

Profiling

1. EPISODES:

Variable show_id:

show_id es una foreign key referenciada a la tabla de shows, este permite conocer cuál es el show que más emisiones de episodios genera, siendo este The Young Turk, con 38 episodios, lo que representa el 1.2% del total de episodios. También se puede evidenciar que se cuenta con episodios de 625 shows diferentes, de los cuales, 124 generan el 50% de los episodios transmitidos.

Variable episode_id:

Es la primary key de la tabla de episodios, por esta razón se espera que no tenga valores nulos y que a su vez, cada uno de ellos sea único. Esto se puede evidenciar al tener 3044 registros, sin valores nulos y 3044 valores diferentes.

name:

La variable name corresponde al nombre del episodio, no se encuentran valores nulos. Se pueden observar valores repetidos para diferentes episodios, sin embargo, en un mayor porcentaje tienen nombres únicos.

url:

La variable url es el enlace a cada uno de los episodios presentados; todos los url son únicos y no hay valores nulos.

season:

La variable season hace referencia a la temporada a la cuál pertenece cada episodio. Se evidencia que hay una gran predominancia por la primera temporada los shows, sin embargo, se encuentran números de temporadas elevados, superiores a las 20 temporadas. Existe un valor atípicamente elevado de temporada, 2020, este puede hacer referencia a la temporada del año 2020 de diferentes shows.

number:

La columna number hace referencia al número de cada episodio; predominan los números de episodios no muy altos (Entre 1 y 15), sin embargo, se encuentran episodios con números superiores a los 300, lo que da a entender que existen shows con más de 300 episodios en una temporada.

type:

Hace referencia al tipo de episodio emitido, este puede pertenecer a 3 categorías: regular, significant_special y insignificant_special. El 97% de los episodios presentes son de tipo regular, mientras que las otras categorías representan aproximadamente un 1% de los episodios cada uno.

airdate:

Representa la fecha en la cuál se emitió el episodio, de esta se puede evidenciar que todos los episodios fueron emitidos en el año 2020 y en el mes de diciembre.

airtime:

Representa la hora a la cuál fue transmitido un episodio. Se presenta un gran número de valores nulos para esta columna, lo que quiere decir que no se conoce la hora de transmisión de los episodios. Para los valores presentes, se tiene que la hora de transmisión más frecuente son las 20:00 horas, esto puede deberse al elevado rating presentado a esa hora.

runtime:

Hace referencia a la duración de cada episodio en minutos; se puede evidenciar que en su mayoría, estos tienen una duración inferior a una hora, sin embargo, se encuentran un número considerable de episodios de 120 minutos (81 episodios) y un número bajo de episodios con duración superior a los 130 minutos.

2. CHANNELS:

id:

Es la primary key de los canales existentes. Se evidencia que no hay valores nulos ni repetidos en esta columna.

name:

Hace referencia al nombre del canal, se tiene un 99% de nombres de canales diferentes, sin embargo, hay 2 repidos con identificadores diferentes, estos son Sky Go y discovery+.

country:

Hace referencia al país al cuál pertenece el canal. Se presentan 32 países diferentes, siendo United States (Estados Unidos) el más frecuente de todos, seguido por Norway (Noruega) y China. La columna no presenta valores nulos, por tanto se conoce el país de pertenencia de todos los canales.

code:

Hace referencia al código del país de pertenencia de cada canal, por esta razón se presentan las mismas proporciones que en la columna country. No se presentan valores nulos.

timezone:

Hace referencia a la zona horaria que utiliza cada uno de los canales, esta está completamente relacionada con el país al que pertenece cada uno de los canales. Se presentan 32 zonas diferentes y ningún valor nulo.

3. SHOWS:

how_id:

Es la primary key de la tabla de shows. Presenta 625 registros en total, ninguno es nulo y todos son únicos.

name:

Representa el nombre de cada uno de los shows presentes en la tabla. Se cuenta con 623 valores distintos y 2 registros repetidos 2 veces, estos son Mermaid Prince y Klassen.

url:

Representa el enlace a cada uno de los shows presentes en la tabla. Se tienen 625 registros distintos y ninguno de ellos es nulo.

type:

Hace referencia al tipo de show. Se cuentan con 11 tipos de shows diferentes, siendo Scripted, Documentary y Animation los predominantes.

Language:

Hace referencia al lenguaje de cada show, siendo English, Russian y Chinese los predominantes, a pesar de contar con 36 lenguajes diferentes en total y 8 registros nulos.

status:

Representa el estado actual del show, el cuál puede tomar 3 valores distintos (Running, Ended, To Be Determined). La mayoría de los registros se encuentran en estado Running (316 registros), seguido por Ended (220 registros) y finalmente 89 registros se encuentran en estado To Be Determined. No se presentan registros nulos.

averageRuntime:

Hace referencia al tiempo promedio de duración en minutos de cada episodio de un show un específico. Se tiene un promedio de 36.05 minutos de duración, con un mínimo de 1 minuto y un máximo de 300. Se presentan 32 valores nulos.

premiered:

Hace referencia a la fecha de estreno del show, siendo la más antigua el 10 de noviembre de 1969 y el más reciente el 31 de diciembre del 2020. Sin embargo, existe una notoria oblicuidad de la distribución hacia las fechas más recientes.