

# PROYECTO FINAL MIN

Candela Espinosa





# Obtención de los datos

Al principio intentamos obtener los datos de una API que se proporcionaba desde la [página de datos abiertos de la Comunidad de Madrid](https://datos.comunidad.madrid/catalogo/api/action/datastore_search?id=6f9f3bb1-ec4f-4f89-872d-89b3527c2b31):

[https://datos.comunidad.madrid/catalogo/api/action/datastore\\_search?id=6f9f3bb1-ec4f-4f89-872d-89b3527c2b31](https://datos.comunidad.madrid/catalogo/api/action/datastore_search?id=6f9f3bb1-ec4f-4f89-872d-89b3527c2b31)

	provincia	municipio	estacion	magnitud	punto_muestreo	ano	mes	dia	h01	v01	...	h20	v20	h21	v21	h22	v22	h23	v23	h24	v24
0	28	102	1	1	28102001_1_38	2019	1	1	1,0	V	...	1,0	V	2,0	V	2,0	V	2,0	V	2,0	V
1	28	102	1	6	28102001_6_48	2019	1	1	0,2	V	...	0,2	V	0,2	V	0,2	V	0,2	V	0,2	V
2	28	102	1	7	28102001_7_8	2019	1	1	1,0	V	...	1,0	V	1,0	V	1,0	V	1,0	V	1,0	V
3	28	102	1	8	28102001_8_8	2019	1	1	9,0	V	...	7,0	V	9,0	V	10,0	V	16,0	V	17,0	V
4	28	102	1	10	28102001_10_49	2019	1	1	13,0	V	...	19,0	V	15,0	V	15,0	V	13,0	V	15,0	V

Había un problema, no teníamos los datos para el municipio 79 (Madrid Centro)

Tuvimos que sacarlos de otra [página de la Comunidad](#), pero esta vez los datos estaban más limpios, en formato csv.

	PROVINCIA	MUNICIPIO	ESTACION	MAGNITUD	PUNTO_MUESTREO	ANO	MES	D01	V01	D02	...	D27	V27	D28	V28	D29	V29	D30	V30	D31	V31
0	28	79	4	1	28079004_1_38	2019	1	18.0	V	20.0	...	15.0	V	15.0	V	15.0	V	14.0	V	14.0	V
1	28	79	4	1	28079004_1_38	2019	2	13.0	V	13.0	...	19.0	V	20.0	V	0.0	N	0.0	N	0.0	N
2	28	79	4	1	28079004_1_38	2019	3	18.0	V	18.0	...	6.0	V	7.0	V	6.0	V	4.0	V	4.0	V
3	28	79	4	1	28079004_1_38	2019	4	3.0	V	4.0	...	3.0	V	2.0	V	2.0	V	2.0	V	0.0	N
4	28	79	4	1	28079004_1_38	2019	5	1.0	V	1.0	...	2.0	V	1.0	V	1.0	V	1.0	V	2.0	V

Resulta que nos faltaba el mes de Diciembre de 2022 que por alguna razón no habían querido meter ya que los datos sí que se encontraban en un fichero que era por horas en lugar de la media por días.

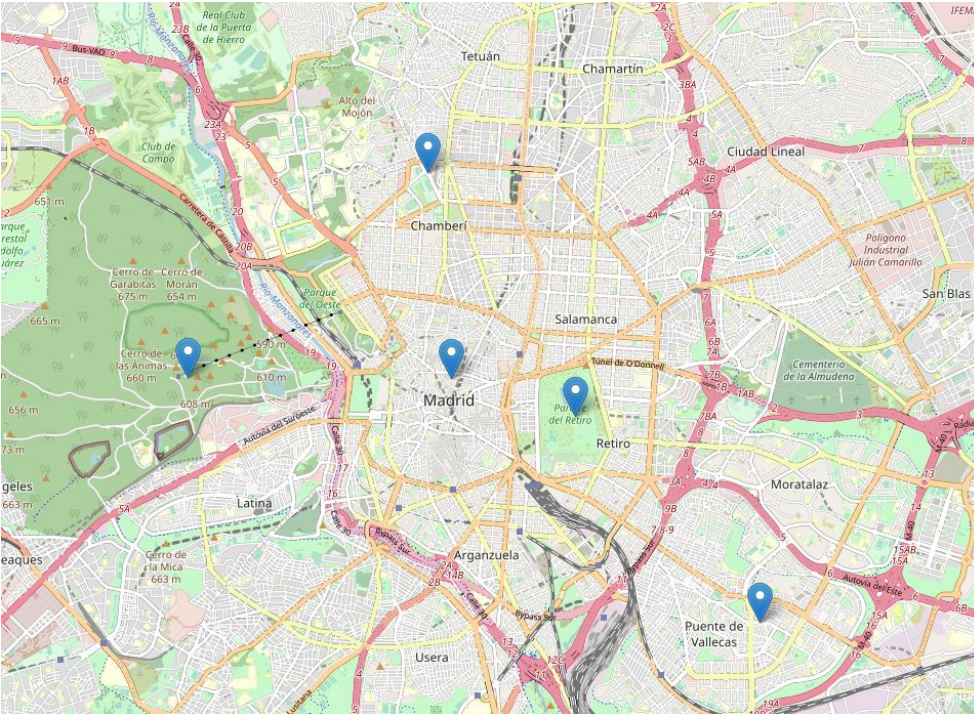


```
df_all['MEDIA'] = None
```

```
for i, row in df_all.iterrows():
    values = []
    for j in range(1, 32):
        d_col = f'D{j:02d}'
        v_col = f'V{j:02d}'
        d_val = row[d_col]
        v_val = row[v_col]
        if v_val == 'V':
            values.append(d_val)
    if len(values) > 0:
        media = sum(values) / len(values)
        df_all.at[i, 'MEDIA'] = media
```

```
df_med = pd.DataFrame({'ANO': df_all['ANO'], 'MES': df_all['MES'], 'ESTACION': df_all['ESTACION'], 'MAGNITUD': df_all['MAGNITUD'], 'MEDIA': df_all['MEDIA']})
```

# Datos a analizar



Magnitudes vamos a analizar:

- Dióxido de Azufre
- PM<sub>2,5</sub>
- PM<sub>10</sub>
- Dióxido de Nitrógeno
- Ozono

¿COVID19?

¿Madrid Central?

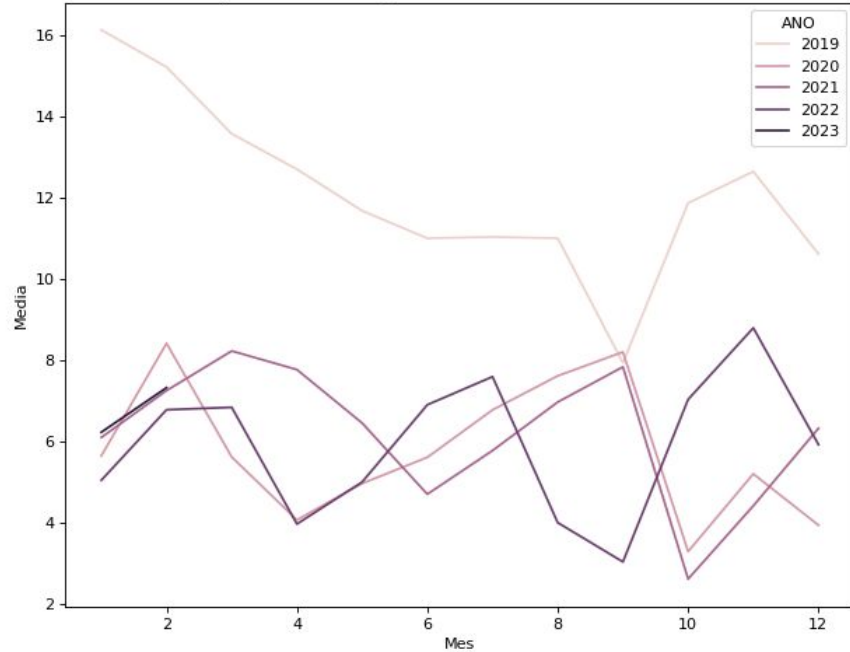


# Dióxido de Azufre

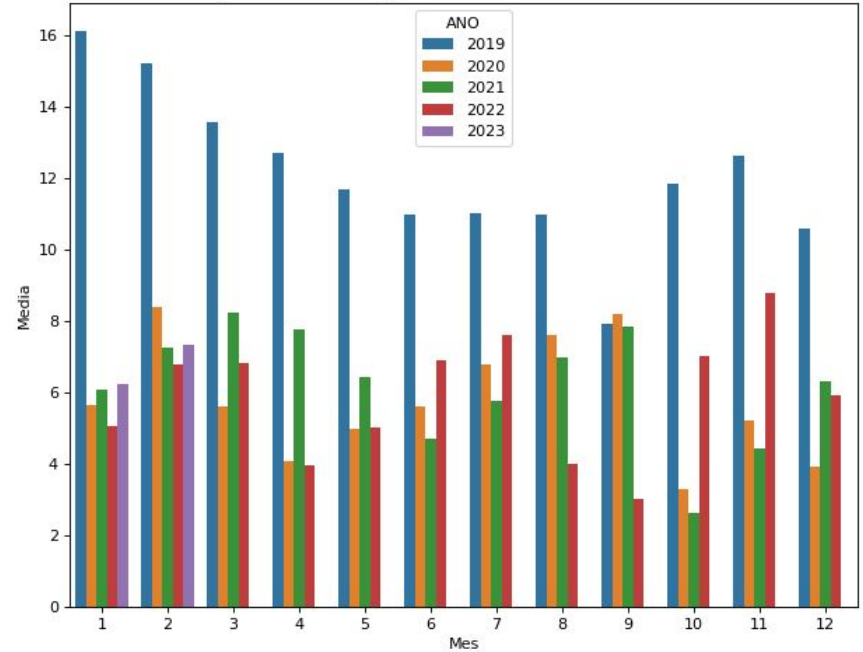
- El dióxido de azufre se produce debido a la combustión de combustibles fósiles que contienen azufre, como el carbón y el petróleo.
- Los riesgos para la salud incluyen irritación de las vías respiratorias, dificultad para respirar, tos, sibilancias... Además, la exposición a largo plazo al dióxido de azufre puede contribuir al desarrollo de enfermedades respiratorias crónicas.
- El valor límite permitido por la OMS es de 20 microgramos por metro cúbico ( $\mu\text{g}/\text{m}^3$ ) como media anual.

# Estación de Plaza del Carmen (Sol)

Media por año de la magnitud 1 en la estación Plaza del Carmen



Media por año de la magnitud 1 en la estación Plaza del Carmen





# Dióxido de Nitrógeno

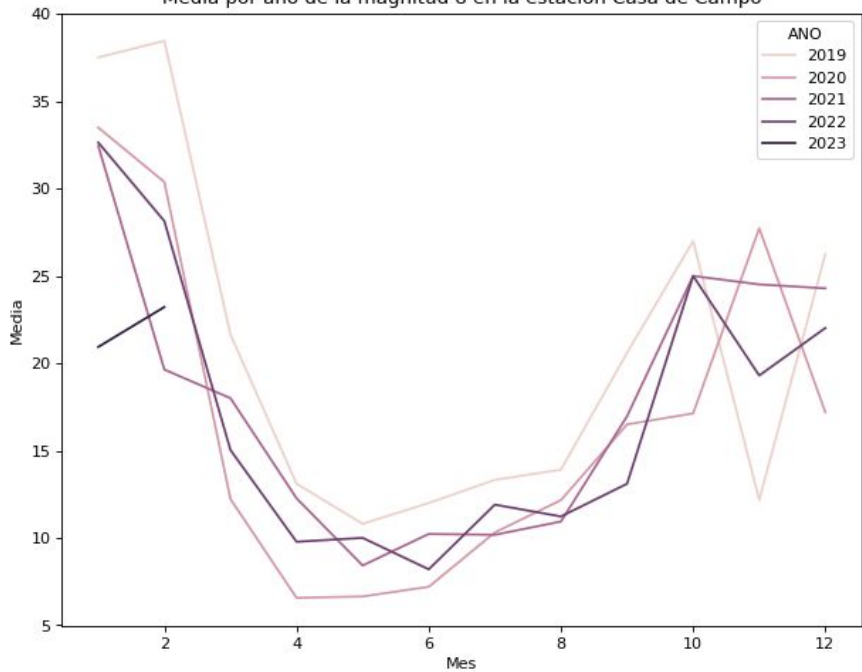
- El dióxido de nitrógeno se produce debido a las emisiones de vehículos motorizados, procesos industriales y la combustión de combustibles fósiles.
- Los **riesgos para la salud** incluyen irritación de las vías respiratorias, aumento de condiciones respiratorias preexistentes como el asma, y aumento del riesgo de infecciones respiratorias.
- La Organización Mundial de la Salud (OMS) ha establecido un valor límite de 40 microgramos por metro cúbico ( $\mu\text{g}/\text{m}^3$ ) como concentración media anual.



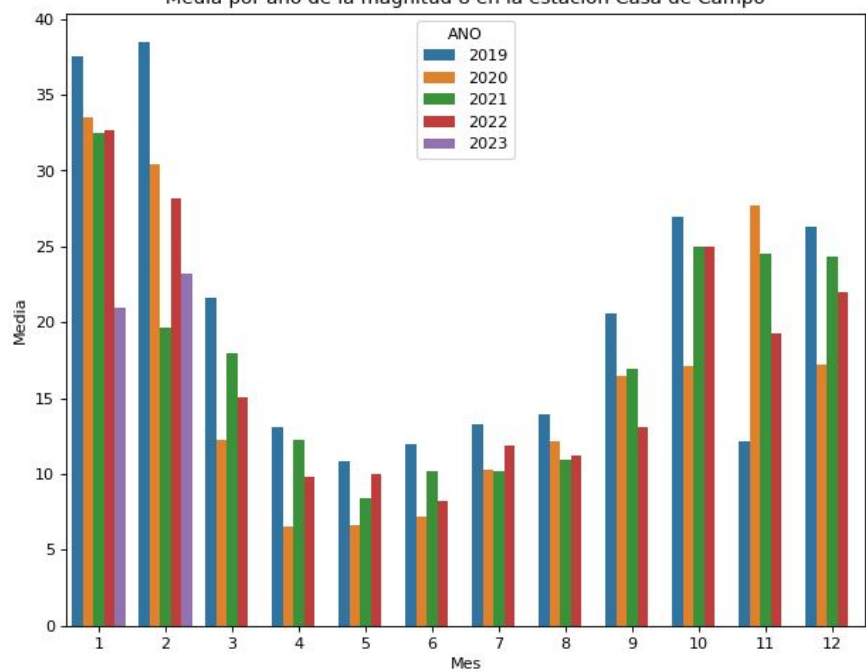


# Estación de Casa de Campo

Media por año de la magnitud 8 en la estación Casa de Campo



Media por año de la magnitud 8 en la estación Casa de Campo





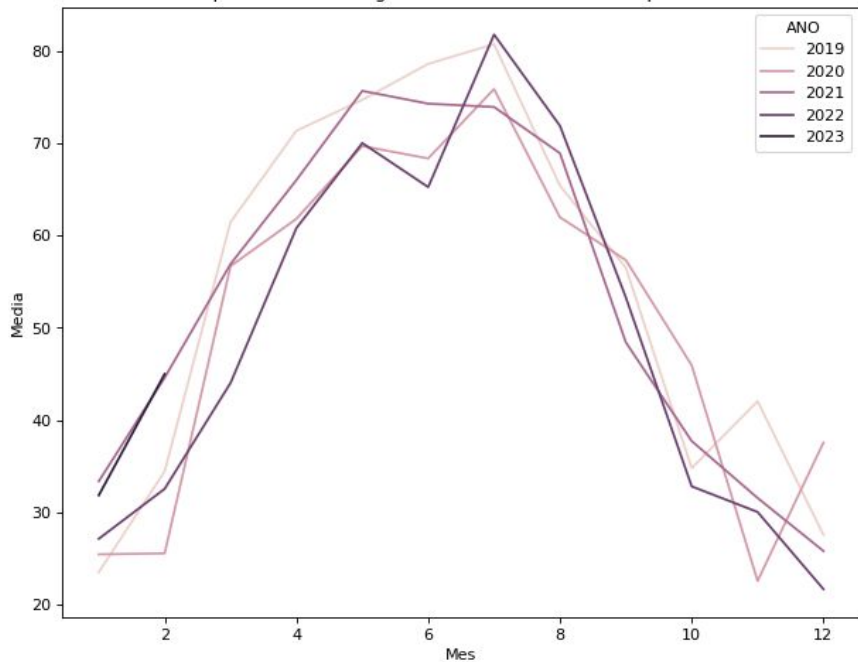
# Ozono

- El ozono **troposférico** se forma como resultado de reacciones químicas, en presencia de la luz solar, a partir de los contaminantes emitidos por automóviles, centrales térmicas, refinerías, procesos industriales diversos etc.
- Irritación de las vías respiratorias, tos, dificultad para respirar. Además, el ozono troposférico puede afectar negativamente los cultivos, los bosques y otros ecosistemas, y contribuir al cambio climático.
- La Organización Mundial de la Salud (OMS) ha establecido un valor objetivo de 100 microgramos por metro cúbico ( $\mu\text{g}/\text{m}^3$ ).

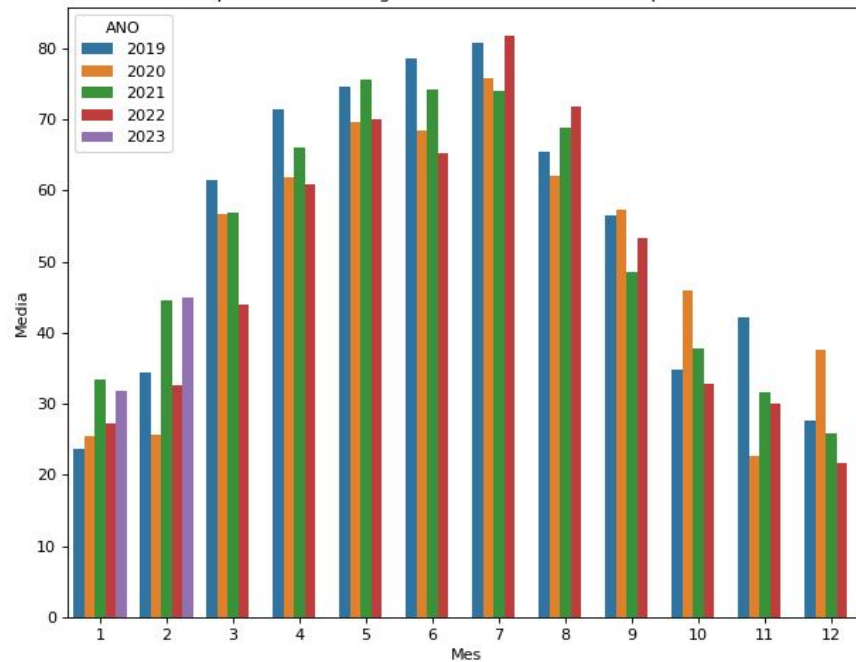


# Estación del Retiro

Media por año de la magnitud 14 en la estación Parque del Retiro



Media por año de la magnitud 14 en la estación Parque del Retiro





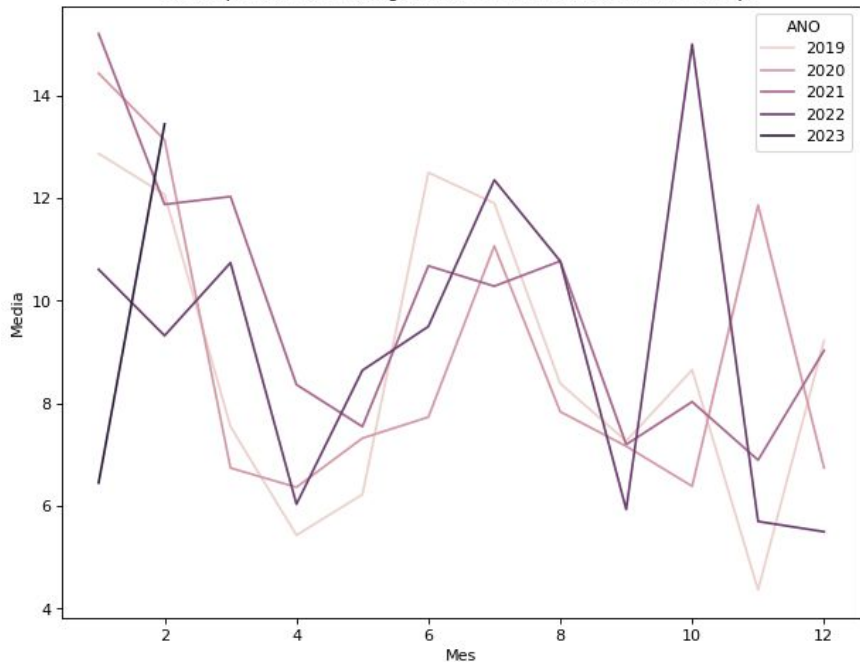
# PM<sub>2,5</sub>

- PM<sub>2,5</sub>, llamadas así por su tamaño **inferior** a 2,5 micras.
- Proceden tanto de **fuentes naturales** (tormentas de arena, erupciones volcánicas, incendios forestales, etc.) como de la **actividad humana** (tráfico, especialmente vehículos diésel, incineradoras, calefacciones de carbón, minería, procesos industriales, etc.).
- La OMS ha recomendado un valor límite de 10 microgramos por metro cúbico ( $\mu\text{g}/\text{m}^3$ ) anual.

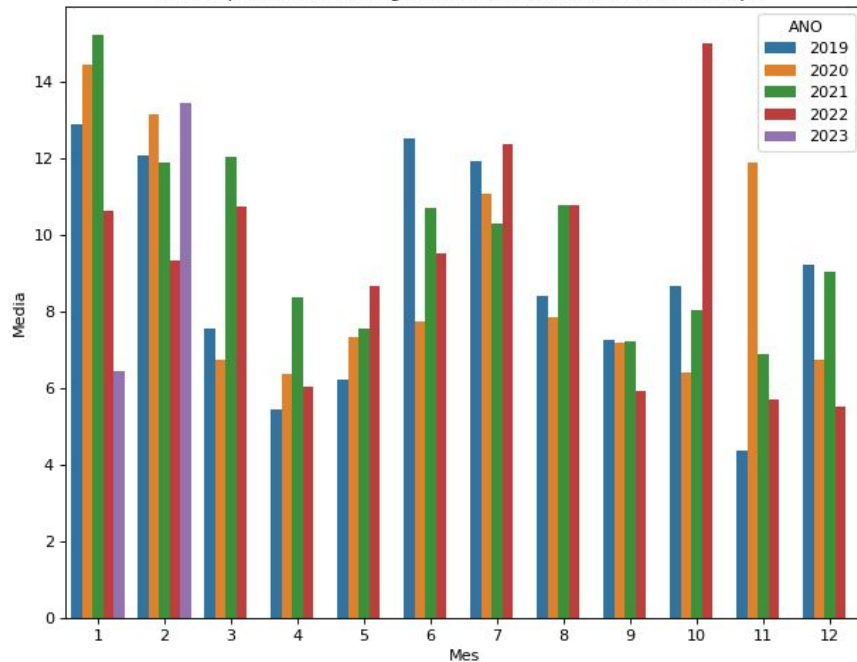


# Estación de Casa de Campo

Media por año de la magnitud 9 en la estación Casa de Campo



Media por año de la magnitud 9 en la estación Casa de Campo





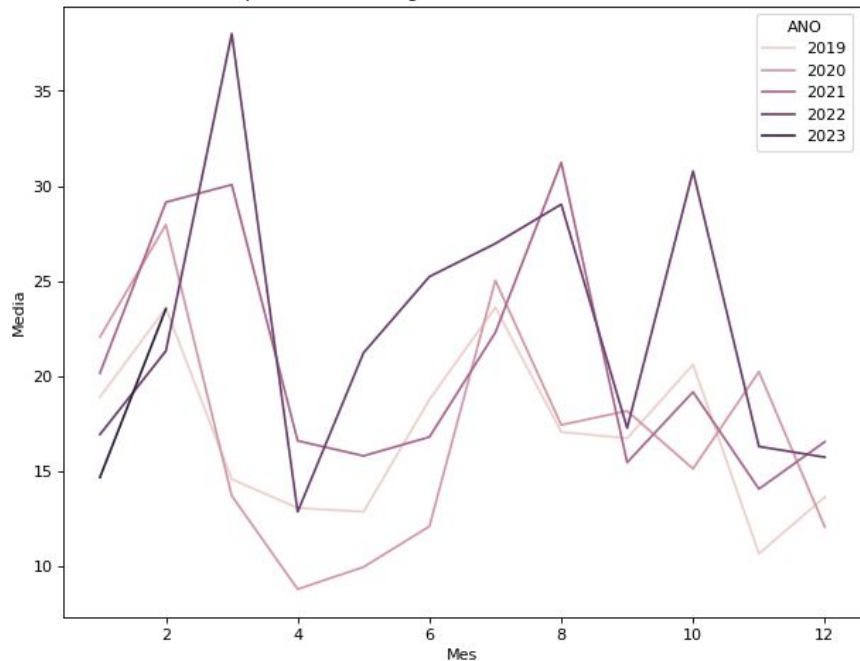
# PM10

- PM10, con un tamaño inferior a 10 micras, suelen tener mayor componente natural. Resultan menos perjudiciales para la salud que las PM2,5, pero no son inocuas.
- Proceden tanto de **fuentes naturales** (tormentas de arena, erupciones volcánicas, incendios forestales, etc.) como de la actividad humana (tráfico, especialmente vehículos diésel, incineradoras, calefacciones de carbón, minería, procesos industriales, etc.).
- La OMS ha recomendado un valor límite de 20 microgramos por metro cúbico ( $\mu\text{g}/\text{m}^3$ ) anual.

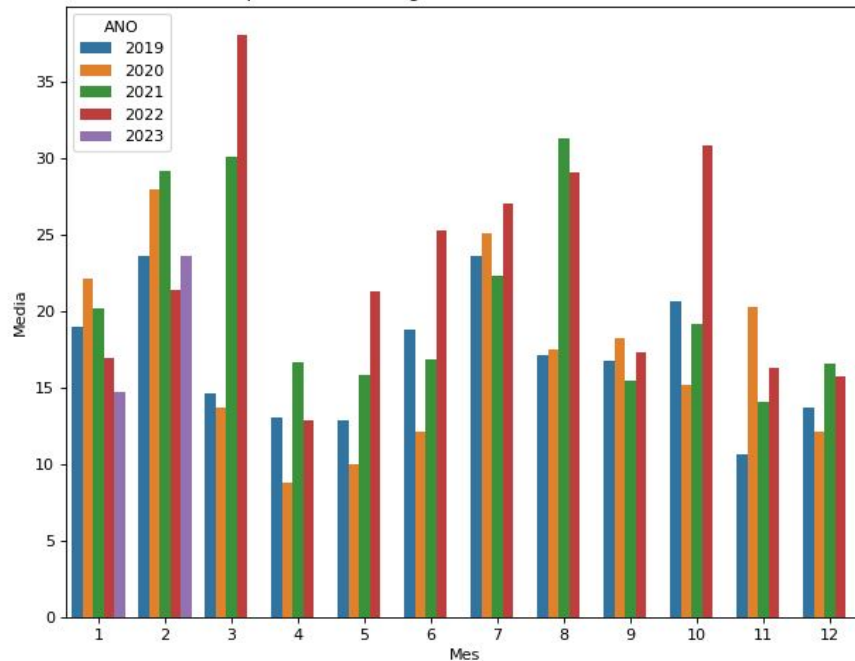


# Estación de Vallecas

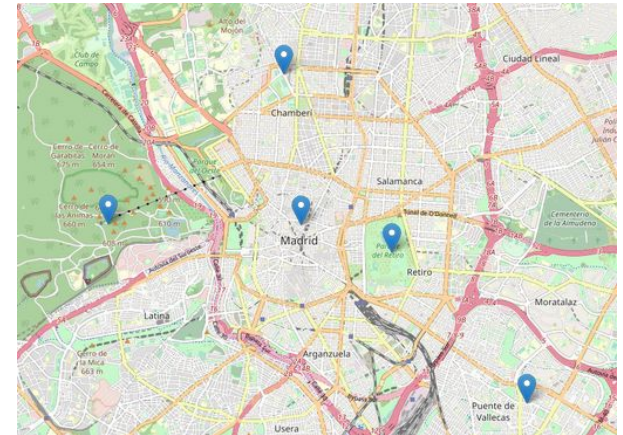
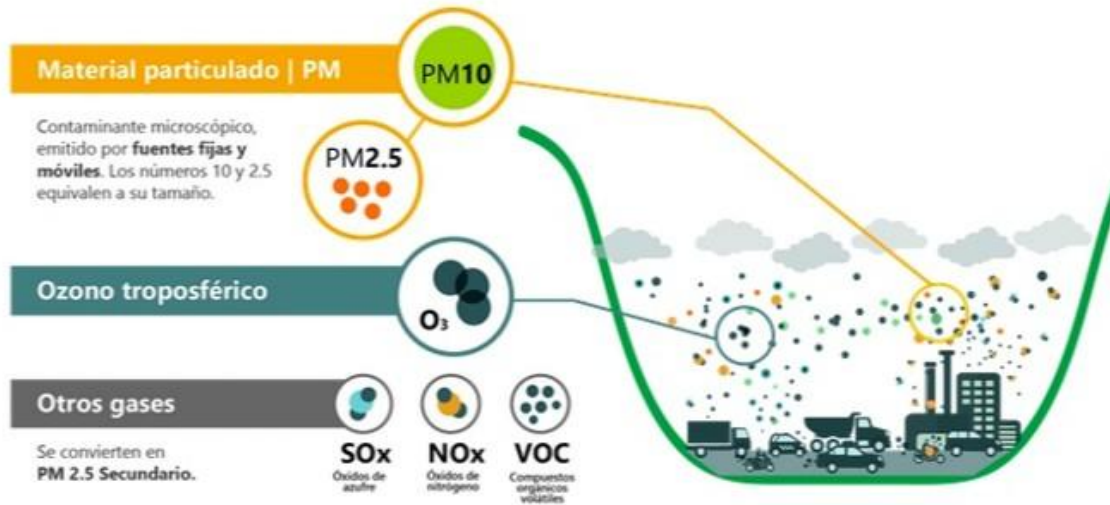
Media por año de la magnitud 10 en la estación Castellana



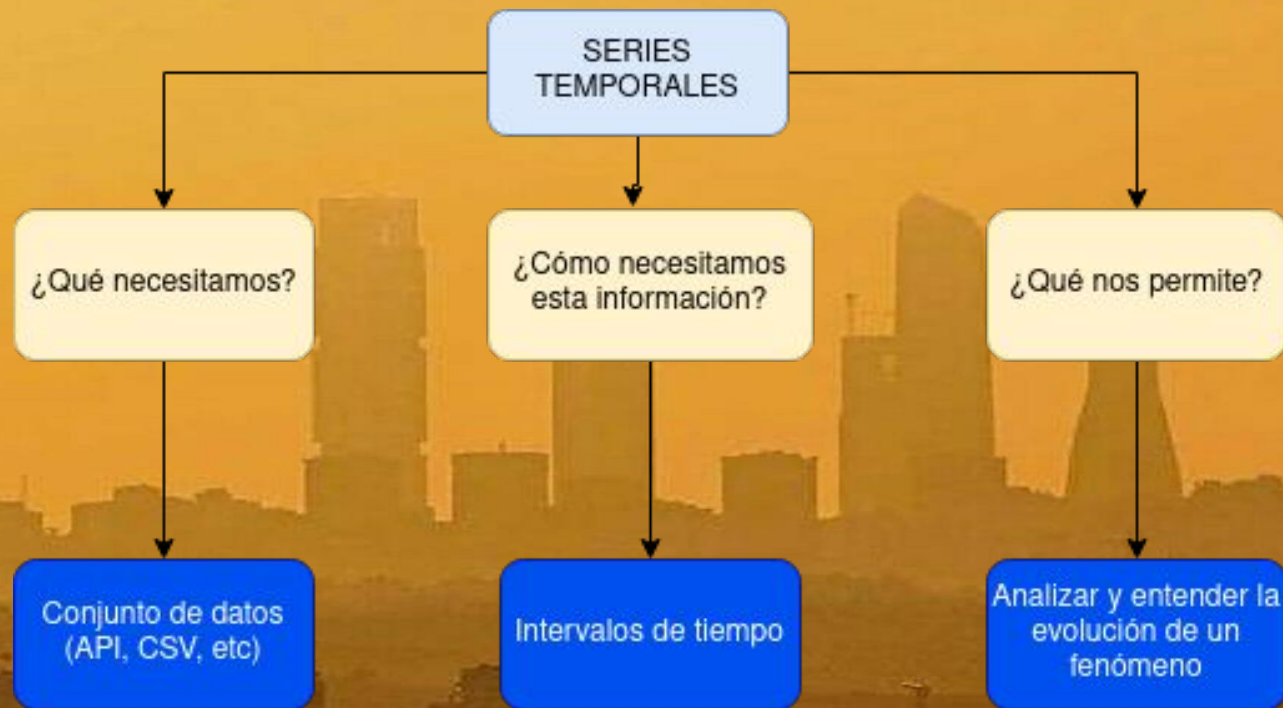
Media por año de la magnitud 10 en la estación Castellana

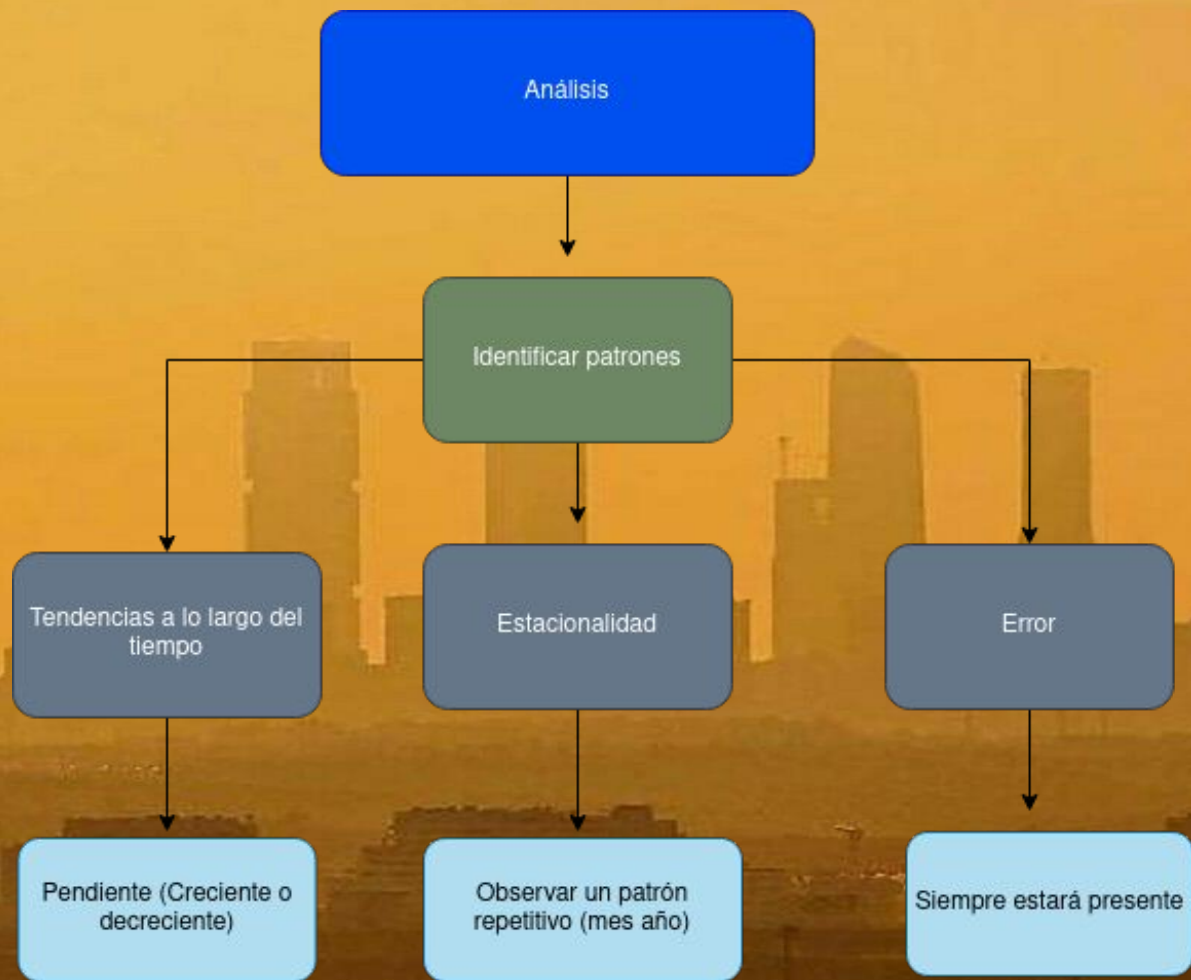


# Series temporales

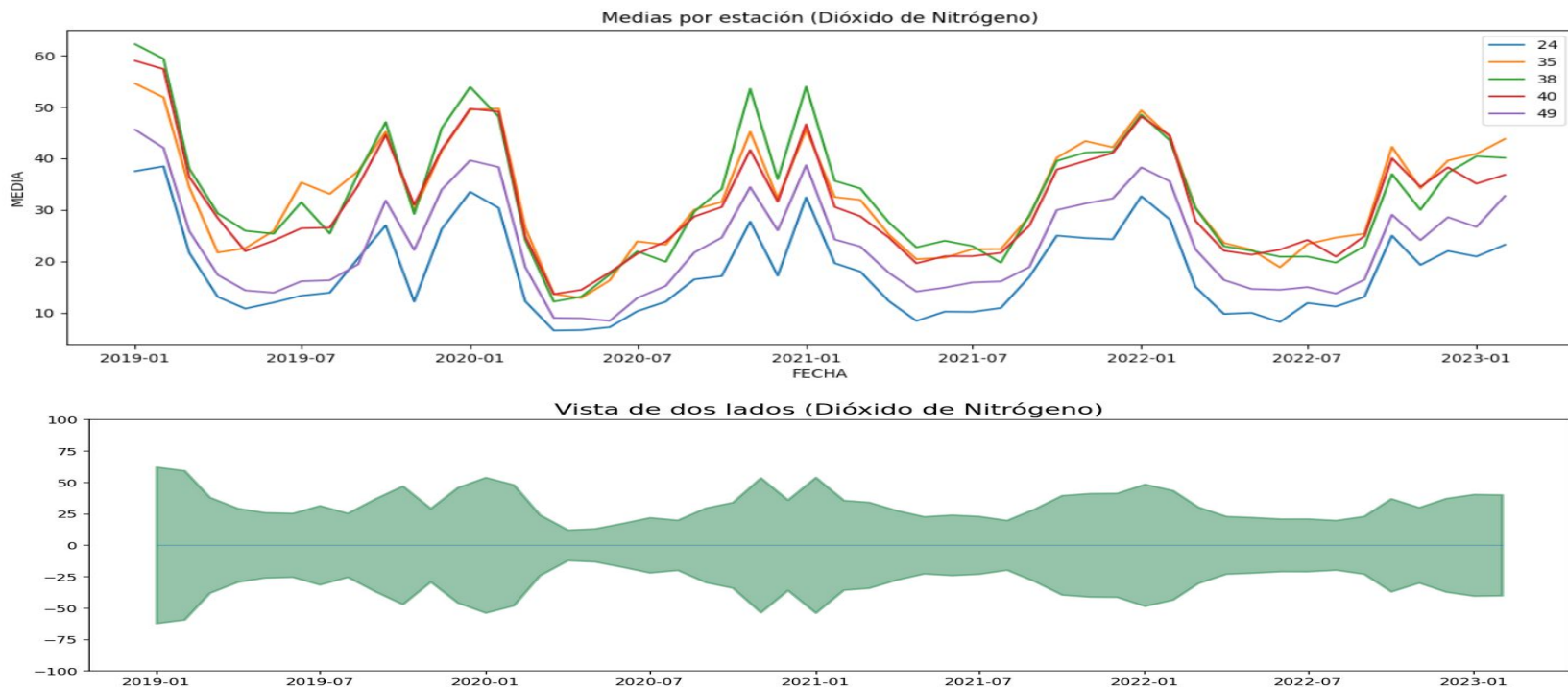


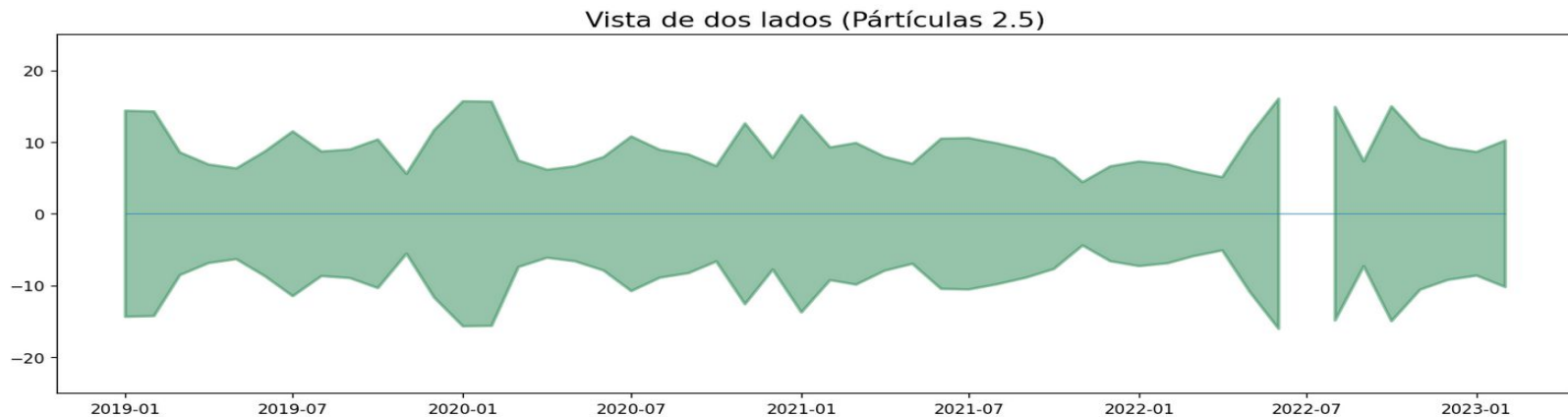
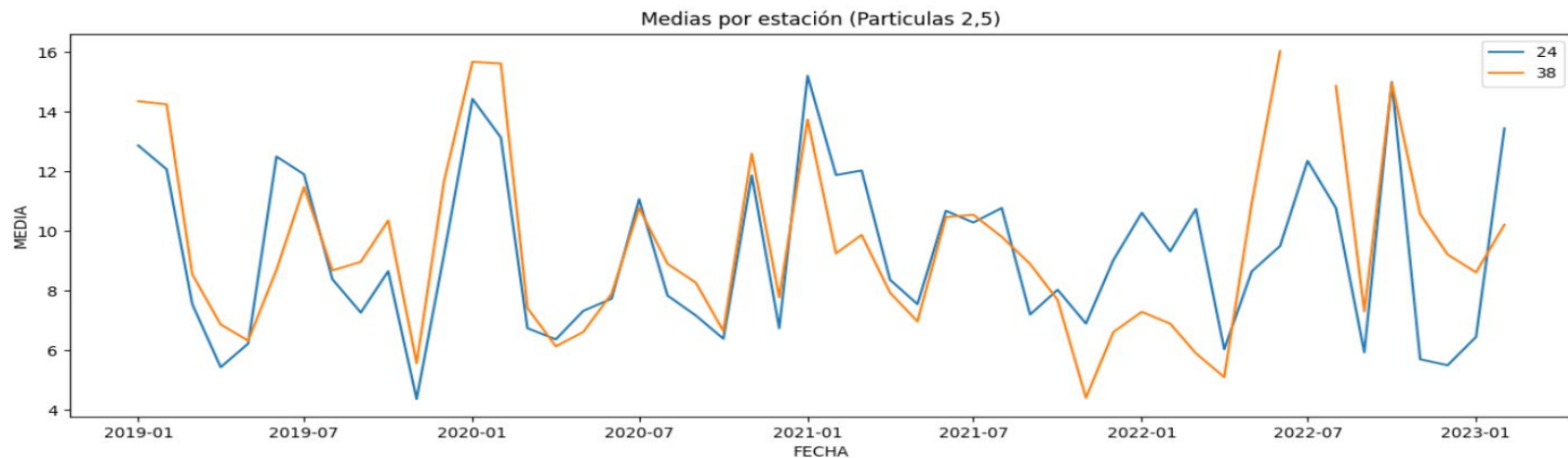


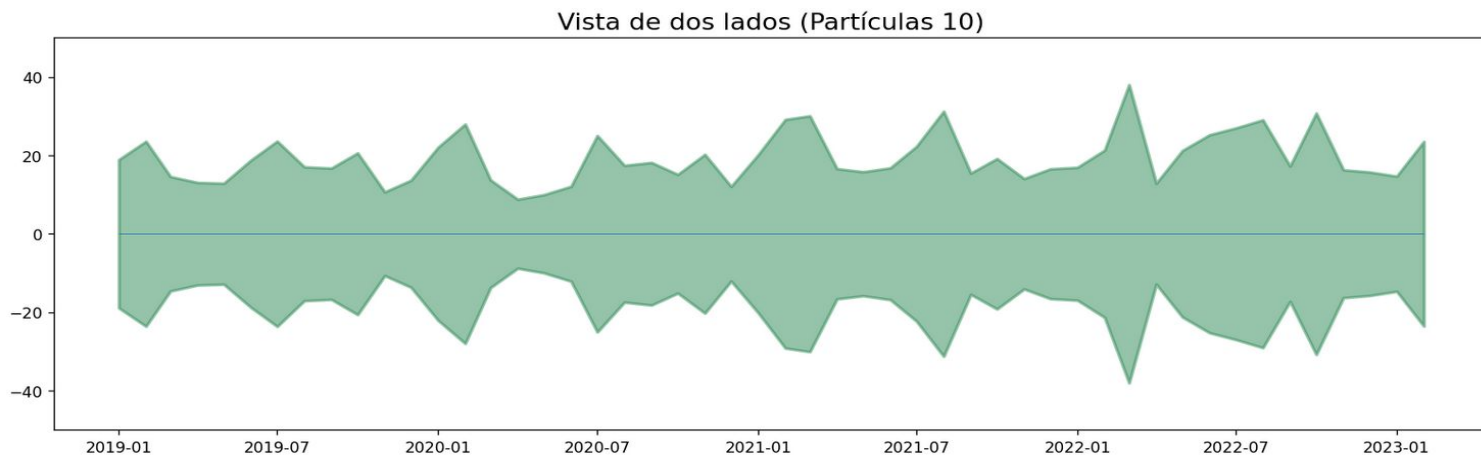
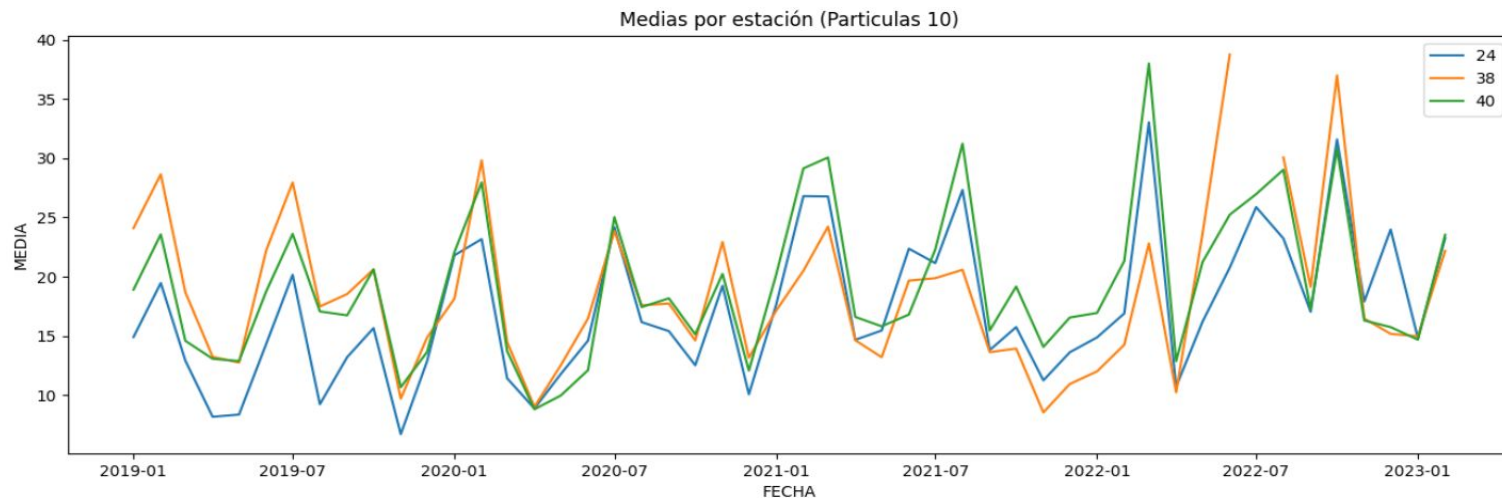




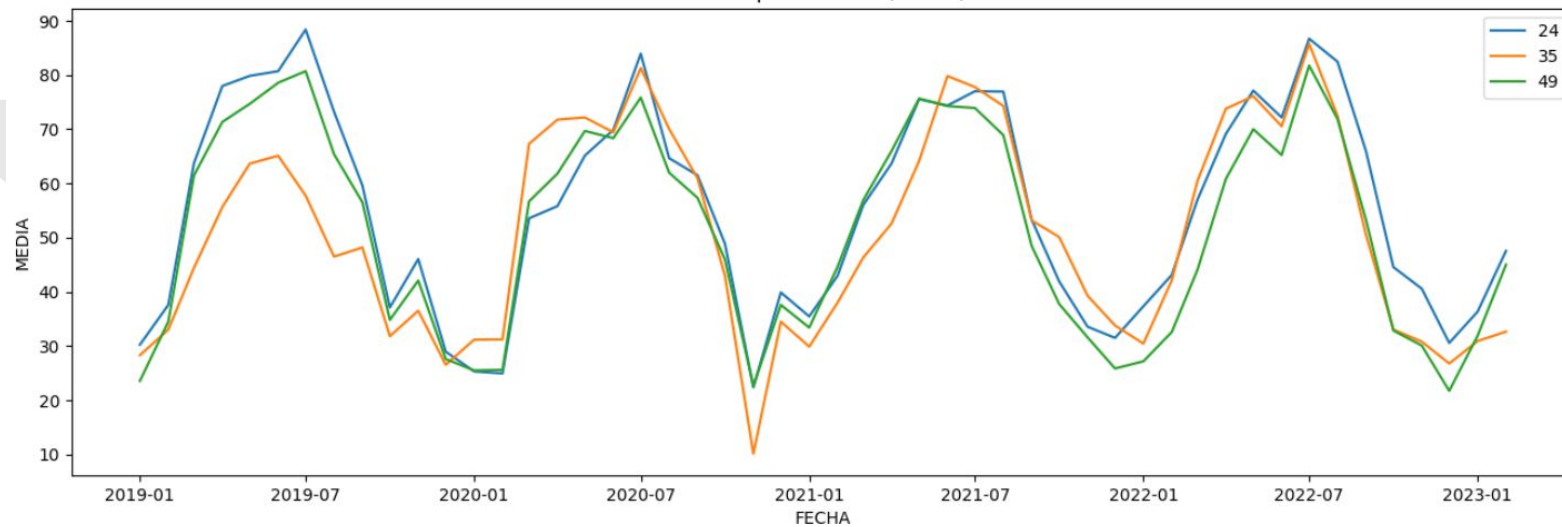
# Observación de los datos



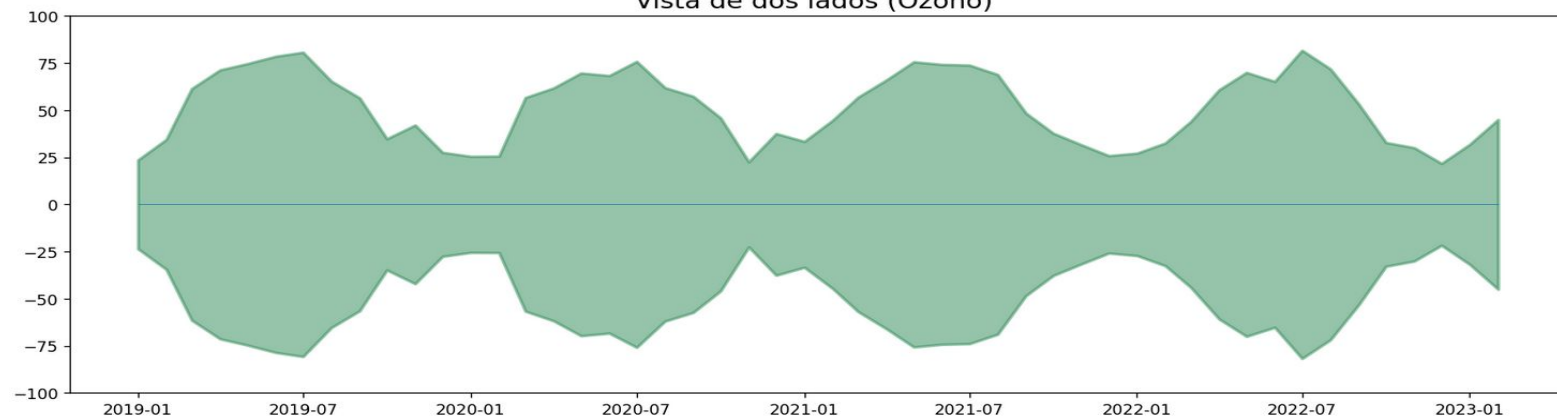




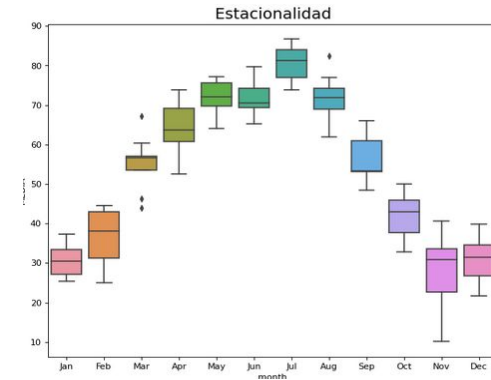
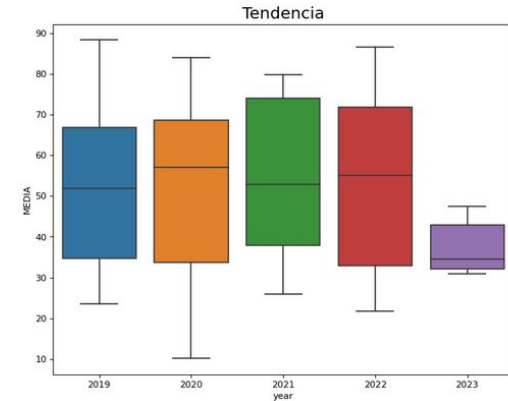
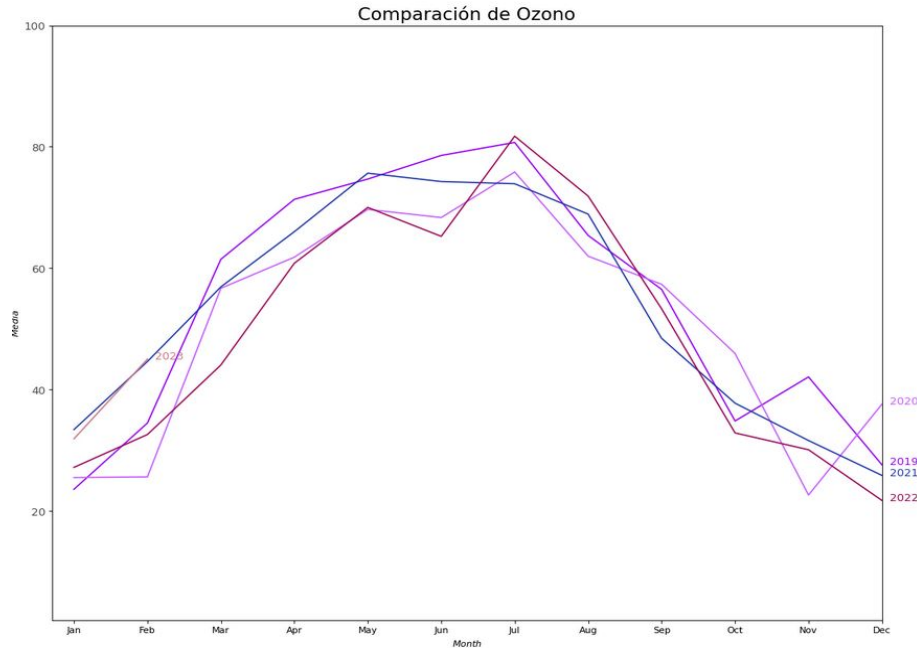
Medias por estación (Ozono)



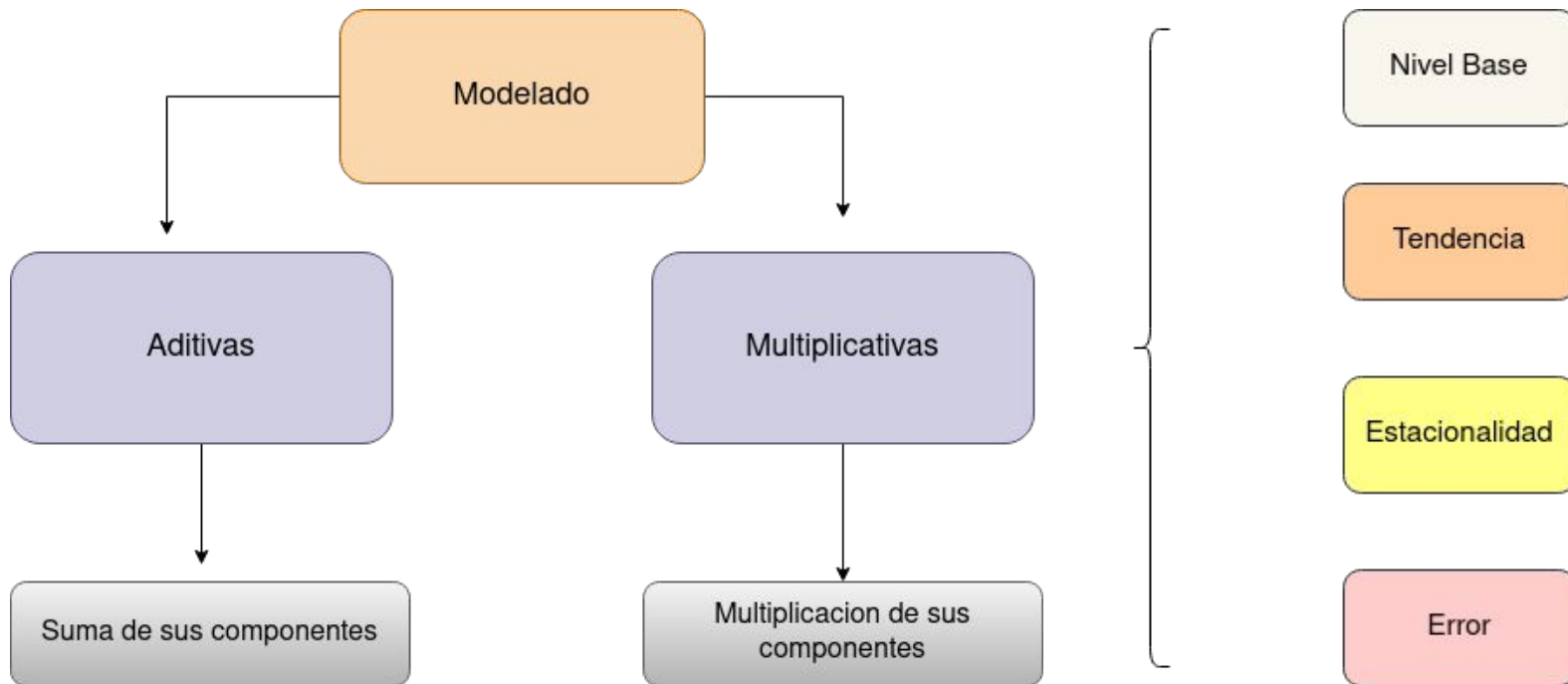
Vista de dos lados (Ozono)



# Comprobar que la magnitud ozono sigue una serie temporal (hipótesis)



# Modelado de la serie

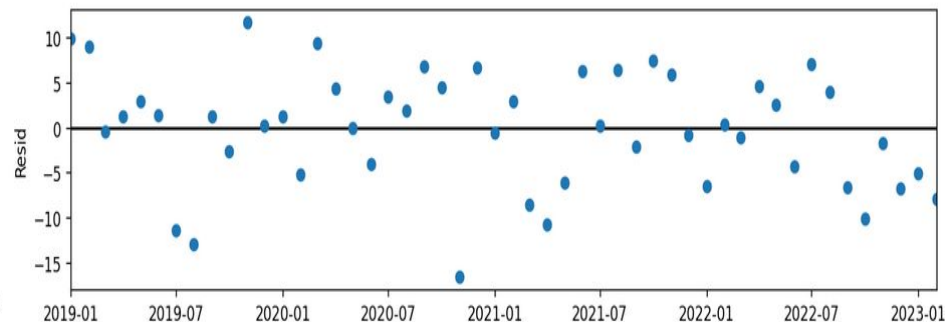
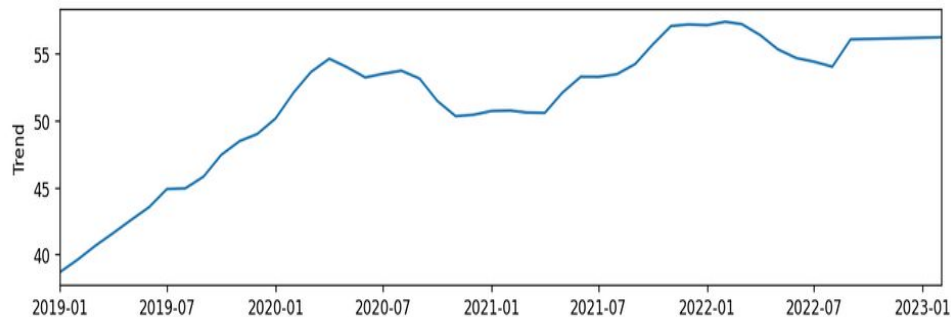
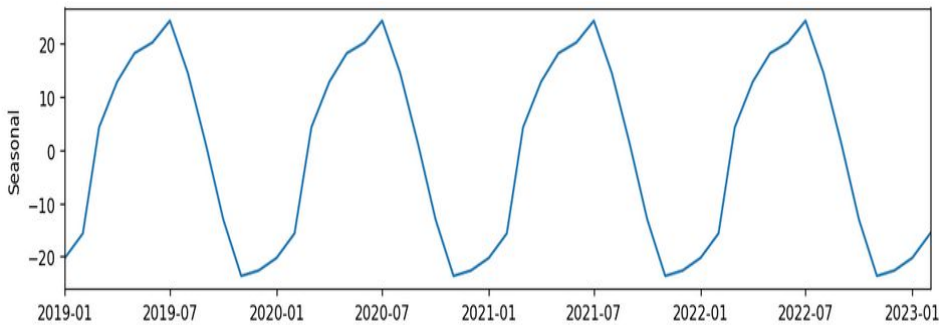
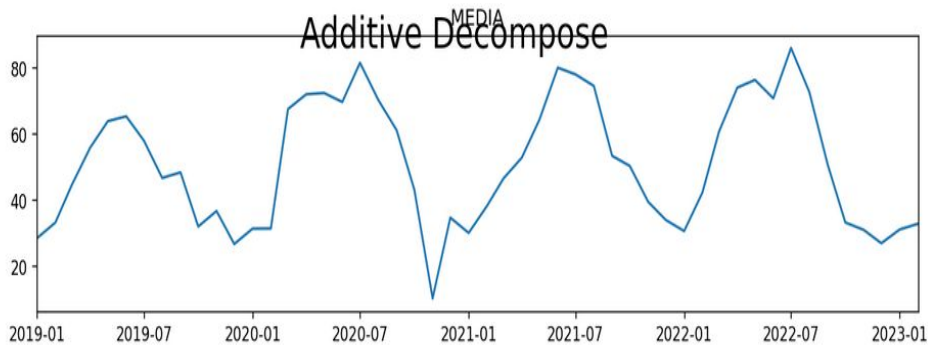






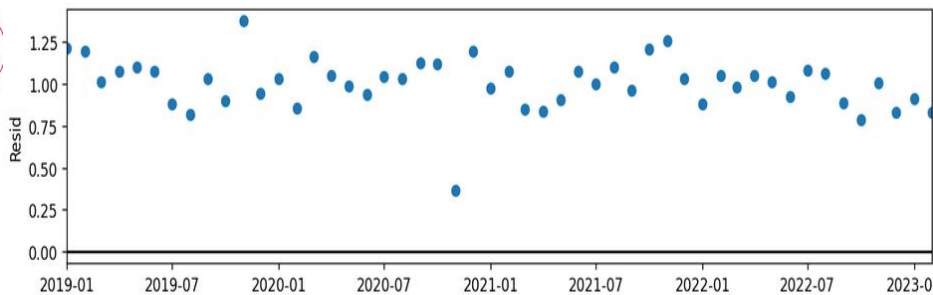
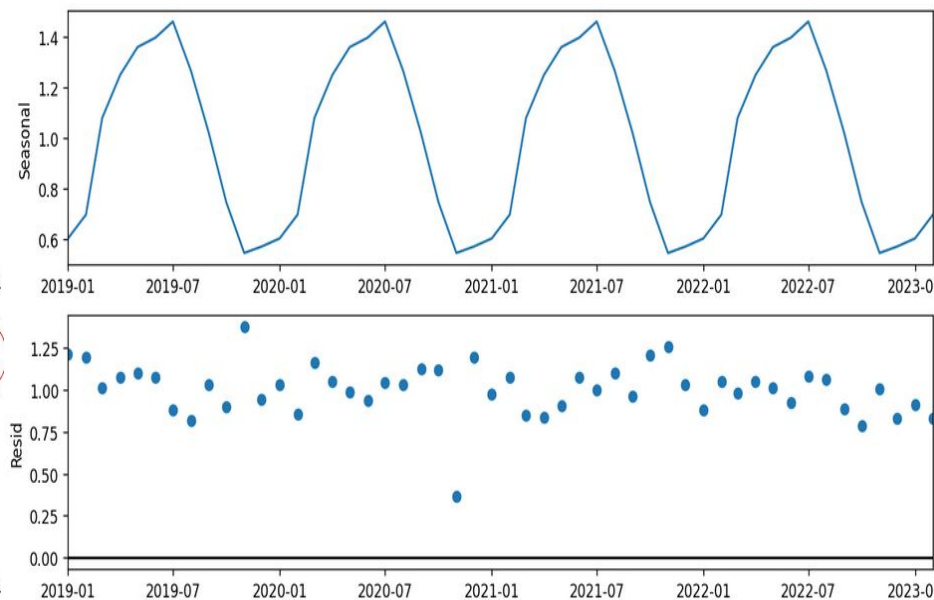
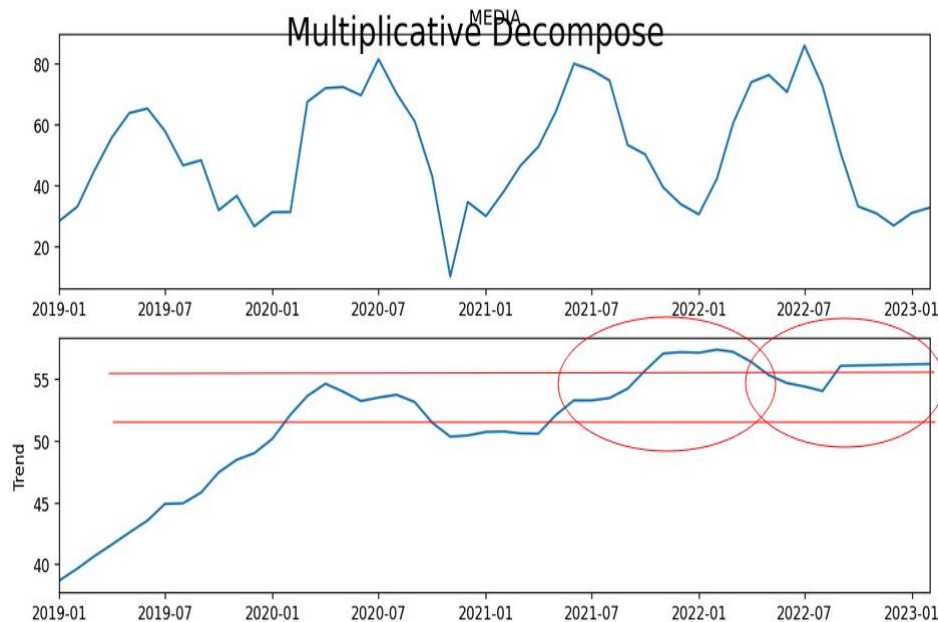
# Modelado de la serie (Aditiva)

Additive Decompose





# Modelado de la serie (Multiplicativa)



# Comprobar la estacionalidad

```
from statsmodels.tsa.stattools import adfuller, kpss

# ADF Test
result = adfuller(df.MEDIA.values, autolag='AIC')
print(f'ADF Statistic: {result[0]}')
print(f'p-value: {result[1]}')
for key, value in result[4].items():
    print('Critical Values:')
    print(f'...{key}, {value}')

# KPSS Test
result = kpss(df.MEDIA.values, regression='c')
print(f'\nKPSS Statistic: {result[0]}')
print(f'p-value: {result[1]}')
for key, value in result[3].items():
    print('Critical Values:')
    print(f'    {key}, {value}')
```

ADF -> valor P (0.05)

KPPS -> valor P (0.05)

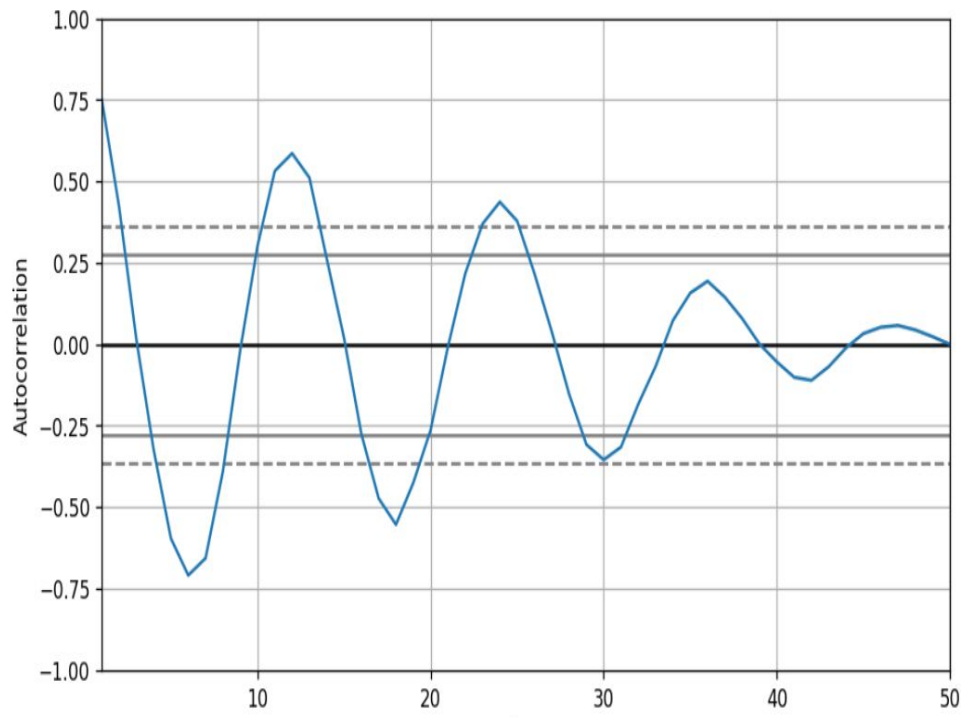
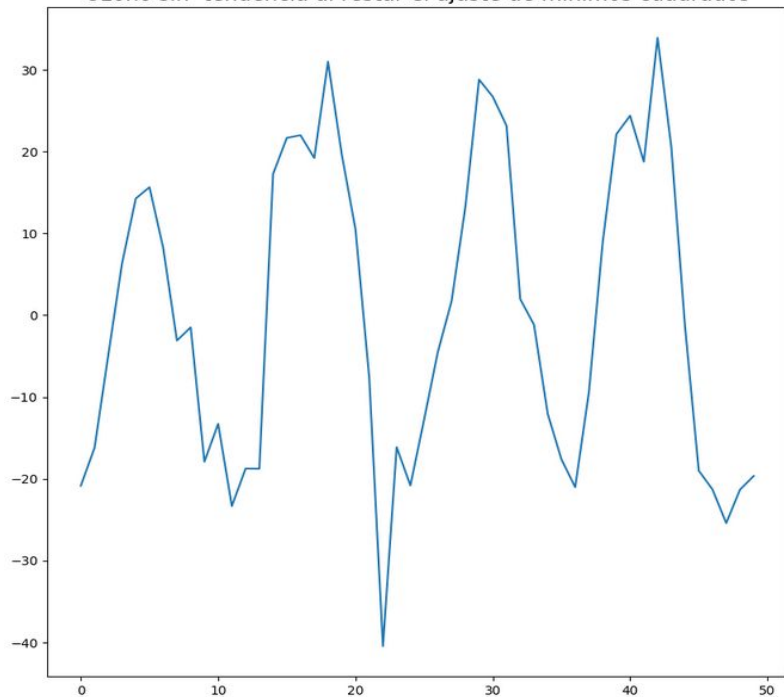
```
KPSS Statistic: 0.055682
p-value: 0.100000
Critical Values:
  10%, 0.347
Critical Values:
   5%, 0.463
Critical Values:
  2.5%, 0.574
Critical Values:
   1%, 0.739
```

En la prueba KPSS, la hipótesis nula ( $H_0$ ) es que la serie temporal es estacionaria en torno a una tendencia. Dado que el estadístico KPSS (0.055682) es menor que los valores críticos en todos los niveles de significancia, **En consecuencia, la prueba KPSS sugiere que la serie temporal es estacionaria en torno a una tendencia.**



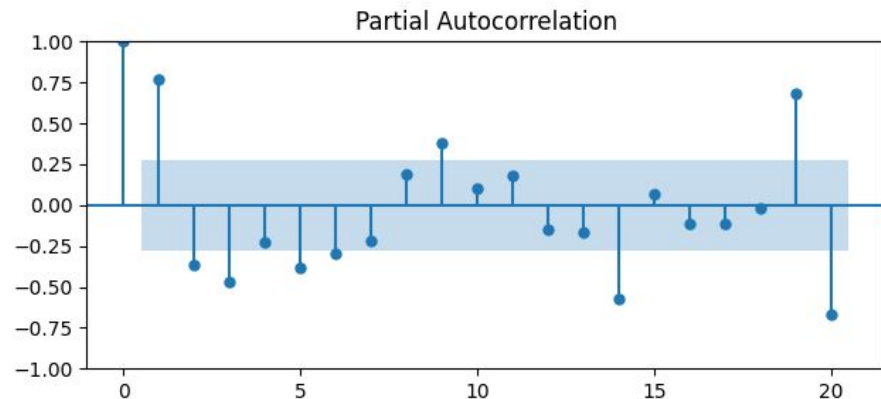
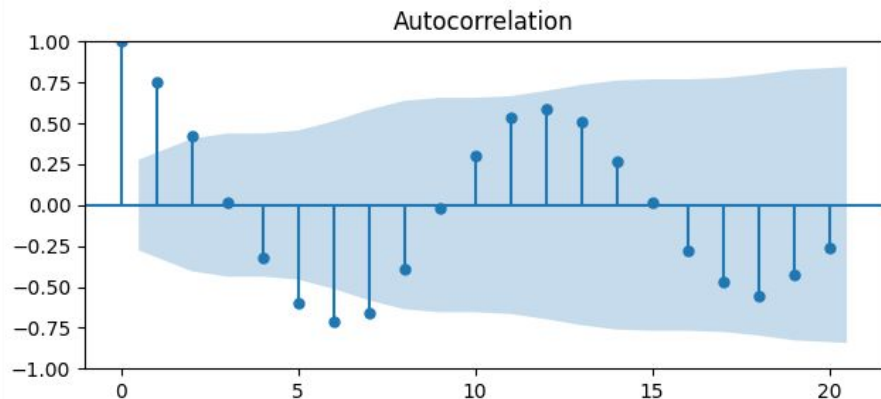
# Quitamos la tendencia y observamos su correlación

Ozono sin tendencia al restar el ajuste de mínimos cuadrados





# Verificación de hipótesis y conclusión



**Tendencias:** Si existe una tendencia en los datos, se observará un decaimiento gradual en la ACF y la PACF.

**Estacionalidad:** La estacionalidad puede identificarse a partir de picos regulares y significativos en la ACF y la PACF en intervalos regulares.

Relaciones autorregresivas (AR): Si una serie temporal tiene una estructura autorregresiva, es probable que la PACF muestre una disminución brusca después de un cierto número de rezagos, mientras que la ACF decaerá de manera más gradual.

Relaciones de promedio móvil (MA): Si una serie temporal tiene una estructura de promedio móvil, es probable que la ACF muestre una disminución brusca después de un cierto número de rezagos, mientras que la PACF decaerá de manera más gradual.

Ruido blanco: Si la serie temporal es ruido blanco, no se observarán correlaciones significativas en la ACF y la PACF.

