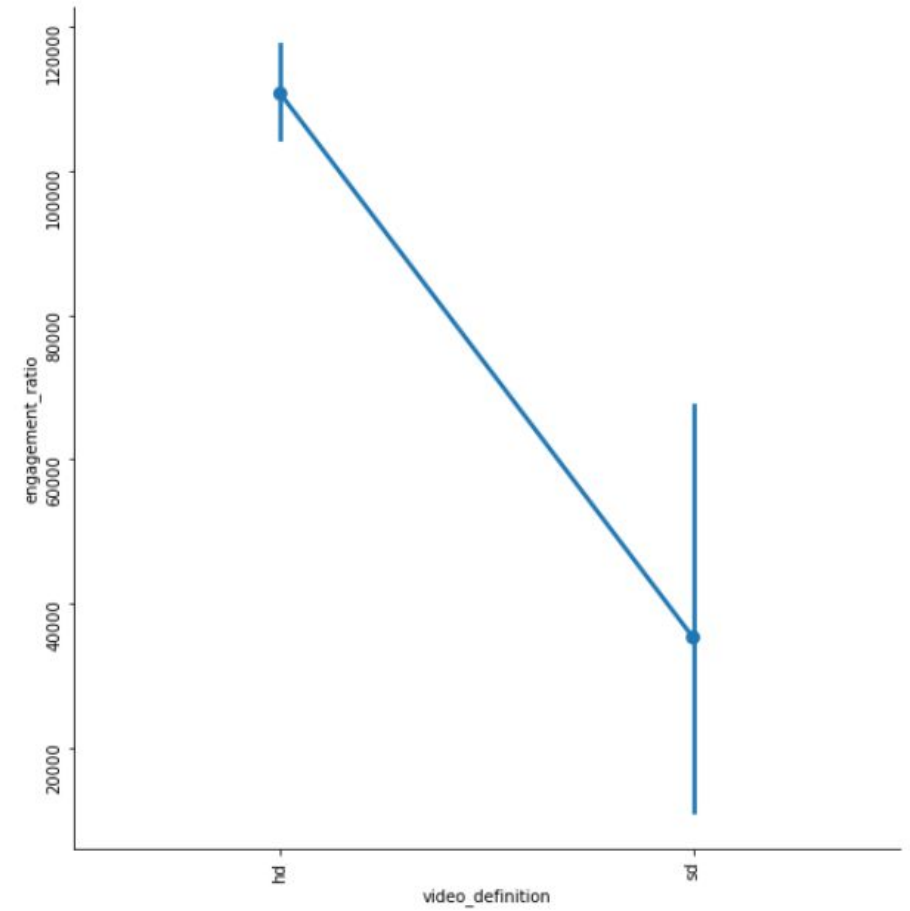
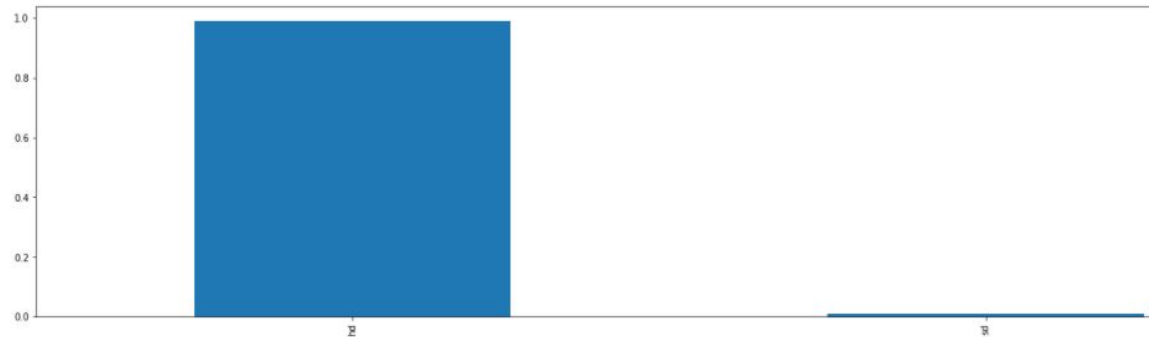
Gaming	3392
Entertainment	2151
People & Blogs	2080
Sports	1294
Comedy	976
Education	814
Howto & Style	796
Science & Technology	722
Music	499
Film & Animation	448
News & Politics	152
Autos & Vehicles	3
Pets & Animals	3
Travel & Events	2
Nonprofits & Activism	1

Name: video_category, dtype: int64

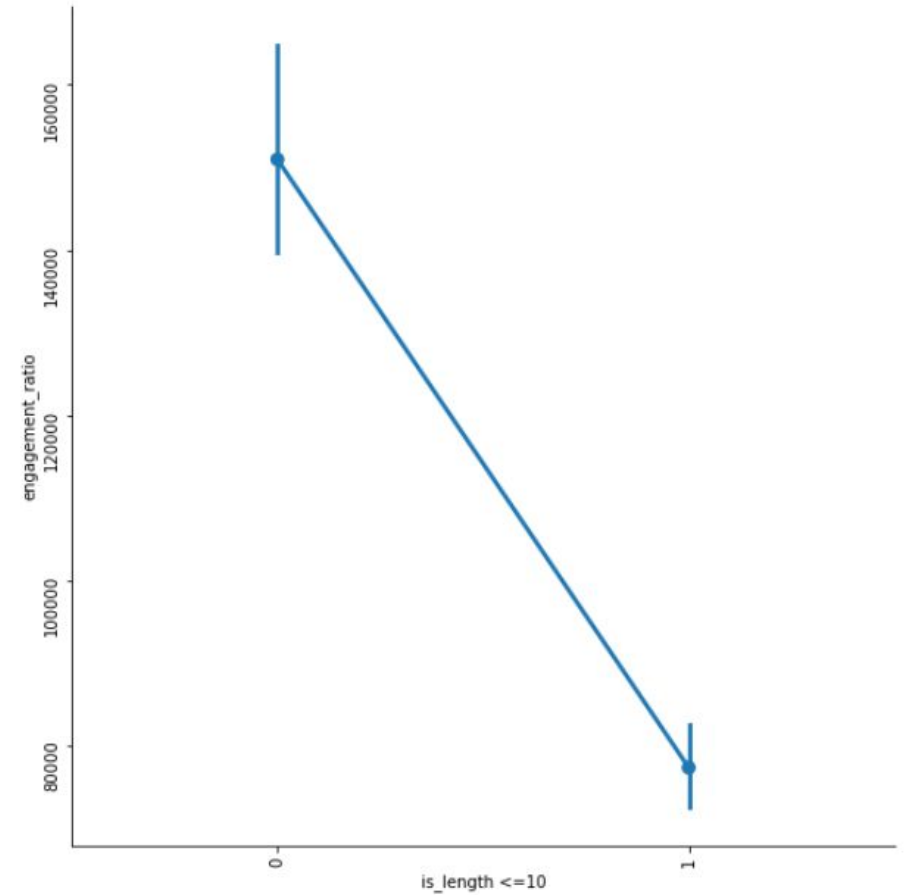
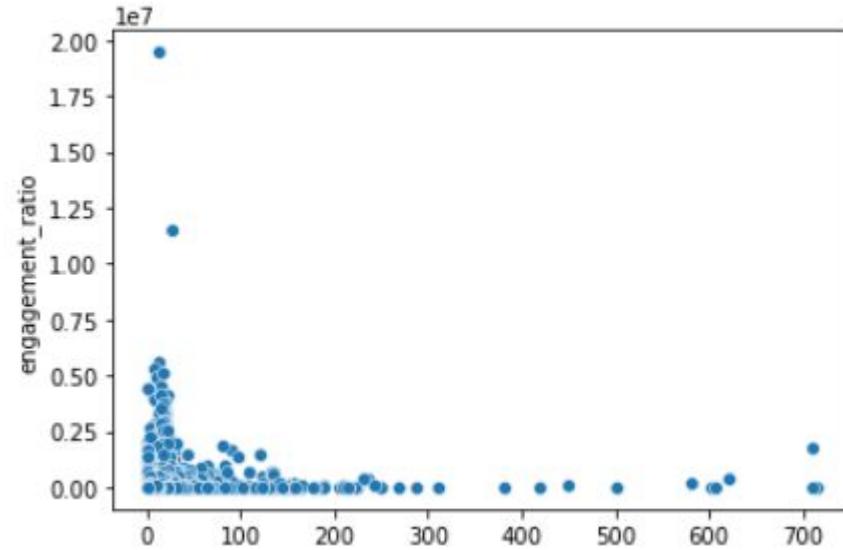


Data Exploration

```
hd    13194  
sd     139  
Name: video_definition, dtype: int64
```

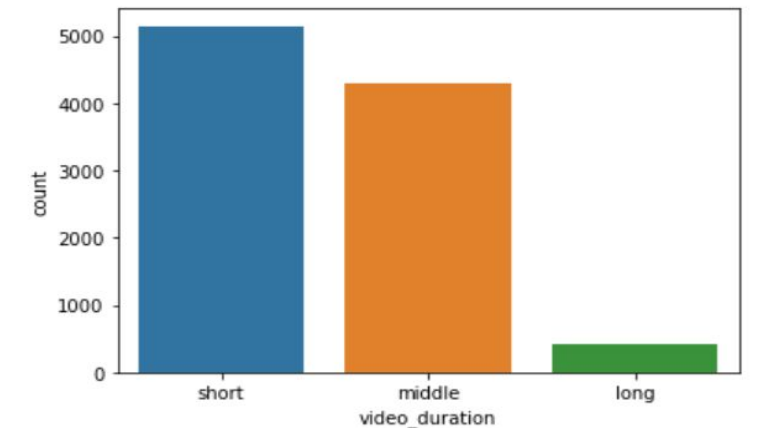
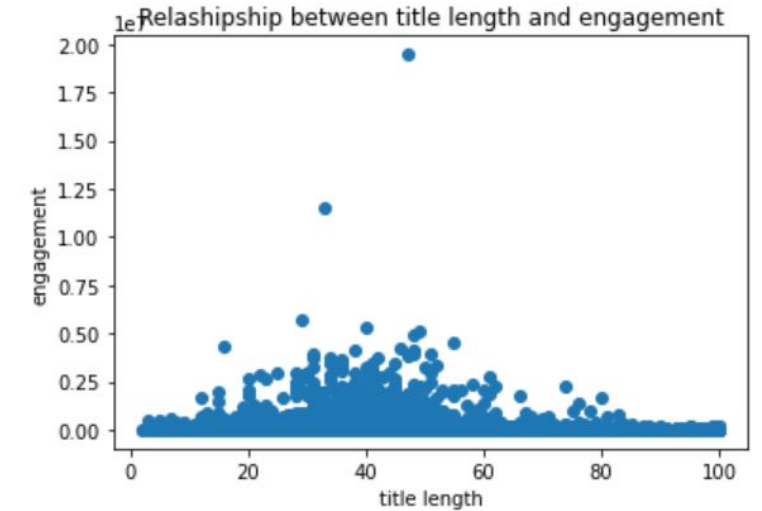
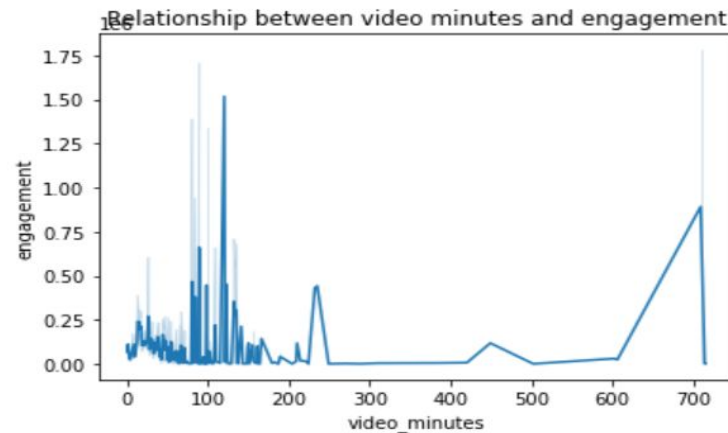
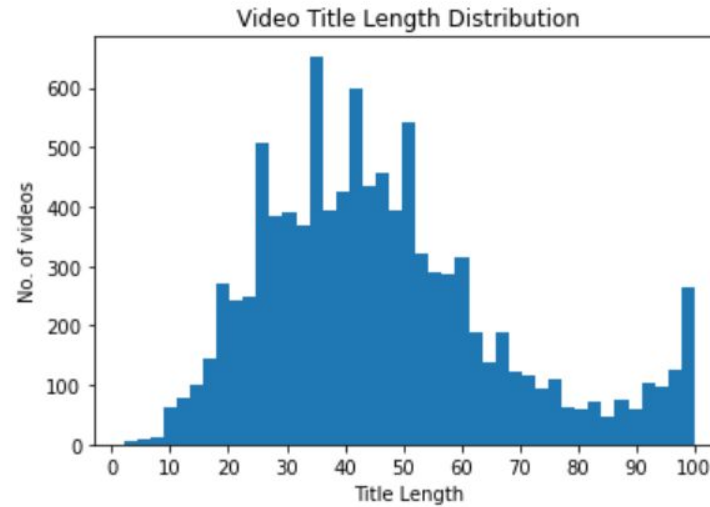
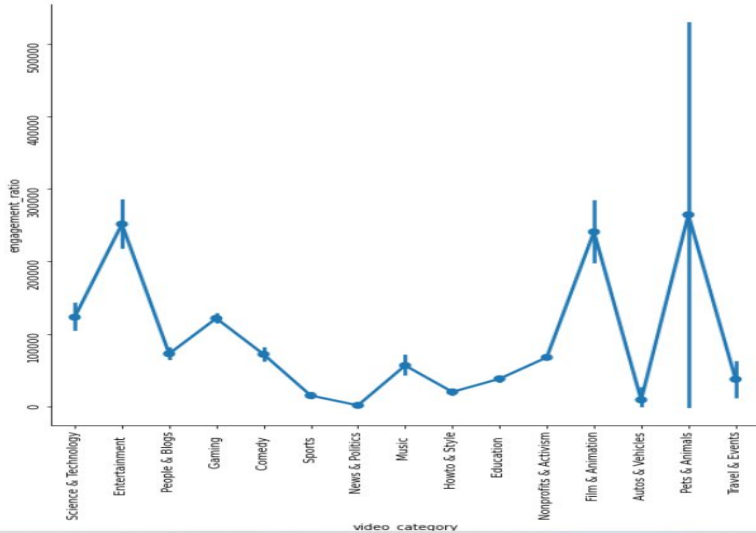


Data – Video Length



Feature Mining

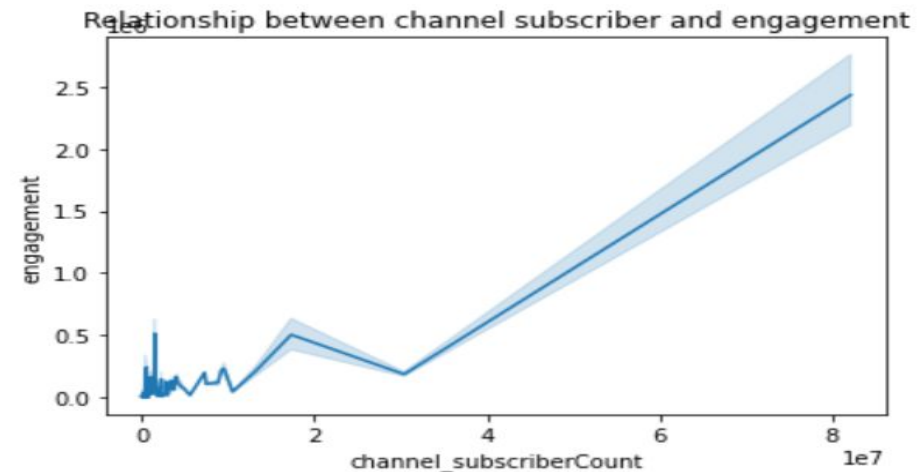
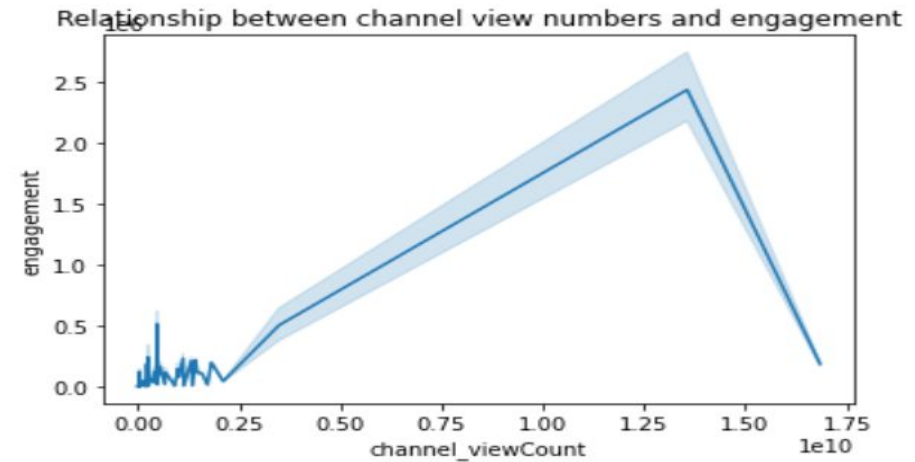
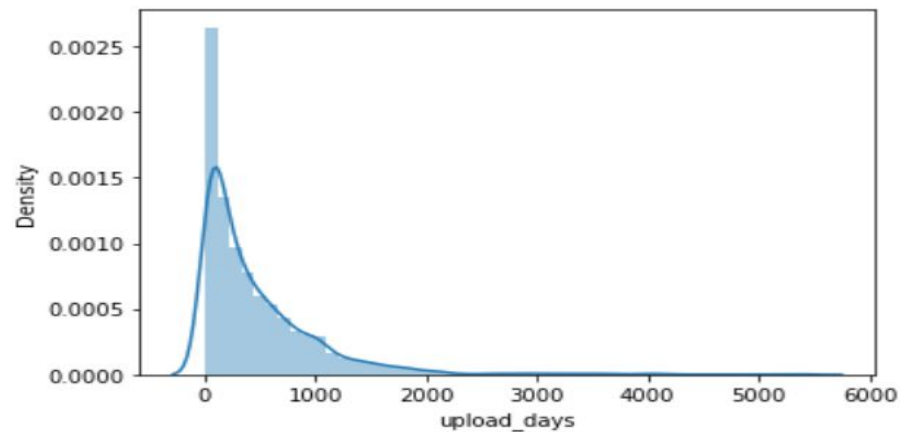
Categorical & Dummy Variables



Feature Mining

Numerical Variables

- **Views**
- **Subscribers**
- **Publish time:**
 - frequency of uploading videos every month
 - how many days after being uploaded



Feature Mining

Text Analysis

Text Pre-processing

- Remove non-words(characters, symbols, etc.)
- Extract all keywords from description using NLP

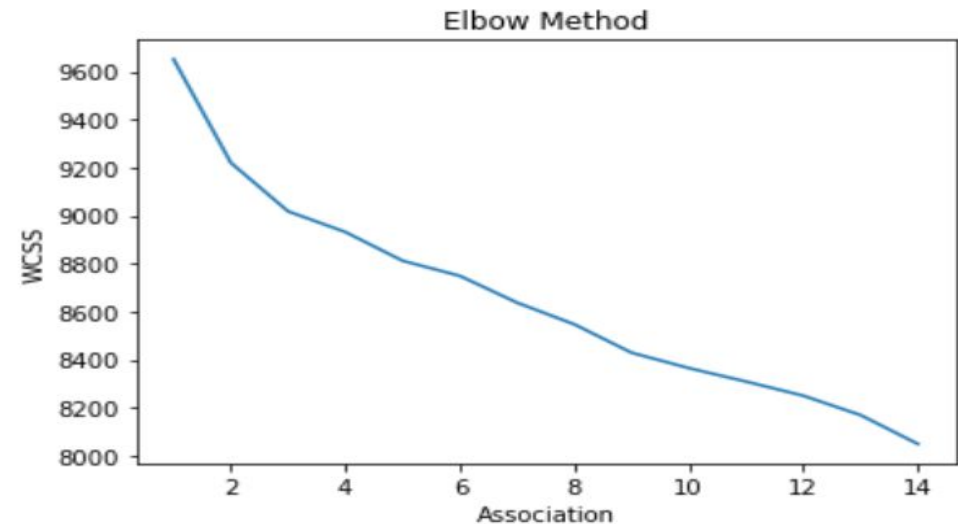
Text Encoding

- Transform words into vectorizers using **TF-IDF** Vectorizer, so similar words are mapped near each other

Clustering

- Elbow Method
- **Kmeans**: gave 9 clusters
- Get labels for every video based on the word similarity
- Tried **DBSCAN**, gave 2 labels for all data

```
dict_keys(['tether', 'printer', 'gone', 'overdrive', 'made', 'want', 'reflect',  
osing', 'video', 'printed', 'another', '2', 'billion', 'since', 'big', 'deal', '  
edits', '3d', 'artist', 'ed', 'leszczynski', 'editor', 'harry', 'bagg', 'opinion  
'business', 'opportunities', 'fake', 'gurus', 'subjective', 'terms', 'mean', 'di  
s', '100k', 'month', 'upfront', 'fee', '2k', 'scam', 'others', 'would', 'call',
```



Feature Mining

Text Analysis

Is TF-IDF a good encoding method for clustering?

After using TF-IDF and getting all the words, we found that those similar words cluster because they look similar alphabetically, not because they have the same meaning.

```
'admin', 'administer', 'administration', 'administrative', 'admins', 'admire', 'admirin  
mitting', 'ado', 'adobe', 'adobo', 'adoboli', 'adolescence', 'adolf', 'adolfo', 'adolph  
ing', 'adoption', 'adoptive', 'adopts', 'adorable', 'adorablebirbpon', 'adorably', 'ado  
riangranados', 'adrien', 'ads', 'adult', 'adulthood', 'adulthood', 'adulthood', 'adults', 'advance', 'a  
s', 'advancing', 'advantage', 'advantaged', 'advantages', 'advent', 'adventure', 'advent  
'adversaries', 'adversary', 'adverse', 'adversity', 'advertise', 'advertised', 'advertisi  
ng', 'advice', 'advised', 'adviser', 'advising', 'advisor', 'advisors', 'advocacy', 'adv  
e', 'aea', 'aeneid', 'aerial', 'aero', 'aerobic', 'aerobatics', 'aerodynamic', 'aerodyn  
spyke', 'aeryn', 'aeschliman', 'aesthetic', 'aesthetically', 'aesthetics', 'aesuno', 'a  
dark', 'aewdarkerelevation', 'aewdynamite', 'aewfullgear', 'aewrampage', 'af', 'afaik', 'a  
ct', 'affected', 'affectent', 'affecting', 'affection', 'affectionate', 'affects', 'aff  
ffinity', 'affion', 'affioncrocket', 'affirming', 'affleck', 'afflict', 'affluence', 'a
```



Text Analysis

We set key words as labels manually, like “minecraft”, “show”, “film”, “play”, “sports”, “movie”, “talk”, “music”. We let similar videos cluster around those keys.



Generate more values from text

-
- | tax_hypertonic | employment |
|----------------|------------|
| 0 | 100000 |
| 1 | 110000 |



Modeling



Model Performance

Model	Train RMSE	Test RMSE	Cross Validation	
Random Forest	84,103	104,971	236,226	.89
Linear Regression	225,549	151,184	226,017	.77
Decision Tree	189,455	112,099	216,817	.87

While Random Forest had the highest accuracy, the current apparent over fitting leads the team to select Linear Regression the purpose of this exercise

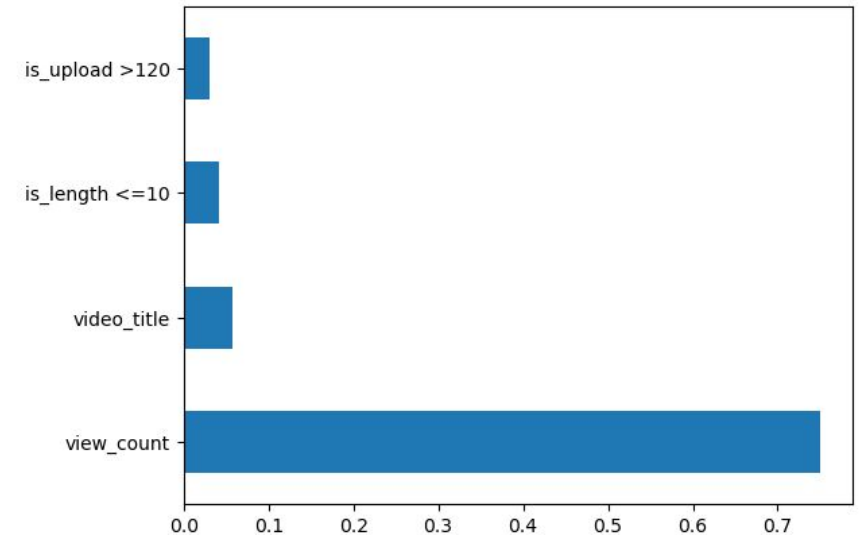
As additional channels/videos were added, RMSE, prediction, and over-fitting improved

Additional models that were tested were Gradient Boosting and Bagging Classifier

Models were ran on 3 different sets of features while optimizing

The team believes that the low Test RMSE associated with test data was related to the sample size and characteristics of this dataset

Feature Importance



Model Performance

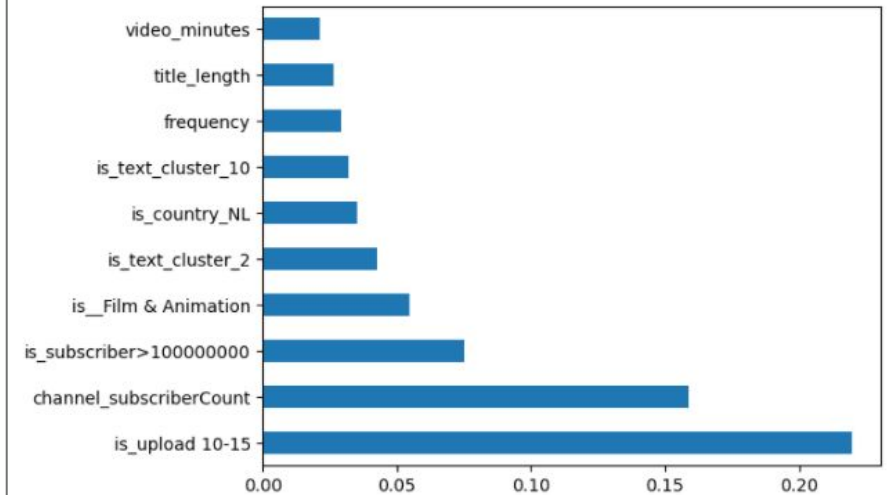
<u>Model</u>	<u>Train RMSE</u>	<u>Test RMSE</u>	<u>Cross Validation</u>	
Random Forest	97449.11	97449.11	244221	.67
Linear Regression	241899	152735	244441	.61
Decision Tree	229753	136638	232276	.69
XGradient Boost	69350	157254	295522	.67

Based on feedback received on Tuesday, following the presentation, the team attempted to re-run the model by using word2vec, and without using view_count as a predictor

With enhancements and modifications, the team was able to still achieve ~.7 accuracy

The updated model is optimized when using Decision Tree, but still suffers from some level of overfitting and would benefit from a larger sample of channels and videos

Feature Importance



Conclusion

- The higher the views, the higher the engagement
- Don't make your video too long! Or too short!

Feature Mining:

- KMeans: the number of clusters needs to be determined in advance. When the data set is relatively large, it is difficult to give an appropriate value in advance

Modeling:

- The current model is heavily reliant on the linear relationship with number of views to determine engagement

Next Steps

- Generate more data!
- Continued feature refinement
 - Text Clustering: Word2Vec + TSNE
- Experimenting with building views into the predicted variable
(likes + comments) / views