

1921
— 2021

UN SECOLO
DI STORIA
D'AVANTI A NOI



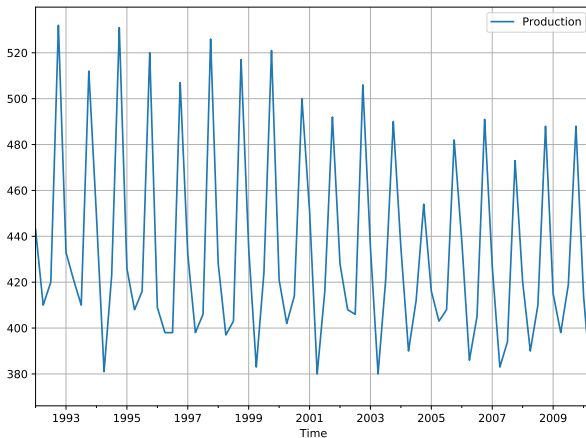
UNIVERSITÀ
CATTOLICA
del Sacro Cuore

Time Series Analysis and Forecasting

MGO962

Lab 2: Data visualization in Python

Australian Beer Production Plot > 1992



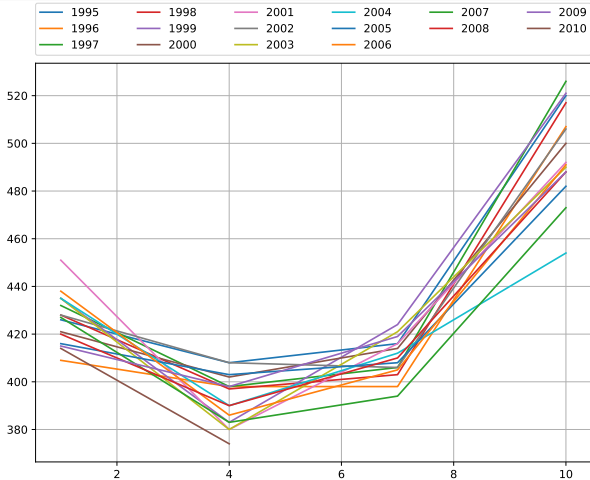
Australian Beer Production

Seasonal Plot

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("data/ausbeer.csv", header=0,
                 parse_dates=True, index_col=0)
df['year'] = df.index.year
df['month'] = df.index.month
df = df.loc['1995-01-01':'2020-01-01']
years = df['year'].unique()
for year in years:
    plt.plot('month', 'Production',
             data=df.loc[df.year == year, :], label=year)
```

Australian Beer Production

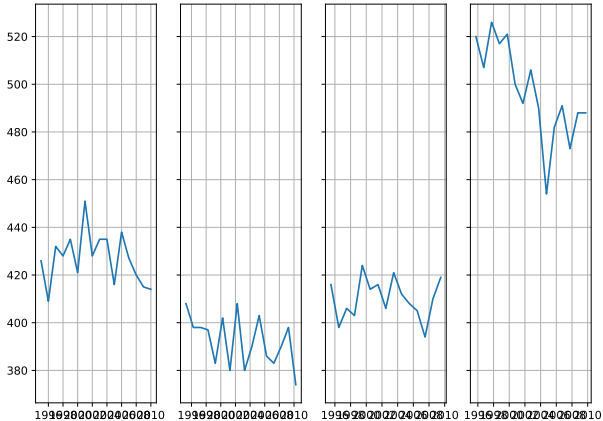
Seasonal Plot



Australian Beer Production Seasonal Subplot

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("data/ausbeer.csv", header=0,
                 parse_dates=True, index_col=0)
df['year'] = df.index.year
df['month'] = df.index.month
df = df.loc['1995-01-01':'2020-01-01']
months = df['month'].unique()
fig, axs = plt.subplots(1, 4, sharey=True)
for idx, month in enumerate(months):
    axs[idx].plot('Production', data=df.loc[df.month == month, :],
```

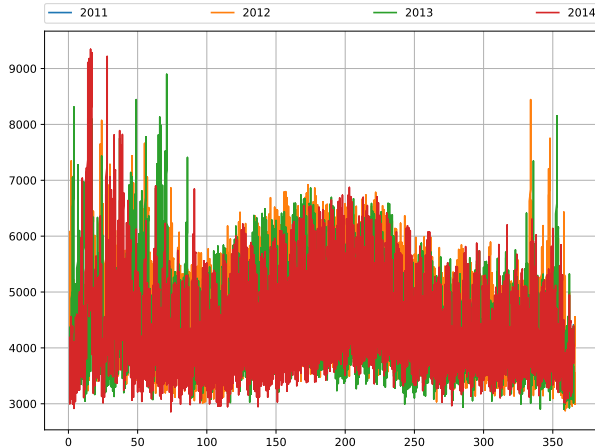
Australian Beer Production Seasonal Subplot



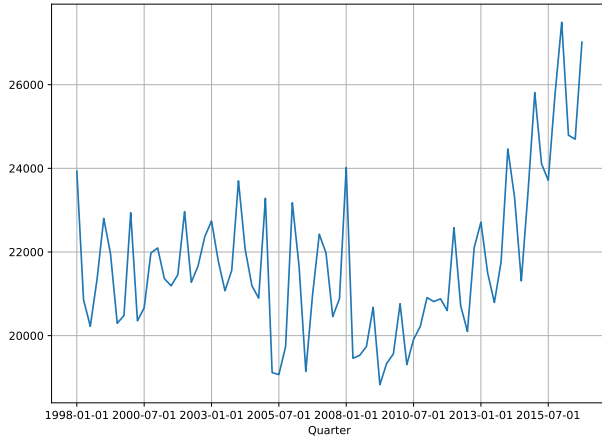
Electricity Production Seasonal Plot by Year

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("data/vic_elec.csv", header=0,
                 parse_dates=True, index_col=0)
df['year'] = df.index.year
df['day'] = df.index.dayofyear
years = df['year'].unique()
for year in years:
    dfYear = df.loc[df.year == year, :]
    plt.plot('day', 'Demand', data=dfYear, label=year)
```

Electricity Production Seasonal Plot by Year



Australian Tourism Total Trips



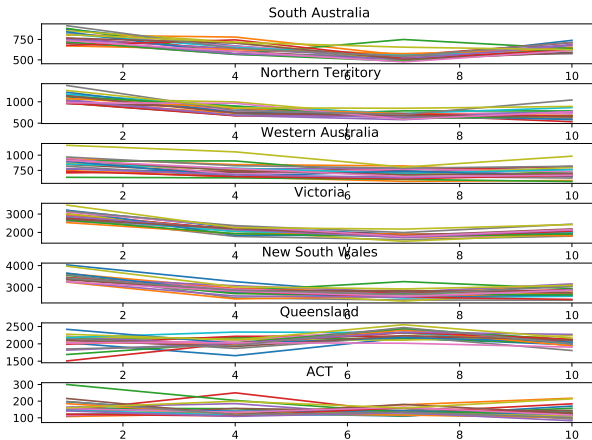
Australian Tourism

Seasonal Subplot per State by Year

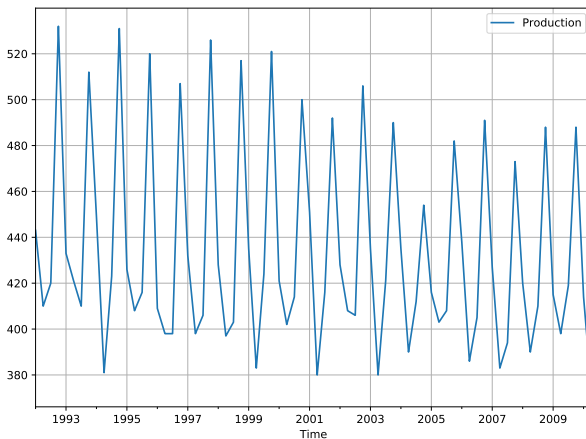
```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("data/tourism.csv", parse_dates=True,
                 index_col=1).drop(columns=["Unnamed: 0"])
df.columns = [col_name.lower() for col_name in df.columns]
fig, axs = plt.subplots(7, 1)
df['quarter'] = pd.DatetimeIndex(df.index)
df['year'] = pd.DatetimeIndex(df['quarter']).year
df = df.query('purpose == "Holiday"')
for idx, state in enumerate(df['state'].unique()):
    for idy, year in enumerate(df['year'].unique()):
        dfs = df.query('state == "' + state + '"')
        dfs = dfs.loc[str(year) + '-01-01': str(year) + '-12-31']
        dfy = dfs.groupby('quarter', as_index=False)
        dfy = dfy['trips'].agg(sum)
        dfy['month'] = pd.DatetimeIndex(dfy['quarter']).month
        dfy = dfy.sort_values(by=['month'], ascending=True)
        axs[idx].plot('month', 'trips', data=dfy, label=year)
```

Australian Tourism

Seasonal Subplot per State by Year



Australian Beer Production Plot > 1992

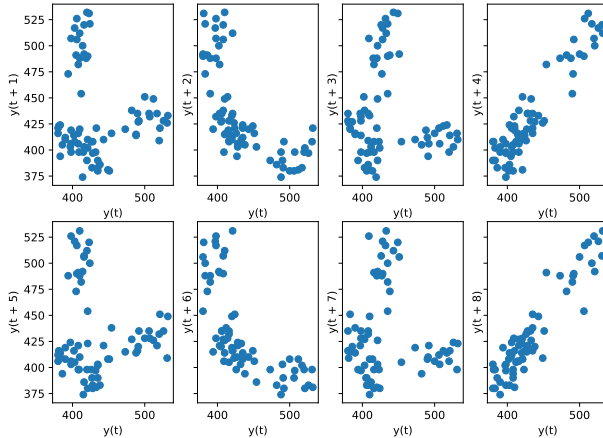


Australian Beer Production Lag Plot

```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("data/ausbeer.csv",
                  header=0, parse_dates=True, index_col=0)
df = df.loc['1992-01-01':'2020-01-01']
fig, axs = plt.subplots(2, 4, sharey=True)
pd.plotting.lag_plot(df, lag=1, ax=axs[0][0])
pd.plotting.lag_plot(df, lag=2, ax=axs[0][1])
pd.plotting.lag_plot(df, lag=3, ax=axs[0][2])
pd.plotting.lag_plot(df, lag=4, ax=axs[0][3])
pd.plotting.lag_plot(df, lag=5, ax=axs[1][0])
pd.plotting.lag_plot(df, lag=6, ax=axs[1][1])
pd.plotting.lag_plot(df, lag=7, ax=axs[1][2])
pd.plotting.lag_plot(df, lag=8, ax=axs[1][3])
plt.show()
```

Australian Beer Production Lag Plot

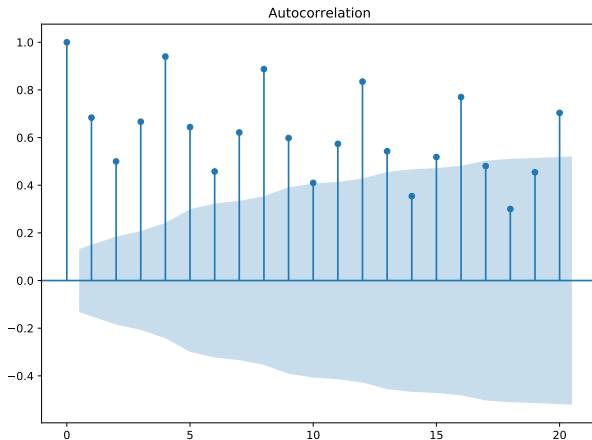


Australian Beer Production ACF Plot

```
from pandas import read_csv, DataFrame
from matplotlib import pyplot
from statsmodels.graphics.tsaplots import plot_acf

series = read_csv('data/ausbeer.csv', header=0,
                  parse_dates=True, index_col=0, squeeze=True)
plot_acf(series, lags=20)
pyplot.show()
```

Australian Beer Production ACF Plot



Australian Tourism ACF Plot

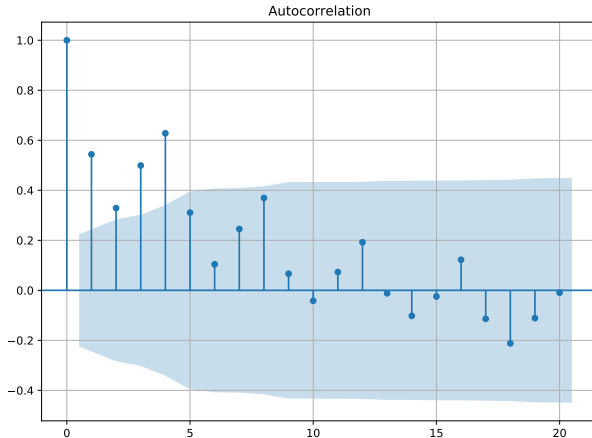
```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf

# load the data
df = pd.read_csv("data/tourism.csv",
                 parse_dates=True).drop(columns=["Unnamed: 0"])

# Total
dft = df.groupby("Quarter")
dft = dft['Trips'].agg(sum)
plot_acf(dft, lags=20)

plt.grid()
plt.show()
```

Australian Tourism ACF Plot



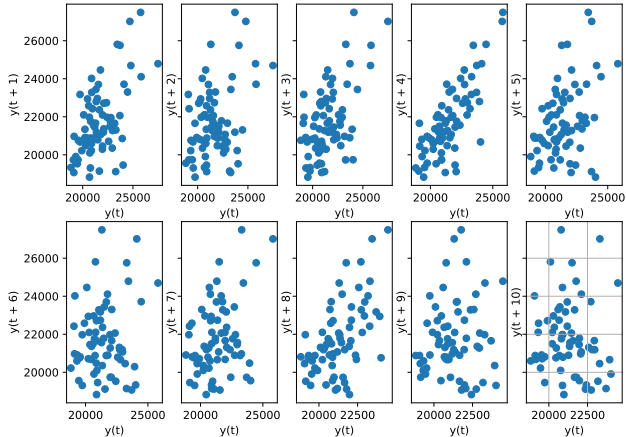
Australian Tourism Lag Plot

```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf

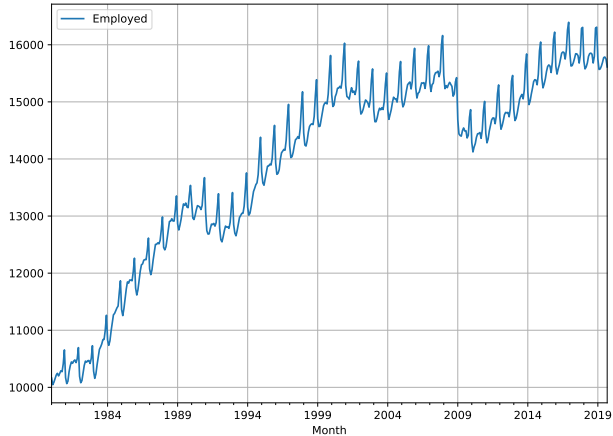
# load the data
df = pd.read_csv("data/tourism.csv",
                 parse_dates=True).drop(columns=["Unnamed: 0"])

# Total
dft = df.groupby("Quarter")
dft = dft['Trips'].agg(sum)
fig, axs = plt.subplots(2, 5, sharey=True)
pd.plotting.lag_plot(dft, lag=1, ax=axs[0][0])
pd.plotting.lag_plot(dft, lag=2, ax=axs[0][1])
pd.plotting.lag_plot(dft, lag=3, ax=axs[0][2])
pd.plotting.lag_plot(dft, lag=4, ax=axs[0][3])
pd.plotting.lag_plot(dft, lag=5, ax=axs[0][4])
pd.plotting.lag_plot(dft, lag=6, ax=axs[1][0])
pd.plotting.lag_plot(dft, lag=7, ax=axs[1][1])
pd.plotting.lag_plot(dft, lag=8, ax=axs[1][2])
pd.plotting.lag_plot(dft, lag=9, ax=axs[1][3])
```

Australian Tourism Lag Plot



Us Employment Retail Plot



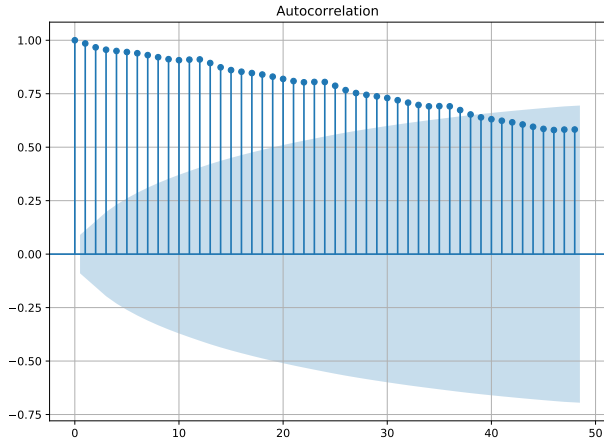
Us Employment ACF Plot

```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf

# load the data
df = pd.read_csv("data/usemployment.csv",
                 header=0, parse_dates=True, index_col=0)
dfs = df.query('Title == "Retail Trade"')
dfs = dfs.loc['1980-01-01':'2020-01-01']
plot_acf(dfs['Employed'], lags=48)

plt.grid()
plt.show()
```

Us Employment ACF Plot

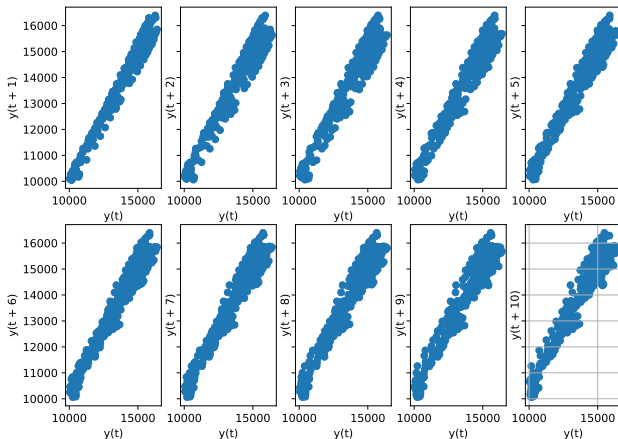


Us Employment Lag Plot

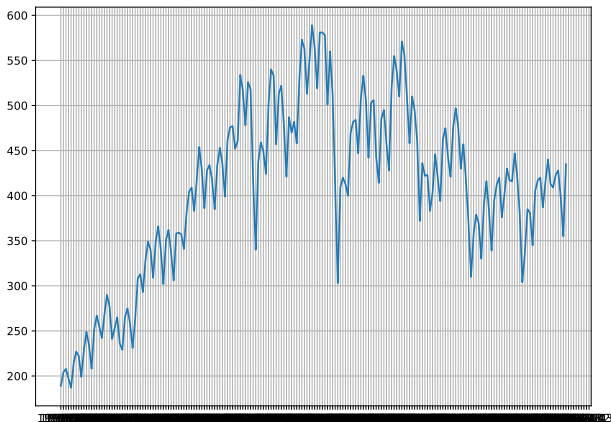
```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("data/usemployment.csv", header=0, parse_dates=True)
dfs = df.query('Title == "Retail Trade"')
dfs = dfs.loc['1980-01-01':'2020-01-01']
fig, axs = plt.subplots(2, 5, sharey=True)
pd.plotting.lag_plot(dfs['Employed'], lag=1, ax=axs[0][0])
pd.plotting.lag_plot(dfs['Employed'], lag=2, ax=axs[0][1])
pd.plotting.lag_plot(dfs['Employed'], lag=3, ax=axs[0][2])
pd.plotting.lag_plot(dfs['Employed'], lag=4, ax=axs[0][3])
pd.plotting.lag_plot(dfs['Employed'], lag=5, ax=axs[0][4])
pd.plotting.lag_plot(dfs['Employed'], lag=6, ax=axs[1][0])
pd.plotting.lag_plot(dfs['Employed'], lag=7, ax=axs[1][1])
pd.plotting.lag_plot(dfs['Employed'], lag=8, ax=axs[1][2])
pd.plotting.lag_plot(dfs['Employed'], lag=9, ax=axs[1][3])
pd.plotting.lag_plot(dfs['Employed'], lag=10, ax=axs[1][4])
plt.grid()
```


Us Employment Lag Plot



Australian Production Bricks Plot



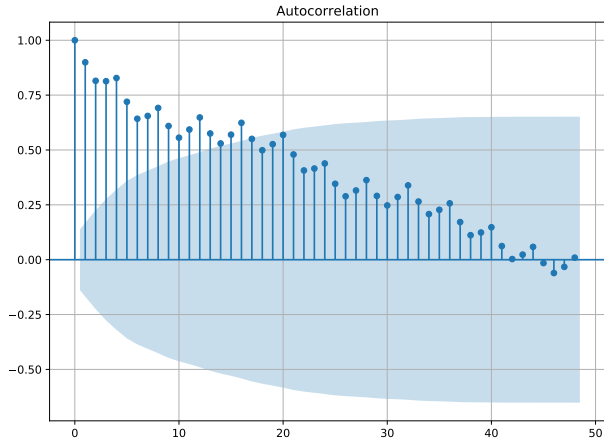
Australian Production Bricks ACF Plot

```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf

# load the data
df = pd.read_csv("data/aus_production.csv",
                 header=0, parse_dates=['Quarter'], index_col=0)
dfs = df.loc['1956 Q1':'2005 Q2']
plot_acf(dfs['Bricks'], lags=48)

# Total
plt.grid()
plt.show()
```

Australian Production Bricks ACF Plot

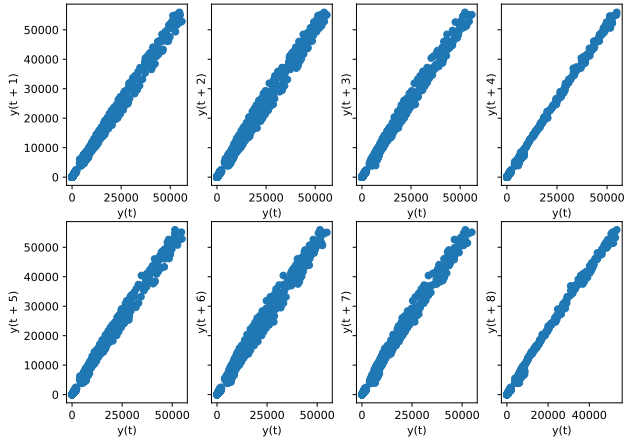


Australian Production Bricks Lag Plot

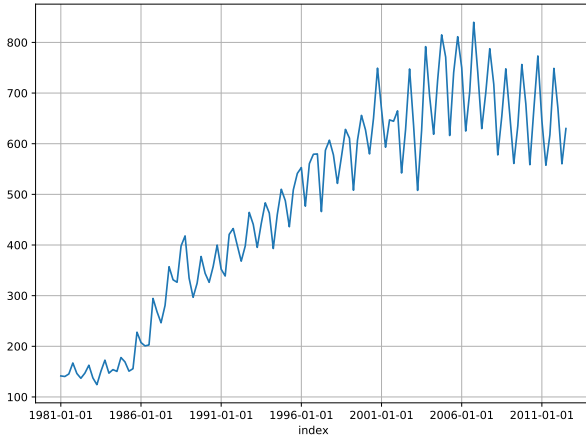
```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("data/aus_production.csv",
                  header=0, parse_dates=True, index_col=0)
dfs = df.loc['1956 Q1':'2005 Q2']
fig, axs = plt.subplots(2, 4, sharey=True)
pd.plotting.lag_plot(dfs, lag=1, ax=axs[0][0])
pd.plotting.lag_plot(dfs, lag=2, ax=axs[0][1])
pd.plotting.lag_plot(dfs, lag=3, ax=axs[0][2])
pd.plotting.lag_plot(dfs, lag=4, ax=axs[0][3])
pd.plotting.lag_plot(dfs, lag=5, ax=axs[1][0])
pd.plotting.lag_plot(dfs, lag=6, ax=axs[1][1])
pd.plotting.lag_plot(dfs, lag=7, ax=axs[1][2])
pd.plotting.lag_plot(dfs, lag=8, ax=axs[1][3])
plt.show()
```

Australian Production Bricks Lag Plot



Australian Arrivals Plot



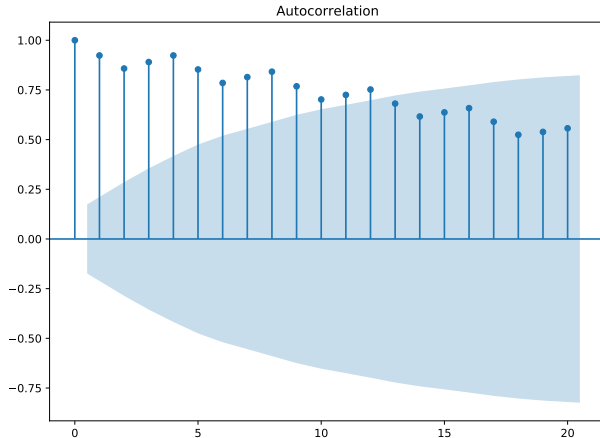
Australian Arrivals ACF Plot

```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf

# load the data
df = pd.read_csv("data/arrivals.csv", parse_dates=True)

# Total
dft = df.groupby("index")
dft = dft['value'].agg(sum)
plot_acf(dft, lags=20)
plt.show()
```

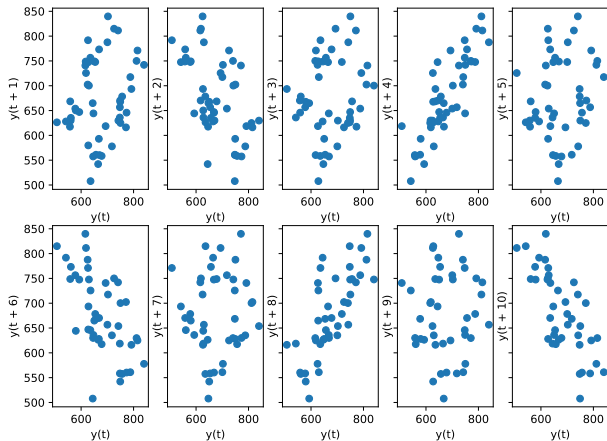

Australian Arrivals ACF Plot



Australian Arrivals Lag Plot

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("data/arrivals.csv",
                  header=0, parse_dates=True, index_col=0)
dft = df.groupby("index")
dft = dft['value'].agg(sum)
dft = dft.loc['2000-01-01':'2020-01-01']
fig, axs = plt.subplots(2, 5, sharey=True)
pd.plotting.lag_plot(dft, lag=1, ax=axs[0][0])
pd.plotting.lag_plot(dft, lag=2, ax=axs[0][1])
pd.plotting.lag_plot(dft, lag=3, ax=axs[0][2])
pd.plotting.lag_plot(dft, lag=4, ax=axs[0][3])
pd.plotting.lag_plot(dft, lag=5, ax=axs[0][4])
pd.plotting.lag_plot(dft, lag=6, ax=axs[1][0])
pd.plotting.lag_plot(dft, lag=7, ax=axs[1][1])
pd.plotting.lag_plot(dft, lag=8, ax=axs[1][2])
pd.plotting.lag_plot(dft, lag=9, ax=axs[1][3])
pd.plotting.lag_plot(dft, lag=10, ax=axs[1][4])
```

Australian Arrivals Lag Plot



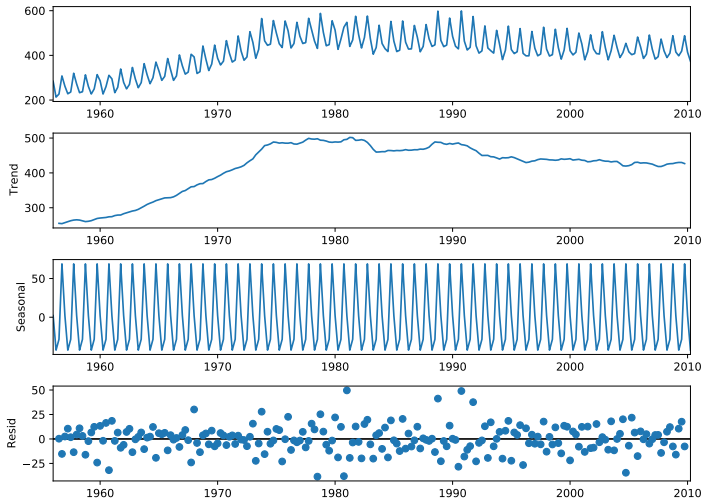
Australian Beer STL Decomposition

```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.seasonal import seasonal_decompose

df = pd.read_csv("data/ausbeer.csv",
                 header=0, parse_dates=True, index_col=0)

decomposition = seasonal_decompose(df, model='additive')
fig = decomposition.plot()
plt.show()
```

Australian Beer STL Decomposition



Australian Tourism STL Decomposition

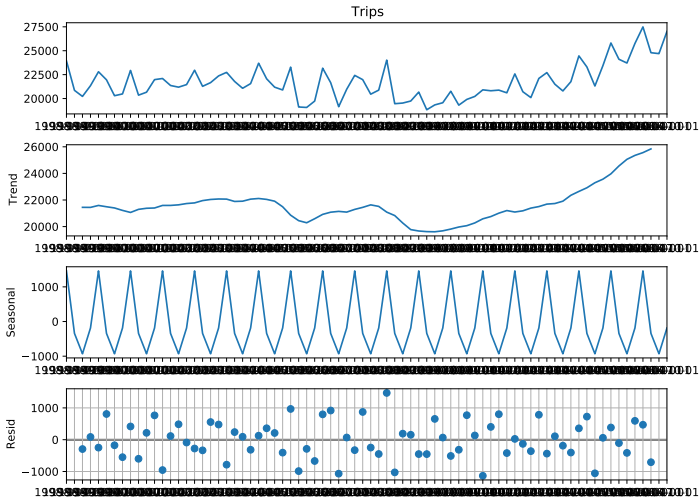
```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.seasonal import seasonal_decompose

# load the data
df = pd.read_csv("data/tourism.csv",
                 parse_dates=True).drop(columns=["Unnamed: 0"])

# Total
dft = df.groupby("Quarter")
dft = dft['Trips'].agg(sum)
decomposition = seasonal_decompose(dft, model='additive', period=4)
fig = decomposition.plot()

plt.grid()
plt.show()
```

Australian Tourism STL Decomposition



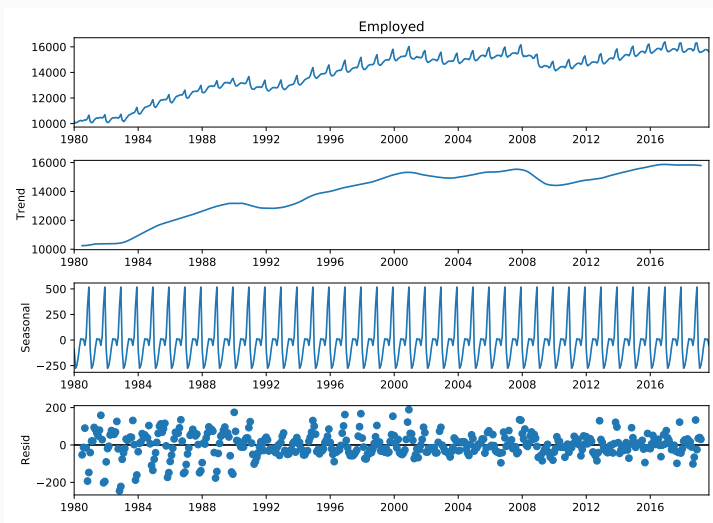
US Employment STL Decomposition

```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.seasonal import seasonal_decompose

# load the data
df = pd.read_csv("data/usemployment.csv",
                 header=0, parse_dates=True, index_col=0)

dfs = df.query('Title == "Retail Trade"')
dfs = dfs.loc['1980-01-01':'2020-01-01']
decomposition = seasonal_decompose(dfs['Employed'], model='additive')
fig = decomposition.plot()
plt.show()
```


US Employment STL Decomposition

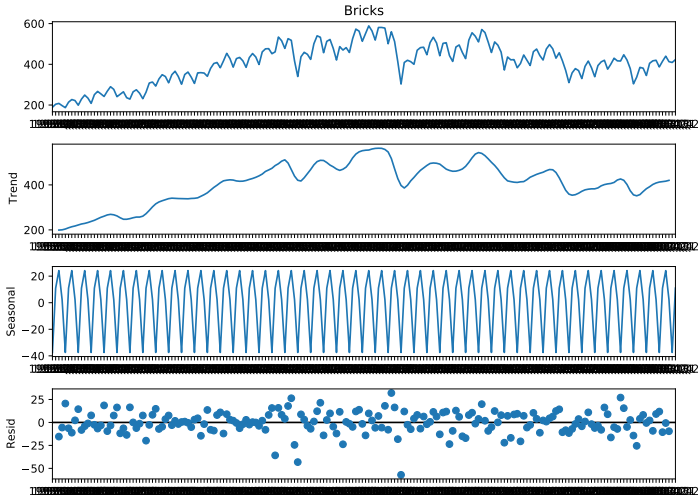


Australian Brick Production STL Decomposition

```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.seasonal import seasonal_decompose

# load the data
df = pd.read_csv("data/aus_production.csv",
                 header=0, parse_dates=True, index_col=0)
dfs = df.loc['1956 Q1':'2004 Q2']
# Total
decomposition = seasonal_decompose(dfs['Bricks'],
                                   model='additive', period=4)
fig = decomposition.plot()
plt.show()
```

Australian Brick Production STL Decomposition



Australian Arrivals STL Decomposition

```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.seasonal import seasonal_decompose
# load the data
df = pd.read_csv("data/arrivals.csv", parse_dates=True)
# Total
dft = df.groupby("index")
dft = dft['value'].agg(sum)
decomposition = seasonal_decompose(dft, model='additive', period=4)
fig = decomposition.plot()
plt.show()
```

Australian Arrivals STL Decomposition

