
ICC-321-T Inteligencia de Negocios

Lisibonny Beato

1-2024-2025

Proyecto Final














(Dashboard Interactivo y Modelo de Minería de Datos Descriptivo)

El proyecto final de la asignatura se corresponde con el desarrollo de un **dashboard interactivo** y la construcción de un **modelo de minería de datos descriptivo**, para ayudar al canal de Youtube [The Quiz Challenge](#).

En los siete meses que tiene el canal creado, los dueños del canal tienen una gran dificultad para saber qué pueden hacer para mejorar el *engaging* con la audiencia y finalmente poder monetizarlo.

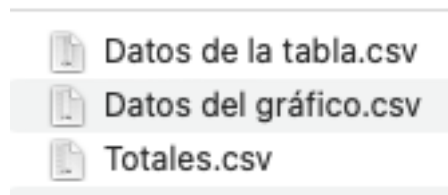
Entre las cosas que pueden notar es que algunos videos son muy populares a nivel de vistas, mientras que otros no lo son, pero no encuentran explicación al respecto. También se dan cuenta de que, a pesar de la popularidad de algunos videos, esto no se traduce en un número significativo de suscriptores para el canal. Otra preocupación que tiene es la de saber qué temáticas, tópicos o tipos de quiz debe seguir desarrollando, ya que en la actualidad toma esta decisión por corazonadas o eventos que ocurren en el momento.

Se les pide a ustedes que ayuden a *The Quiz Challenge* a obtener algunas respuestas, basado en las estadísticas que provee Youtube de distintos parámetros que se miden. El conjunto de datos lo encontrarán en un archivo comprimido en la carpeta de *Datasets* de la PVA. Al descomprimirlo verán la siguiente estructura de carpetas:

- >  Área geográfica
- >  Contenido
- >  Edad
- >  Estado de suscripción
- >  Fecha
- >  Fuente de tráfico
- >  Origen de las suscripciones
- >  Sexo
- >  Tipo de contenido
- >  Tipo de dispositivo
- >  Tipo de elemento de pantalla final
- >  Tipo de reproductor
- >  Usuarios nuevos y recurrentes

Pontificia Universidad Católica Madre y Maestra
Escuela de Ingeniería en Computación y Telecomunicaciones

La cantidad de archivos dentro de cada carpeta puede variar, pero la mayoría presentan la siguiente estructura:



El archivo *Datos de la tabla.csv* muestra información resumida del parámetro:

Datos de la tabla

Ubicación geográfica	Suscriptores	Vistas	Tiempo de reproducción (horas)	Duración promedio de vistas
Total	125	22476	410.8421	0:01:05
US	0	3870	90.7612	0:01:24
GB	0	2251	41.7824	0:01:06
IN	0	980	20.3077	0:01:14
AU	0	419	9.3463	0:01:20
ID	0	381	2.5692	0:00:24
PH	0	335	3.1569	0:00:33
ES	0	328	2.2172	0:00:24
BR	0	274	2.8471	0:00:37
VN	0	225	1.1223	0:00:17
CA	0	209	1.9202	0:00:33
RU	0	140	0.5436	0:00:13
DO	0	132	1.2271	0:00:33
ZA	0	122	1.0469	0:00:30
NZ	0	117	2.562	0:01:18
MY	0	100	0.7274	0:00:26

El archivo *Datos del gráfico.csv* muestra información del parámetro día por día (En la mayoría de los casos las Vistas y Suscriptores, pero hay casos en donde solo se presentan las vistas):

Datos del gráfico

Fecha	Ubicación geográfica	Vistas	Suscriptores
2024-11-04	AE	0	0
2024-11-04	AR	0	0
2024-11-04	AT	0	0
2024-11-04	AU	0	0
2024-11-04	BD	0	0
2024-11-04	BG	0	0
2024-11-04	BR	16	0
2024-11-04	CA	0	0
2024-11-04	CL	0	0
2024-11-04	CO	0	0

Pontificia Universidad Católica Madre y Maestra
Escuela de Ingeniería en Computación y Telecomunicaciones

El archivo *Totales.csv* tiene los datos totales por día de las vistas y las suscripciones al canal. Es el mismo en todas las carpetas en donde aparece. Se le sugiere que lo ignore.

No obstante a lo ya descrito, se le pide que investigue a fondo sobre las estadísticas de los canales de Youtube utilizando distintas fuentes confiables, para ampliar su conocimiento de las mismas.

Tome en cuenta que el principal problema es saber los factores que tienen efecto en la cantidad de vistas de los videos y los suscriptores del canal, sin embargo en casi ninguno de los archivos de estas carpetas los datos están directamente relacionados al contenido. Por ejemplo, con este dataset no es posible saber de manera explícita cuánta gente de Estados Unidos se suscribió al canal porque vio un video específico, por lo que la labor de usted es ver si es posible buscar una forma alternativa de saberlo.

Para esto tiene usted que realizar las siguientes tareas:

1. **Diseñar e Implementar un Dashboard Interactivo:** este debe permitir a *The Quiz Challenge* observar las métricas Vistas y Suscriptores, explorando los elementos de los videos que tienen efecto sobre estas. Se espera que usted estudie en detalle los datos, indique cuales son de interés, desarrolle un flujo para la preparación y transformación de los mismos en Tableau, según sea necesario, y finalmente implemente un dashboard interactivo en la herramienta que desee. Debe documentar todo el proceso de manera apropiada y proveer un enlace público para poder visualizar el dashboard **(10 puntos)**.
2. **Construcción de Modelo de Minería de Datos Descriptivo:** este debe permitir extraer información no explícita desde el conjunto de datos que ayude a *The Quiz Challenge* a saber qué tipo, tópico y/o temática de video debe hacer para sus quices, de forma tal que sea exitoso a nivel de vistas y suscriptores que ayuda a conseguir al canal. Para esto usted debe elegir una tarea (Asociación o Clustering) y debe indicar el objetivo que busca conseguir esta tarea y las métricas que usará para evaluar la calidad de dicho modelo. Tome en cuenta que es probable que deba integrar otras fuentes de datos para enriquecer lo que tiene. Debe documentar paso a paso todo lo que realice e interpretar los hallazgos que va obteniendo. Debe proveer el notebook documentado, en donde se recopila el trabajo realizado en la construcción del modelo. **(30 puntos)**

Tome en cuenta que, adicionalmente, debe realizar una presentación oral de ambos trabajos el último día de clases, para la cual contará con 15 min.