

TERESA JOVEN

Sección #10 – Limpieza, Procesado y Análisis

Individual.

Puntuación: 3 puntos

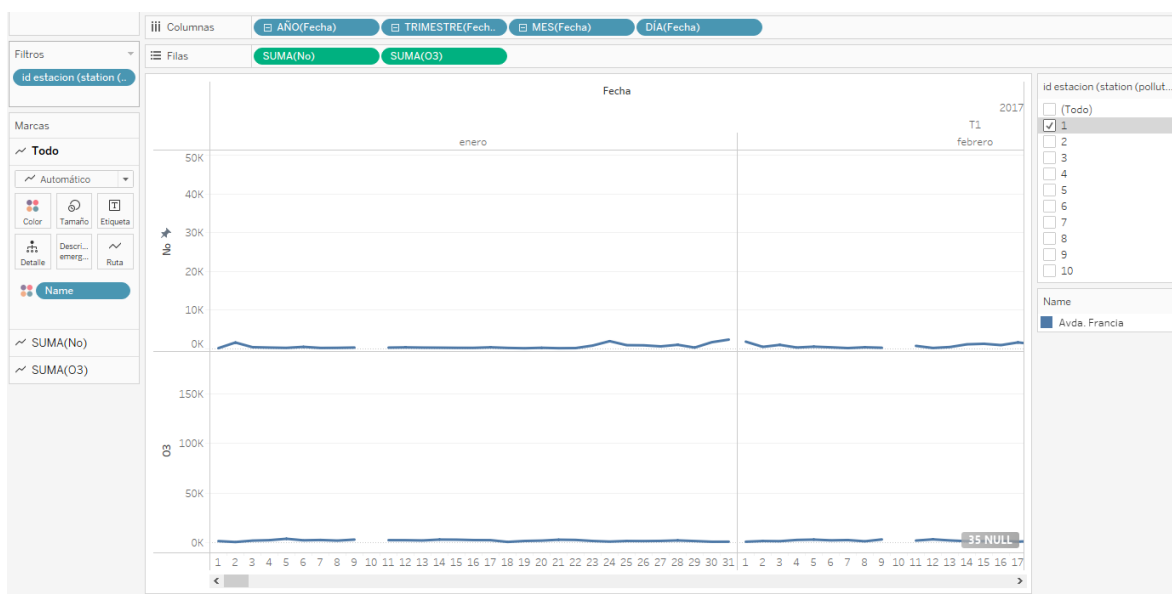
En esta sección haremos uso de los siguientes contaminantes: NO, O3.

Data Cleansing

14. Muestra estos 2 contaminantes en función del tiempo (en uno o más gráficos), y filtra por estación. Utiliza esta visualización para eliminar los valores nulos.

[CS09] Capturas de pantalla del gráfico sin nulos: (0,5 puntos)

El primer gráfico que mostramos es en función del año, el trimestre, el mes y el día.



Observamos que el contaminante O3 indica 35 nulos para la estación 1 (Avda. Francia). Filtramos los datos.

Valores especiales para [No]



El eje **No** tiene 35 valores NULL.

¿Qué desea hacer?

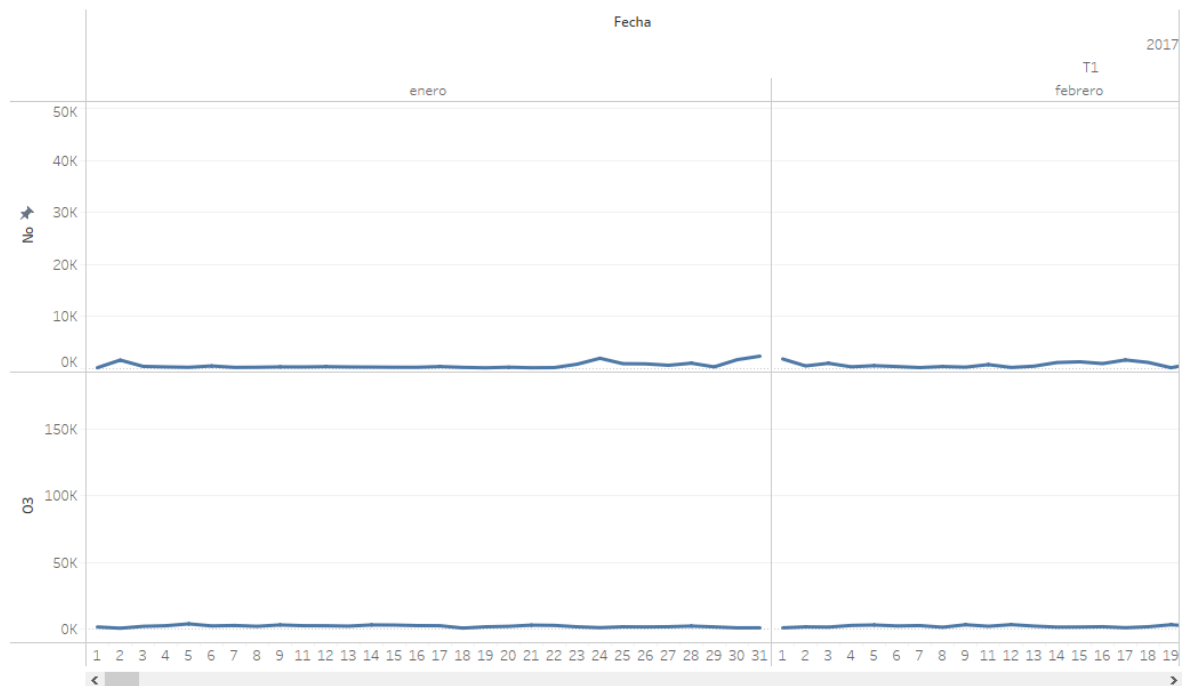
Filtrar datos

Excluya los valores especiales desde la vista y cálculos.

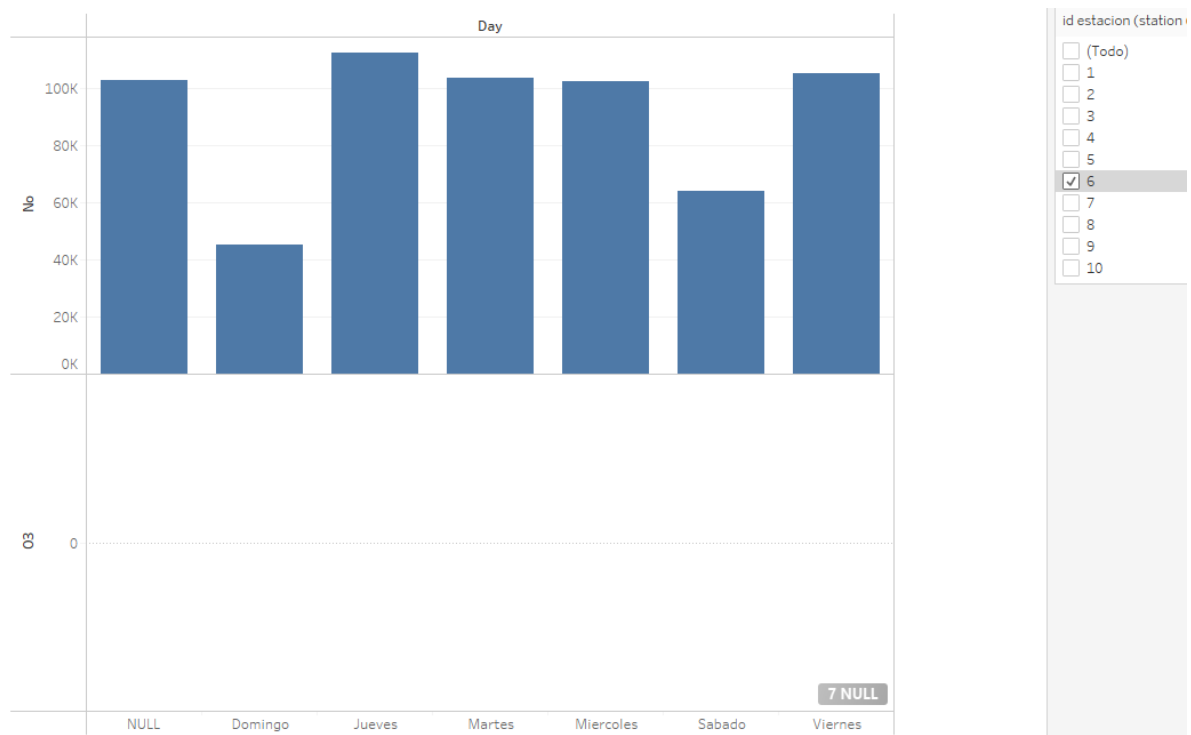
Mostrar datos en la posición predeterminada

Mostrar los valores especiales en una posición predeterminada en el eje. Por ejemplo, los valores NULL se muestran en 0.

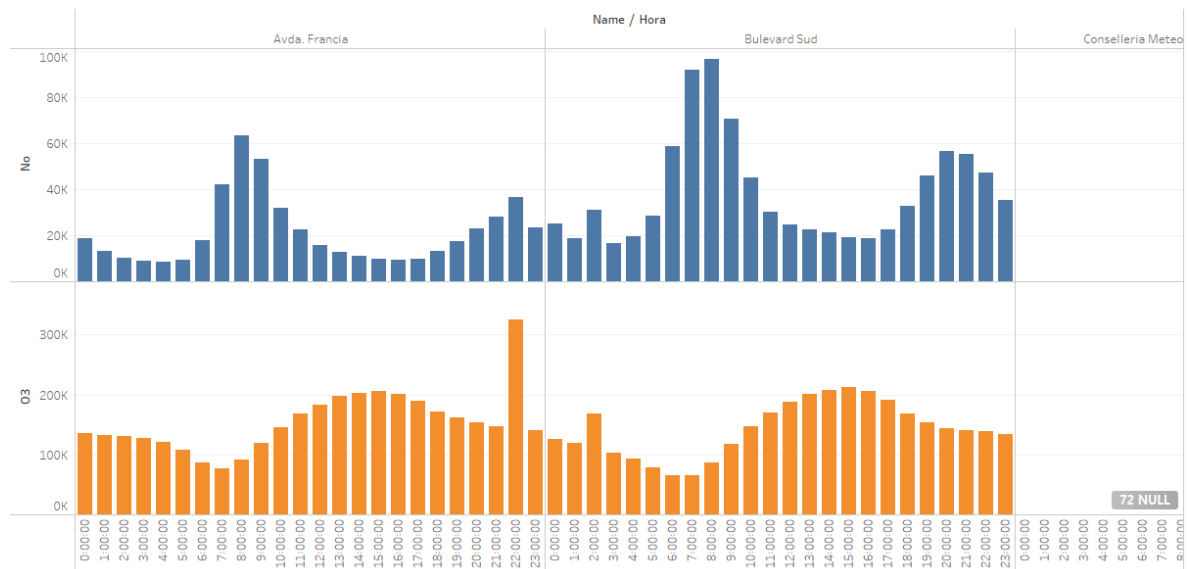
Vemos en la gráfica inferior que la del O3 cómo se ha eliminado los valores nulos.



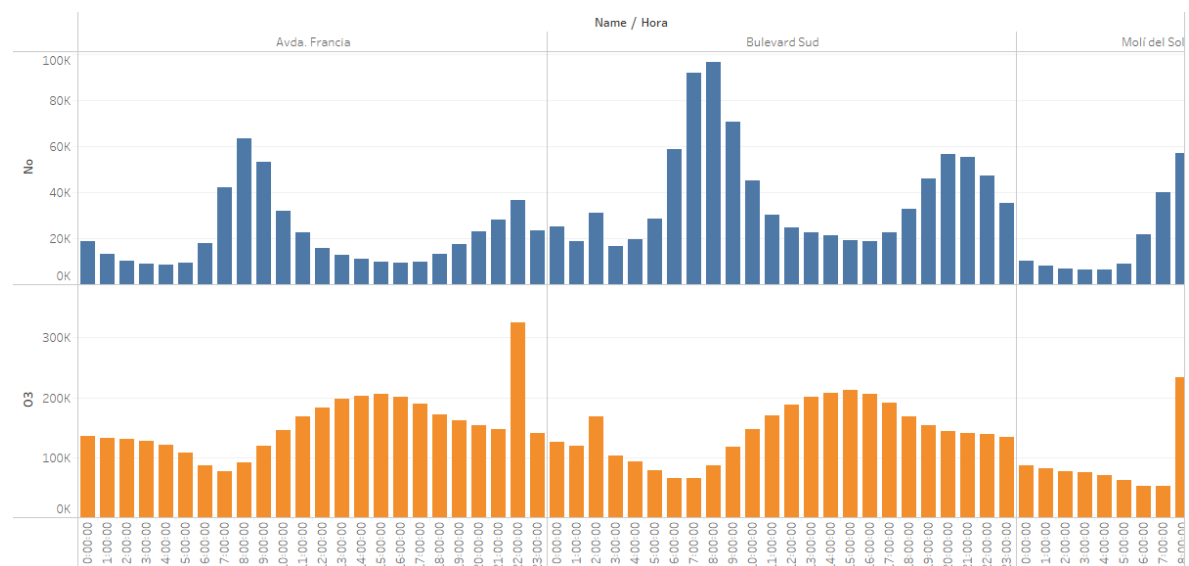
Utilizamos un segundo gráfico para representar los contaminantes en función del día de la semana. Filtrando por la estación 6 de “València Centre” aparecen 7 nulos que procedemos a eliminar.



Por último, un tercer gráfico que muestra los contaminantes por hora y por nombre de la estación.



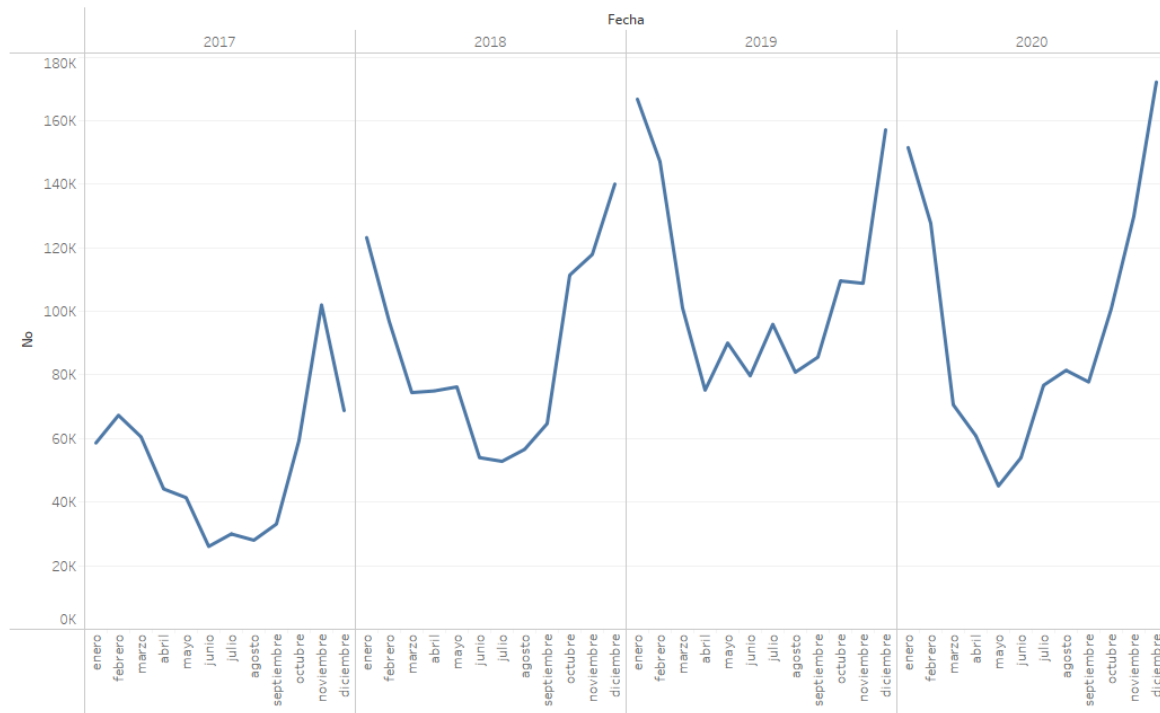
Eliminamos los valores nulos y vemos que desaparece la estación “Conselleria Meteo” con respecto a la imagen anterior.



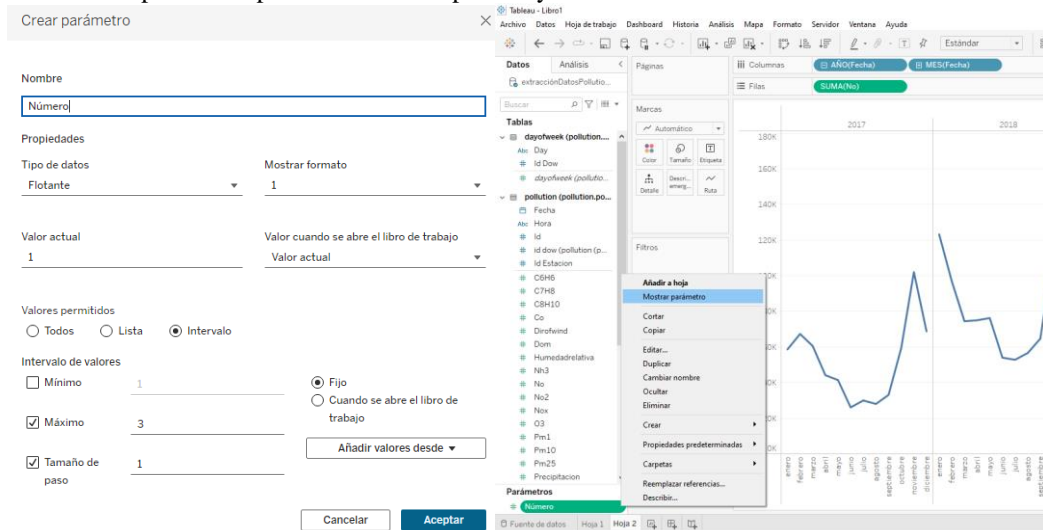
15. **EXTRA:** (Este apartado es optativo, su realización beneficiará a la nota de forma aditiva) Detecta los Outliers. Indica cuál es límite escogido y da algunos ejemplos de los casos (día/hora/estación). Una vez detectados elimina los Outliers directamente desde la fuente de datos.

[CS10 EXTRA] Detección y limpieza de outliers: (0,5 puntos)

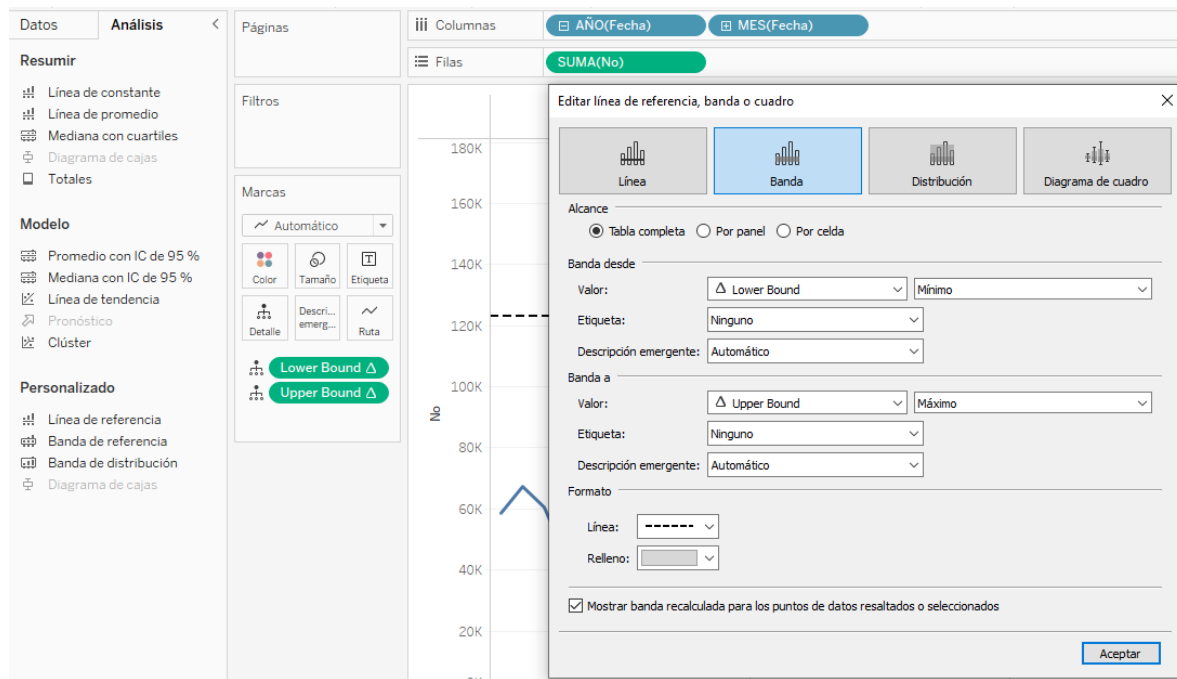
Para este apartado opcional sigo el vídeo que se presentó en la videoconferencia. Partimos del contaminante NO en función del mes y el año.



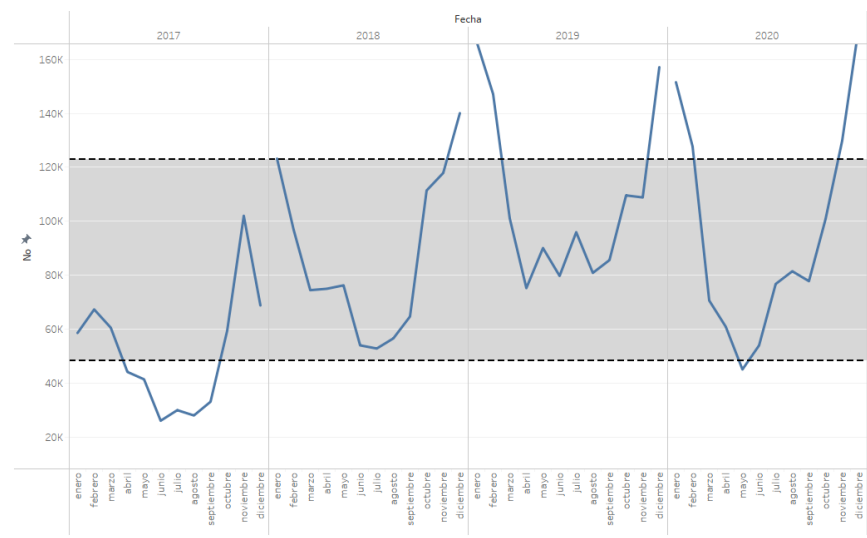
Creamos un parámetro para las series temporales y lo mostramos.



Ahora en la siguiente página creamos un campo calculado para la cota superior y hacemos lo mismo para la cota inferior.



Aceptamos y vemos la banda sobre el gráfico.



Creo un campo calculado llamado “Outliers” que detecte los valores no comprendidos en esta banda.

Outliers

×

```

IF SUM([No]) > [Upper Bound]
OR SUM([No]) < [Lower Bound ]
THEN "Outliers"
ELSE "In bounds"
END

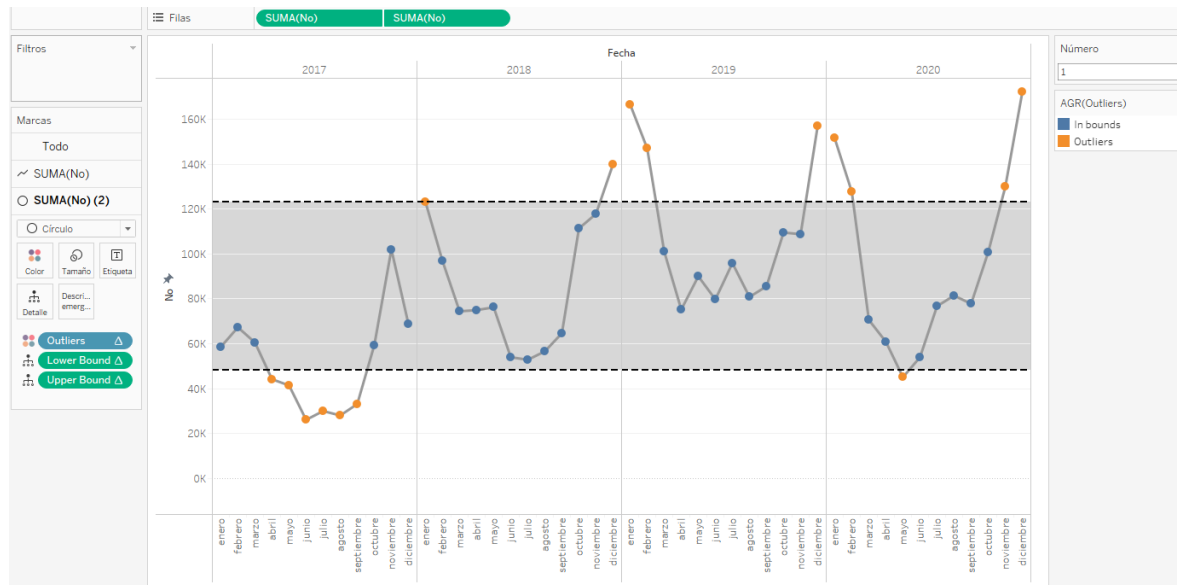
```

El cálculo es válido.

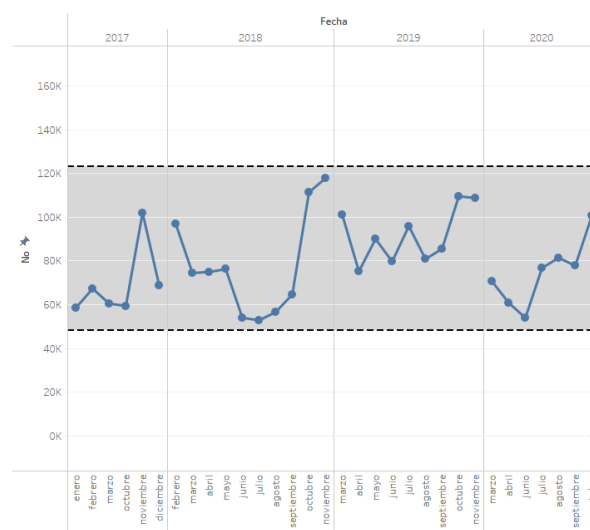
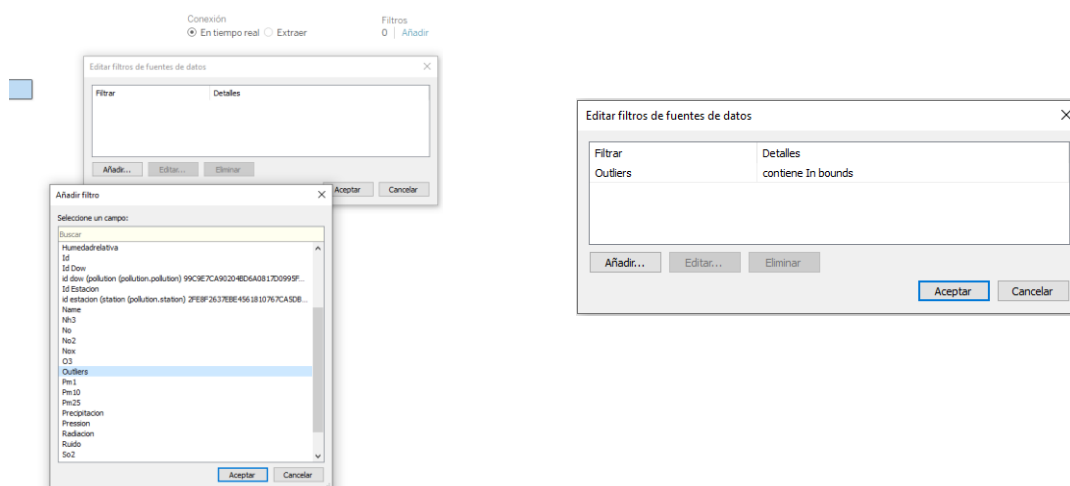
Aplicar

Aceptar

Los represento en color naranja sobre el gráfico. Para ello, señalo eje doble, sincronizo el eje y muestro el encabezado.



Por último, elimino los “Outliers” directamente sobre la fuente de datos.



Data Processing

En esta sección procesaremos los datos creando un campo calculado:

16. Crea un campo flag: Imagina que tenemos que alertar la población en caso de alta contaminación por NO y O3. Crea un campo que avise cuando esto suceda.

Este campo podrá valer los valores “ALERTA” o bien “-” según el siguiente criterio:

- Daremos alerta cuando uno de los 2 contaminantes alcance el 90% de su valor máximo histórico. Busca el valor máximo histórico de cada contaminante y aplícalo en el cálculo a mano.

[CS11] Capturas de pantalla de la fórmula: (0,5 puntos)

Buscamos el valor máximo histórico del contaminante NO, que es 7404, y aplicamos el 90% en el cálculo a mano.

Año de Fecha	
2017	7.404
2018	431
2019	4.936
2020	3.702

```
Flag_no  
  
IF [No]>=6663.6  
THEN "ALERTA"  
ELSE "-"  
END
```

El cálculo es válido y le damos a Aceptar.

Por otro lado, buscamos el valor máximo histórico del contaminante O3, que es 104890, y aplicamos el 90% en el cálculo a mano.

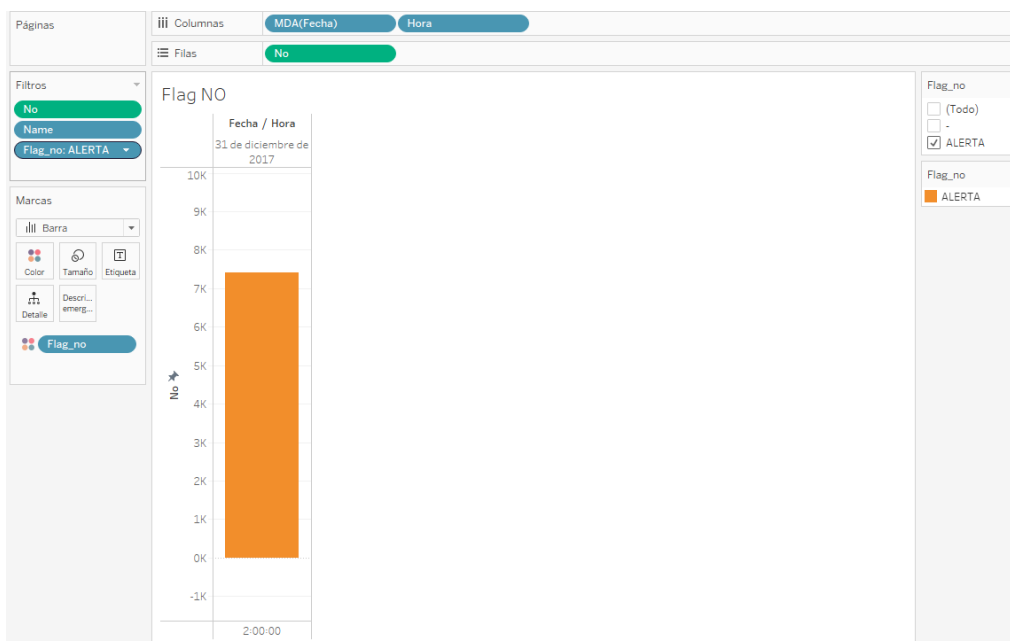
Año de Fecha	
2017	90.082
2018	165
2019	82.678
2020	104.890

```
Flag_o3  
  
IF [O3]>=94401  
THEN "ALERTA"  
ELSE "-"  
END
```

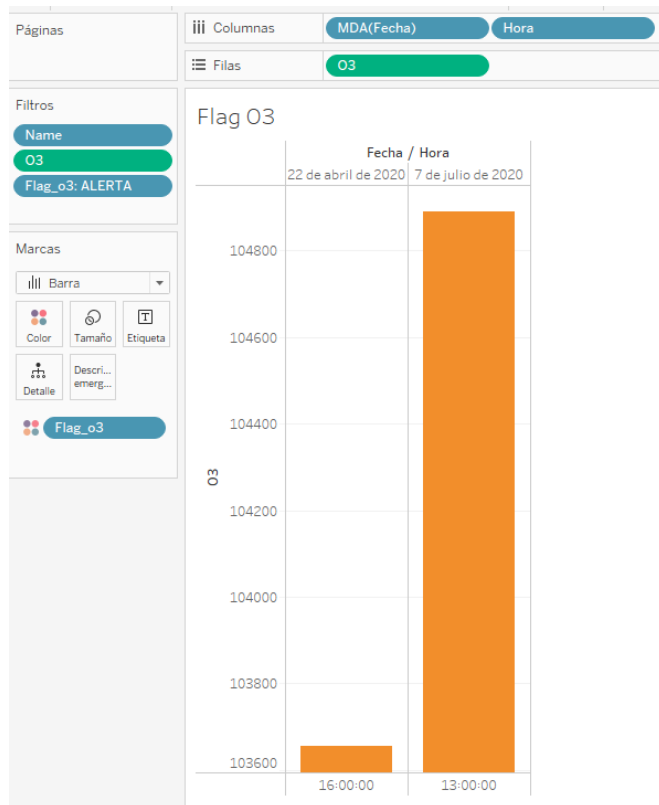
17. Muestra la flag en un gráfico. Hay que poder ver claramente en que momentos ha saltado el aviso.

[CS12] Capturas de pantalla del gráfico de la flag: (0,5 puntos)

Para el contaminante NO el aviso salta el 31 de diciembre de 2017 a las 2:00:00.



Para el contaminante O3 el aviso salta el 22 de abril de 2020 a las 16:00:00 y el 7 de julio de 2020 a las 13:00:00.

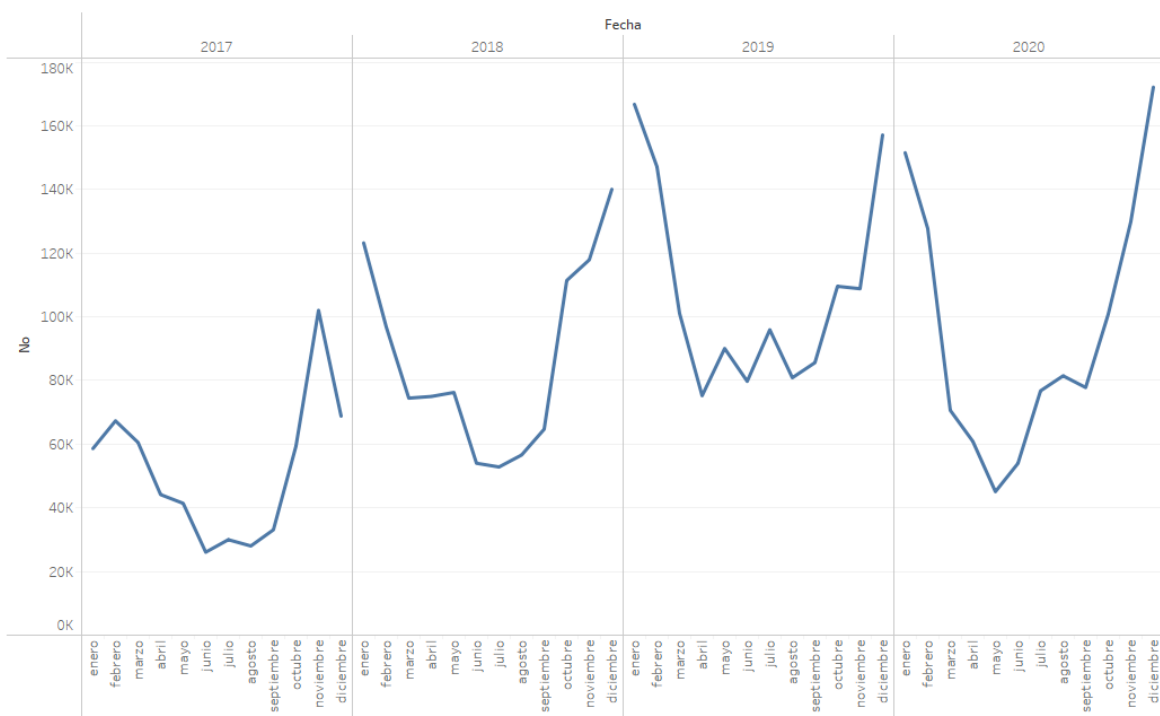


Data Analytics y conclusion

18. Muestra los contaminantes NO y O3 en función del tiempo a escala mensual. ¿Puedes observar algún patrón de estacionalidad? Si es así, ¿a qué se debe?

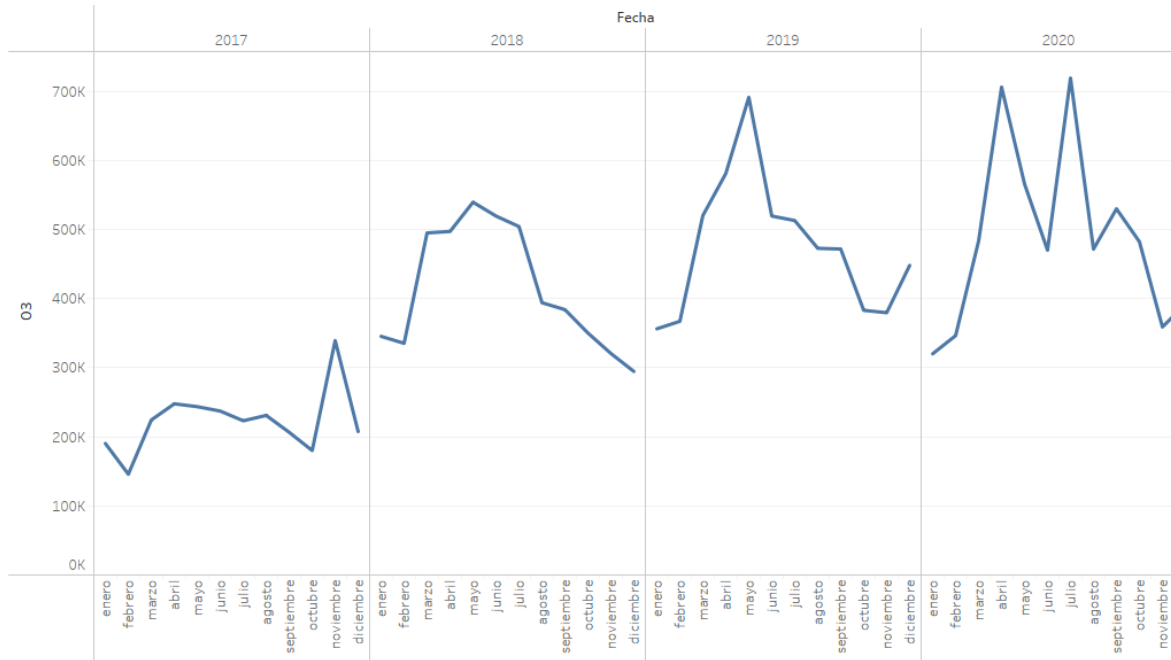
[CS13] Capturas de pantalla de la estacionalidad: (0,25 puntos) Explicación:

(0,25 puntos)



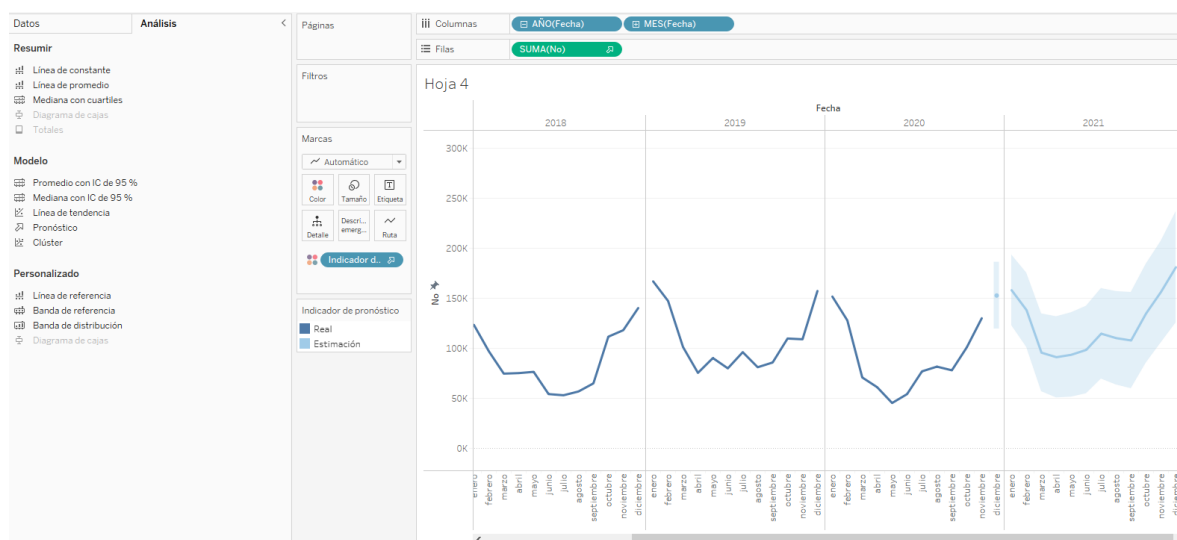
Se observa que el contaminante NO toma valores más bajos en los meses más cálidos y valores más altos en los meses más fríos. Se emite en los procesos de combustión que se dan más en los meses de invierno.

Por el contrario, con el contaminante O3 el patrón de estacionalidad es a la inversa (ver imagen inferior). Se observa que generalmente toma valores superiores en los meses cálidos. Esto es debido a que su formación empieza a partir de la emisión del dióxido de nitrógeno (NO2) e hidrocarburos (compuestos que reaccionan en la presencia de calor y luz solar para producir ozono).

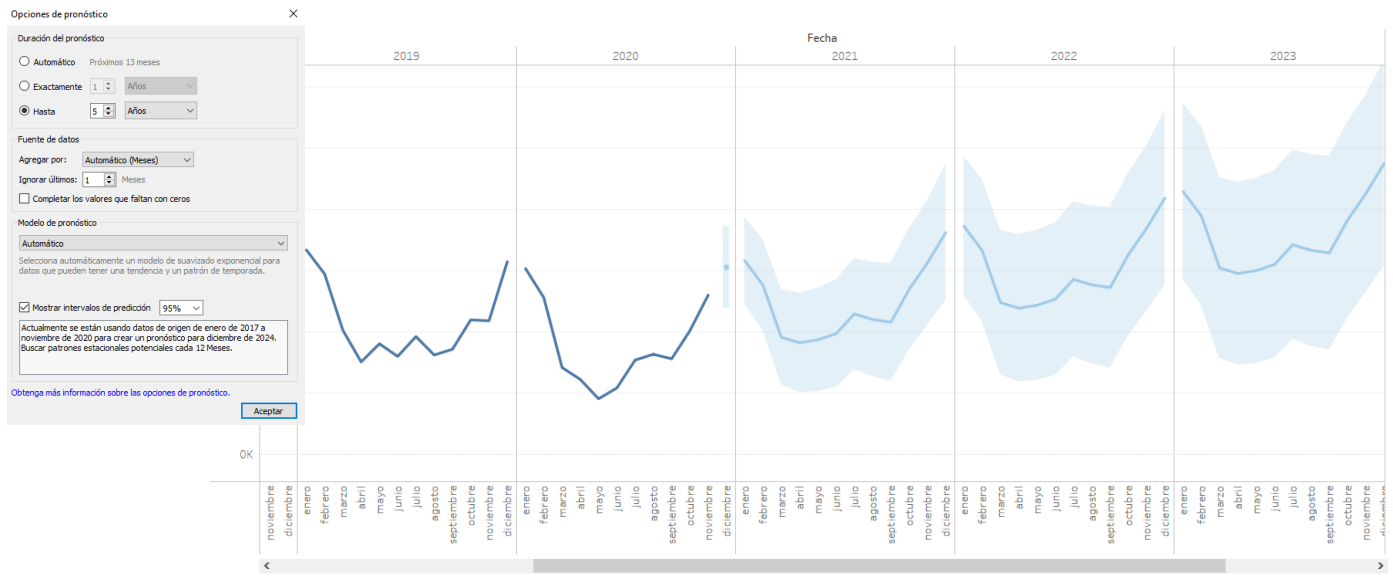


19. Haz un pronóstico a futuro (Forecast) para NO.

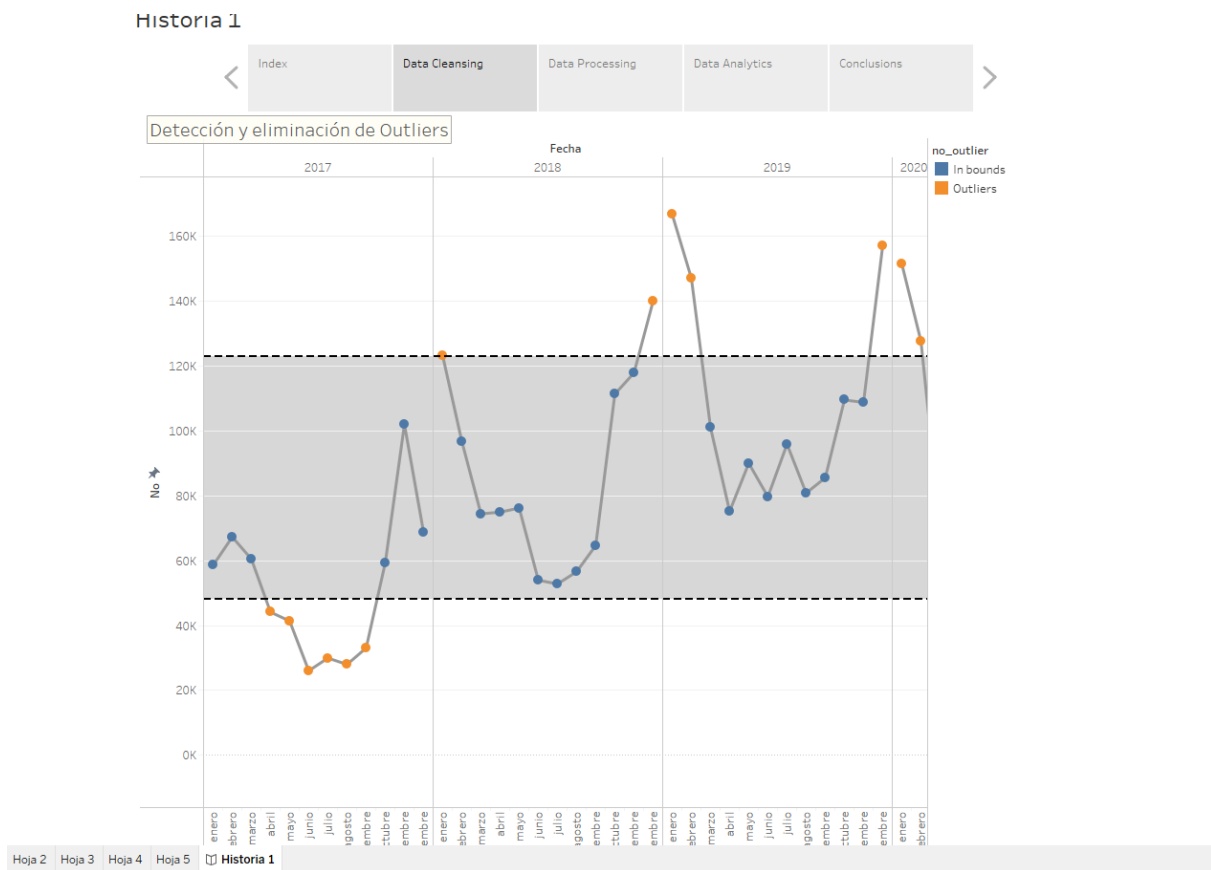
[CS14] Capturas de pantalla del pronóstico: (0,25 puntos)



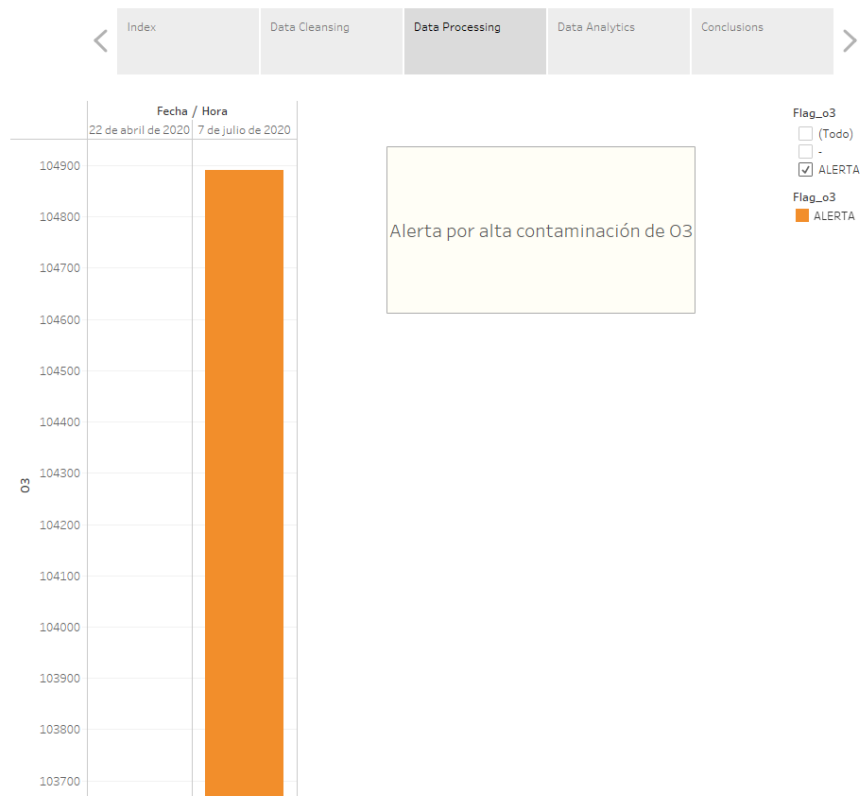
Ampliamos el pronóstico 5 años. Con los años se observa que la tendencia es a aumentar los valores de NO.



20. Añade los gráficos anteriores a la story.

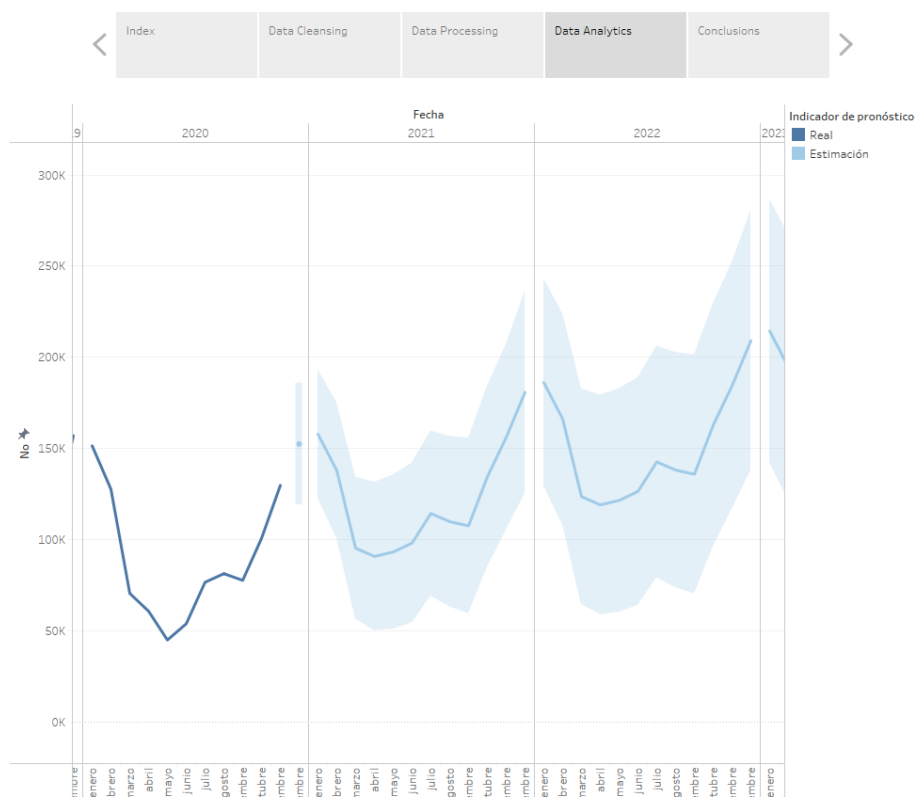


Historia 1



Hoja 2 Hoja 3 Hoja 4 Hoja 5 **Historia 1**

Historia 1



Hoja 2 Hoja 3 Hoja 4 Hoja 5 **Historia 1**

21. Añade una página de inicio y conclusiones propias.

[CS15] Capturas de pantalla de las conclusiones: (0,25 puntos)

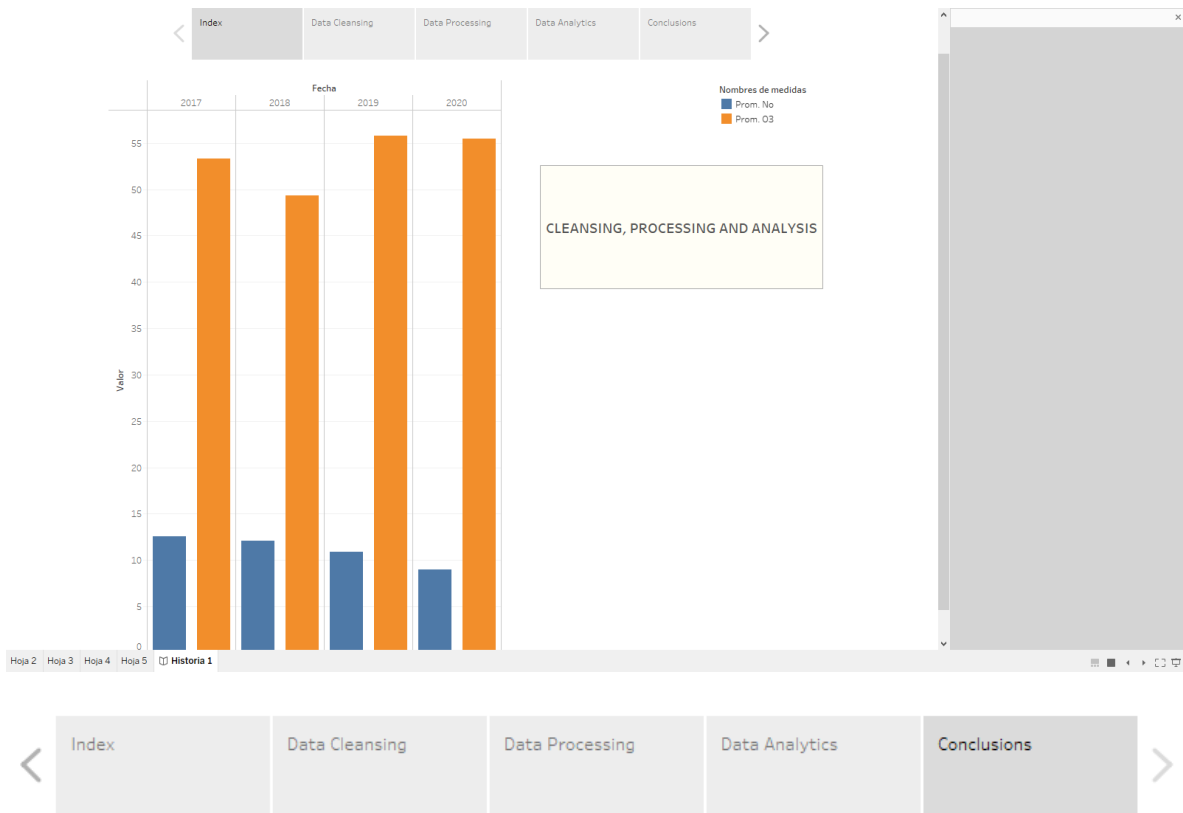


Tableau permite:

1. Eliminar valores nulos y outliers.
2. Procesar los datos creando un campo calculado.
3. Analizar los datos por estacionalidad y mediante líneas de tendencia y forecasting.