In [1]:

```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.tsa.stattools import adfuller
from statsmodels.tsa.holtwinters import SimpleExpSmoothing
from statsmodels.tsa.holtwinters import ExponentialSmoothing
from statsmodels.tsa.ar_model import AutoReg
from statsmodels.graphics.tsaplots import plot_acf
from statsmodels.graphics.tsaplots import plot_pacf
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error
import numpy as np
from pandas.plotting import register_matplotlib_converters
register_matplotlib_converters()
```

In [2]:

```
co2 = pd.read_csv("co2.csv", sep = '\t')
co2.head()
```

Out[2]:

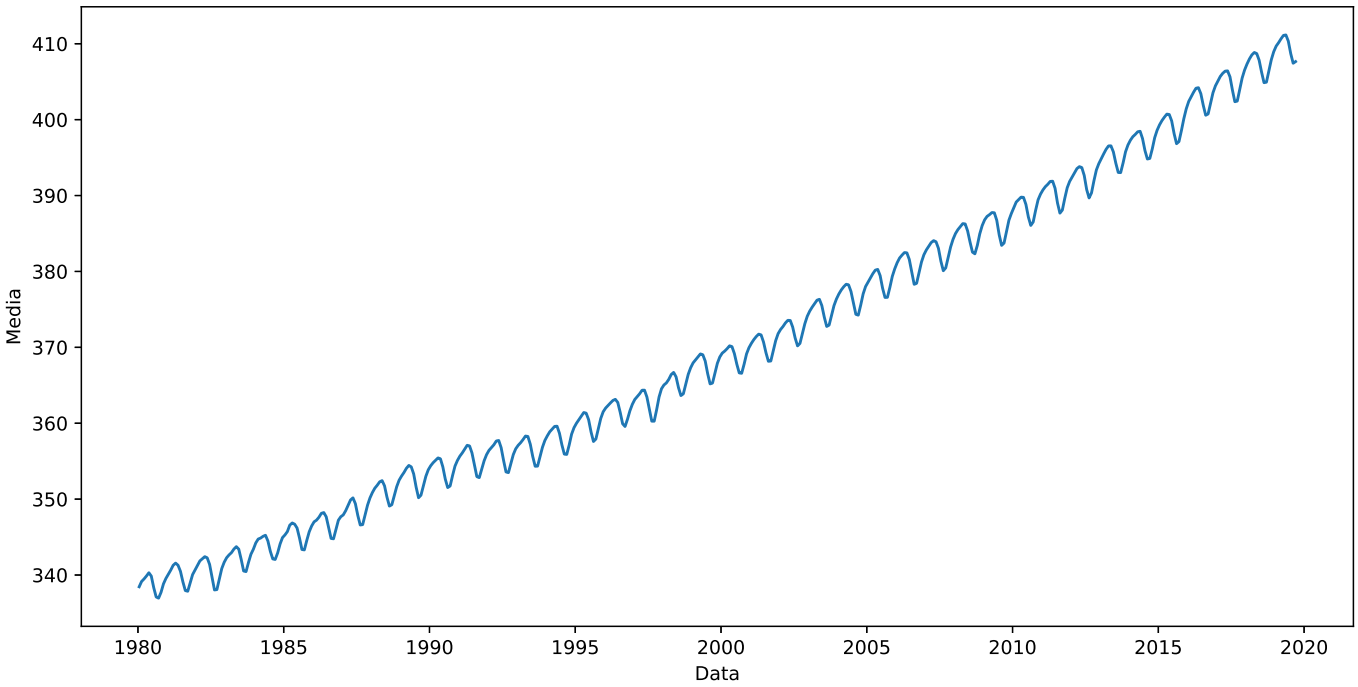
| | data | media |
|---|---|---|
| **0** | 1980.042 | 338.45 |
| **1** | 1980.125 | 339.15 |
| **2** | 1980.208 | 339.48 |
| **3** | 1980.292 | 339.87 |
| **4** | 1980.375 | 340.30 |

In [3]:

```
plt.figure(figsize=(12,6))
plt.plot(co2['data'], co2['media'])
plt.ylabel('Media')
plt.xlabel('Data')
plt
```

Out[3]:

```
<module 'matplotlib.pyplot' from 'c:\\users\\ed\\anaconda3\\envs\\data_science\\lib\\site-packages\\matplotlib\\pyplot.py'>
```




In [4]:

```
nasc = pd.read_csv('nascimentos.csv')
nasc.head()
```

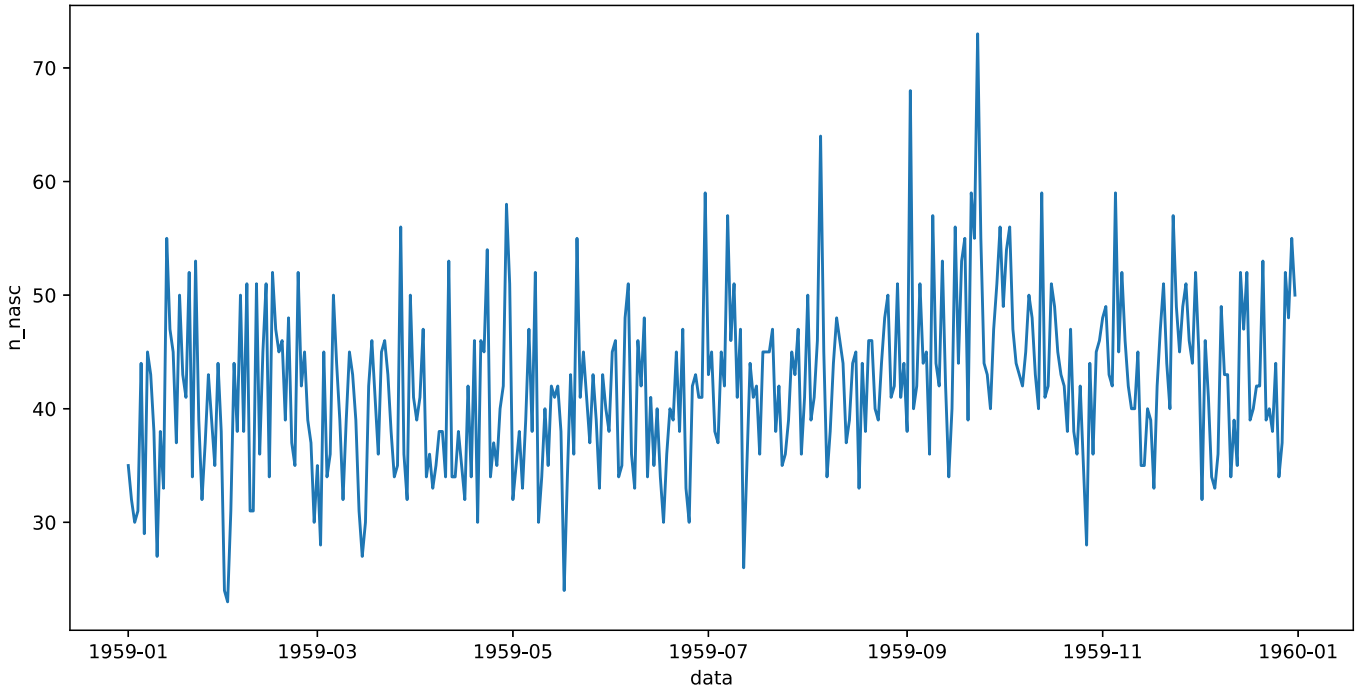
Out[4]:

| | data | n_nasc |
|---|---|---|
| **0** | 1959-01-01 | 35 |
| **1** | 1959-01-02 | 32 |
| **2** | 1959-01-03 | 30 |
| **3** | 1959-01-04 | 31 |
| **4** | 1959-01-05 | 44 |

In [5]:
```
nasc['data'] = pd.to_datetime(nasc['data'])
```

In [6]:
```
plt.figure(figsize=(12,6))
plt.plot(nasc['data'], nasc['n_nasc'])
plt.ylabel('n_nasc')
plt.xlabel('data')
plt
```

Out[6]:
```
<module 'matplotlib.pyplot' from 'c:\\users\\ed\\anaconda3\\envs\\data_science\\lib\\site-packa
ges\\matplotlib\\pyplot.py'>
```




In [7]:
```
media_carb = co2['media'].mean()
media_carb
```

Out[7]:
```
370.10314465408806
```

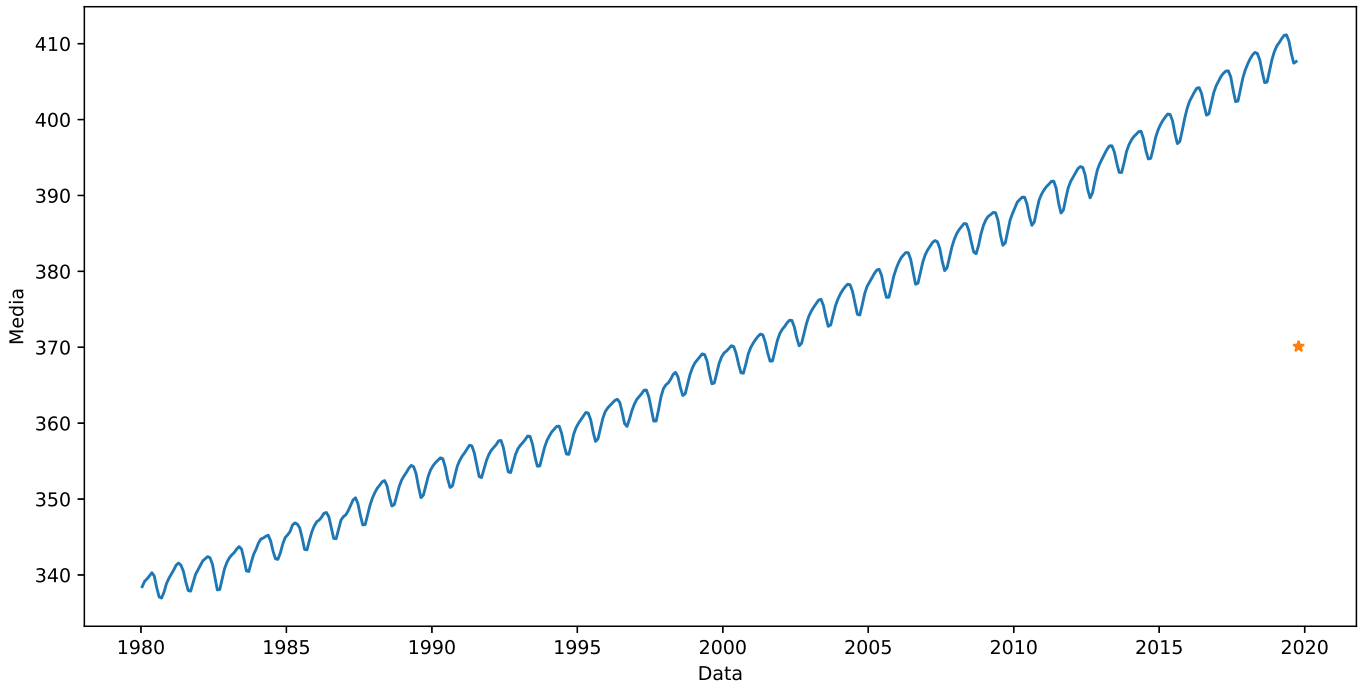
In [8]:
```
dt_carb = co2['data'][1] - co2['data'][0]
dt_carb
```

Out[8]:
```
0.08300000000008367
```

In [9]:
```
plt.figure(figsize=(12,6))
plt.plot(co2['data'], co2['media'])
plt.plot(co2.iloc[-1, 0] + dt_carb, media_carb, "*")
plt.ylabel('Media')
plt.xlabel('Data')
plt
```

```
<module 'matplotlib.pyplot' from 'c:\\users\\ed\\anaconda3\\envs\\data_science\\lib\\site-packa
```

Out[9]: ges\\matplotlib\\pyplot.py'>




In [10]:
```
media_nasc = nasc['n_nasc'].mean()
media_nasc
```

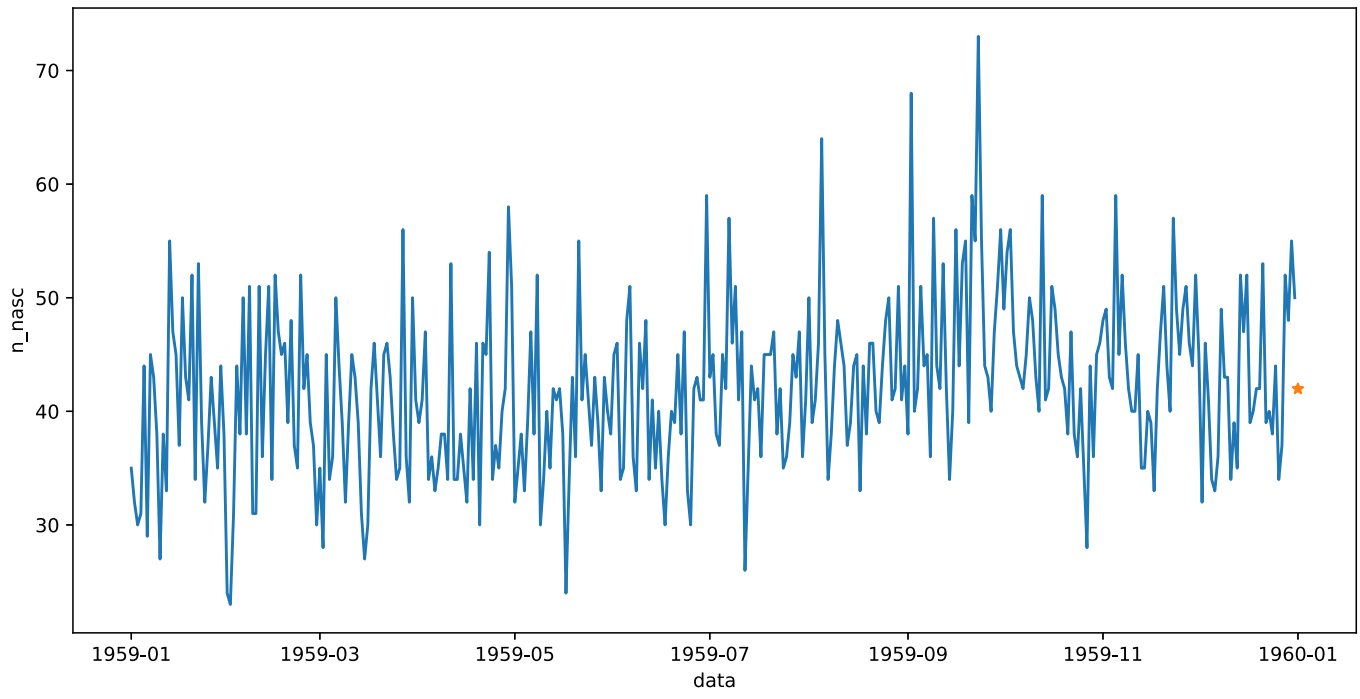
Out[10]: 41.98082191780822

In [11]:
```
dt_nasc = nasc['data'][1] - nasc['data'][0]
dt_nasc
```

Out[11]: Timedelta('1 days 00:00:00')

In [12]:
```
plt.figure(figsize=(12,6))
plt.plot(nasc['data'], nasc['n_nasc'])
plt.plot(nasc.iloc[-1, 0] + dt_nasc, media_nasc, "*")
plt.ylabel('n_nasc')
plt.xlabel('data')
plt
```

Out[12]: <module 'matplotlib.pyplot' from 'c:\\users\\ed\\anaconda3\\envs\\data_science\\lib\\site-packages\\matplotlib\\pyplot.py'>

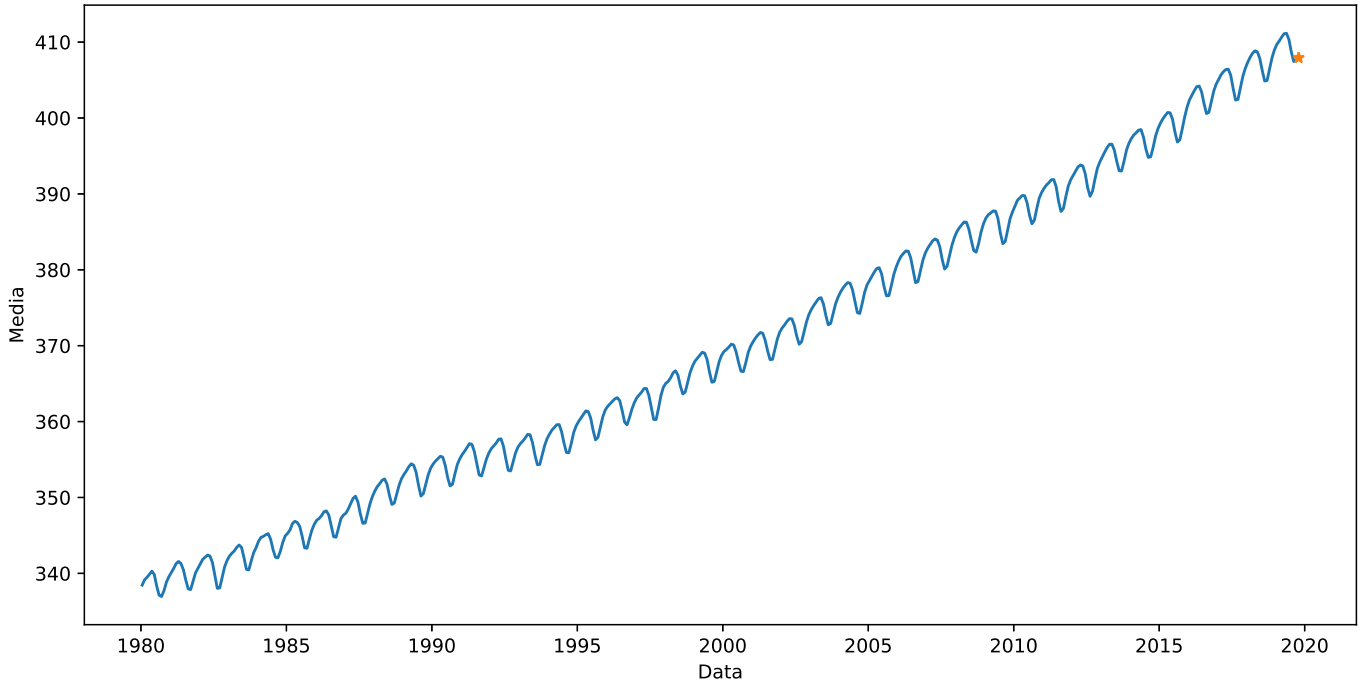


In [13]:
```
media_tres_pontos_co2 = co2['media'].tail(3).mean()
media_tres_pontos_co2
```

Out[13]:
```
407.92333333333335
```

In [14]:
```
plt.figure(figsize=(12,6))
plt.plot(co2['data'], co2['media'])
plt.plot(co2.iloc[-1, 0] + dt_carb, media_tres_pontos_co2, "*")
plt.ylabel('Media')
plt.xlabel('Data')
plt
```

Out[14]:
```
<module 'matplotlib.pyplot' from 'c:\\users\\ed\\anaconda3\\envs\\data_science\\lib\\site-packages\\matplotlib\\pyplot.py'>
```

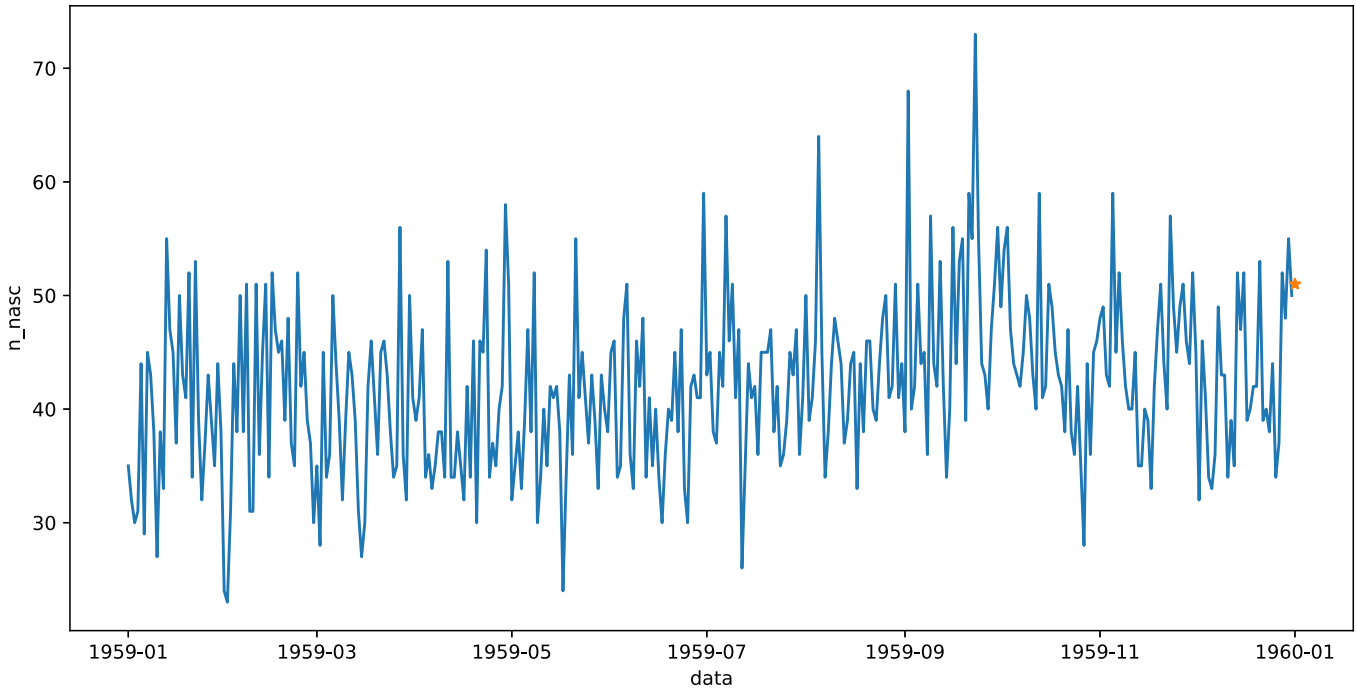



In [15]:
```
media_tres_pontos_nasc = nasc['n_nasc'].tail(3).mean()
media_tres_pontos_nasc
```

Out[15]:
```
51.0
```

In [16]:
```
plt.figure(figsize=(12,6))
plt.plot(nasc['data'], nasc['n_nasc'])
plt.plot(nasc.iloc[-1, 0] + dt_nasc, media_tres_pontos_nasc, "*")
plt.ylabel('n_nasc')
plt.xlabel('data')
plt
```

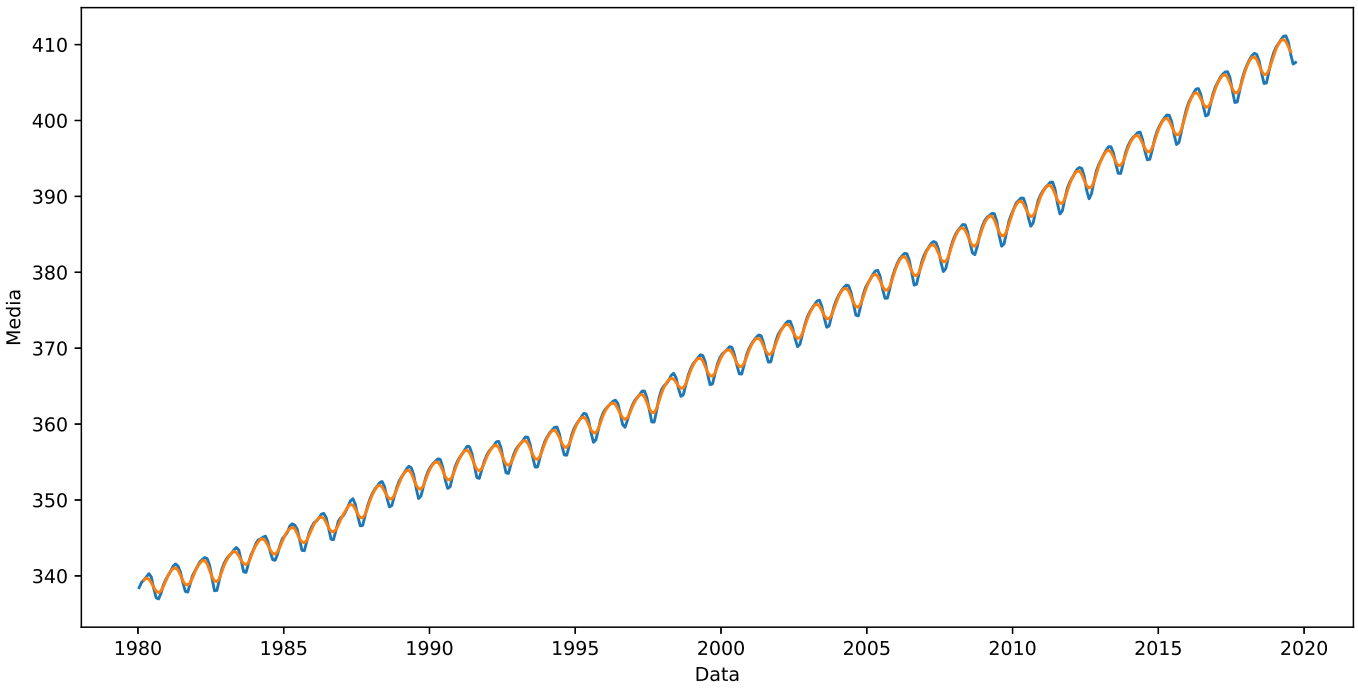
Out[16]:
```
<module 'matplotlib.pyplot' from 'c:\\users\\ed\\anaconda3\\envs\\data_science\\lib\\site-packages\\matplotlib\\pyplot.py'>
```




In [17]:
```
media_movel_co2 =  co2.rolling(5).mean()
```

In [18]:
```
plt.figure(figsize=(12,6))
plt.plot(co2['data'], co2['media'])
plt.plot(media_movel_co2['data'], media_movel_co2['media'])
plt.ylabel('Media')
plt.xlabel('Data')
plt
```

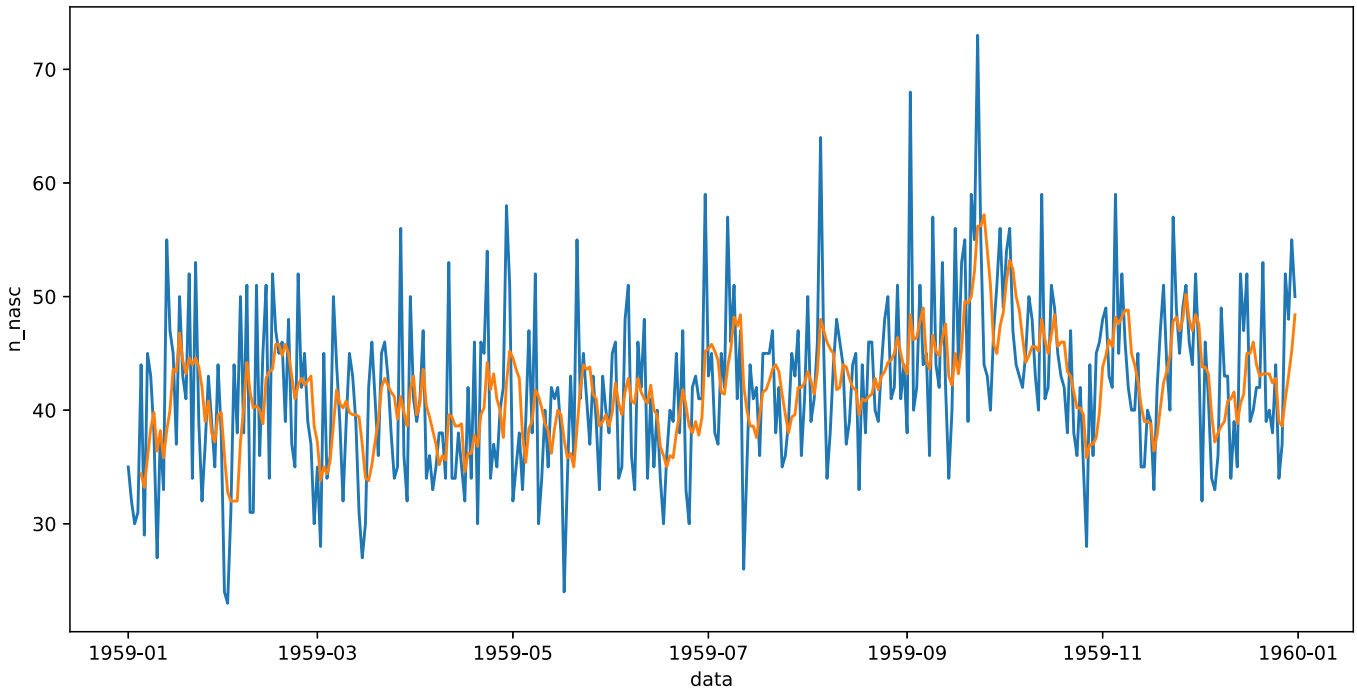
Out[18]:
```
<module 'matplotlib.pyplot' from 'c:\\users\\ed\\anaconda3\\envs\\data_science\\lib\\site-packages\\matplotlib\\pyplot.py'>
```




In [19]:
```
media_movel_nasc = nasc.rolling(5).mean()
```

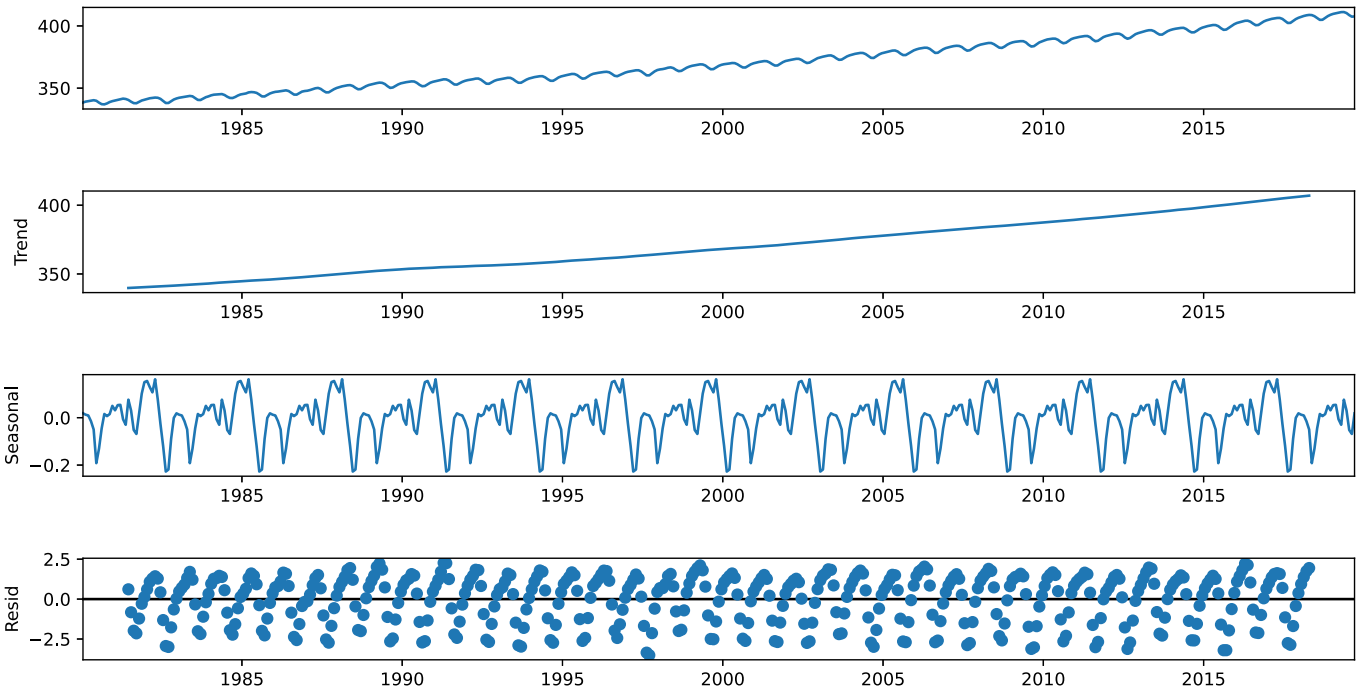
In [20]:
```
plt.figure(figsize=(12,6))
plt.plot(nasc['data'], nasc['n_nasc'])
plt.plot(nasc['data'], media_movel_nasc['n_nasc'])
plt.ylabel('n_nasc')
plt.xlabel('data')
plt
```

Out[20]:
```
<module 'matplotlib.pyplot' from 'c:\\users\\ed\\anaconda3\\envs\\data_science\\lib\\site-packages\\matplotlib\\pyplot.py'>
```



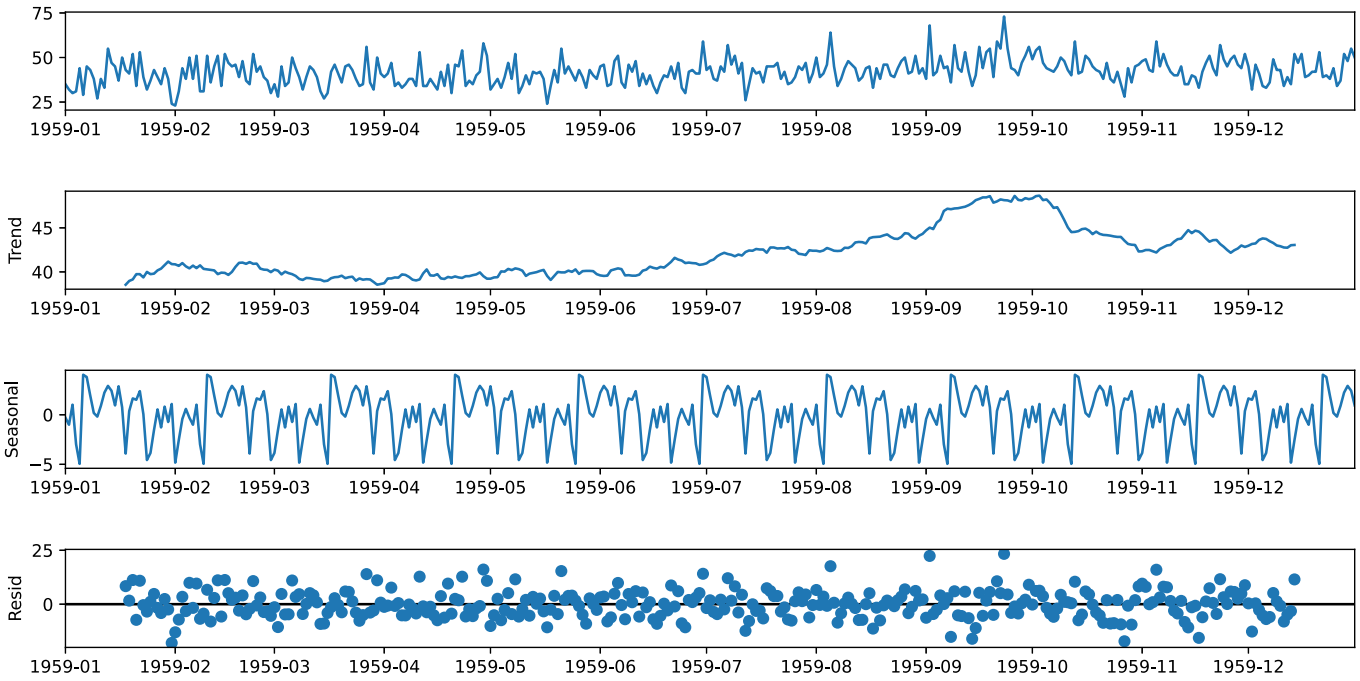

In [21]:
```
result_co2 = seasonal_decompose(co2.set_index('data'), period = 35)
```

In [22]:
```
ax = result_co2.plot();
ax.set_size_inches(12,6)
```

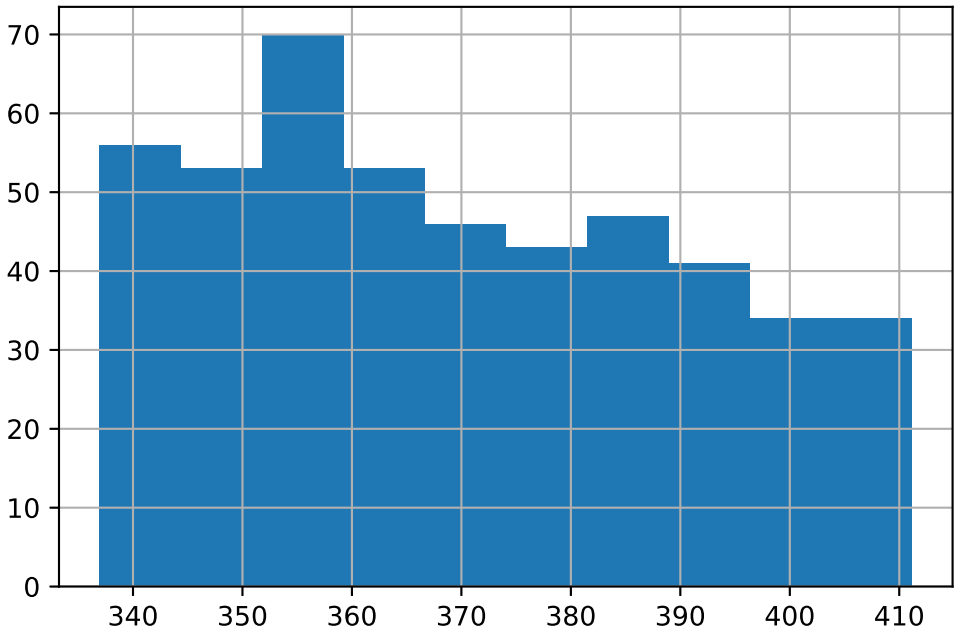



In [23]:
```
result_nasc = seasonal_decompose(nasc.set_index('data'), period = 35)
```

In [24]:
```
ax = result_nasc.plot();
ax.set_size_inches(12,6)
```




In [25]:
```
ax = co2['media'].hist()
```



In [26]:
```
divide = int(len(co2)/2)
x = co2['media'].values
```

In [27]:
```
c1, c2 = x[0:divide], x[divide:]
```

In [28]:
```
c1.mean()
```

Out[28]: 352.31550420168065

In [29]:
```
c2.mean()
```

Out[29]: 387.816359832636

In [30]:
```
c1.var()
```

Out[30]: 73.90219533401597

In [31]:
```
c2.var()
```

Out[31]: 151.542490933282

In [32]:
```
divide = int(len(nasc)/2)
```

In [33]:
```
x = nasc['n_nasc'].values
```

In [34]:
```
n1, n2 = x[0:divide], x[divide:]
```

In [35]:
```
n1.mean()
```

Out[35]: 39.76373626373626

In [36]:
```
n2.mean()
```

Out[36]: 44.185792349726775

In [37]:
```
n1.var()
```

Out[37]: 49.21341021615746

In [38]:
```
n2.var()
```

Out[38]: 48.708650601690096

In [39]:
```
resultado_co2 = adfuller(co2['media'].values)
```

In [40]:
```
print(f'Estatística ADF {resultado_co2[0]}')
print(f'p-valor {resultado_co2[1]}')
```

```
Estatística ADF 3.036828829166984
p-valor 1.0
```

In [41]:
```
resultado_nasc = adfuller(nasc['n_nasc'].values)
```

In [42]:
```
print(f'Estatística ADF {resultado_nasc[0]}')
print(f'p-valor {resultado_nasc[1]}')
```

```
Estatística ADF -4.808291253559764
p-valor 5.2434129901498554e-05
```

In [43]:
```
co2.set_index('data', inplace = True)
```

In [44]:
```
co2_treino = co2[1980.042:2015]
co2_teste = co2[2015:]
```

In [45]:
```
modelo_ajustado = SimpleExpSmoothing(co2_treino).fit(smoothing_level=0.5)
```

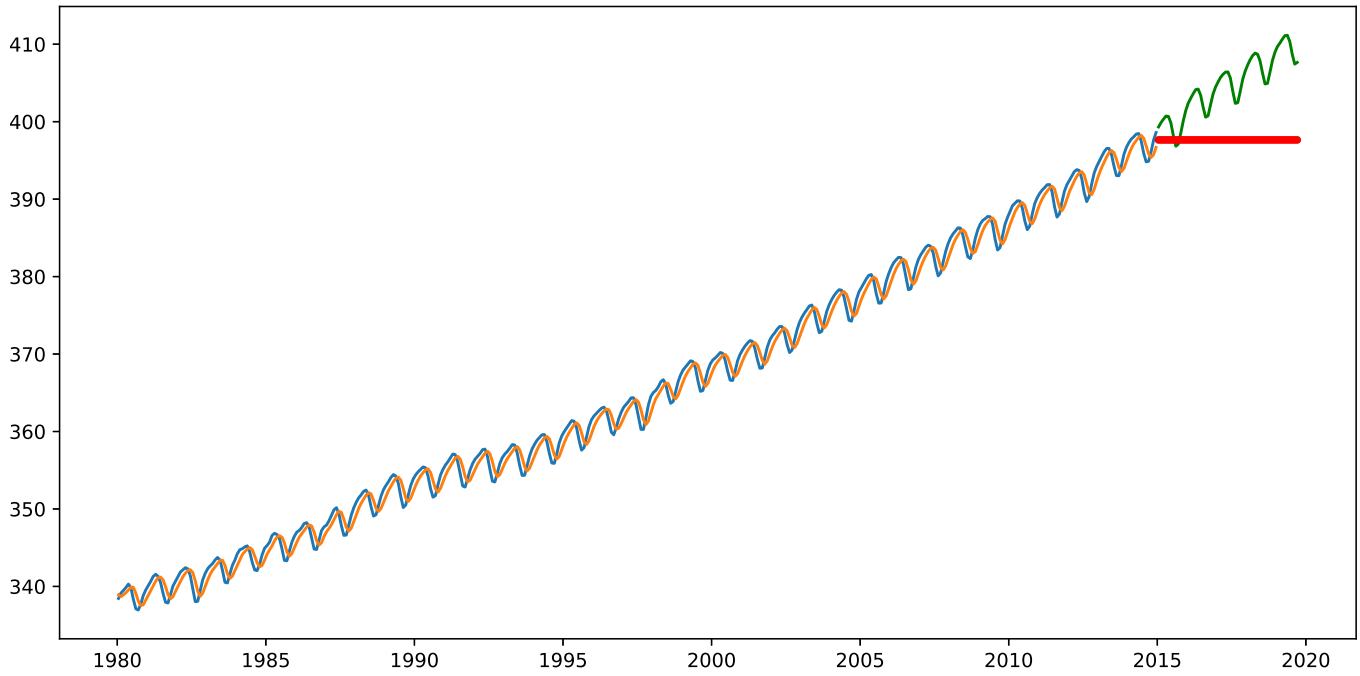
```
c:\users\ed\anaconda3\envs\data_science\lib\site-packages\statsmodels\tsa\base\tsa_model.py:57
8: ValueWarning: An unsupported index was provided and will be ignored when e.g. forecasting.
  warnings.warn('An unsupported index was provided and will be'
```

In [46]:
```
modelo_previsto = modelo_ajustado.forecast(57)
```

```
c:\users\ed\anaconda3\envs\data_science\lib\site-packages\statsmodels\tsa\base\tsa_model.py:37
6: ValueWarning: No supported index is available. Prediction results will be given with an inte
ger index beginning at `start`.
  warnings.warn('No supported index is available.'
```

In [47]:
```
plt.figure(figsize=(12,6))
plt.plot(co2_treino)
plt.plot(co2_treino.index,modelo_ajustado.fittedvalues.values)
plt.plot(co2_teste, 'g')
plt.plot(co2_teste.index, modelo_previsto, 'r.')
```

Out[47]:
```
[<matplotlib.lines.Line2D at 0xcb067f0>]
```




In [48]:
```
nasc.set_index('data', inplace = True)
```

In [49]:
```
nasc_treino = nasc['1959-01-01':'1959-12-01']
nasc_teste = nasc['1959-12-01':]
```

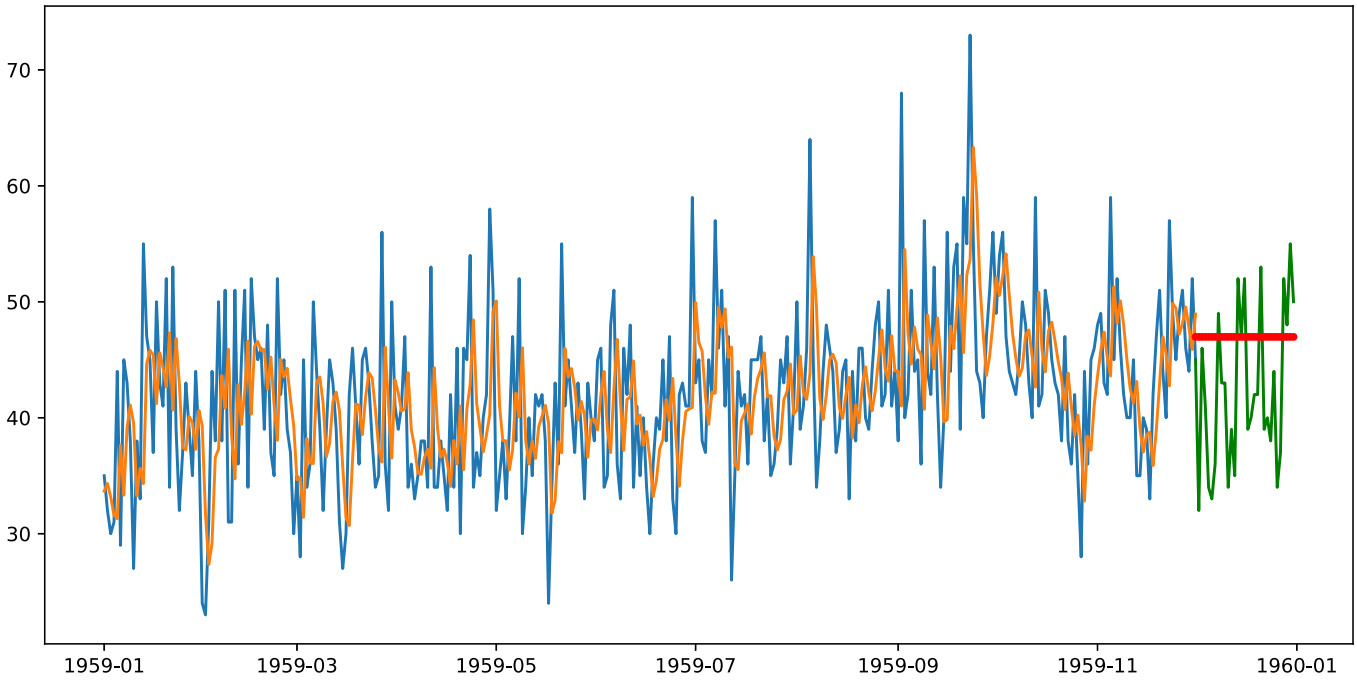
In [50]:
```
modelo_ajustado = SimpleExpSmoothing(nasc_treino).fit(smoothing_level = 0.5)
```

```
c:\users\ed\anaconda3\envs\data_science\lib\site-packages\statsmodels\tsa\base\tsa_model.py:52
4: ValueWarning: No frequency information was provided, so inferred frequency D will be used.
  warnings.warn('No frequency information was'
```

In [51]:
```
modelo_previsto = modelo_ajustado.forecast(31)
```

In [52]:
```
plt.figure(figsize=(12,6))
plt.plot(nasc_treino)
plt.plot(nasc_treino.index, modelo_ajustado.fittedvalues.values)
plt.plot(nasc_teste,'g')
plt.plot(nasc_teste.index,modelo_previsto,'r.')
```

Out[52]:
```
[<matplotlib.lines.Line2D at 0xcb71f40>]
```

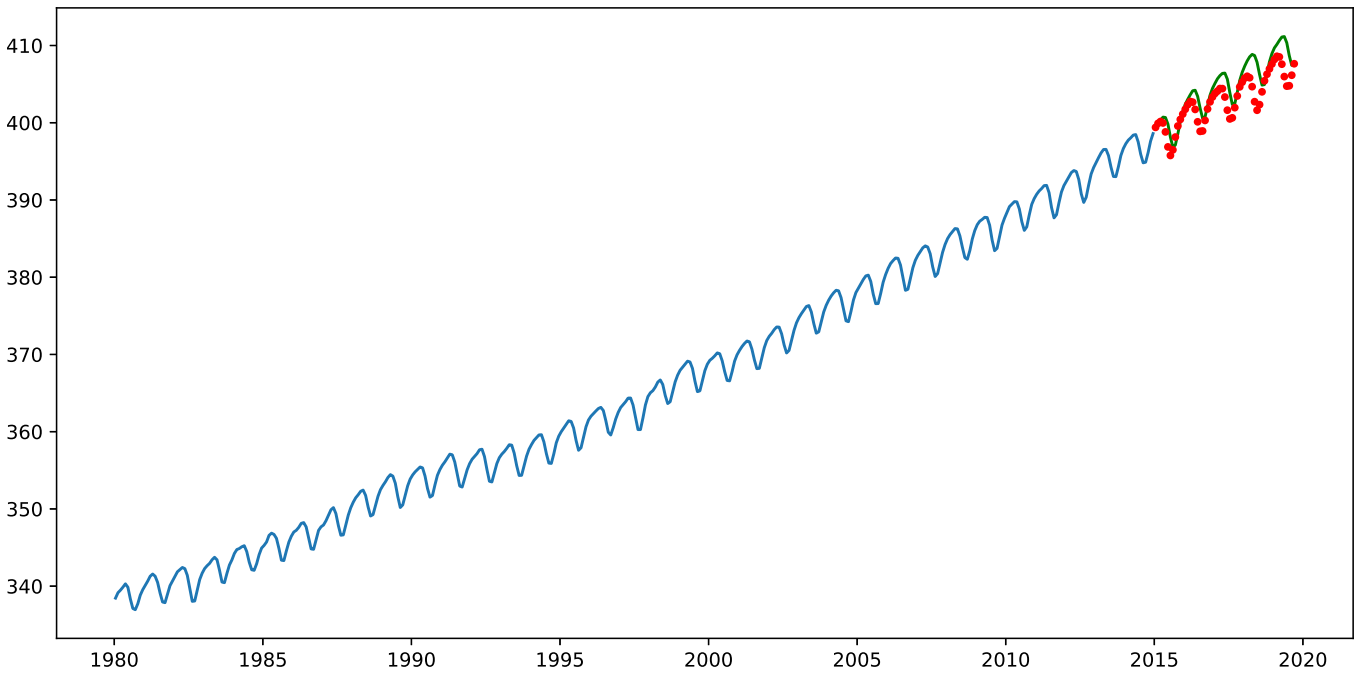



In [53]:
```
modelo_ajustado = ExponentialSmoothing(co2_treino.values, trend='additive', seasonal='additive'
```

In [54]:
```
modelo_previsto = modelo_ajustado.predict(start=420, end=476)
```

In [55]:
```
plt.figure(figsize=(12,6))
plt.plot(co2_treino)
plt.plot(co2_teste, 'g')
plt.plot(co2_teste.index, modelo_previsto, 'r.')
```

Out[55]:
```
[<matplotlib.lines.Line2D at 0xcbd1a30>]
```



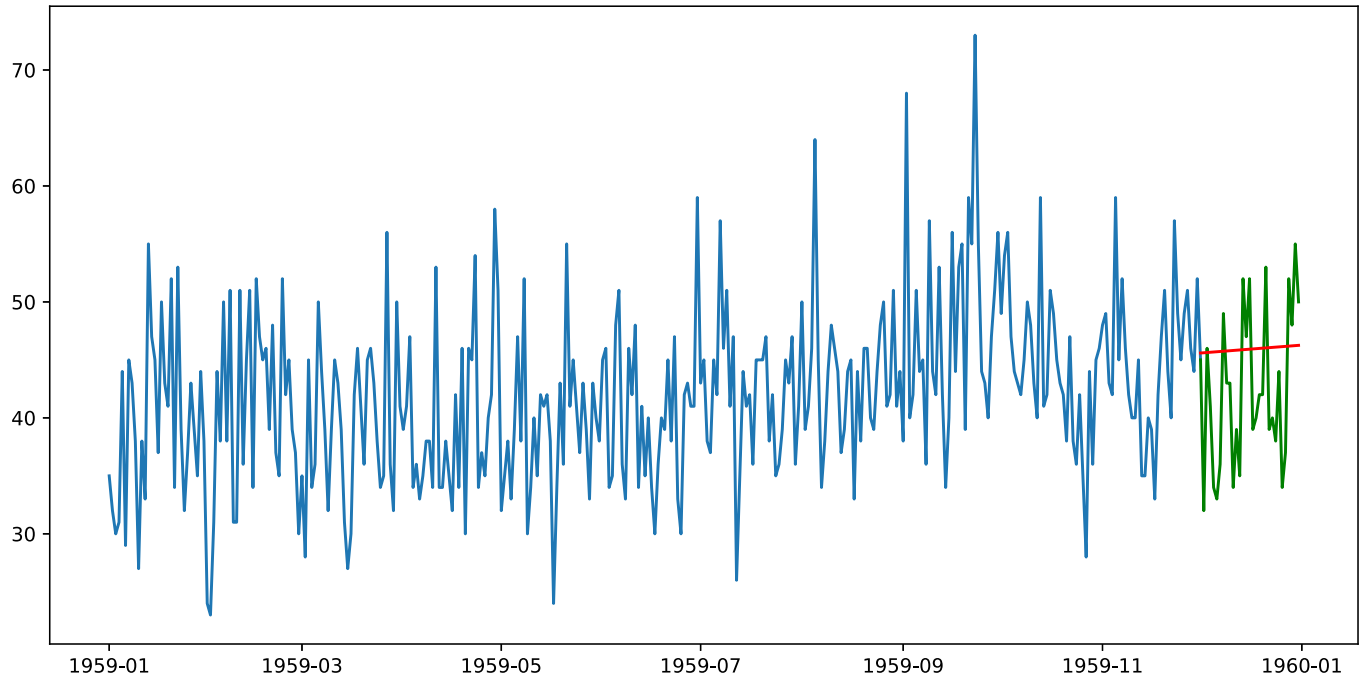

In [56]:
```
modelo_ajustado = ExponentialSmoothing(nasc_treino.values, trend='multiplicative', seasonal=Nor
```

In [57]:
```
modelo_previsto_suave = modelo_ajustado.predict(start=335,end=365)
```

In [58]:
```
plt.figure(figsize=(12,6))
plt.plot(nasc_treino)
```

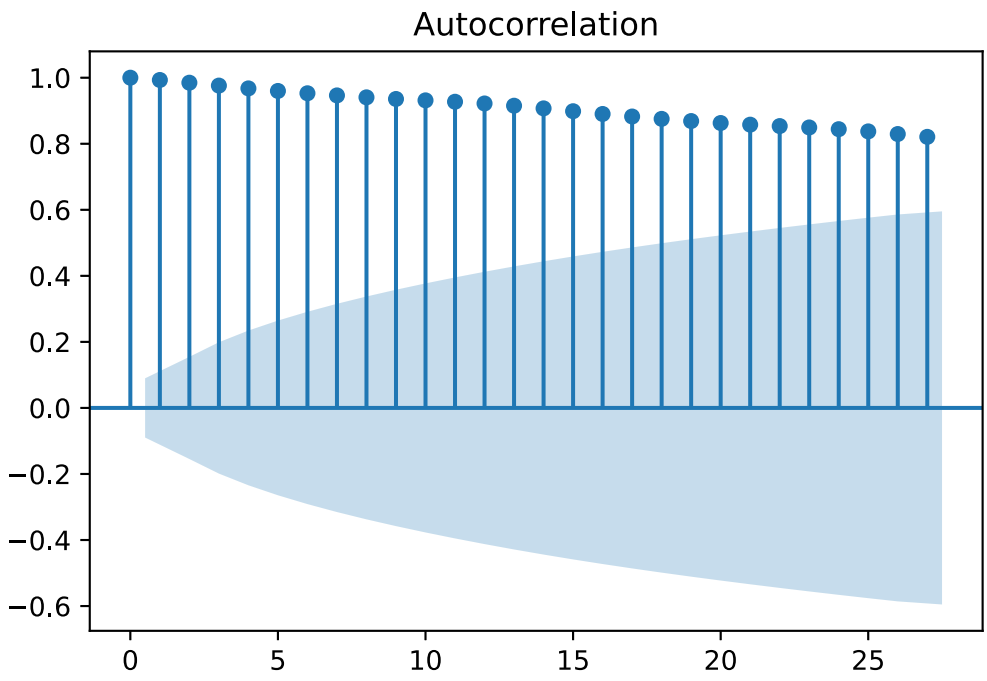
```
plt.plot(nasc_teste, 'g')
plt.plot(nasc_teste.index, modelo_previsto_suave, 'r')
```

Out[58]: [<matplotlib.lines.Line2D at 0xdbaaa60>]

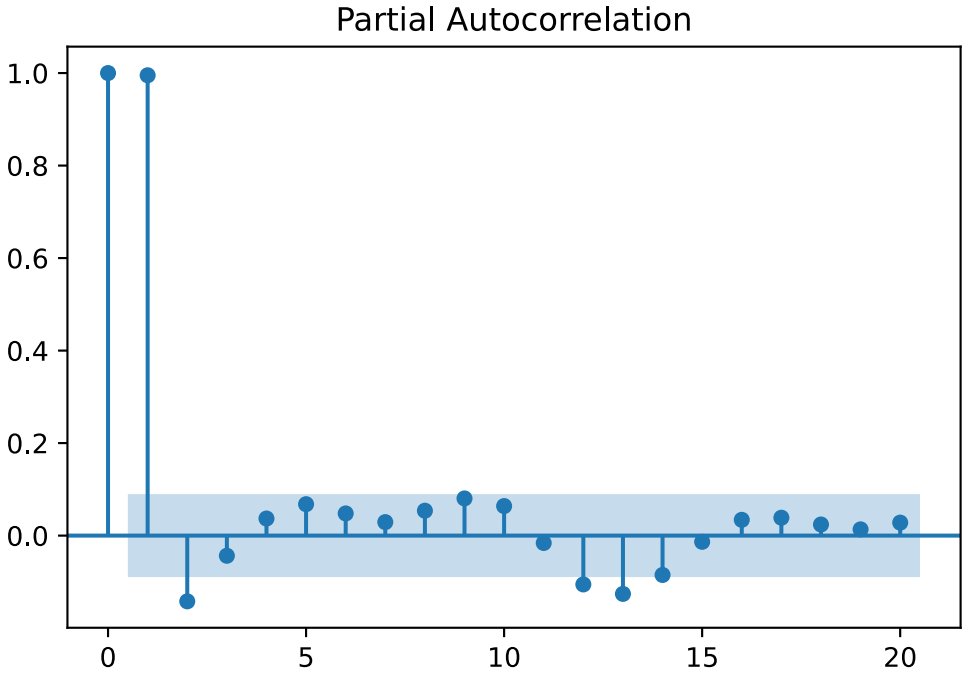



---

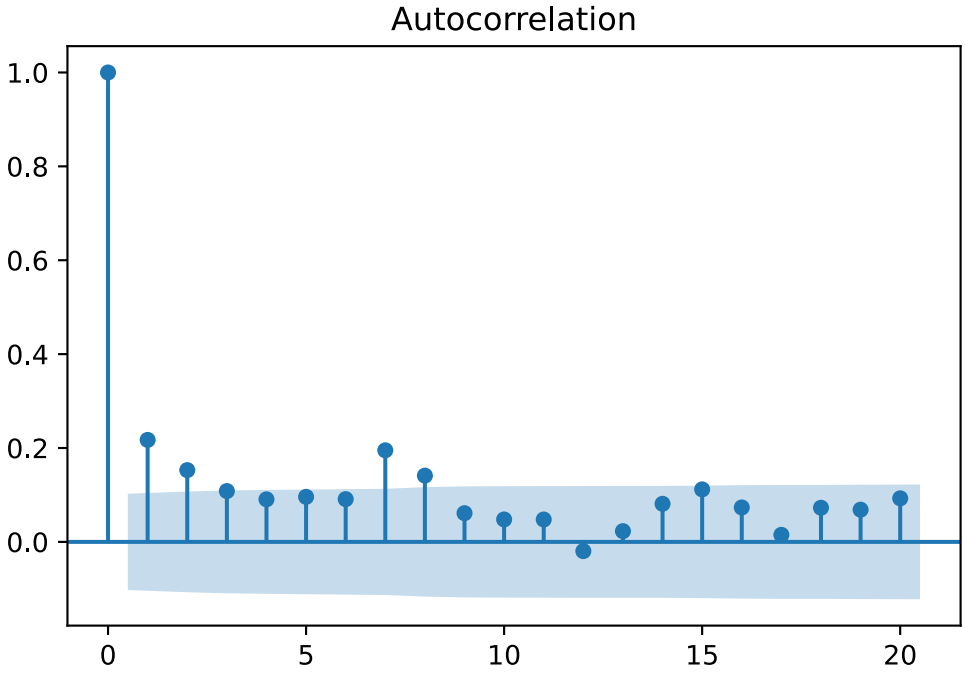
In [59]:
```
plot_acf(co2['media']);
```



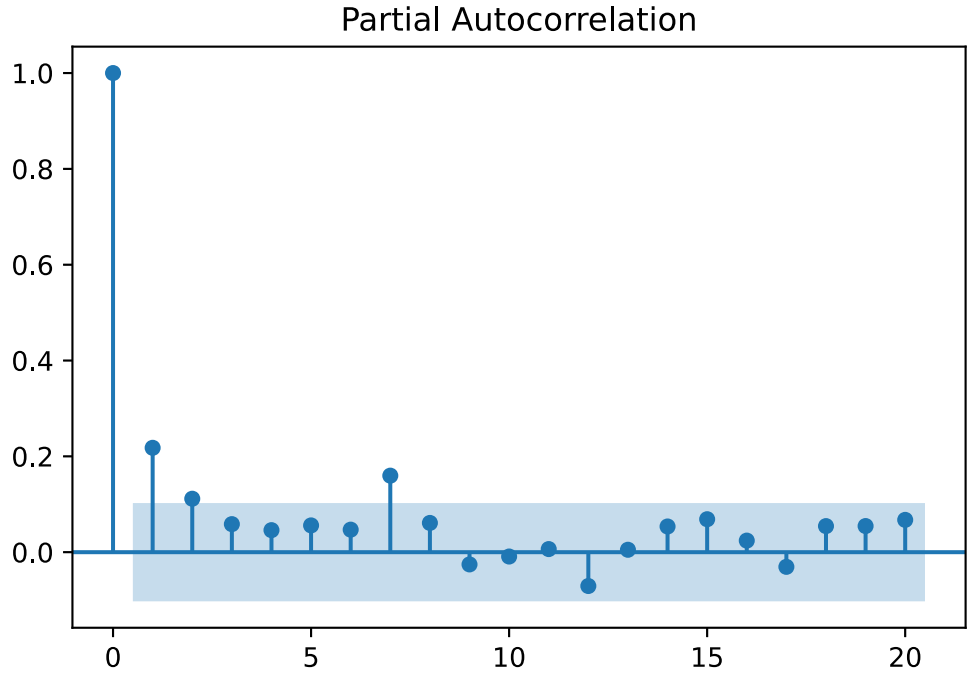

In [60]:
```
plot_pacf(co2['media'], lags = 20);
```




In [61]:
```
plot_acf(nasc['n_nasc'], lags = 20);
```




In [62]:
```
plot_pacf(nasc['n_nasc'], lags = 20);
```



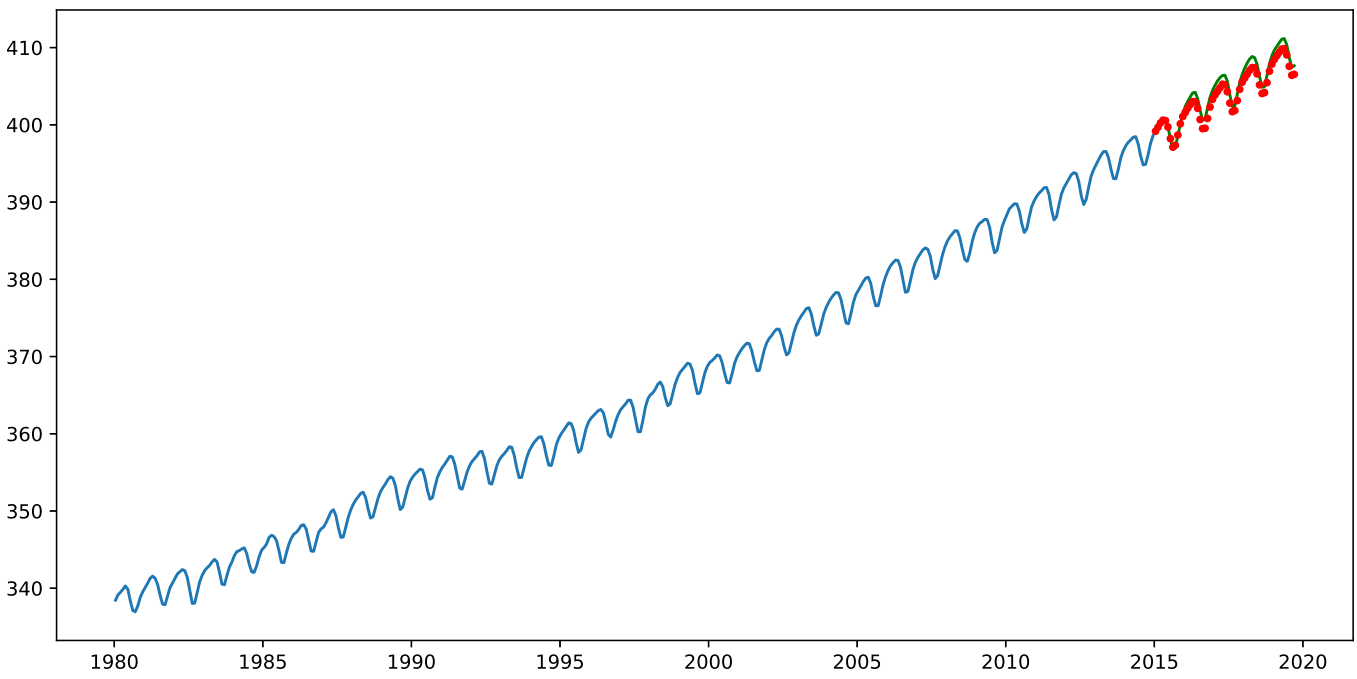

In [63]:
```
lista = np.linspace(1, 40, 40)
```

In [64]:
```
modelo_ajustado = AutoReg(co2_treino.values, lags = lista, trend = 'c', seasonal = True, perioc
```

In [65]:
```
modelo_previsto = modelo_ajustado.predict(start = 420, end = 476)
```

In [66]:
```
plt.figure(figsize=(12,6))
plt.plot(co2_treino)
plt.plot(co2_teste, 'g')
plt.plot(co2_teste.index, modelo_previsto, 'r.')
```

Out[66]:
```
[<matplotlib.lines.Line2D at 0xc8e22b0>]
```

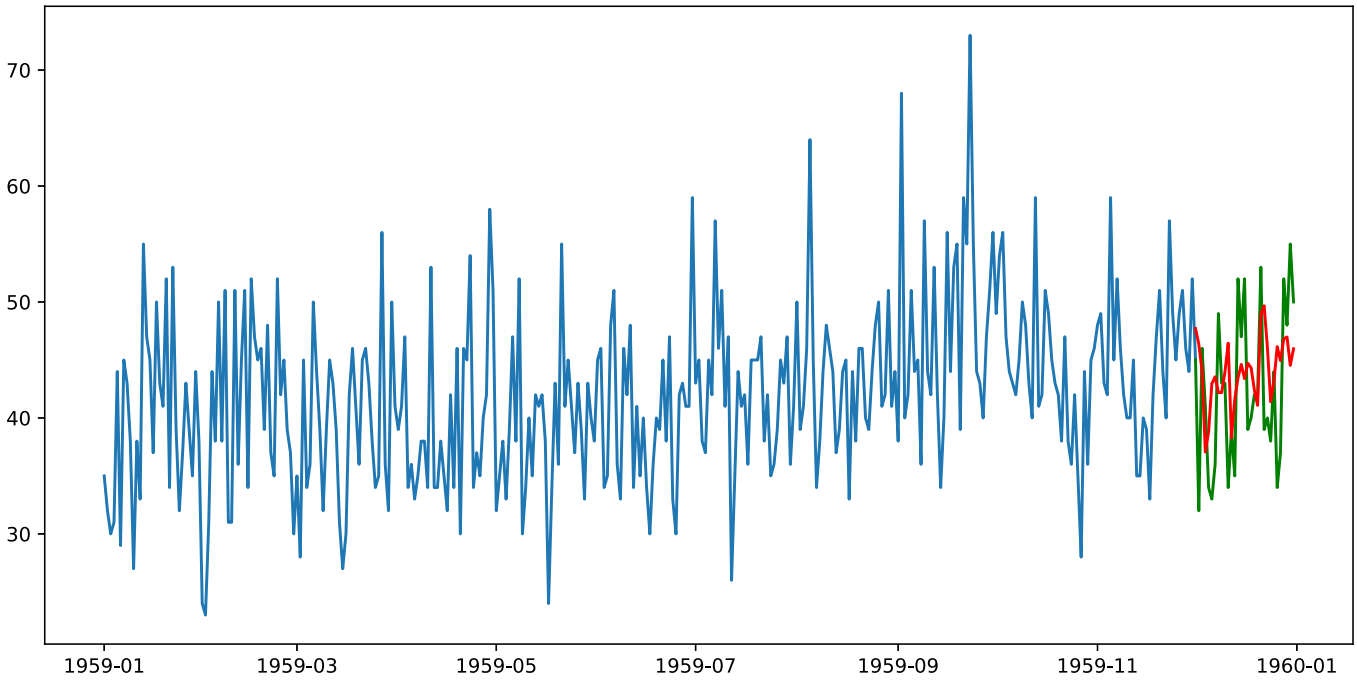



In [67]:
```
modelo_ajustado = AutoReg(nasc_treino.values, lags = lista, trend = 'c', seasonal = True, perio
```

In [68]:
```
modelo_previsto_ar = modelo_ajustado.predict(start = 335, end = 365)
```

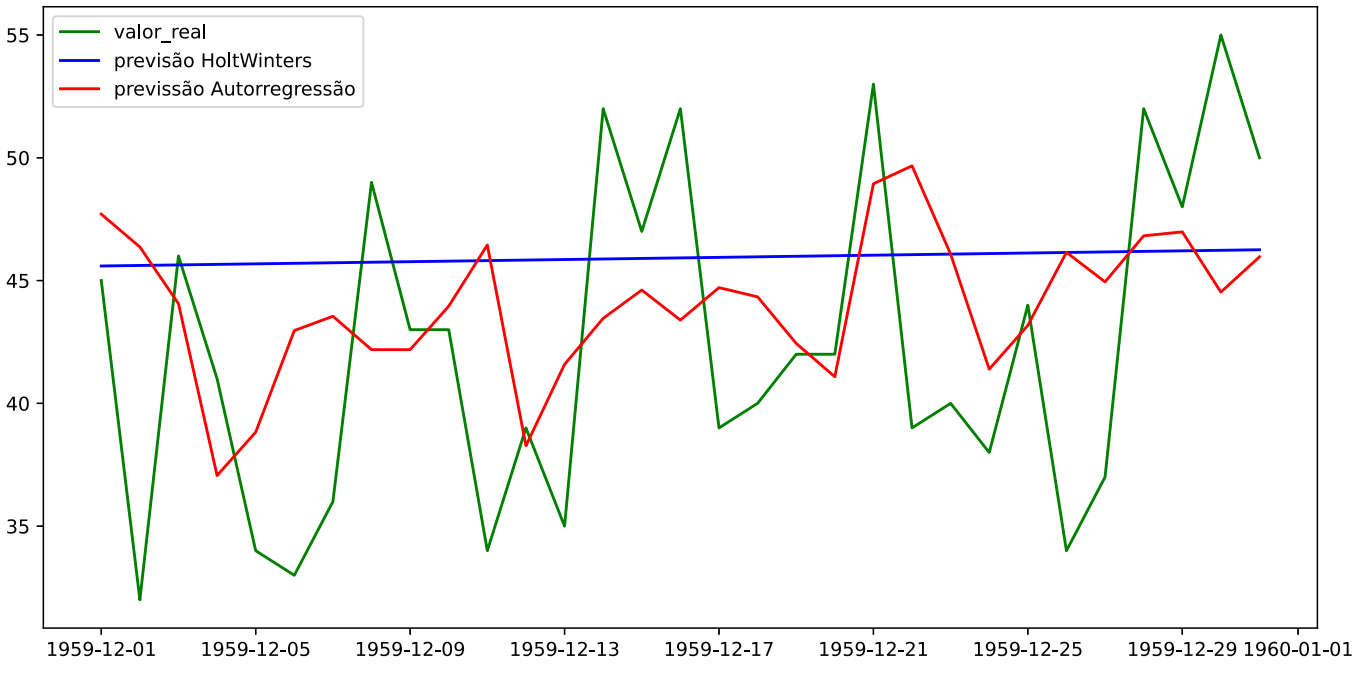
In [69]:
```
plt.figure(figsize=(12,6))
plt.plot(nasc_treino)
plt.plot(nasc_teste, 'g')
plt.plot(nasc_teste.index, modelo_previsto_ar, 'r')
```

Out[69]:
```
[<matplotlib.lines.Line2D at 0xc5896a0>]
```




In [70]:
```
plt.figure(figsize=(12,6))
plt.plot(nasc_teste, 'g', label = 'valor_real')
plt.plot(nasc_teste.index, modelo_previsto_suave, 'b-', label = 'previsão HoltWinters')
plt.plot(nasc_teste.index, modelo_previsto_ar, 'r', label = 'previssão Autorregressão')
plt.legend()
```

Out[70]:
```
<matplotlib.legend.Legend at 0xc97f790>
```



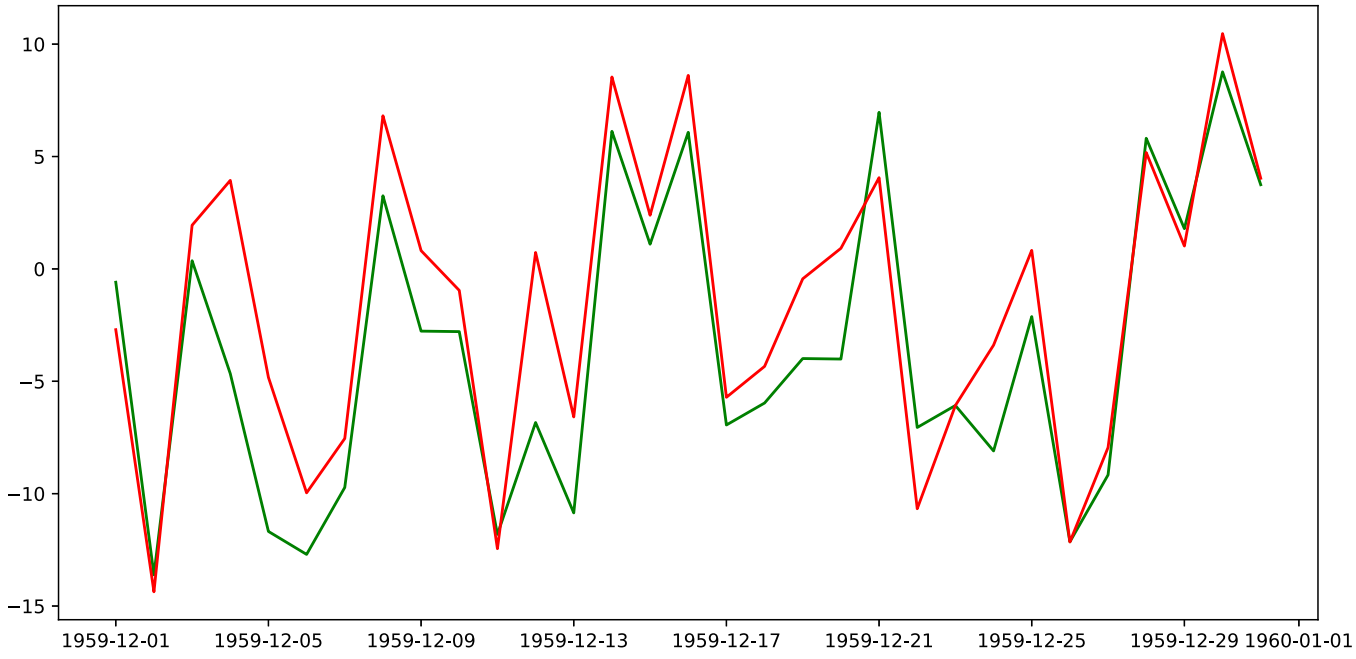

In [71]:
```
erro_prev_s = nasc_teste['n_nasc'] - modelo_previsto_suave
```

In [72]:
```
erro_prev_ar = nasc_teste['n_nasc'] - modelo_previsto_ar
```

In [73]:
```
plt.figure(figsize=(12,6))
```

```
plt.plot(erro_prev_s, 'g')
plt.plot(erro_prev_ar, 'r')
```

Out[73]: [<matplotlib.lines.Line2D at 0xc8e4a90>]




In [74]:
```
np.mean(erro_prev_s)
```

Out[74]: -3.5361160668765756

In [75]:
```
np.mean(erro_prev_ar)
```

Out[75]: -1.6083080804502645

In [76]:
```
np.mean(np.abs(erro_prev_s))
```

Out[76]: 6.374533894721303

In [77]:
```
np.mean(np.abs(erro_prev_ar))
```

Out[77]: 5.496720548092109

In [78]:
```
mae = mean_absolute_error(nasc_teste['n_nasc'], modelo_previsto_ar)
mae
```

Out[78]: 5.496720548092109

In [79]:
```
mse = mean_squared_error(nasc_teste['n_nasc'], modelo_previsto_suave)
mse
```

Out[79]: 54.39579283969049

In [80]:
```
mse = mean_squared_error(nasc_teste['n_nasc'], modelo_previsto_ar)
mse
```

Out[80]: 45.43739422181397

In [ ]:
```

```