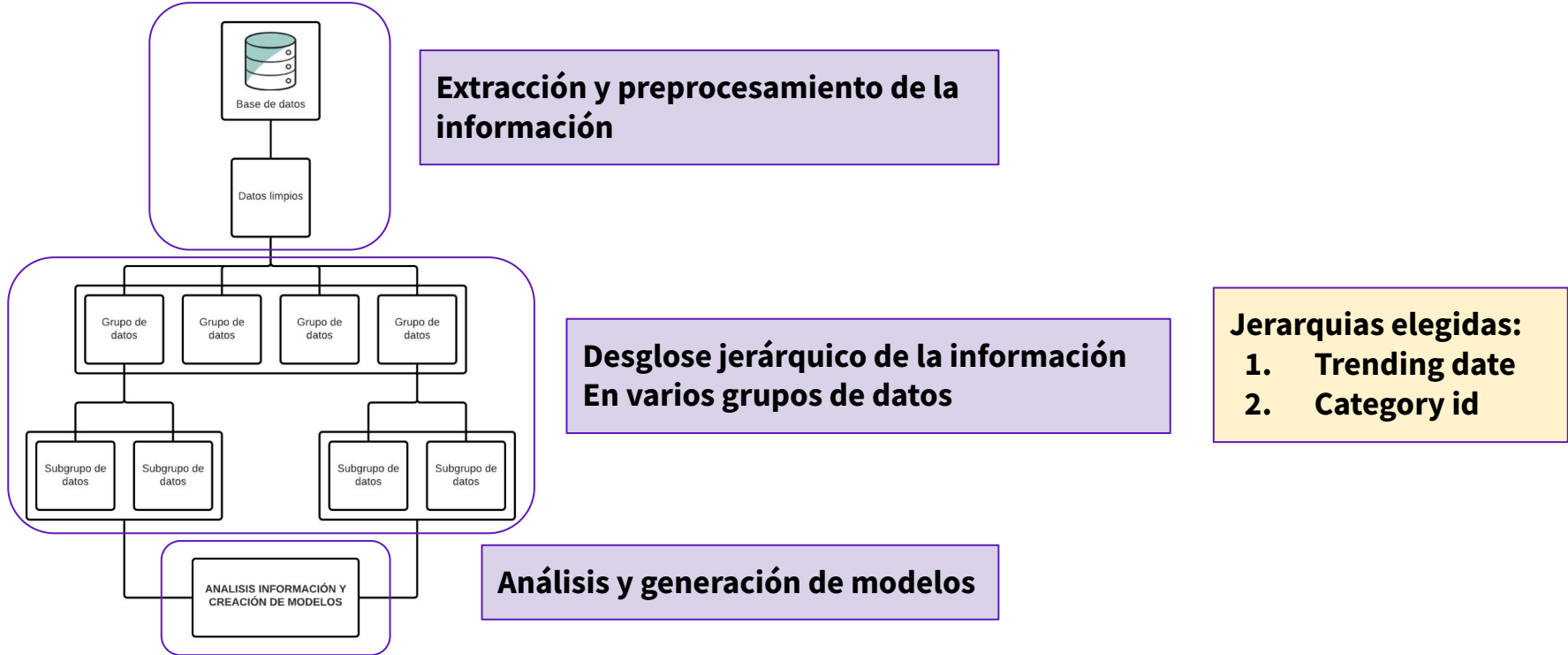


Prueba práctica Científico de Datos **pragma**

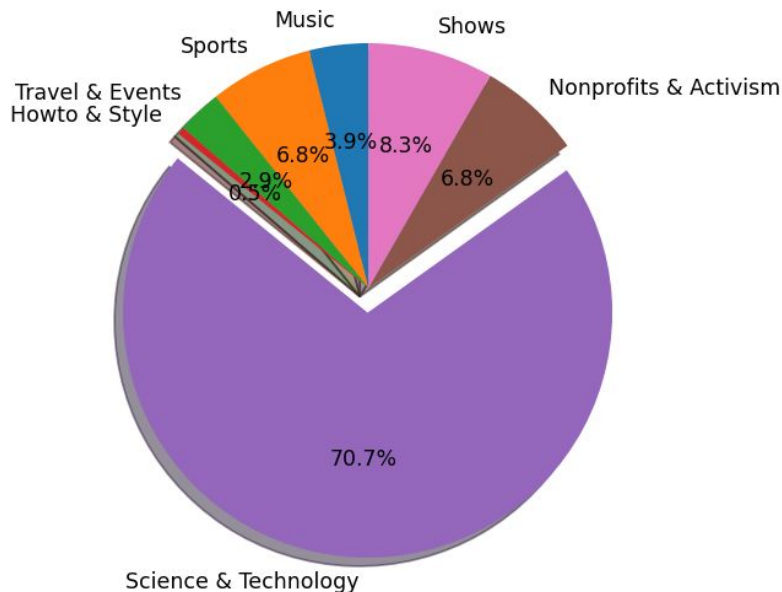
Elkin Javier Guerra Galeano



Metodología usada para el descubrimiento de información



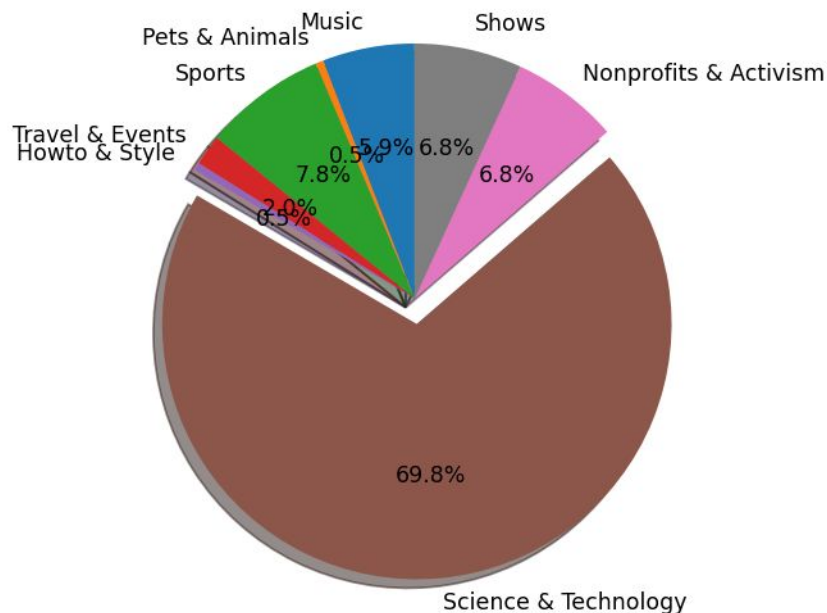
¿Cuáles son las categorías de vídeos que reciben mayores vistas?



- En el dataset se tiene un total de 205 trending dates diferentes.
- En cada una de estas fechas se establece una jerarquía de las categorías según el número de vistas y se escoge la que tiene la mayor cantidad.
- Se establece la frecuencia de las categorías que alcanzan el mayor número de vistas por fecha.
- Se analizan y grafican los resultados.

Resultados del análisis: Los videos de ciencia y tecnología son los más vistos

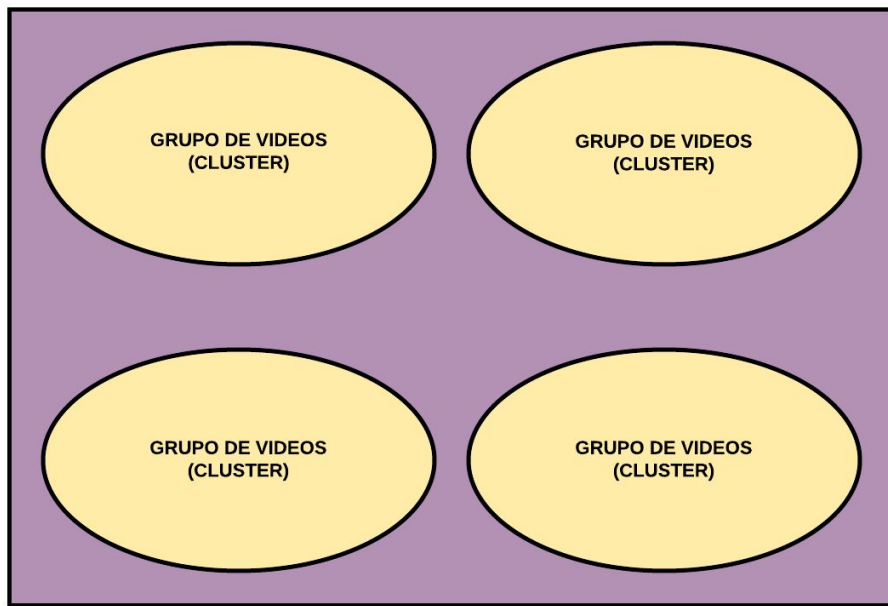
¿Cuáles son las categorías de vídeos que reciben mayores likes?



- En el dataset se tiene un total de 205 trending dates diferentes.
- En cada una de estas fechas se establece una jerarquía de las categorías según el número de likes y se escoge la que tiene la mayor cantidad.
- Se establece la frecuencia de las categorías que alcanzan el mayor número de likes por fecha.
- Se analizan y grafican los resultados.

Resultados del análisis: Los videos de ciencia y tecnología son los que tienen más likes

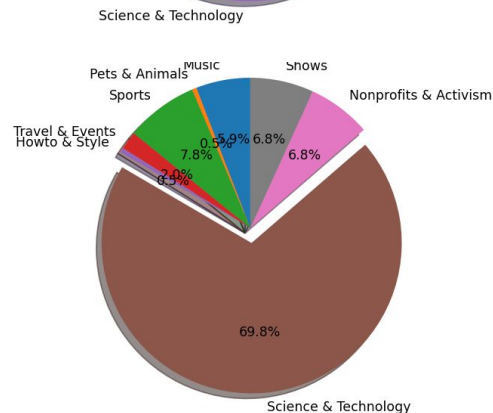
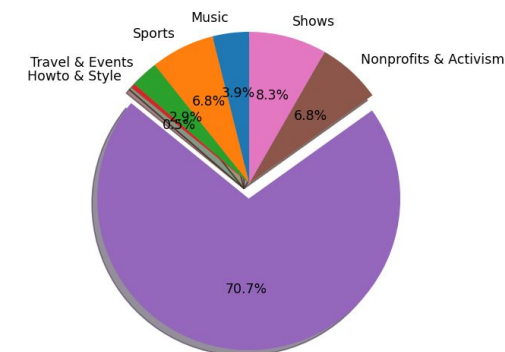
¿Es posible encontrar agrupaciones o clasificaciones de videos?



Dataset videos

- Tomando como base los análisis de vistas y likes, se evidencia que los videos se pueden clasificar utilizando variables relevantes.
- Se pueden clasificar por la categoría a la que pertenecen, y a su vez se pueden establecer más grupos dependiendo de la cantidad de likes y vistas que tenga cada video.
- Es posible la creación de un modelo de clustering para hacer esta tarea.

¿Cuál es la combinación de características o atributos más importantes que hacen de un video tendencia?

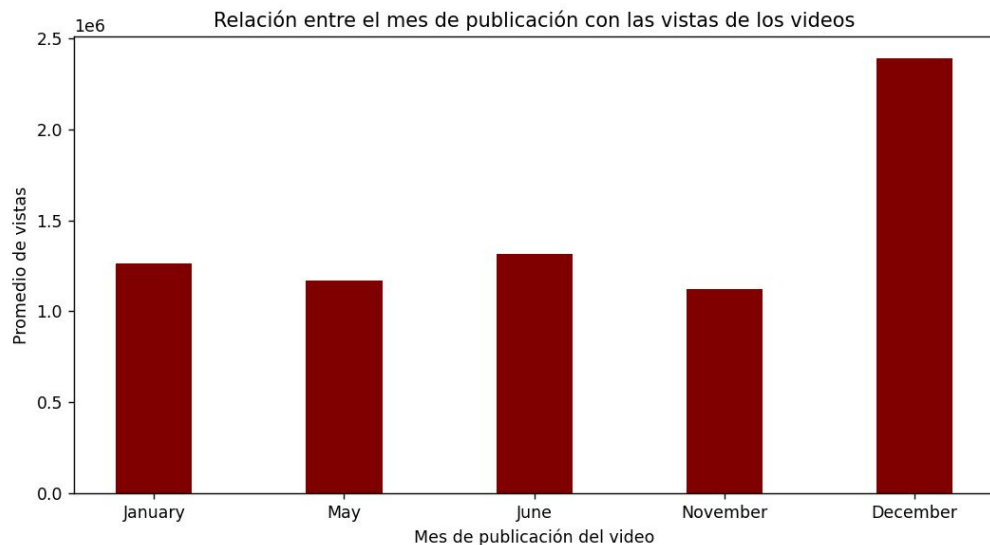


Tomando como base el análisis sobre la cantidad de vistas y likes se pueden deducir las siguientes premisas:

- La categoría del video repercute directamente en el número de vistas del mismo y por lo tanto en su capacidad de volverse tendencia.
- El número de likes esta fuertemente relacionado con la categoría de los videos, y debido a lo mencionado anteriormente también se relacionan con las vistas.

Resultados del análisis: Estas características son la categoría del video y el número de likes

¿La temporada o fecha en el que el video es publicado tiene alguna influencia?



Analizando el dataset, dividiéndolo en grupos según el mes de publicación del video se tiene:

- Se puede apreciar claramente como en el mes de diciembre el promedio de vistas es mucho más elevado de lo normal.
- Se puede deducir que al publicar videos en el mes de diciembre se podrían alcanzar mayores vistas.

Resultados del análisis: Al publicar videos en el mes de diciembre es posible alcanzar más vistas y por lo tanto hacer que los videos sean tendencia

¿Es posible predecir cuantos likes o visitas tendrá un video?

Variables de entrada

category_id	X1
trending day	X2
trending month	X3
trending year	X4
dislikes	X5

Variables de salida

views	Y1
likes	Y2

Utilizando regresiones lineales se pueden deducir modelos matemáticos para predecir el número de vistas y likes de un video.

Se utiliza una regresión lineal debido a que es un modelo muy fácil de implementar y construir.

Para la construcción de los modelos se utilizan variables importantes como la categoría del video y el trending date.

$$Y1 = -1.41158689e+05X1 + 7.53451531e+05X2 + 2.29480497e+04X3 + 7.63962619e+06X4 + 1.05606267e+02X5$$

$$Y2 = -4.64831229e+03X1 + 2.05721215e+04X2 + 6.72515425e+02X3 + 2.05843912e+05X4 + 3.17566759X5$$

Insights y conclusiones finales

- De acuerdo al análisis realizado, tenemos certeza de que en Estados Unidos, de manera general los videos de ciencia y tecnología son los más vistos y los que más les gustan a las personas ver.
- Es posible clasificar los videos utilizando diversas variables importantes como lo son la categoría, el número de vistas, el número de likes y la fecha de tendencia.
- Es posible desarrollar diversos modelos para tratar de predecir el número de vistas y de likes de un video, utilizando herramientas como la regresión lineal, sin embargo es posible llegar a modelos mucho más acertados utilizando herramientas más complejas como las redes neuronales.