

```
In [ ]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from statsmodels.stats.diagnostic import acorr_ljungbox
from statsmodels.graphics.tsaplots import plot_acf
from statsmodels.tsa.stattools import adfuller
```

```
In [ ]: df = pd.read_csv('data/electricity.csv', index_col=0, parse_dates=True)
df.dropna(inplace=True)
df.head()
```

C:\Users\user\AppData\Local\Temp\ipykernel\_12416\1662925816.py:1: UserWarning: Could not infer format, so each element will be parsed individually, falling back to `dateutil`. To ensure parsing is consistent and as-expected, please specify a format.

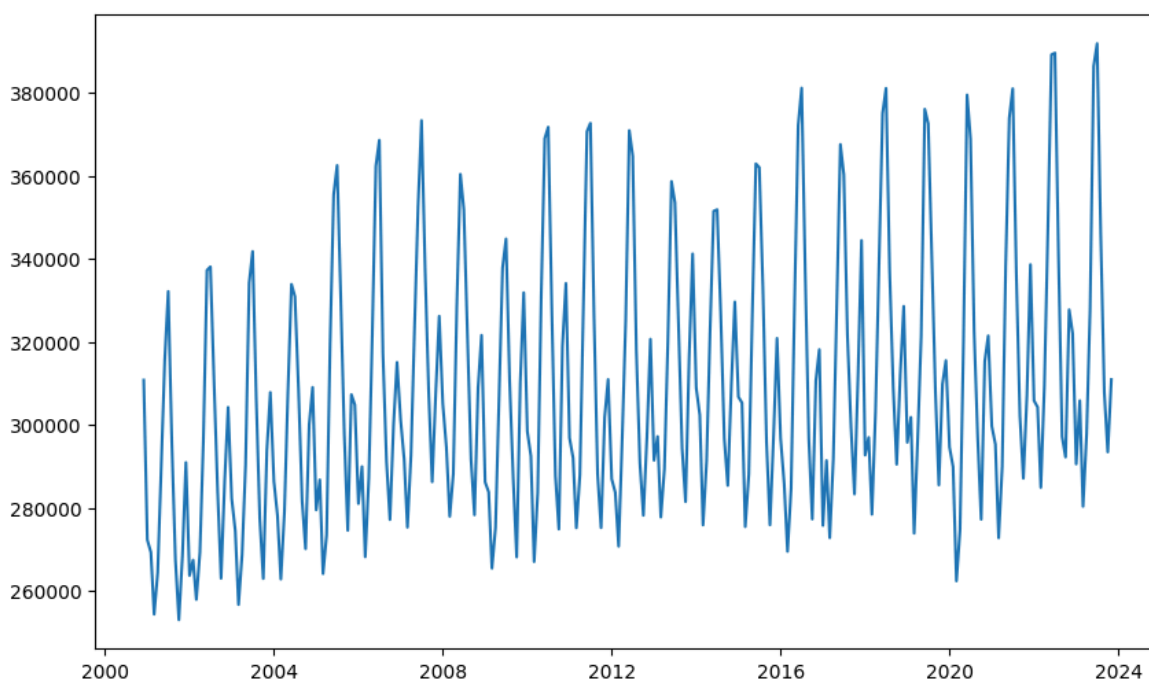
```
df = pd.read_csv('data/electricity.csv', index_col=0, parse_dates=True)
```

Out [ ]: **Sales**

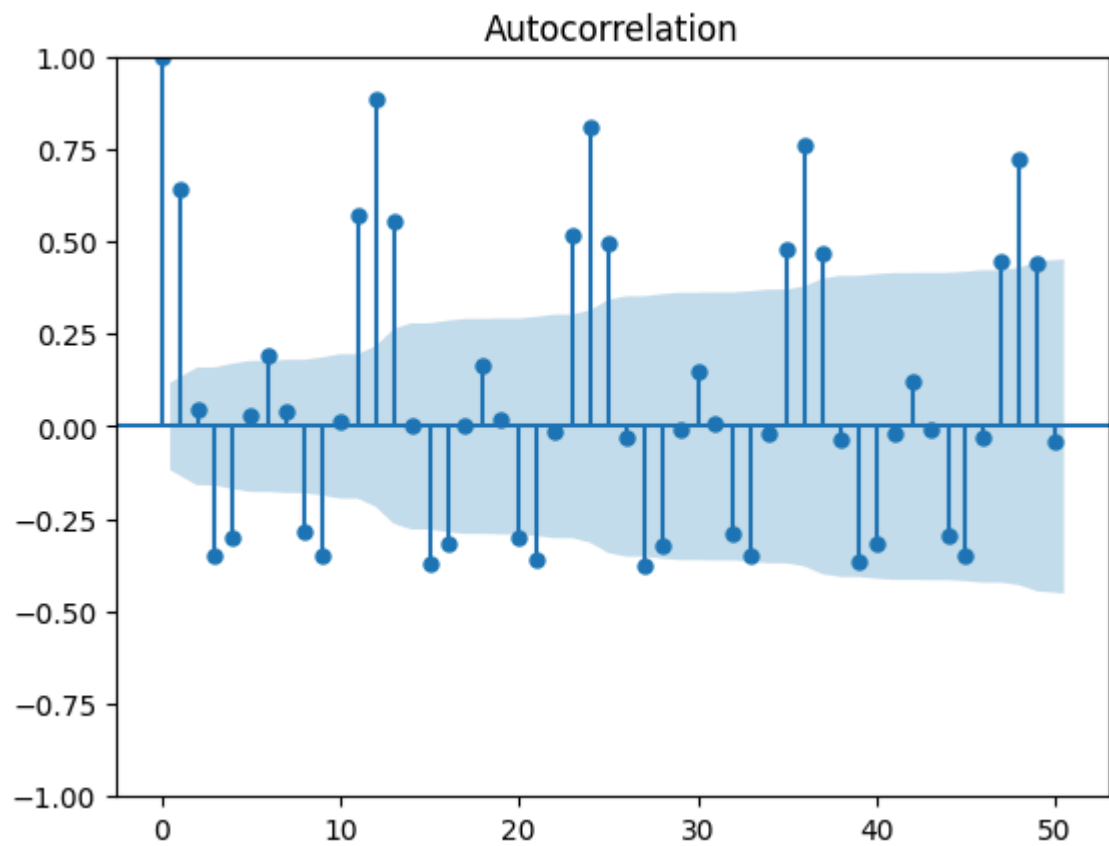
	Month
<b>2023-11-01</b>	310959.15193
<b>2023-10-01</b>	293487.06424
<b>2023-09-01</b>	307874.44273
<b>2023-08-01</b>	346128.54938
<b>2023-07-01</b>	391900.00897

```
In [ ]: plt.figure(figsize=(10, 6))
plt.plot(df)
```

Out [ ]: [



```
In [ ]: fig = plot_acf(df['Sales'], lags=50)
```



```
In [ ]: acorr_ljungbox(df['Sales'], lags=20, return_df=True)
```

```
Out[ ]:
```

	<b>lb_stat</b>	<b>lb_pvalue</b>
<b>1</b>	114.006841	1.298597e-26
<b>2</b>	114.572372	1.321069e-25
<b>3</b>	148.874907	4.607484e-32
<b>4</b>	174.417991	1.177951e-36
<b>5</b>	174.669255	7.355873e-36
<b>6</b>	185.273144	2.572518e-37
<b>7</b>	185.802908	1.157733e-36
<b>8</b>	209.358584	6.796681e-41
<b>9</b>	244.002678	1.839325e-47
<b>10</b>	244.070019	9.570758e-47
<b>11</b>	338.129345	7.855465e-66
<b>12</b>	566.420498	1.557483e-113
<b>13</b>	656.939637	5.447370e-132
<b>14</b>	656.940722	3.951490e-131
<b>15</b>	697.662504	5.854965e-139
<b>16</b>	727.498647	1.808323e-144
<b>17</b>	727.505434	1.236168e-143
<b>18</b>	735.785105	1.432397e-144
<b>19</b>	735.887304	8.846116e-144
<b>20</b>	763.073918	9.660727e-149

The ljungbox test p-val is below 5% for all lags, this means data is not white noise.

```
In [ ]: adfuller(df['Sales'])
```

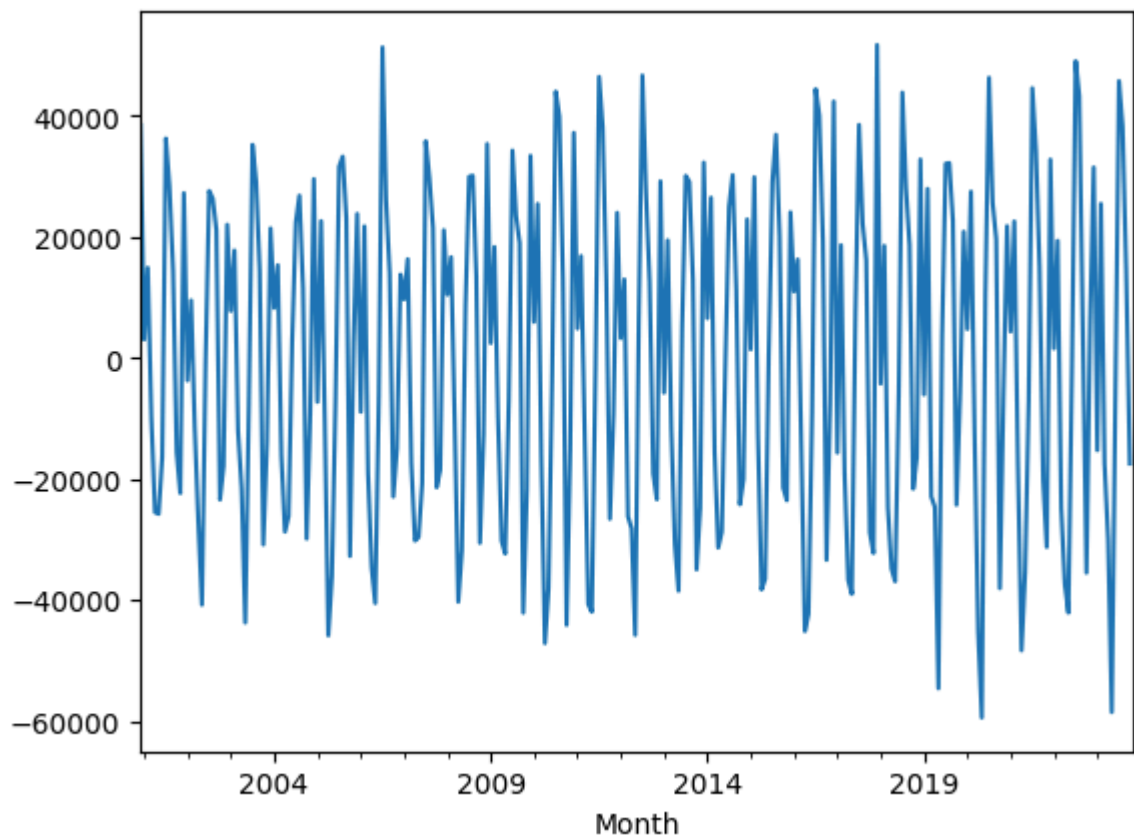
```
Out[ ]: (-0.8526613529300668,
0.8031619350195995,
12,
263,
{'1%': -3.4554613060274972,
'5%': -2.8725931472675046,
'10%': -2.5726600403359887},
5425.966646793114)
```

The adfuller test p-val is above 5%, this means data is not stationary.

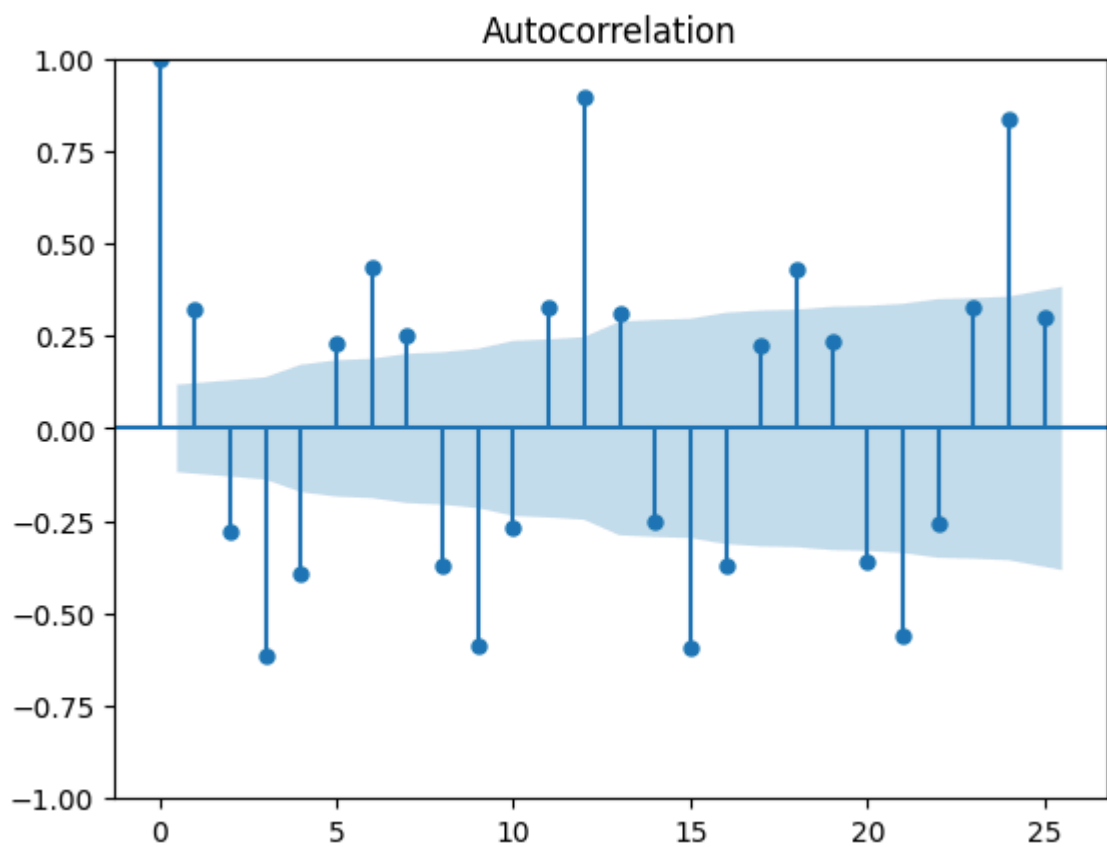
```
In [ ]: df['Sales_diff'] = df['Sales'].diff()
```

```
In [ ]: df['Sales_diff'].plot()
```

```
Out[ ]: <Axes: xlabel='Month'>
```



```
In [ ]: fig = plot_acf(df['Sales_diff'].dropna())
```



```
In [ ]: acorr_ljungbox(df['Sales_diff'].dropna(), lags=20, return_df=True)
```

Out[ ]:

	lb_stat	lb_pvalue
1	29.222818	6.451496e-08
2	50.679955	9.885249e-12
3	156.217561	1.200456e-33
4	199.366837	5.140428e-42
5	214.514720	2.222205e-44
6	268.050660	5.667128e-55
7	285.736178	6.707361e-58
8	325.442381	1.567823e-65
9	423.802868	1.136055e-85
10	444.152190	3.691738e-89
11	475.551984	5.216273e-95
12	707.278954	1.219313e-143
13	735.882366	7.315752e-149
14	754.259854	6.654470e-152
15	857.638924	4.077020e-173
16	898.519002	5.766772e-181
17	913.127572	3.335118e-183
18	968.003218	4.800085e-194
19	984.835733	9.149738e-197
20	1023.546869	3.708925e-204

In [ ]: `adfuller(df['Sales_diff'].dropna())`

Out[ ]: `(-8.078575897138567,  
1.4824386282429989e-12,  
11,  
263,  
{'1%': -3.4554613060274972,  
'5%': -2.8725931472675046,  
'10%': -2.5726600403359887},  
5403.715281749693)`

The Adfuller test p-val is below 5% it means that changes in the data are stationary.