

PROYECTO SPRINT 12

Descripción del proyecto

Trabajas como analista de vídeos publicitarios en la agencia de publicidad Sterling & Draper. Dedicas mucho tiempo a analizar tendencias de vídeos en YouTube para determinar qué contenido merece atención para la mercadotecnia.

Cada video tiene una categoría específica (entretenimiento, música, noticias y política, etc.), una región y una fecha en que se hace tendencia.

Un video puede estar en la sección de tendencias durante varios días seguidos.

Cada semana, las nuevas empleadas Melanie y Ashok te preguntan esto:

¿Qué categorías estaban en las tendencias de la semana pasada? ¿Cómo se distribuyeron en diversas regiones? ¿Qué categorías fueron particularmente populares en los Estados Unidos? En tu sexta semana en el trabajo, decides que ya es momento de que el proceso sea automático. Vas a crear un dashboard para Melanie y Ashok.

Después de hablar con los administradores de la base de datos, reuniste unos requisitos técnicos:

- Objetivo de negocios: analizar el historial de tendencias de videos en YouTube
- Con qué frecuencia se usará el dashboard: al menos una vez al día
- Usuario objetivo del dashboard: gerentes de planificación de videos publicitarios
- Contenido de los datos del dashboard:
 - Tendencias pasadas de videos, ordenadas por día y categoría
 - Tendencias de videos, ordenadas por país
 - Una tabla de correspondencia entre categorías y países
- Parámetros para agrupar los datos:
 - Fecha y hora de tendencia
 - Categoría de video
 - País
- Los datos:
 - Historial de tendencias — valores absolutos ordenados por día (dos gráficos: números absolutos y proporción de porcentaje)
 - Eventos, ordenados por país — valores relativos (% de eventos)
 - La correspondencia entre las categorías y los países — valores absolutos (una tabla)
- Importancia: todos los gráficos son igualmente importantes
- Fuentes de datos para el dashboard: los ingenieros prometieron crear una tabla de agregación llamada `trending_by_time`. Esta es la estructura:

- record_id: la clave primaria
- region: país/región geográfica
- trending_date: fecha y hora
- category_title: categoría del video
- videos_count: número de videos en la sección de tendencias
- La tabla está en la base de datos youtube, que se creó especialmente para tus necesidades.
- Intervalo de actualización de los datos: una vez cada 24 horas, a la media noche UTC
- Gráficos, controles de dashboard y su orden:

Título y descripción del dashboard	Filtro de fecha y hora	Filtro de país
Gráfico "Historial de tendencias": tendencias de videos divididas por tiempo de tendencia y categoría (valores absolutos, gráfico de área)	Gráfico "Tendencias de videos por país": tendencias de videos dividas por país (valores relativos, gráfico de pastel)	
Gráfico "Historial de tendencias": tendencias de videos divididas por tiempo de tendencia y categoría (% del total, gráfico de área)	Tabla "Tendencias de videos por país y categoría". Las columnas solo deben ser para países, las filas deben indicar categorías de tendencias. Las celdas deben contener números absolutos de videos en tendencia. Las celdas deben estar resaltadas según sus valores (tabla de resaltado).	

El filtro de fecha y hora y el de país debería modificar todos los gráficos del dashboard. Fíjate que los gráficos de historial de interacción "Historial de tendencias" e "Historial de tendencias, %" deben tener fecha y hora en el eje X. "Historial de tendencias" debería tener el número de videos en la sección de tendencias (el campo videos_count) en el eje Y y el otro gráfico debería tener el porcentaje ahí.

Para crear el dashboard, sigue los siguientes pasos:

1. En Tableau Public, usa trending_by_time.csv (ve al final de esta página para descargarlo) para crear un dashboard modelado en el borrador.
2. Publica el dashboard en el servidor de Tableau Public. Asegúrate de que todos puedan acceder a él; intenta abrirlo en varios navegadores. Si no se puede, el revisor no podrá hacer la revisión.

3. Usa tu dashboard para responder a las preguntas que te hicieron:
 - ¿Qué categorías de videos estuvieron en tendencia más frecuentemente?
 - ¿Cómo se distribuyeron en las regiones?
 - ¿Qué categorías fueron particularmente populares en los Estados Unidos?
¿Hubo diferencias entre las categorías populares en Estados Unidos y en otros lugares?

Prepara una presentación breve con un informe (las respuestas a estas preguntas y gráficos).

Paso 1. Abrir el archivo `trending_by_time` y leer la información general

Importar librerías

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import datetime as dt
```

Leer los datos

```
videos = pd.read_csv('../notebooks/trending_by_time.csv')
videos.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12343 entries, 0 to 12342
Data columns (total 5 columns):
#   Column          Non-Null Count  Dtype
---  -
0   record_id       12343 non-null  int64
1   region          12343 non-null  object
2   trending_date   12343 non-null  object
3   category_title  12343 non-null  object
4   videos_count    12343 non-null  int64
dtypes: int64(2), object(3)
memory usage: 337.6+ KB
```

```
videos.sample(10)
```

	record_id	region	trending_date	category_title	videos_count
6282	6283	Japan	2018-06-07 00:00:00.000000	Pets & Animals	32
5514	5515	Japan	2018-04-14 00:00:00.000000	Film & Animation	24
9768	9769	United States	2018-04-21 00:00:00.000000	Comedy	38
4661	4662	India	2018-04-06 00:00:00.000000	Science & Technology	2
1446	1447	France	2017-12-02 00:00:00.000000	Music	50
4656	4657	India	2018-03-29 00:00:00.000000	Science & Technology	4
898	899	France	2018-02-20 00:00:00.000000	Film & Animation	14
5796	5797	Japan	2018-05-29 00:00:00.000000	Howto & Style	20
10365	10366	United States	2018-03-28 00:00:00.000000	Film & Animation	14
9000	9001	Russia	2017-11-17 00:00:00.000000	Shows	8

Para conocer un poco más sobre los datos y tipo de datos de los datos utilizados en la creación del Dashboard "Tendencias de videos en Youtube", realicé una lectura de los datos de `treding_by_time` en Jupyter Notebook.

De la información obtenida del método `info()` se puede observar que los datos cuentan con 12343 filas y que todas sus columnas tienen esta misma cantidad de valores no nulos, por lo tanto se puede asumir que no hay valores ausente.

Las columnas `'record_id'` y `'videos_count'` se presentan con tipos de datos como números enteros, mientras que las columnas `'region'`, `'trending_date'` y `'category_title'` se presentan con

tipos de datos object. La columna 'trending_date' al referirse a una fecha y hora en que se realizó el registro de los videos entendencia, lo correcto es cambiar el tipo de datos de esta columna a tipo de datos datetime. Aunque el Dashboard ya fue creado en la plataforma de Tableau Public, como una buena práctica presento a continuación el cambio.

Cambio de tipo de datos de la columna 'trending_date' de object a datetime

```
videos['trending_date'] = pd.to_datetime(videos['trending_date'])
```

```
videos.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12343 entries, 0 to 12342
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype
---  -
0   record_id             12343 non-null  int64
1   region                12343 non-null  object
2   trending_date         12343 non-null  datetime64[ns]
3   category_title       12343 non-null  object
4   videos_count          12343 non-null  int64
dtypes: datetime64[ns](1), int64(2), object(2)
memory usage: 385.8+ KB
```

Analisis de duplicados

```
videos.duplicated()
```

```
0      False
1      False
2      False
3      False
4      False
...
12338  False
12339  False
12340  False
12341  False
12342  False
Length: 12343, dtype: bool
```

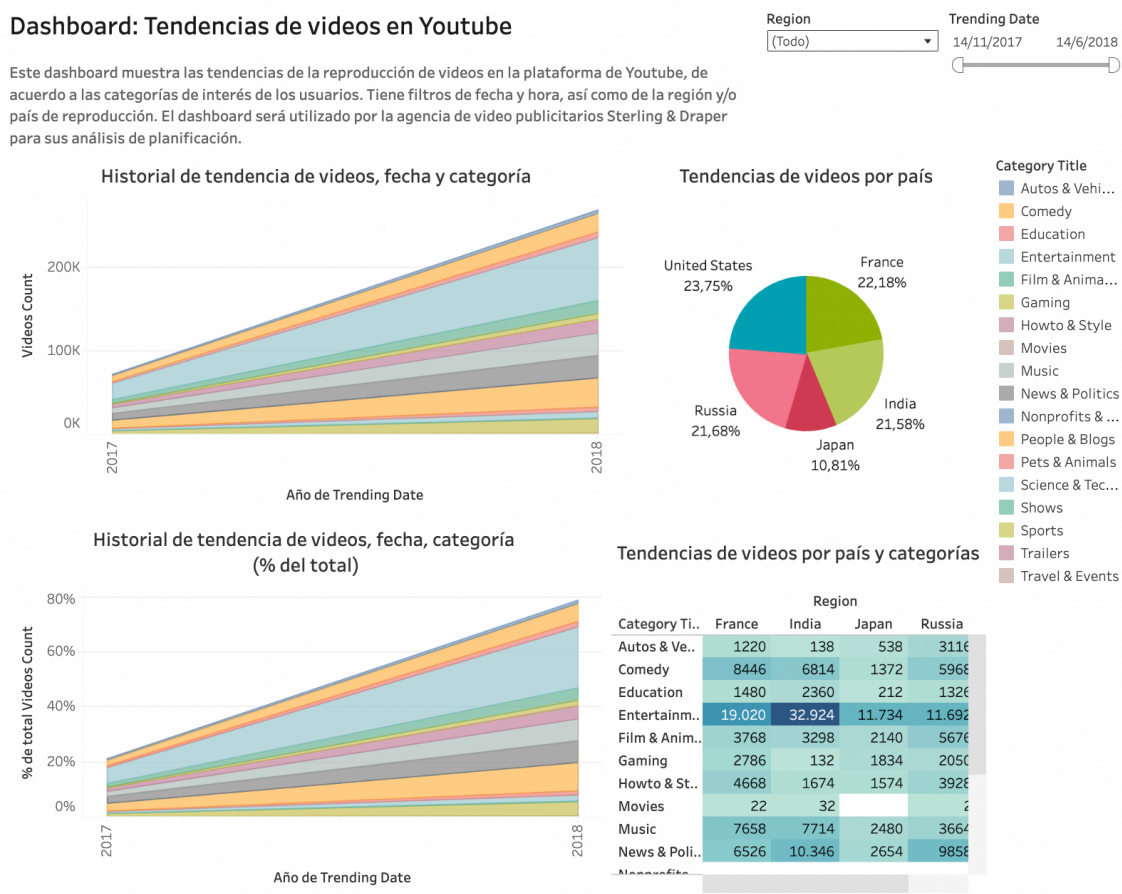
También realicé una búsqueda de filas duplicadas, como se puede observar no se encontraron valores duplicados y nuevamente se reportan 12343 filas de valores importantes para este análisis.

Paso 2. Link al Dashboard "Tendencia de videos en Youtube" en el servidor de Tableau Public

https://public.tableau.com/views/DashboardProyectoSprint12/Dashboard1?:language=es-ES&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

Dashboard: Tendencias de videos en Youtube

Este dashboard muestra las tendencias de la reproducción de videos en la plataforma de Youtube, de acuerdo a las categorías de interés de los usuarios. Tiene filtros de fecha y hora, así como de la región y/o país de reproducción. El dashboard será utilizado por la agencia de video publicitarios Sterling & Draper para sus análisis de planificación.



Paso 3. Respuesta a las preguntas del proyecto en relación al Dashboard

3.1. ¿Qué categorías de videos estuvieron en tendencia más frecuentemente?

Gráfico "Historial de tendencias": tendencias de videos divididas por tiempo de tendencia y categoría (valores absolutos)

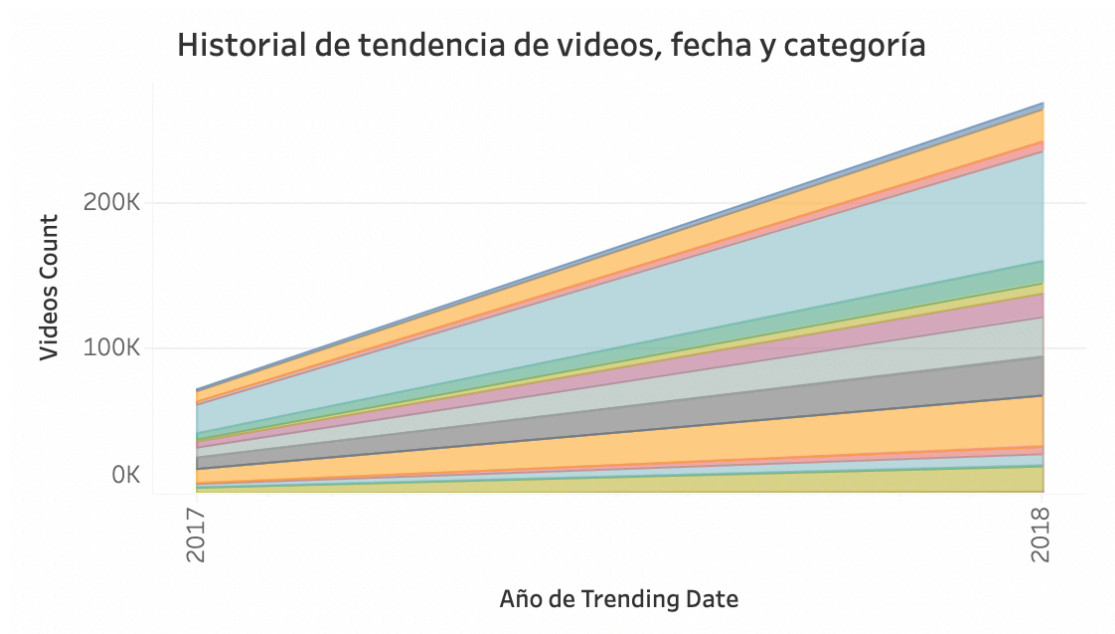
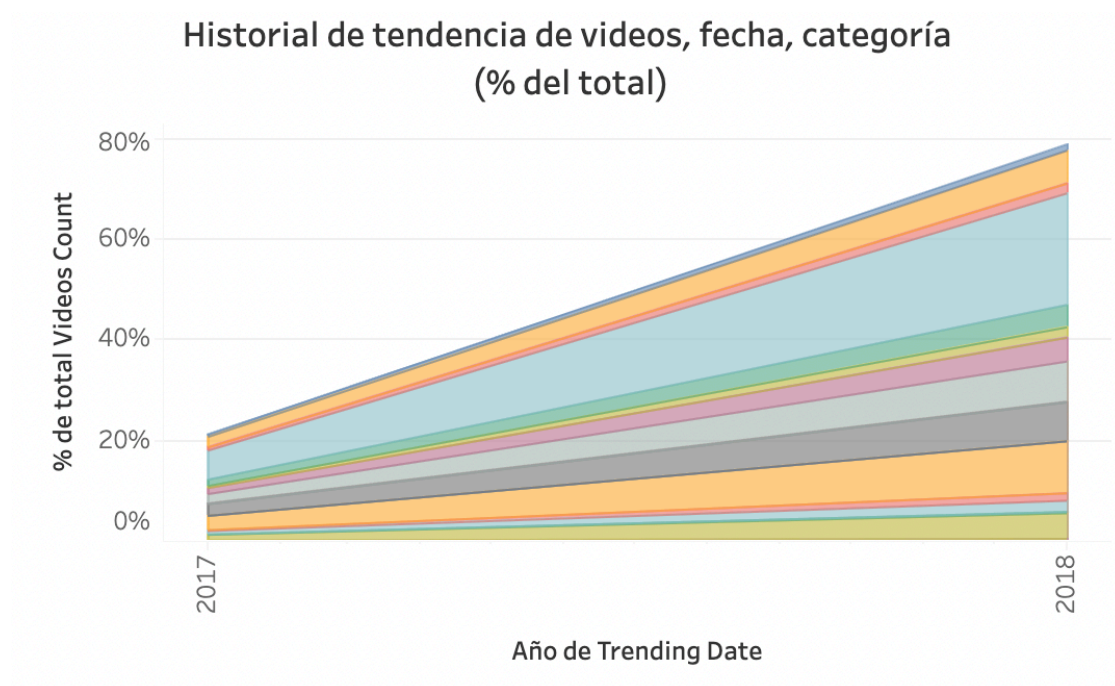
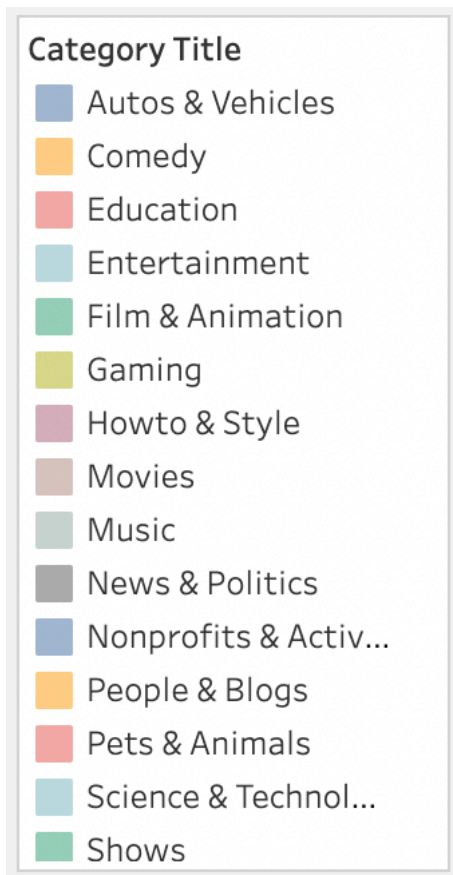


Gráfico "Historial de tendencias": tendencias de videos divididas por tiempo de tendencia y categoría (% del total)





Ambas gráficas son gráficos de área acumulada y son importantes para analizar los videos que estuvieron en tendencia y la proporción de cada categoría de video con respecto al total de los videos a los largo del tiempo de análisis.

Inicialmente se observa que durante el tiempo de análisis de noviembre - 2017 a junio - 2018, la altura total del área aumenta consistentemente indicando que el número total de videos que se vuelven tendencia en las reproducciones de Youtube se ha ido incrementado con el tiempo.

En el Dashboard se puede ver claramente que las 5 categorías con mayor reproducción de videos en Youtube son:

1. Entretenimiento (Entertainment – azul celeste)
2. Personas y Blogs (Peple & Blos – amarillo claro)
3. Música (Music – gris claro)
4. Noticias y Política (News & Politic –gris oscuro)
5. Comedia (Comedy – naranja claro)

Otras categorías de interés para los que reproducen videos son:

1. Deportes (Sport – verde claro)
2. Tutoriales (Howto & Style – rosado claro)
3. Películas y animación (Film & animation – verde turquesa)

La gráfica de "Historial de tendencias" de valores absolutos, muestra una clara tendencia al alza en el número de videos que se encuentran en tendencia, con las categorías de Entretenimiento, Personas & Blogs y Música en los primeros lugares.

En la gráfica de "Historial de tendencias" % del total, se puede confirmar lo observado en la gráfica de valores absolutos, es un complemento entre ambas gráficas porque así se puede analizar la proporción de cada categoría con respecto al total de videos en tendencia.

Al igual que en el gráfico de valores absolutos, las categorías Entretenimiento y Música siguen teniendo la mayor proporción del total. Sin embargo, al observar en el dashboard la gráfica de porcentaje, queda mucho más claro que la categoría Entretenimiento tiene una dominancia muy significativa.

1. Entretenimiento (Entertainment – azul celeste) 22%
2. Personas y Blogs (People & Blogs – amarillo claro) 10.33%
3. Música (Music – gris claro) 8.12%
4. Noticias y Política (News & Politics –gris oscuro) 7.74%
5. Comedia (Comedy – naranja claro) 6.56%

La gráfica muestra que la proporción de las categorías principales se mantiene relativamente constante a lo largo del tiempo.

Las otras categorías que también parece mantener una proporción estable, sugiere que su popularidad crece en proporción con el número total de videos en tendencia.

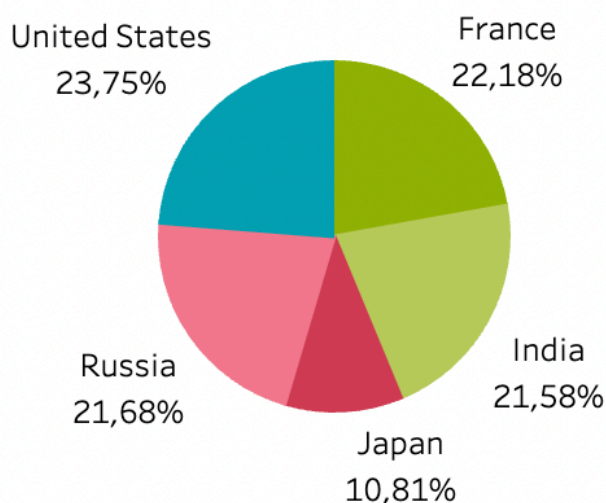
1. Deportes (Sport – verde claro) 5.08%
2. Tutoriales (Howto & Style – rosado claro) 4.72%
3. Películas y animación (Film & animation – verde turquesa) 4.47%

Al análisis de estos gráficos puede ofrecer a las gerentes de planificación de videos de la agencia Sterling & Draper una visión rápida de los tipos de contenido que más éxito se esta teniendo en la plataforma de YouTube. Además de ver cuales categorías están en tendencia y en aumento de reproducción, también permite visualizar las categorías que van perdiendo interés y relevancia en comparación las demás.

3.2. ¿Cómo se distribuyeron en las regiones?

Gráfico "Tendencias de videos por país": tendencias de videos divididas por país (valores relativos)

Tendencias de videos por país



En este gráfico de pastel se muestra la distribución de las tendencias de videos por país presenta la distribución de los videos en tendencia en cinco regiones, usando valores relativos es decir, la proporción de los videos reproducidos en cada región con respecto al total de reproducciones.

En Estados Unidos lidera la cantidad de videos en tendencia con porcentaje de 23.75%, seguido muy cerca por Francia con el 22.18% del total de reproducciones.

En otros países como Rusia con un 21.68% e India con un 21.58%, se muestran un número casi idéntico de videos en tendencia, lo que indica una popularidad similar.

En último lugar se encuentra Japón con un 10.81% del total de reproducciones de videos en tendencia en la plataforma de Youtube, esto indica una menor actividad con respecto a los otros países.

Se puede concluir que las tendencias de videos se distribuyen de manera relativamente equitativa entre Estados Unidos, Francia, Rusia e India, pues tienen un número similar de proporción de videos en tendencia. Por otro por otra parte, Japón presenta una distribución significativamente menor, prácticamente la mitad de la proporción para Rusia.

3.3. ¿Qué categorías fueron particularmente populares en los Estados Unidos?
¿Hubo diferencia entre las categorías populares en Estados Unidos y en otros lugares?

Tabla "Tendencias de videos por país y categoría" (valores absolutos)

Category Title	Region				
	France	India	Japan	Russia	United States
Autos & Vehicles	1220	138	538	3116	758
Comedy	8446	6814	1372	5968	6870
Education	1480	2360	212	1326	3284
Entertainment	19.020	32.924	11.734	11.692	19.638
Film & Animation	3768	3298	2140	5676	4680
Gaming	2786	132	1834	2050	1606
Howto & Style	4668	1674	1574	3928	8280
Movies	22	32		2	
Music	7658	7714	2480	3664	12.874
News & Politics	6526	10.346	2654	9858	4818
Nonprofits & Activism					106
People & Blogs	9346	4988	5792	18.452	6122
Pets & Animals	468	6	2250	1154	1832
Science & Technology	1588	1096	300	2226	4722
Shows	198	410		388	114
Sports	8002	1424	3606	3684	4250
Trailers	4				
Travel & Events	204	16	276	510	804

Esta tabla con los valores resaltados permite identificar rápidamente los valores más altos en cada columna (país) por los tonos de los colores más oscuros, para los valores más bajos los colores claros.

Ahora, para responder a la pregunta: ¿Qué categorías fueron particularmente populares en los Estados Unidos? Se observa en esta columna de Estados Unidos, los valores más altos, que están resaltados con una tonalidad color más oscura, las categorías más populares:

1. Entretenimiento (Entertainment – 19,638 videos)
2. Música (Music – 12,874 videos)
3. Tutoriales (Howto & Style – 8,280 videos)
4. Comedia (Comedy – 6,870 videos)
5. Personas y Blogs (People & Blogs – 6,122 videos)

Estas cinco categorías generaron el mayor número de videos en tendencia en los Estados Unidos.

Respondiendo a lo otra pregunta: ¿Hubo diferencias entre las categorías populares en Estados Unidos y en otros lugares?

Al comparar los resultados de Estados Unidos con los de otros países, se observan lo siguiente:

1. La categoría de "Entretenimiento" practicamente es la catageroría lider en todos los países Estados Unidos (19,638), Francia (19,020), Japón (11,734). En la India esta categoría es tan popular, que incluso supera en videos de reproducción todos los demás países (32,924).
2. En Rusia este comportamiento es diferente, debido a que la categoría más popular en ese país es "Personas y Blogs" (18,452) mientras que para Estados Unidos esta misma categoría ocupa el quito lugar como se muestra en la respuesta de la pregunta anterior (6,122).
3. Con respecto a Francia, la segunda y tercera categoría más popular en reproducción de videos son "Personas y Blogs" (9,346) y Comedia (8,446); mientras que para Estados Unidos fueron la quinta y cuarta categoría respectivamente.
4. En India, después de la categroría "Entretenimiento", la segunda categoría con más reproducciones es "Noticias y Política" (10,346), con un número de videos significativamente mayor que en otros países, incluido Estados Unidos (4,818) que ni si quiera figura entra las 5 categorías más populares en este país.

5. Para Japón la segunda categoría en reproducción de videos más popular es "Personas y Blogs" (5,792), para Estados Unidos esta categoría ocupó el quinto lugar.

En conclusión, se puede decir que la categoría de "Entretenimiento" es un éxito que domina las tendencias en casi todas las regiones, este comportamiento se puede decir que es similar al de Estados Unidos. Sin embargo, con respecto a las otras categorías si hay diferencia en la preferencia de reproducción de las personas en Estados Unidos con respecto a las personas de los otros países.

Este análisis muestra que, aunque hay un gusto general por el "Entretenimiento", las estrategias de marketing deben adaptarse a las preferencias únicas de cada región.