

Metodología Hefesto: Tendencias en Youtube

Integrantes:

CACERES, Martin

MENEL, Angelo

Índice

| | |
|----------------------------|----------|
| Introducción | 3 |
| Paso 1 | 3 |
| Preguntas de negocio | 3 |
| Indicadores y perspectivas | 3 |
| Modelo conceptual | 4 |
| Paso 2 | 4 |
| Hechos e indicadores | 4 |
| Mapeo | 5 |
| Granularidad | 5 |
| Modelo conceptual ampliado | 7 |
| Paso 3 | 7 |
| Tipo de esquema | 7 |
| Tablas de dimensiones | 7 |
| Tablas de hechos | 8 |
| Uniones | 8 |

Introducción

Aquellos usuarios comunes de Youtube conocemos la famosa pestaña de trending, eso no significa que entendemos como está funciona. En su página oficial Youtube nos dice que la sección de tendencias, tiene como objetivo mostrar a luz aquellos videos que podrían ser considerados interesantes o entretenidos para una gran cantidad de usuarios.

La premisa de esta pestaña tiene sentido, pero como muchos conocerán los videos que encontramos ahí normalmente son vídeos musicales, noticias, deportes, trailers de películas, pero no tantos creadores individuales.

Por este detalle decidimos investigar esta pestaña para averiguar si existe algún indicador que nos permita identificar cuáles vídeos van a ser tendencia y cuáles no.

Paso 1

Análisis de Requerimientos

Preguntas de negocio

Como primer paso, nos propusimos indagar sobre los videos en que Youtube los toma como tendencias, es decir, aquellos videos que Youtube los encuentra como interesantes para los usuarios.

El dataset incluye videos de Estados Unidos con su título, nombre del canal, tags, cantidad de likes, dislikes, vistas, comentarios, un link a la miniatura del video y una descripción. Con estos datos nos propusimos realizar un BI que nos pueda responder o acercar al como Youtube elige estos videos.

Propusimos las siguientes incógnitas que el datawarehouse nos brinde información:

- Cantidad de likes por canal y categoría en un tiempo determinado.
- Cantidad de dislikes por canal y categoría en un tiempo determinado.
- Cantidad de vistas por canal y categoría en un tiempo determinado.
- Saber si influye la diferencia de likes y dislikes por canal y categoría.

Indicadores y perspectivas

Una vez analizados los datos del dataset definimos los siguientes indicadores y perspectivas para analizar cada video. Estos son:

Indicadores:

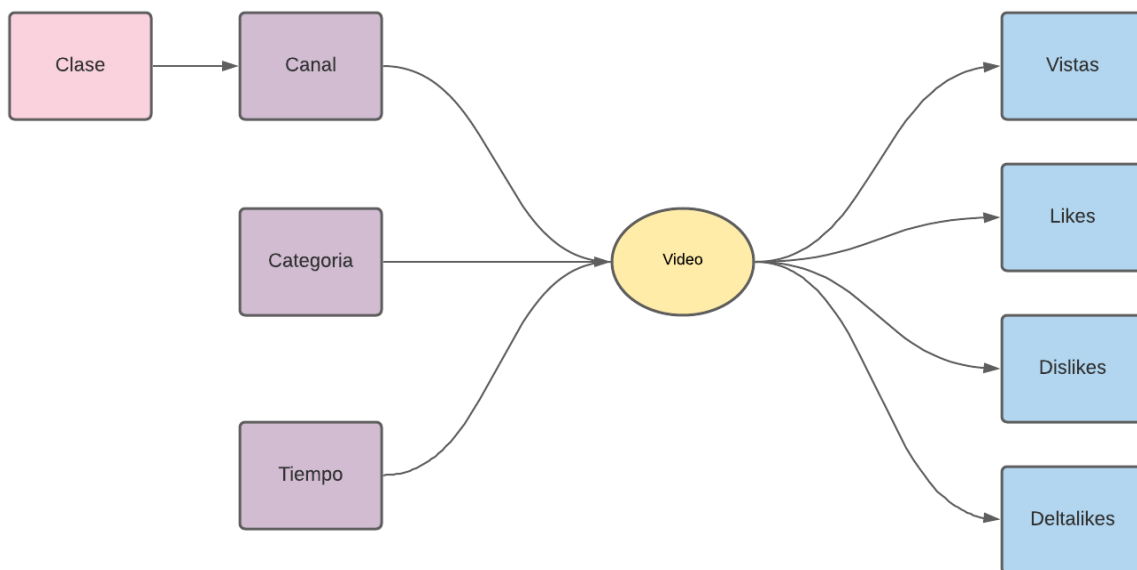
- Likes
- Dislikes

- Diferencia de likes y dislikes
- Vistas

Perspectivas:

- Canal
- Categoría
- Tiempo
- Clases (Esta tabla proviene de una normalización en la perspectiva Canal)

Modelo conceptual



Paso 2

Análisis de data sources

Hechos e indicadores

Likes:

- Hechos: Likes
- Función de agregación: SUM

Dislikes:

- Hechos: Dislikes.
- Función de agregación: SUM

Diferencia de Likes:

- Hechos: Likes - Dislikes
- Función de agregación: SUM

Views:

- Hechos: Vistas
- Función de agregación: SUM

Aclaración: Los indicadores Likes, Dislikes y Views representan la sumatoria de likes, dislikes y views respectivamente, de cada video por canal o categoría. El indicador Diferencia de Likes representa la sumatoria de Likes menos Dislikes de cada video por canal o categoría.

Mapeo

Luego de analizar el Dataset, elegimos los datos con los cual se relaciona cada tabla de dimensiones y la tabla de hechos. Estos son los siguientes:

- La Perspectiva Channel se relaciona con:
 - Channel Title
- La Perspectiva Category se relaciona con:
 - snippet.title
- La Perspectiva Date se relaciona con:
 - Trending_Date
- La Tabla Class se relaciona con:
 - class_name
- El Indicador Likes se relaciona con:
 - Likes -> SUM (likes)
- El Indicador Dif_Likes se relaciona con:
 - Likes and Dislikes -> SUM (likes-dislikes)
- El indicador Views se relaciona con:
 - Views -> SUM(views)

Granularidad

De acuerdo al mapeo realizado, se analizaron los diferentes campos, y si bien son pocos y explícitos se describirán a continuación:

- Video id: es la clave primaria del video, representa unívocamente un video en particular.
- Trending date: fecha en la cual el video ingresó a la sección Trending de YouTube.
- Title: nombre del video.
- Channel title: nombre del canal del video.

- Category id: clave foránea de la categoría del video.
- Publish time: fecha en la cual se publicó el video en YouTube.
- Tags: indica diferentes tags que se le agregan al video para luego buscarlo por diferentes nombres.
- Views: indica la cantidad de vistas que tuvo el video.
- Likes: indica la cantidad de likes que tuvo el video.
- Dislikes: indica la cantidad de dislikes que tuvo el video.
- Comment Count: indica la cantidad de comentarios que tuvo el video.
- Thumbnail Link: link a la imagen miniatura del video.
- Comments disabled: indica si los comentarios están activados o no.
- Ratings disabled: indica si el video está calificado o no.
- Description: descripción del video.
- Category name: nombre de la categoría, se refiere al tipo de video que sube el canal.
- Class name: nombre de la clase del canal, se refiere al tipo de canal.

Los resultados obtenidos fueron los siguientes:

Perspectiva Channel:

- Channel title, ya que hace referencia al nombre del canal.

Perspectiva Category:

- Category name, ya que hace referencia al nombre de la categoría. Este dato se encuentra en el JSON con todas las categorías que existen.

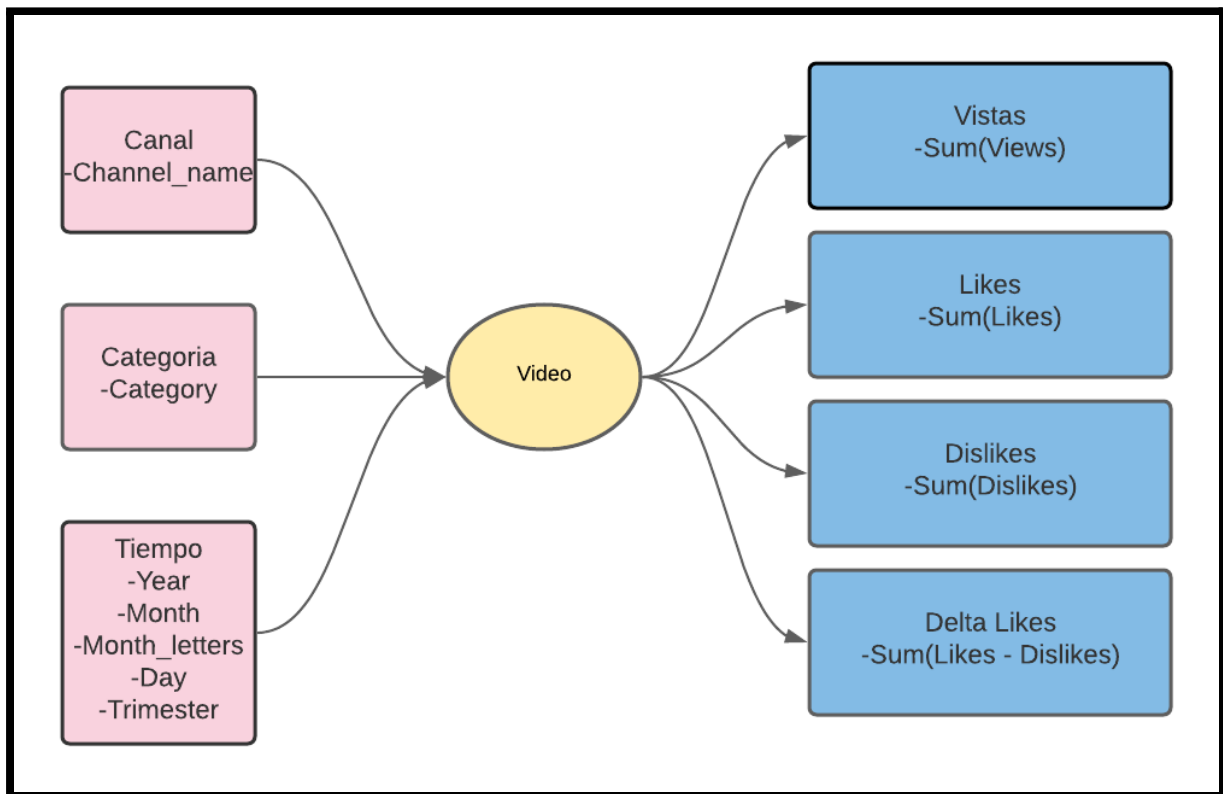
Perspectiva Date:

- Año.
- Trimestre.
- Mes, hace referencia al número del mes.
- Mes en letras, hace referencia al nombre del mes.
- Día.

Tabla Class:

- Class_name, ya que hace referencia al tipo de canal.

Modelo conceptual ampliado



Paso 3

Modelo Lógico del DW

Tipo de esquema

Viendo el Dataset y las perspectivas seleccionadas llegamos a la conclusión de utilizar el Esquema Copo de Nieve, esto es debido a que teníamos que normalizar la perspectiva Channel, creando una nueva tabla con el atributo `class_name`.

Tablas de dimensiones

A continuación se diseñan las tablas dimensiones:

Perspectiva Channel:

- La tabla de Dimensión tendrá el nombre **Channel**.
- Se le agrega una clave principal con el nombre **idChannel**.
- Se modifica el nombre del campo **Channel_title** por **channel_name**.

Perspectiva Category:

- La tabla de Dimensión tendrá el nombre **Category**.

- Se le agrega una clave principal con el nombre **idCategory**.
- El nombre del campo es **category_name**.

Tabla Class:

- La tabla de Dimensión tendrá el nombre **Class**.
- Se le agrega una clave principal con el nombre **idCategory**.
- El nombre del campo es **class_name**.

Perspectiva Date:

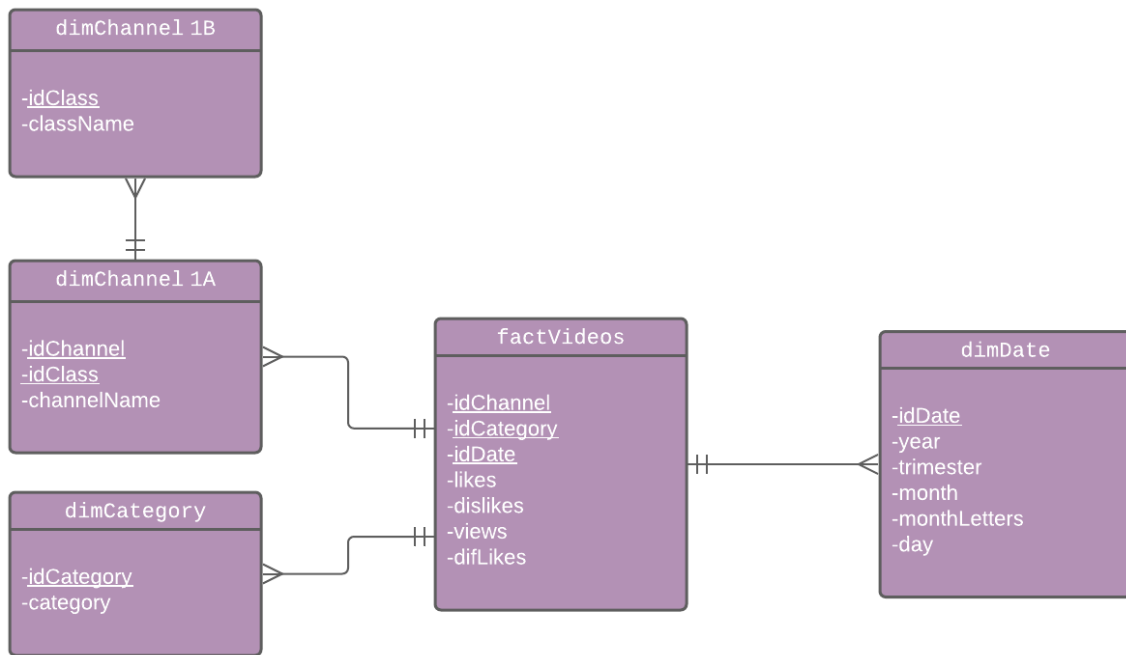
- La tabla de Dimensión tendrá el nombre **Date**.
- Se le agrega una clave principal con el nombre **idDate**.
- Se modifica el nombre del campo **Año** por **year**.
- Se modifica el nombre del campo **Trimestre** por **trimester**.
- Se modifica el nombre del campo **Mes** por **month**.
- Se modifica el nombre del campo **Mes en letras** por **month_letters**.
- Se modifica el nombre del campo **Día** por **day**.

Tablas de hechos

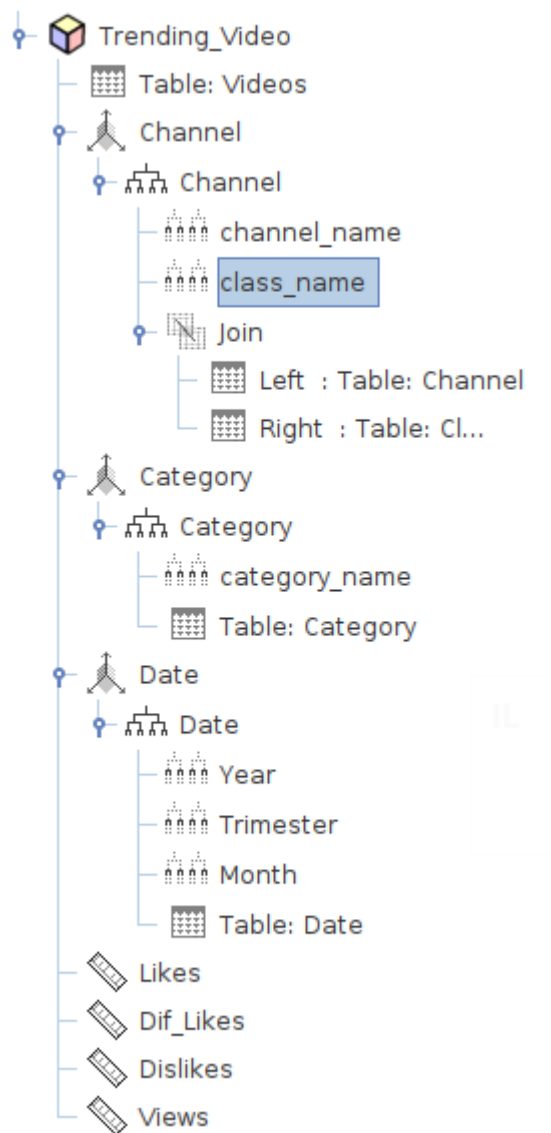
A continuación se realiza la tabla de Hechos:

- La tabla de Hechos tendrá el nombre **Videos**.
- Su clave principal será la combinación de las claves principales de las tablas de Dimensiones antes definidas.
- Se crea 4 Hechos, que corresponden con los indicadores antes definidos:
 - Diferencia de Likes sera renombrado como **dif_likes**.

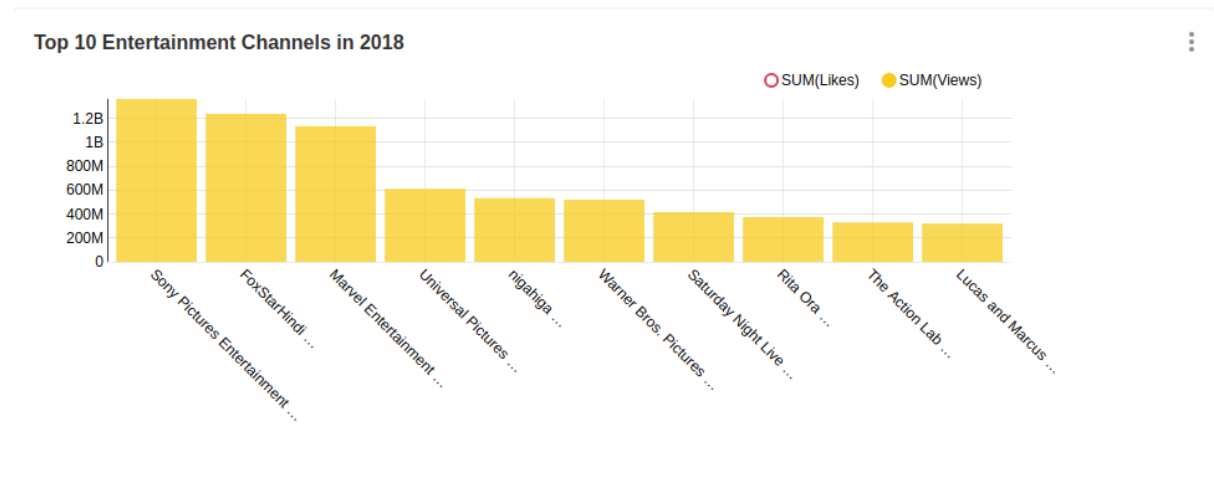
Uniones



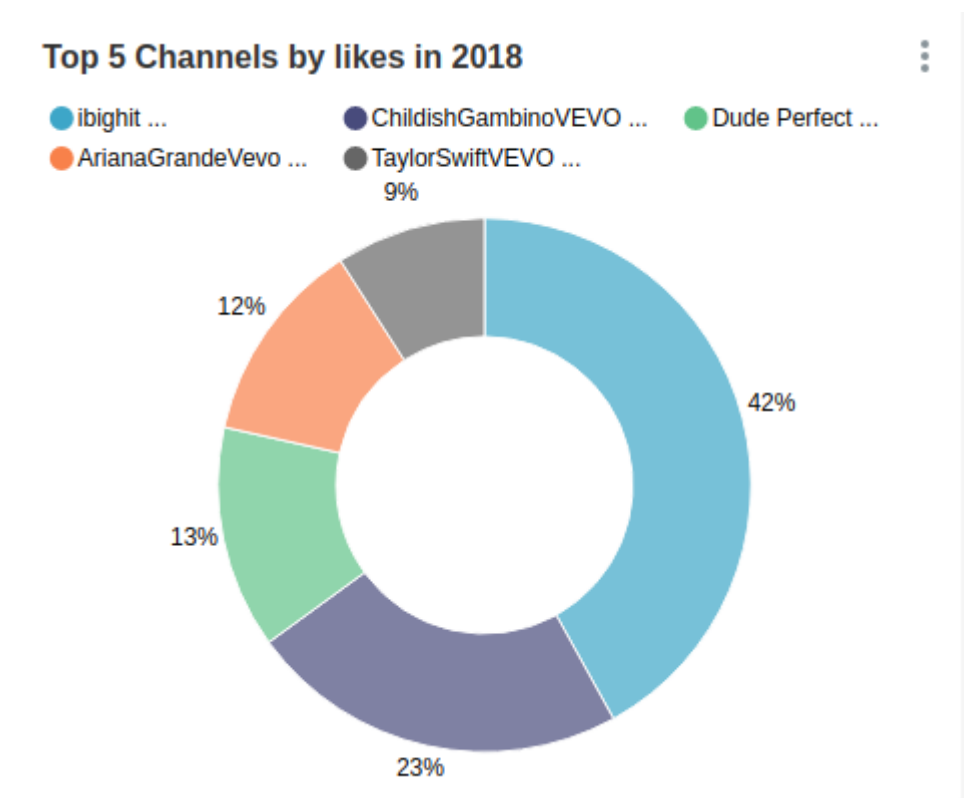
Modelo Multidimensional



Bocetos de Tableros



- Este boceto nos mostrará los 10 canales con mas vistas en el 2018 por categoría, el cual tendrá un parámetro con todas las categorías de YouTube, así por comparar las vistas por categoría.



- Este boceto nos muestra los 5 canales con más likes en el 2018. Un cambio que se le va a hacer es hacerlo que la cantidad de likes sea por mes, por lo tanto tendrá un parámetro que con los meses del 2018.

Reportes Bocetos











Channels per classes




Class: YouTuber

Month: January

| Channel | Amount | Avg. Likes |
|------------------|-----------|--------------------|
| Life Noggin | 29 | 19.840,931 |
| First We Feast | 26 | 24.851,1154 |
| Bon Appétit | 26 | 13.535,7692 |
| TED-Ed | 25 | 11.920,88 |
| Tom Scott | 25 | 15.632,96 |
| CollegeHumor | 25 | 26.060,2 |
| Brave Wilderness | 23 | 43.004,0435 |
| Smosh | 21 | 54.612,3333 |
| SciShow | 21 | 9.207,8571 |
| Screen Junkies | 20 | 46.635,6 |

Channels Data per YouTuber Class

| Channel | Amount | Avg. Likes | Avg. Views |
|--|--------|-------------|----------------|
|  Life Noggin | 29 | 19.840,931 | 617.044,2414 |
|  First We Feast | 26 | 24.851,1154 | 663.414,6538 |
|  Bon Appétit | 26 | 13.535,7692 | 800.346,8462 |
|  CollegeHumor | 25 | 26.060,2 | 823.453,72 |
|  TED-Ed | 25 | 11.920,88 | 299.385,2 |
|  Tom Scott | 25 | 15.632,96 | 337.535,44 |
|  Brave Wilderness | 23 | 43.004,0435 | 1.429.429,2174 |
|  Smosh | 21 | 54.612,3333 | 1.647.684,1429 |
|  SciShow | 21 | 9.207,8571 | 273.337,8571 |
|  The Film Theorists | 20 | 131.171,7 | 3.621.025,35 |

 Avg. Views < 500.000
 Avg. Views < 750.000
 Avg. Views > 750.000