

Usando la librería de pandas profiling se realizó un análisis exploratorio en los 5 dataframes principales que componen en el modelo de datos. Además de esto se hará un resumen de los principales hallazgos en cada una de estas.

1. Información general de los shows : Contiene la información de las series.

- a. Se tienen los identificadores únicos de las tablas, los cuales son las identificaciones de las series y los urls de estas.
- b. Se puede apreciar que la mayoría de los shows son scripteados (43%), documentales(12%) o de animación(12%) de las series que están en la base de datos.
- c. Los idiomas predominantes en las series son : ingles con un 38%, y ruso y chino con un 10 % cada uno. Además, se tiene un total de 36 diferentes idiomas en la base de datos.
- d. El 51% de las series están activas, mientras que el resto ya terminaron (34%) o están por ser definidos el estado de esta.
- e. En promedio una serie tiene entre 3 y 71 minutos donde la mayoría de estas tienen un total de 30 minutos. El cual tiene una alta correlación con el tiempo promedio que están al aire, el cual se comporta muy similar al número de minutos que duran estas series.
- f. Las series no tienen un patrón definido el cual las series terminen o que tengan una fecha premier.
- g. El rating promedio que tienen las series es de 6-8 donde la mayoría de estas tiene 7.2, hay que resaltar que hay un total de 85% de datos faltantes.

2. Información general de los shows en emisión: Contiene la información de las series que fueron emitidas.

- a. Igual que la base de datos de los shows, esta base también tiene un id único el cual es uno por show en emisión y el link a este.
- b. La mayoría de los capítulos de las series que fueron emitidos en diciembre se usa la palabra "episodio" (10%) en vez de un nombre específico para el capítulo.
- c. La mayoría de las series están en sus dos primeras temporadas (72%) y 7% de las series tienen valores de los años que fueron emitidas, por esta razón se decidió eliminar estos valores.
- d. Hay una gran varianza en el número del capítulo de las series emitidas en diciembre donde en promedio están en el capítulo 23 pero por la gran

desviación el rango de estas va del primer capítulo de la serie hasta el capítulo 73.

e. Principalmente los capítulos emitidos son de tipo regular (97%) y los no hay casi especiales en el tipo de capítulos emitidos.

f. El tiempo cuando fueron emitidos los capítulos principalmente fue los viernes del mes de diciembre y en año nuevo.

g. La mayoría de las series de en la base de datos no tiene horario definido y en ellos casos que lo tienen, la mayoría están en horarios nocturnos como lo es 8 pm(10 %).

3. Horarios de los shows: Contiene la información de las horas

a. Los días que más se emiten las series son los jueves y los viernes los cuales en promedio emiten el 29% de las series al aire en diciembre. Mientras que los días que menos salen series son los fines de semanas donde se emiten 13% de las series en promedio.

4. Información de los canales de televisión: Se contiene la información principal de los canales de emisión de las series.

a. Los principales países que tienen canales en emisión son estados unidos (16%), Rusia (7%) y Corea del sur (5%), donde las series son producidas 32 países diferentes

b. Hay 27% de datos que no tienen país definido.

c. Las zonas horarias más comunes en los canales siguen el país el cual fue producida, en el caso de Estados unidos es Nueva York, en el caso de Rusia es Kamchatka y en el caso de Corea del sur es Seuol.

5. Géneros de las series:

a. Los géneros que más series comparten son comedia y drama los cuales en promedio el 28% series son de estos géneros.

b. La mayoría de las series no tienen un entre 0-3 géneros definidos, donde el 27% de estas no tienen un género definido.

c. La combinación de géneros más común es de drama y romance cuando son 2 géneros y cuando son drama y romance y la combinación de tres géneros es drama , comedia y romance.