

AUDIENCE FORECASTING SYSTEM



CONTROL CARGA
DE DATOS



VISUALIZACIÓN Y
ANÁLISIS



PREDICCIÓN

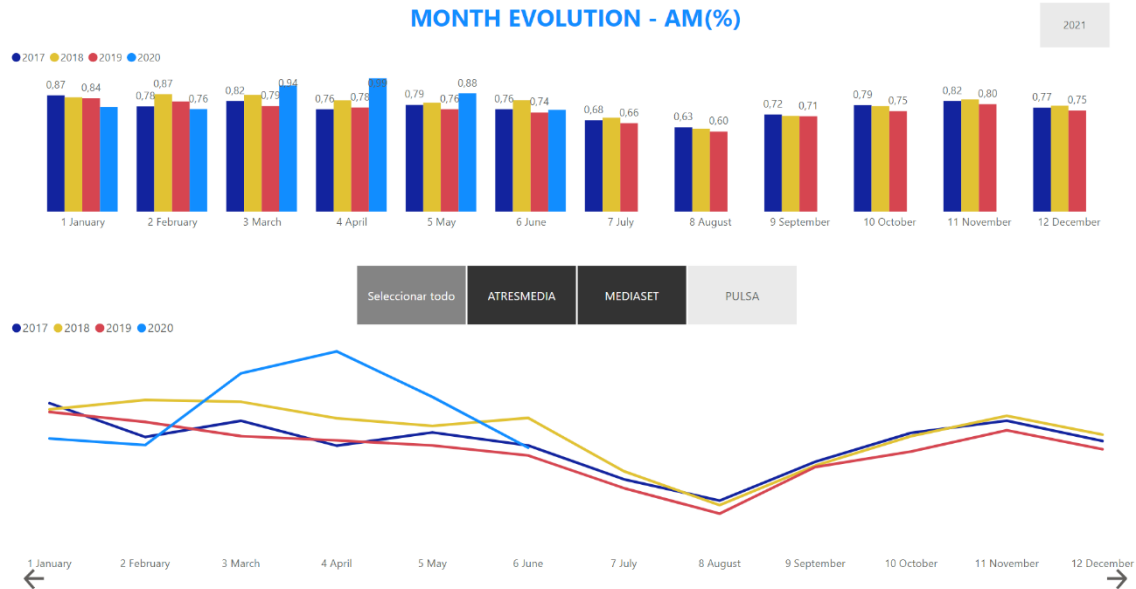
JAVIER CASADO DE AGUINAGA

La herramienta de visualización de los datos cargados se utilizará por parte de planificadores y directores de medios para hacer análisis en profundidad del mercado de la audiencia española, análisis como evolutivos del mercado, estacionalidades entre los meses, incrementos de audiencias por cadenas y grupo comercial, tendencias...

Toda esta información aporta un conocimiento y sensibilidad del mercado español al usuario para poder planificar de una manera más precisa. Con tan solo un click el planificador podrá ver qué días de la semana son los de mayor audiencia, como incrementa o desciende la misma en los meses, como varía por franjas horarias... actualmente su herramienta habitual es Instaranalytics, de Kantar media, la cual permite extracciones en ficheros de diversos formatos, pero no dispone de una herramienta visual. Y lo que es más importante, no tiene la información accesible a un click, ha de calcular cada petición que el usuario solicita por lo que con esta herramienta se evita el tiempo de espera y se accede a la información de forma más visual y más sencilla de entender.

Para la visualización de la herramienta se ha de disponer del dataframe final (df_ml_21.csv) guardado en el repositorio, así como el archivo de power BI llamado VISUALIZATION.PBIX

A continuación, se detallan los análisis más comunes realizados por las agencias de medios que son los que se han procedido a graficar:



Estacionalidad mensual: MONTH

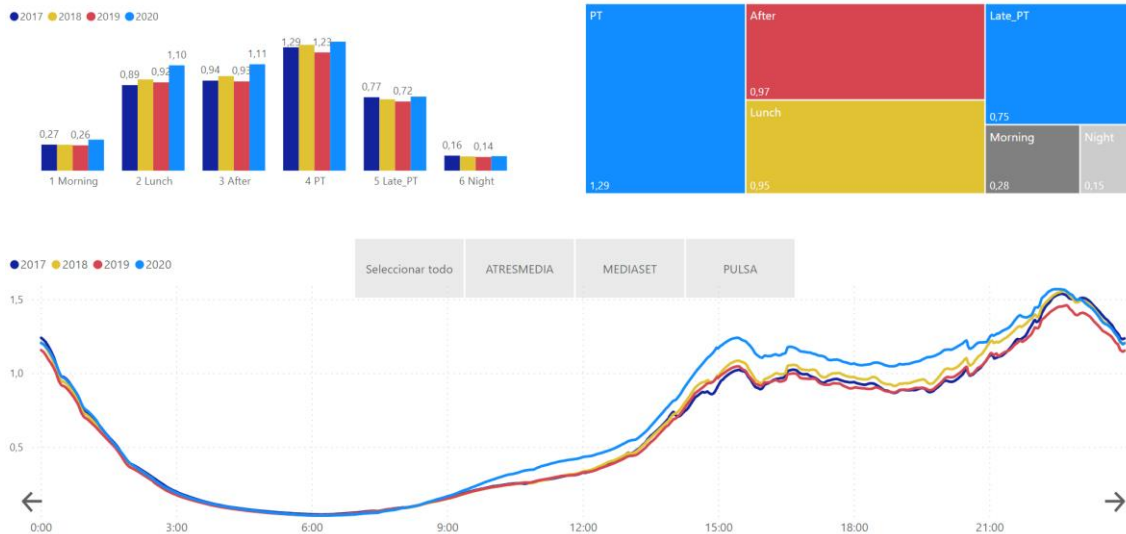
Con este análisis podemos ver como la audiencia de la televisión tiene una tendencia decreciente entre los años y dentro de los mismos, la estacionalidad entre los meses es la misma: los meses fríos tienen más audiencia y los meses más cálidos, durante el verano, esta audiencia media disminuye. Como se aprecia en el gráfico, esta clara estacionalidad entre los meses se repite a pesar de la disminución de audiencia entre los años.

Un detalle importante que se puede observar en la gráfica es la afectación del COVID 19 a la audiencia. Mientras en los años anteriores presentan bajadas de audiencia, en marzo 2020 se observa un brutal despegue de la audiencia llegando a niveles históricos nunca alcanzados. Enero y febrero 2020 mantuvieron la tendencia bajista y en marzo, con el decreto del estado de alarma, las audiencias comienzan a dispararse. El motivo de este incremento de audiencia es evidente, la población tiene la necesidad de estar informado, incrementa su tiempo libre con los ERTES, no disponen de muchas opciones para cubrir las horas del día con el confinamiento y se dedican a ver la televisión. Una vez vuelta a la "nueva normalidad" las audiencias se estabilizan como podemos ver en el mes de junio.

Se incluye el filtro de grupo comercial (Atresmedia, Mediaset y Pulsa) para ver el comportamiento de la audiencia por grupo, ya que se comporta de forma diferente. Si pulsamos en cada uno de los grupos podemos ver que tanto la tendencia bajista entre años, como la estacionalidad de los meses se mantiene. Sin embargo, analizando Mediaset se aprecia una subida de audiencia en junio 2018 cuando la tendencia para este mes respecto a Mayo debería ser descendente; esto se debe a que Mediaset emitió el Mundial de Rusia de 2018 que se celebró en ese mes. De igual manera, se puede apreciar que el COVID 19 ha beneficiado en mayor medida a Pulsa, el gap de la curva de audiencia entre 2019 y 2020 es muy superior al resto de grupos comerciales y esto es así por la programación del grupo comercial que ofrece mayoritariamente series y películas.

TIMEFRAME EVOLUTION - AM(%)

2021



Análisis de franja horaria: TIME FRAME

El objetivo de este análisis es profundizar en los movimientos de audiencia entre franjas principales (Mañana, Comida, Tarde, Prime Time, Noche y Madrugada). Con ello, las agencias de medios pueden ir viendo el comportamiento de la audiencia, los movimientos de una franja a otra.

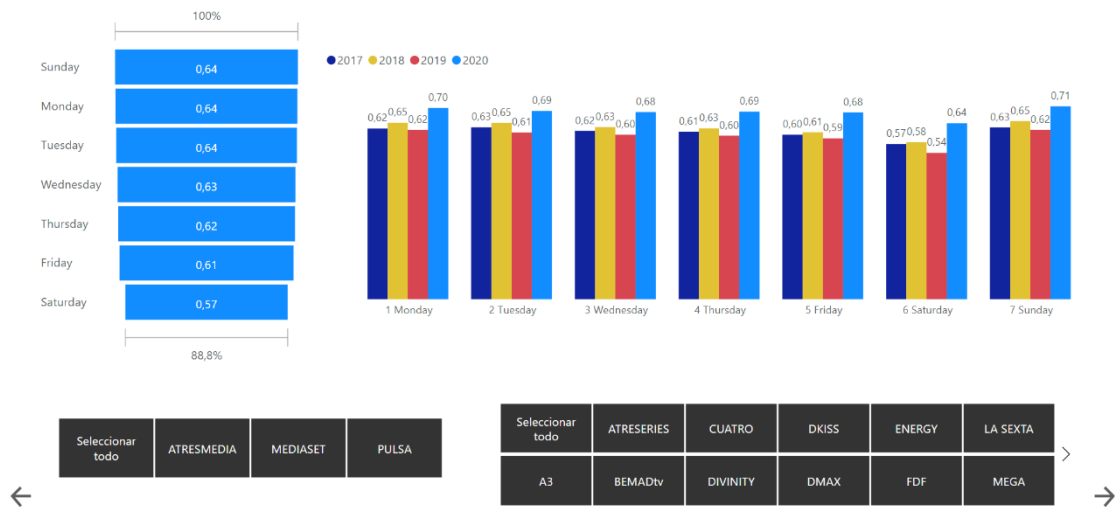
De igual manera que en los meses, las franjas horarias presentan un patrón claramente definido, donde la curva de audiencia del día pinta dos jorobas (los dos picos de mayor audiencia): una en la hora de medio día con la comida, otra en la franja llamada prime time, durante la cena y después de la jornada laboral. Esta curva de las dos jorobas se dibuja prácticamente igual en todos los años, con las bajadas de audiencia media de un año a otro como comentábamos anteriormente, pero siempre con los dos picos perfectamente identificadas (las dos franjas de mayor audiencia: medio día y prime time).

De nuevo el COVID 19 deja su huella en este análisis. En la gráfica de franjas horarias se puede apreciar como en 2020 la curva de audiencia incrementa, pero no lo hace de forma homogénea por franja, lo hace en mayor media en unas horas determinadas del día; durante la madrugada apenas se aprecia este incremento, mientras que, durante la mañana, medio día y tarde, este incremento es muy superior. El motivo es que en situación normal la población no se encuentra en casa en esas horas por lo que no puede ver la televisión, y ahora con el confinamiento, han aprovechado para informarse y distraerse. Por otro lado, el prime time incrementa, pero en menor medida, porque los niveles de los que se partían antes del COVID eran ya muy superiores. Es una franja donde la posibilidad de ver la televisión no queda determinada por situación de confinamiento y donde la audiencia apenas tiene margen de crecimiento.

Se incluye el filtro de grupo comercial de nuevo para ver ambas variaciones, global mercado y por salehouse. Analizando uno a uno, podemos ver que la franja del lunch incrementa notoriamente en Atresmedia y el motivo es porque este grupo comercial es el favorito de la audiencia para mantenerse informado.

WEEKDAY BEHAVIOUR - AM(%)

2021



Análisis del día de la semana: WEEKDAY

Con este análisis se profundiza en el detalle del resultado día a día de la semana para cada una de las cadenas y grupos comerciales. Es un análisis pormenorizado que resulta muy interesante para los planificadores ya que sin saber la programación de cada una de las cadenas pueden ver cuales son los días que mejor funcionan y afinar más las campañas activas.

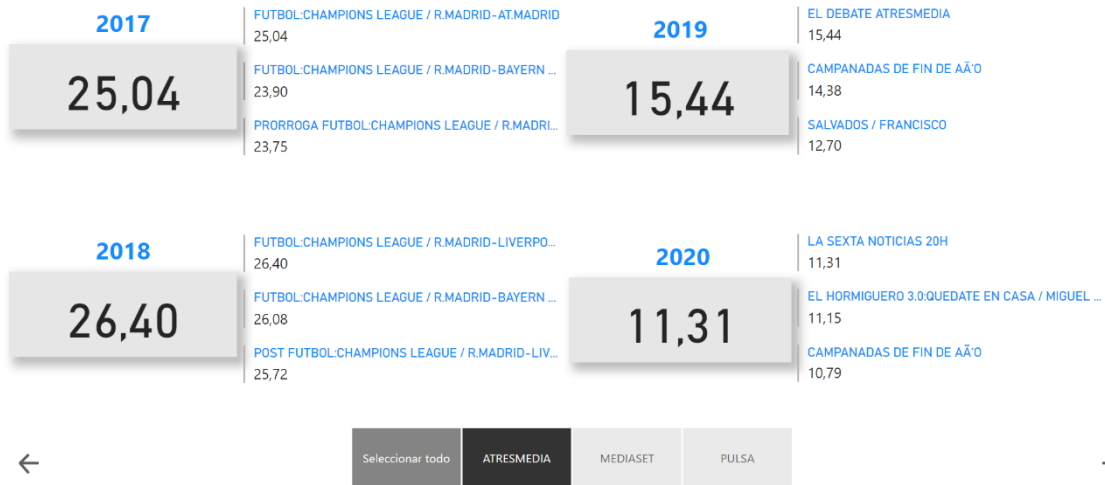
Es otro factor que muestra un patrón que se suele repetir, aunque no es tan claro como en los casos anteriores. Los domingos y los lunes son los días de mayor audiencia, el primero ya asumiendo que vuelve la semana la gente ve más la televisión, así como los lunes cuando vuelven a casa de la jornada laboral. Por el contrario, los viernes y sábados suelen ser los días de menos audiencia porque aprovechan para otro tipo de ocio.

Este patrón varía también según cadena debido a la programación que esta emite, como por ejemplo sucede con fútbol o con realities como gran hermano, la isla de las tentaciones y similar.

Estos tres primeros análisis dejan ver una visual global rápida del mercado para entender como consume la televisión el telespectador en España.

Para estos análisis se ha incluido un botón 2021 en la parte superior derecha de la pantalla para poder ver la predicción del modelo en comparación con lo real cargado y utilizado para dicha estimación.

GOLDEN MINUTE - AM(%)



Análisis de los minutos más vistos: GOLDEN MINUTES

Los minutos más vistos del año es un análisis que se hace para un mayor conocimiento del mercado y no tanto para la utilización en una planificación o estrategia.

Con este análisis, se aprecian los picos de audiencia que se producen tanto de forma anual como por mes y por cadena.

Lo que se puede apreciar en este análisis es la importancia del deporte para la audiencia española, más concretamente el fútbol. Si se analizase en detalle día a día veríamos como momentos como atentados terroristas, catástrofes naturales, debates preelectorales y otros eventos fuera de lo normal, producen también picos en el mes; la población continúa utilizando la televisión para informarse.

TOP FIFTEEN PROGRAMS - AM(%)

TITLE	AM(%)	MINUTES	CHANNEL
POST FUTBOL-CHAMPIONS LEAGUE / R.MADRID-LIVERPOOL	25,08	2	A3
FUTBOL-CHAMPIONS LEAGUE / R.MADRID-LIVERPOOL	22,30	96	A3
FUTBOL-CHAMPIONS LEAGUE / PARIS ST.GERMAIN-R.MADRID	20,77	95	A3
FUTBOL-CHAMPIONS LEAGUE / CHELSEA-BARCELONA	16,34	96	A3
FUTBOL-CHAMPIONS LEAGUE / R.MADRID-BAYERN MUNICH	15,02	289	A3
PREVIO FUTBOL-CHAMPIONS LEAGUE / R.MADRID-LIVERPOOL	14,86	1	A3
PREVIO FUTBOL-CHAMPIONS LEAGUE / PARIS ST.GERMAIN-R.MADRID	14,63	2	A3
FUTBOL-CHAMPIONS LEAGUE / ROMA-BARCELONA	14,24	94	A3
PRORROGA FUTBOL-CHAMPIONS LEAGUE / R.MADRID-BAYERN MUNICH	14,02	68	A3
FUTBOL-CHAMPIONS LEAGUE / JUVENTUS-R.MADRID	13,97	288	A3
EL DEBATE ATRESMEDIA	13,71	125	A3
FUTBOL-CHAMPIONS LEAGUE / R.MADRID-AT.MADRID	13,53	190	A3
POST FUTBOL-CHAMPIONS LEAGUE / R.MADRID-BAYERN MUNICH	12,91	11	A3
FUTBOL-CHAMPIONS LEAGUE / MANCHESTER UNITED-SEVILLA	11,77	98	A3
SALVADOS / FRANCISCO	11,32	63	LA SEXTA
EL DEBATE ATRESMEDIA	11,00	126	LA SEXTA

Seleccionar todo	2018	2020
2017	2019	2021

Seleccionar todo	January	November
April	July	October
August	June	September
December	March	
February	May	

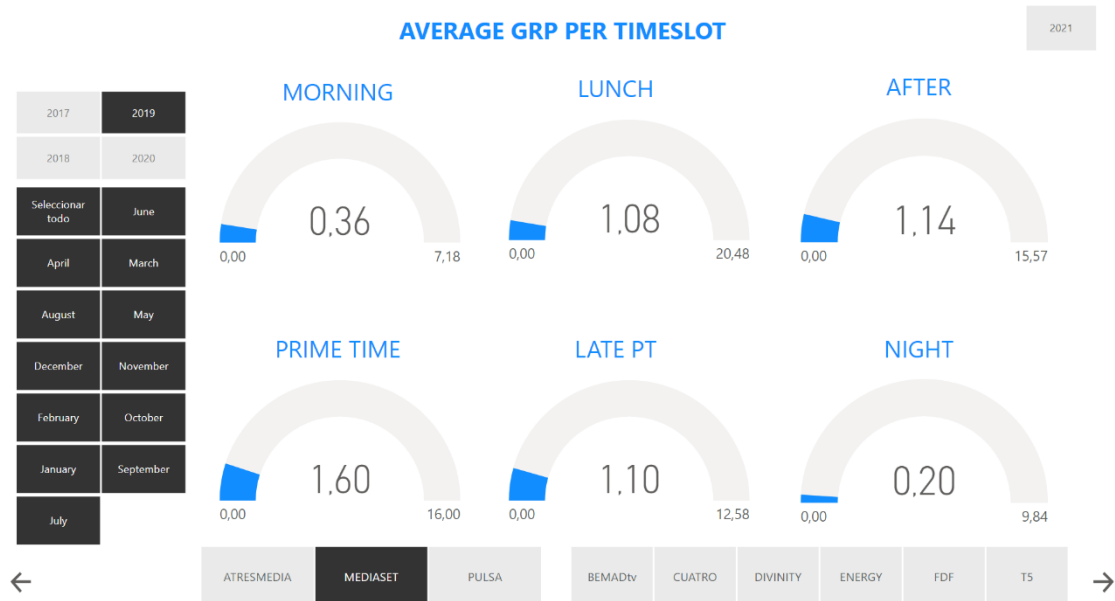
Seleccionar todo	ATRESMEDIA	MEDIASET	PULSA
------------------	------------	----------	-------

← →

Análisis de las emisiones más vistas: TOP TV BROADCASTS

El objetivo de este análisis es entender las emisiones más vistas por la audiencia española. De esta manera se puede saber aquellos programas con los que alcanzaras una audiencia mayor cuando un anunciante solicita recomendación de programas a patrocinar o donde alcanzar a gran número de gente.

Como comentado anteriormente, los contenidos relacionados con deporte son los de más éxito seguidos de realities y el programa El hormiguero, programa estrella que suele contener el minuto de oro del día de forma muy recurrente.



Análisis del grp promedio: AVERAGE GRP

Los planificadores de medios han de controlar de forma diaria el grp promedio que les arrojan cada una de las cadenas ya que de ello dependerá su campaña y sus acciones a tomar, si la cadena planificada incrementa su audiencia tendrán que anular spots para no pasarse de presupuesto, si la audiencia de la cadena disminuye el planificador tendrá que comprar spots para no dejar baja la campaña.

De igual manera con este análisis pueden ver una estimación de lo que serán los próximos días de campaña con esa estimación de grp promedio de los últimos días.

Al departamento de negociación y compra le servirá como arma en la negociación con las cadenas, viendo como van sus resultados de audiencia podrán renegociar costes o presionar con la mejora de cualitativos.

Esta ventana proporciona un minucioso detalle de todo lo visto anteriormente, la variación de audiencia por franja, mes, cadena, año...

Con todos estos insights la agencia de medios puede hacer recomendaciones a sus clientes cara a las planificaciones y estrategias mucho más precisas. Cara a la definición estratégica de la campaña ayuda mucho saber como se distribuye la audiencia en todos los parámetros medidos como meses, cadenas, horas...

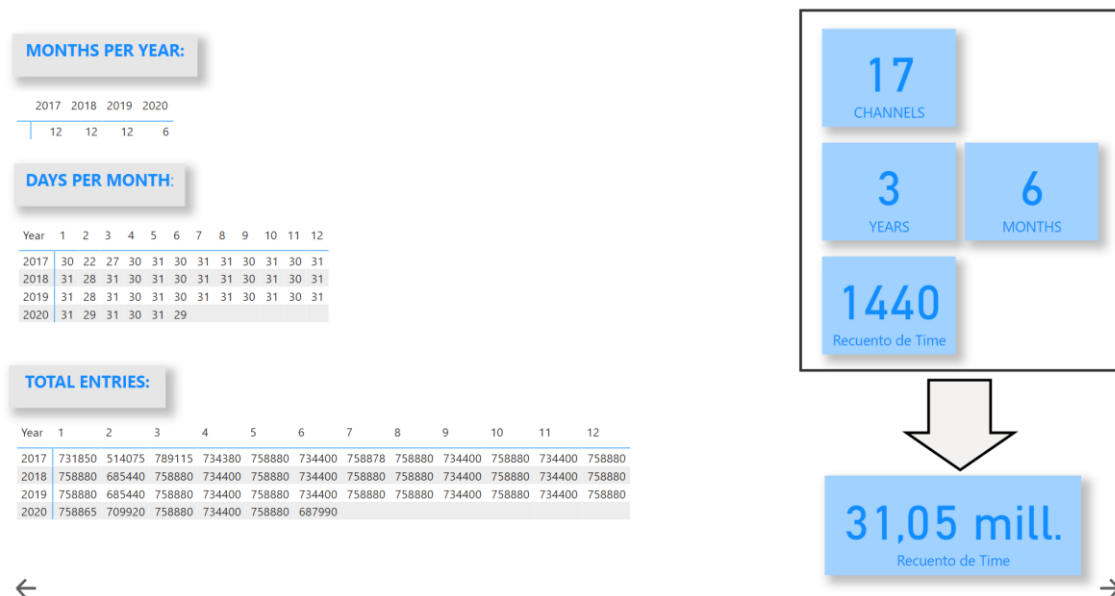
Por otro lado, el módulo de forecast de audiencia es extremadamente relevante para la agencia porque controlando la audiencia controlarán las campañas de sus clientes ya que como comentábamos en la memoria del proyecto, la compra más habitual en el mercado español es la compra a grp.



JAVIER CASADO DE AGUINAGA

Además del desarrollo de la visualización de los análisis realizados sobre los datos, así como la predicción para el 2021, se ha procedido a graficar el control de la carga de datos.

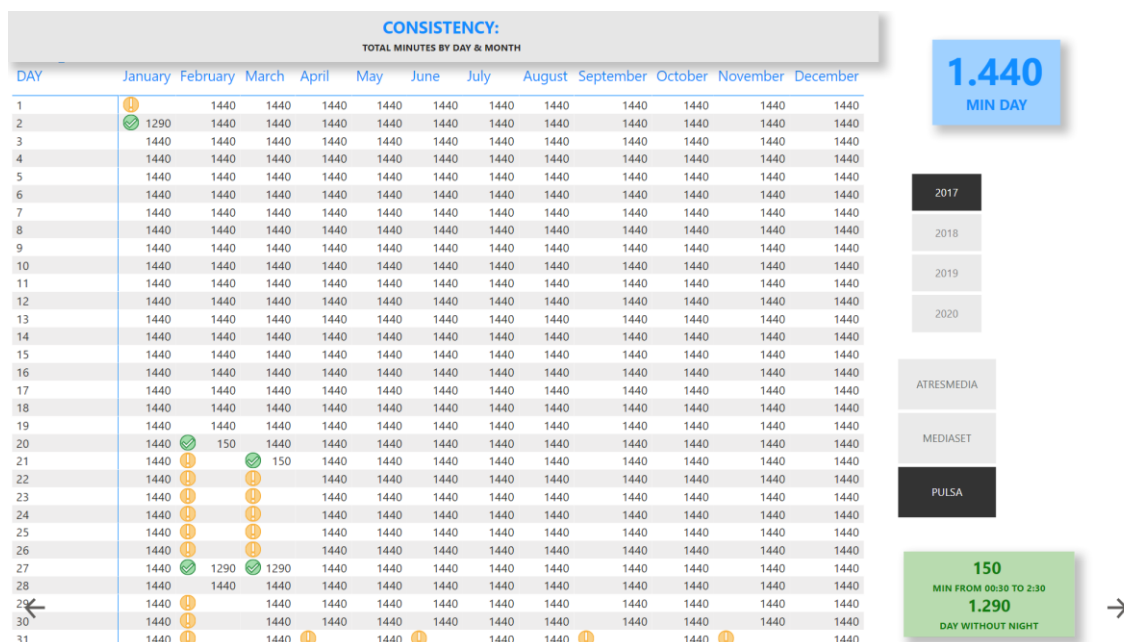
Para ver que las pestañas de control de datos funcionan y detectan los errores en la carga, se ha hecho una carga con una semana duplica, falta de otra semana, minutos en ciertas cadenas no existentes.... Lo iremos viendo en la explicación de cada una de las pestañas.



DATA UPLOADED:

Con esta pestaña se controla:

- Meses cargados para cada uno de los años cargados (donde podemos ver que para 2020 se ha cargado solo hasta Junio, 6 meses)
- Días por cada uno de los meses cargados año a años y así poder detectar si la carga falló y se dejó algún día sin cargar (esto se puede apreciar en Enero de 2017)
- Total de registros donde se puede detectar mes a mes en cada uno de los años si los registros totales son los que debiera y no (se puede ver por ejemplo que Febrero y Marzo de 2017 no corresponden con resto del año por lo comentado anteriormente, se ha cargado una semana de Marzo por duplicado y en Febrero se dejó una sin cargar)
- Para entender un poco la información cargada se detalla número de cadenas, años y meses cargados. Además, se cuantifica el total de registros de la herramienta que superan los 31 millones de registros.



CONTROL BY DAY & MONTH

De un vistazo se puede detectar si se ha dejado sin cargar (por fallo en el proceso o por fallo en la descarga de la propia herramienta de origen) algún minuto de cada uno de los días del mes. Se ha incluido un filtro para poder ver año a año y por cada cadena. Como se puede apreciar en la carga realizada, tenemos 2 semanas sin cargar: del 20 al 26 de Febrero y del 21 al 26 de Marzo, ambas de 2017.

Se incluyen unos iconos que ayudan a la visualización más rápida:

- ! en el caso de ser blanco porque puede ser un mes que no tenga 31 días o el caso de Febrero que tiene 28 o 29 según año bisiesto o no
- Las celdas con registros de 150 se consideran buenos porque significa que se cargo solo la madrugada que va de 00:00 a 2:30 (* 2,5 horas * 60 minutos)
- De igual manera los campos con 1.290 registros se consideran correctos porque significa que el error viene de la semana previa y se quedó sin cargar la madrugada (1.440 registros – 150 registros de la madrugada)



WEEK CONTROL

Al haber hecho la carga por semanas, este análisis muestra rápidamente si alguna de las semanas quedó pendiente de cargar, o incluso si se cargaron doble como es el caso de la semana 10 de 2017.

El control se ha hecho contabilizando los registros totales por semana (17 cadenas * 7 días * 24 horas * 60 minutos = 171.360 entradas) se puede ver incluso el detalle de minutos no cargados: como sucede en 2017 semana 28 donde faltan dos registros y en semana 5 donde faltan 15 registros (veremos más adelante que no es problema de la carga si no de la herramienta de origen)

DUPLICATED:												
TOTAL ENTRIES BY DAY & MONTH - DUPLICATED												
Día	January	February	March	April	May	June	July	August	September	October	November	December
1		24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480
2	21930	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480
3	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480
4	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480
5	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480
6	24480	24480	46410	24480	24480	24480	24480	24480	24480	24480	24480	24480
7	24480	24480	48960	24480	24480	24480	24480	24480	24480	24480	24480	24480
8	24480	24480	48960	24480	24480	24480	24480	24480	24480	24480	24480	24480
9	24480	24480	48960	24480	24480	24480	24480	24480	24480	24480	24480	24480
10	24480	24480	48960	24480	24480	24480	24480	24480	24480	24480	24480	24480
11	24480	24480	48960	24480	24480	24478	24480	24480	24480	24480	24480	24480
12	24480	24480	48960	24480	24480	24480	24480	24480	24480	24480	24480	24480
13	24480	24480	27030	24480	24480	24480	24480	24480	24480	24480	24480	24480
14	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480
15	24480	24475	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480
16	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480
17	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480
18	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480
19	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480
20	24480	2550	24480	24460	24480	24480	24480	24480	24480	24480	24480	24480
21	24480	7710	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480
22	24480	600	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480
23	24480		24480	24480	24480	24480	24480	24480	24480	24480	24480	24480
24	24480		24480	24480	24480	24480	24480	24480	24480	24480	24480	24480
25	24480		24480	24480	24480	24480	24480	24480	24480	24480	24480	24480
26	24480		24480	24480	24480	24480	24480	24480	24480	24480	24480	24480
27	24480	21930	21930	24480	24480	24480	24480	24480	24480	24480	24480	24480
28	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480	24480
29	24480		24480	24480	24480	24480	24480	24480	24480	24480	24480	24480
30	24480		24475	24480	24480	24480	24480	24480	24480	24480	24480	24480
31	24480		24480	24480	24480	24480	24480	24480	24480	24480	24480	24480

24.480
ENTRIES DAY

2017

2018

2019

2020

2.550

FROM 00:30 TO 2:30 (17 CHANNELS)

21.930

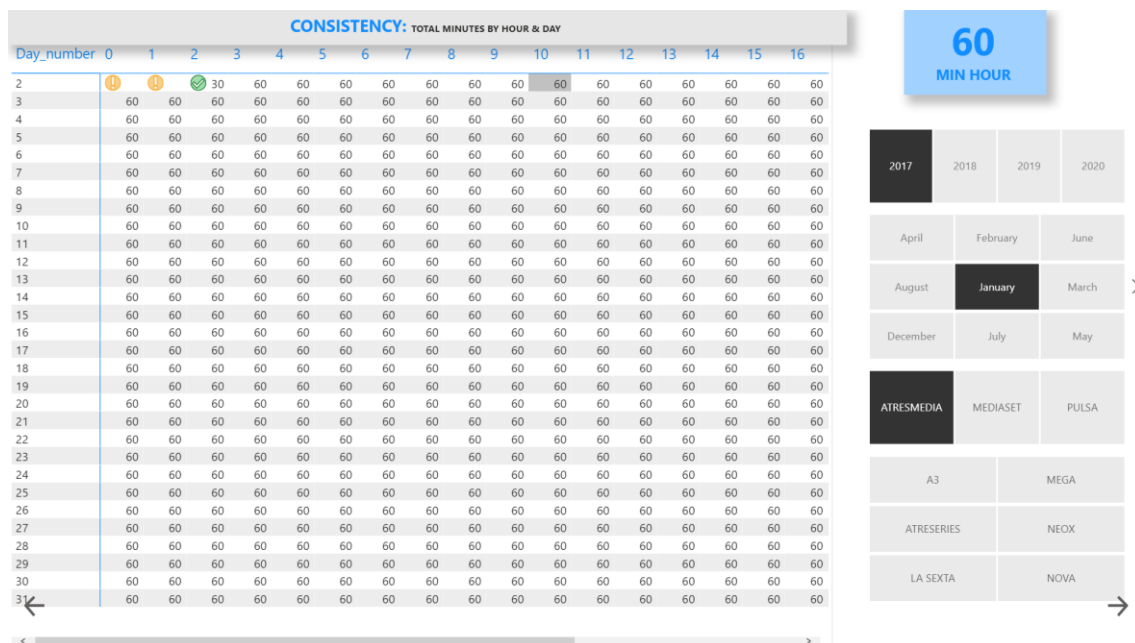
DAY WITHOUT NIGHT OF 17 CHANNELS

DUPLICATED CONTROL:

El objetivo de esta pestaña es el control de aquellos días que se suben de forma duplicada, como es el caso de la semana que va del 6 al 13 de Marzo de 2017. Para ello se incluye un filtro para poder controlar año a año.

El conteo de registros en este caso es el número de cadenas (17) * el número de horas del día (24) * el número de minutos de cada hora (60) = 24.480 registros. Toda aquella celda con número diferente a este número nos dará un icono como aviso:

- ! en el caso de ser blanco porque puede ser un mes que no tenga 31 días o el caso de Febrero que tiene 28 o 29 según año bisiesto o no
- Las celdas con registros de 2.550 se consideran buenos porque significa que se cargo solo la madrugada que va de 00:00 a 2:30 de todas las cadenas (17 cadenas * 2,5 horas * 60 minutos)
- De igual manera los campos con 21.930 registros se consideran correctos porque significa que el error viene de la semana previa y se quedó sin cargar la madrugada (24.480 registros – 2.550 registros de la madrugada)



CONTROL BY HOUR & DAY:

Se controla hora a hora por cada día que no falta ningún minuto de los 60 correspondientes. Para poder ver aquellos registros que nos faltaban en 2017 donde teníamos:

- Semana 28 de 2017 (11 de Julio) donde se detecta la falta de 2 registros. Esto se ha confirmado y viene de la herramienta de origen porque la cadena haya perdido la señal.
- Semana 5 de 2020 (20 de Enero) se detecta que faltan 15 registros que de igual manera viene de la fuente origen.