

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
import statsmodels.tsa.api as smt
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
```

Loading dataset

```
In [2]: df = pd.read_csv('data/owid-covid-data.csv', index_col=['date'], parse_dates=['date'])
df
```

```
Out[2]:
```

	iso_code	continent	location	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_death
date									
2020-01-03	AFG	Asia	Afghanistan	NaN	0.0	NaN	NaN	0.0	
2020-01-04	AFG	Asia	Afghanistan	NaN	0.0	NaN	NaN	0.0	
2020-01-05	AFG	Asia	Afghanistan	NaN	0.0	NaN	NaN	0.0	
2020-01-06	AFG	Asia	Afghanistan	NaN	0.0	NaN	NaN	0.0	
2020-01-07	AFG	Asia	Afghanistan	NaN	0.0	NaN	NaN	0.0	
...	
2023-03-03	ZWE	Africa	Zimbabwe	264127.0	0.0	11.571	5668.0	0.0	
2023-03-04	ZWE	Africa	Zimbabwe	264127.0	0.0	4.571	5668.0	0.0	
2023-03-05	ZWE	Africa	Zimbabwe	264127.0	0.0	2.714	5668.0	0.0	
2023-03-06	ZWE	Africa	Zimbabwe	264127.0	0.0	0.000	5668.0	0.0	
2023-03-07	ZWE	Africa	Zimbabwe	264127.0	0.0	0.000	5668.0	0.0	

293542 rows x 66 columns



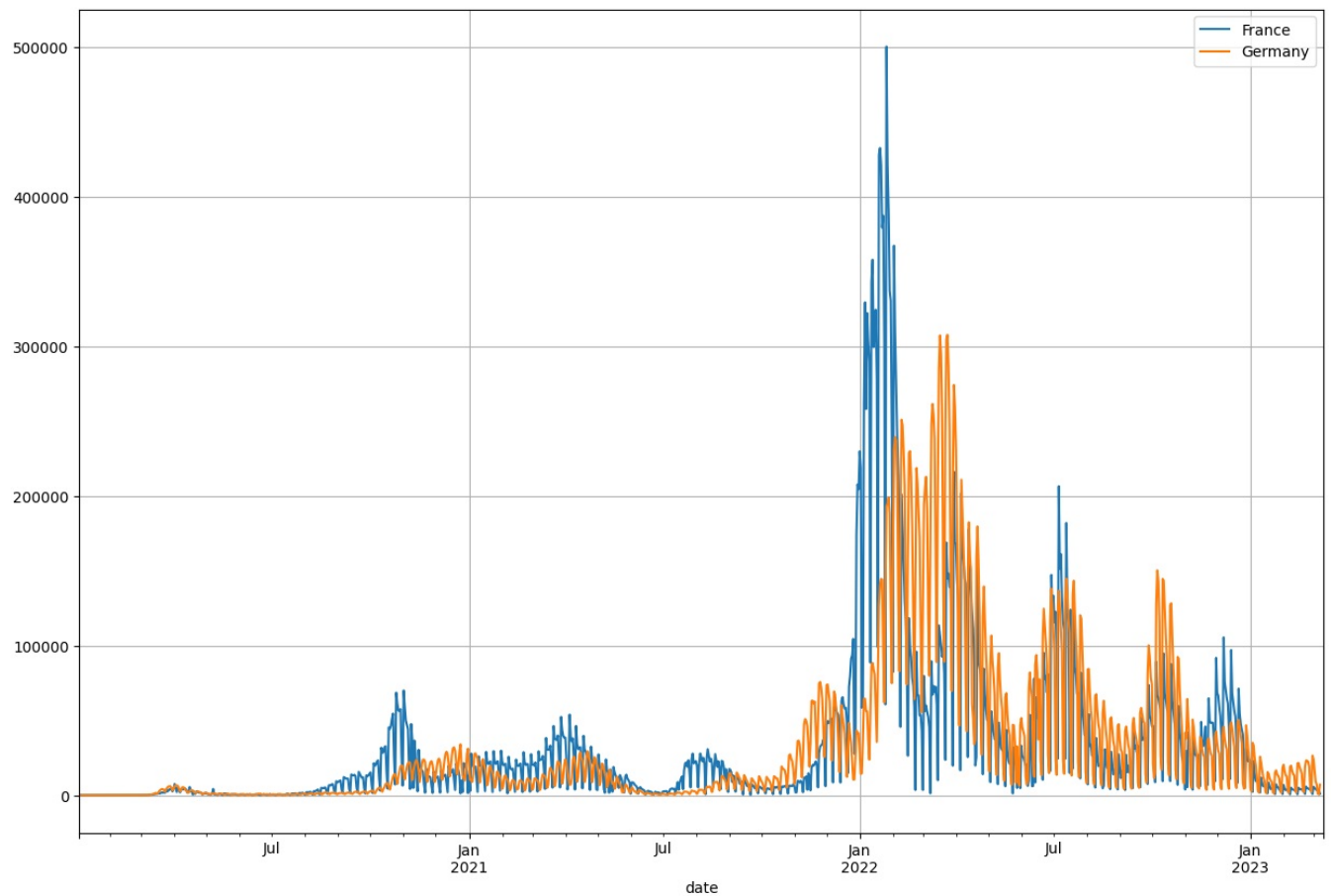
Countries

```
In [3]: countries = ['France', 'Germany']
```

Statistic of COVID-19 emerging cases per day

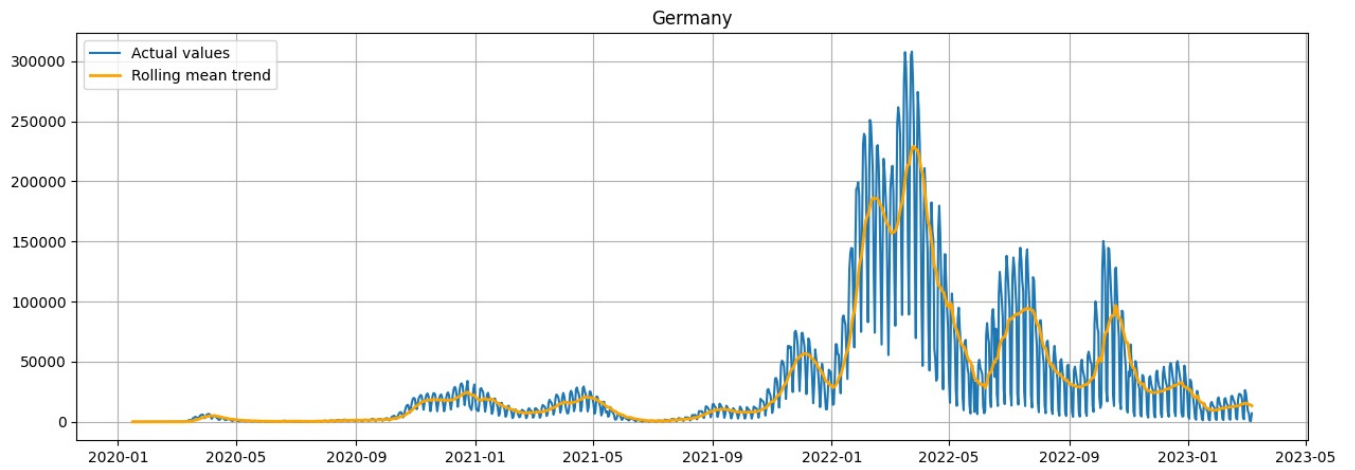
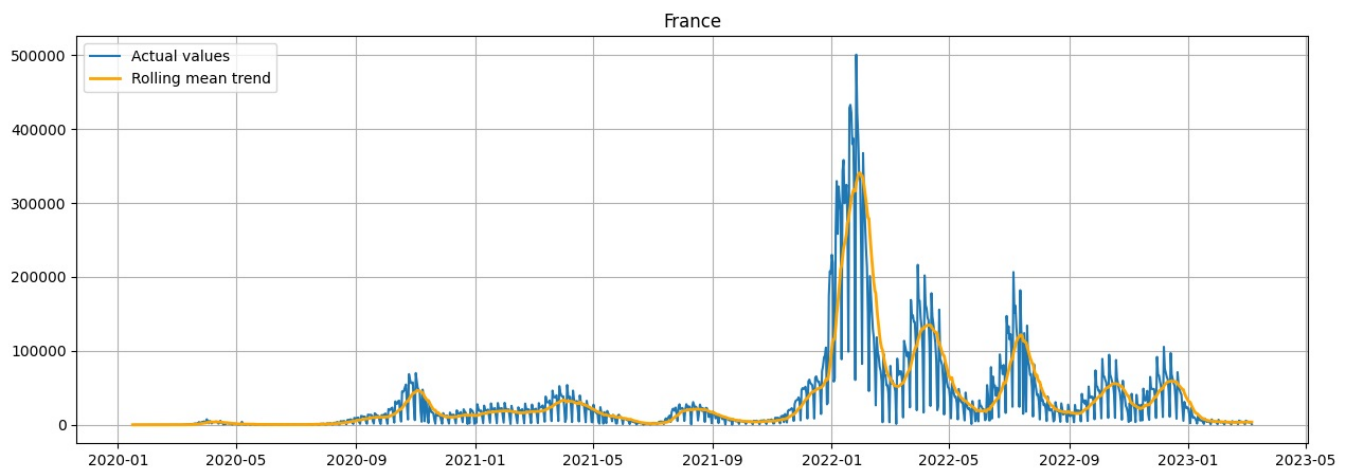
```
In [4]: fig, ax = plt.subplots(figsize=(15, 10))
for name in countries:
    df[df.location == name].new_cases.plot(ax=ax)
ax.grid()
plt.legend(countries)
```

```
Out[4]: <matplotlib.legend.Legend at 0x7fdf00f1b8b0>
```



Graph Smoothing

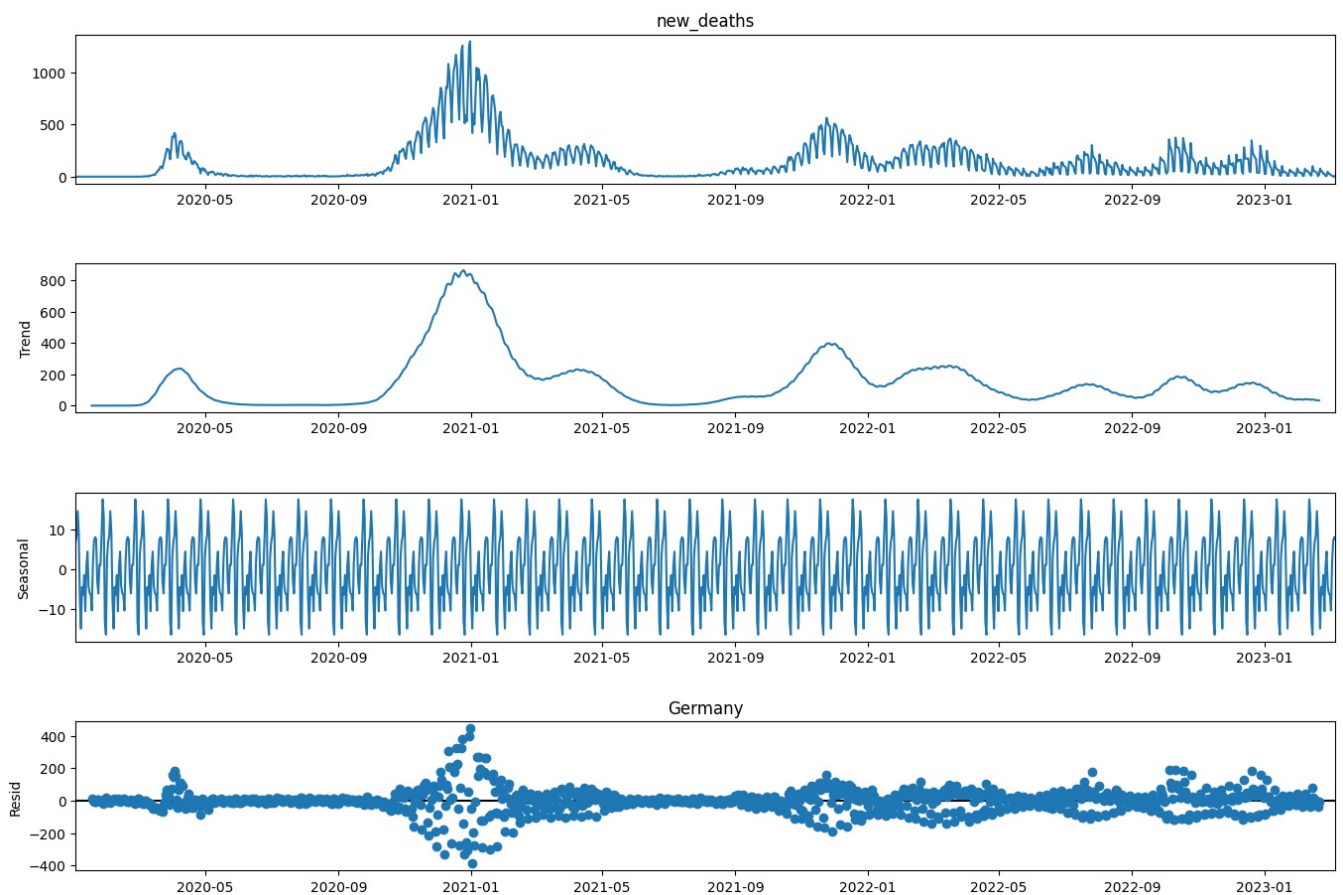
```
In [5]: def plot_moving_average(series, n):
    rolling_mean = series.rolling(window=n).mean()
    plt.figure(figsize=(15, 5))
    plt.title(f'Moving average\n window size = {n}')
    plt.plot(series[n:], label='Actual values')
    plt.plot(rolling_mean, c='orange', linewidth=2, label='Rolling mean trend')
    plt.legend(loc='upper left')
    plt.grid(True)
    for name in countries:
        plot_moving_average(df[df.location == name].new_cases, 14)
    plt.title(name)
```



Decomposition

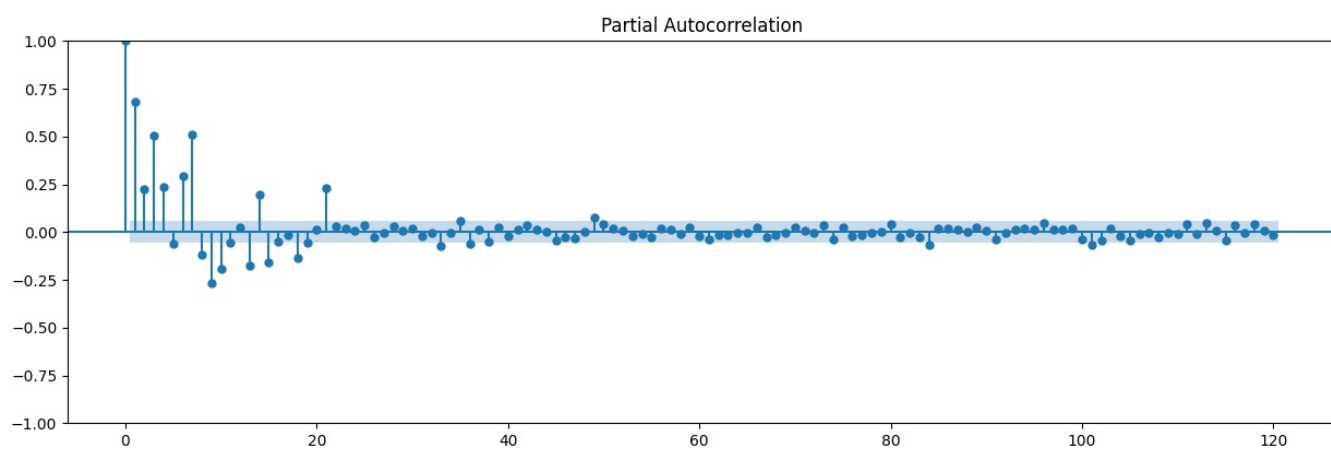
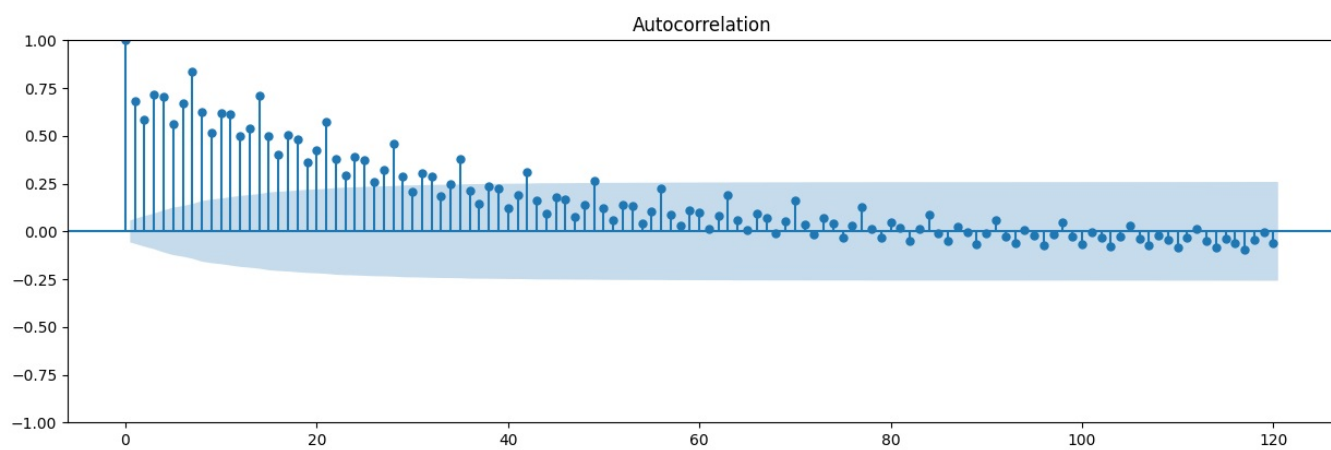
```
In [6]: for name in countries:
smt.seasonal_decompose(df[(~df.new_deaths.isna()) & (df.location == name)].new_deaths, period=30).plot().se
plt.title(name)
plt.show()
```

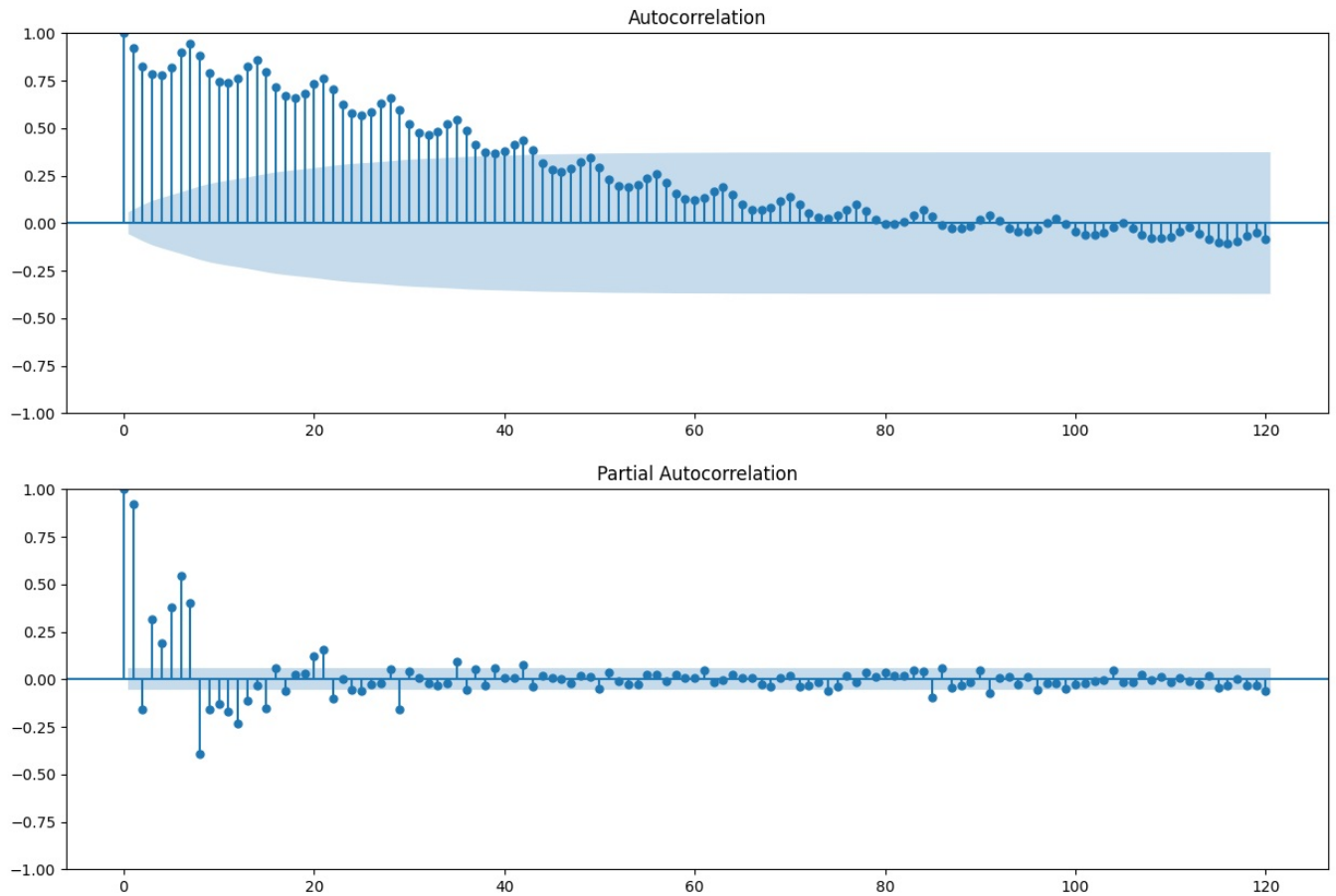




```
In [7]: for name in countries:
    fig, ax = plt.subplots(2, figsize=(15, 10))
    dd = df[(~df.new_deaths.isna()) & (df.location == name)].new_deaths
    ax[0] = plot_acf(dd, ax=ax[0], lags=120)
    ax[1] = plot_pacf(dd, ax=ax[1], lags=120)
```

/usr/local/lib/python3.10/dist-packages/statsmodels/graphics/tsaplots.py:348: FutureWarning: The default method 'yw' can produce PACF values outside of the [-1,1] interval. After 0.13, the default will change to unadjusted Yule-Walker ('ywm'). You can use this method now by setting method='ywm'.
 warnings.warn(





Stationarity

```
In [8]: def dickey_fuller_test(series):
test = smt.adfuller(series, autolag='AIC')
print('adf: ', test[0])
print('p-value: ', test[1])
print('Critical values: ', test[4])
if test[0] > test[4]['5%']:
    print('Навні одиничні корені, ряд не стаціонарний.')
else:
    print('Одиничні корені відсутні, ряд є стаціонарним.')

for name in countries:
    print(name)
    dickey_fuller_test(df[(~df.new_deaths.isna()) & (df.location == name)].new_deaths)
```

France
adf: -3.863007825198076
p-value: 0.002324600221443343
Critical values: {'1%': -3.436114401808766, '5%': -2.8640853428381092, '10%': -2.568125207156112}
Одиничні корені відсутні, ряд є стаціонарним.

Germany
adf: -3.2158987236395693
p-value: 0.019082647020864603
Critical values: {'1%': -3.4361093249345402, '5%': -2.8640831032339706, '10%': -2.5681240143809787}
Одиничні корені відсутні, ряд є стаціонарним.

Correlation

```
In [9]: df[df.location == countries[0]].new_cases.corr(df[df.location == countries[1]].new_cases)
```

```
Out[9]: 0.5911568933928973
```

```
In [10]: import math
from sklearn.preprocessing import MinMaxScaler
from keras.models import Sequential
from keras.layers import Dense, LSTM

df = pd.read_csv('data/usd_uah.csv', index_col='Date', parse_dates=['Date'])[::-1].loc[:, 'Price']
df
```

2023-03-11 22:17:22.439184: I tensorflow/core/platform/cpu_feature_guard.cc:193] This TensorFlow binary is optimized with oneAPI Deep Neural Network Library (oneDNN) to use the following CPU instructions in performance-critical operations: AVX2 FMA

To enable them in other operations, rebuild TensorFlow with the appropriate compiler flags.

2023-03-11 22:17:22.624193: W tensorflow/compiler/xla/stream_executor/platform/default/dso_loader.cc:64] Could not load dynamic library 'libcudart.so.11.0'; dlderror: libcudart.so.11.0: cannot open shared object file: No such file or directory

2023-03-11 22:17:22.624221: I tensorflow/compiler/xla/stream_executor/cuda/cudart_stub.cc:29] Ignore above cuda runtime dlderror if you do not have a GPU set up on your machine.

2023-03-11 22:17:23.737915: W tensorflow/compiler/xla/stream_executor/platform/default/dso_loader.cc:64] Could not load dynamic library 'libnvinfer.so.7'; dlderror: libnvinfer.so.7: cannot open shared object file: No such file or directory

2023-03-11 22:17:23.738013: W tensorflow/compiler/xla/stream_executor/platform/default/dso_loader.cc:64] Could not load dynamic library 'libnvinfer_plugin.so.7'; dlderror: libnvinfer_plugin.so.7: cannot open shared object file: No such file or directory

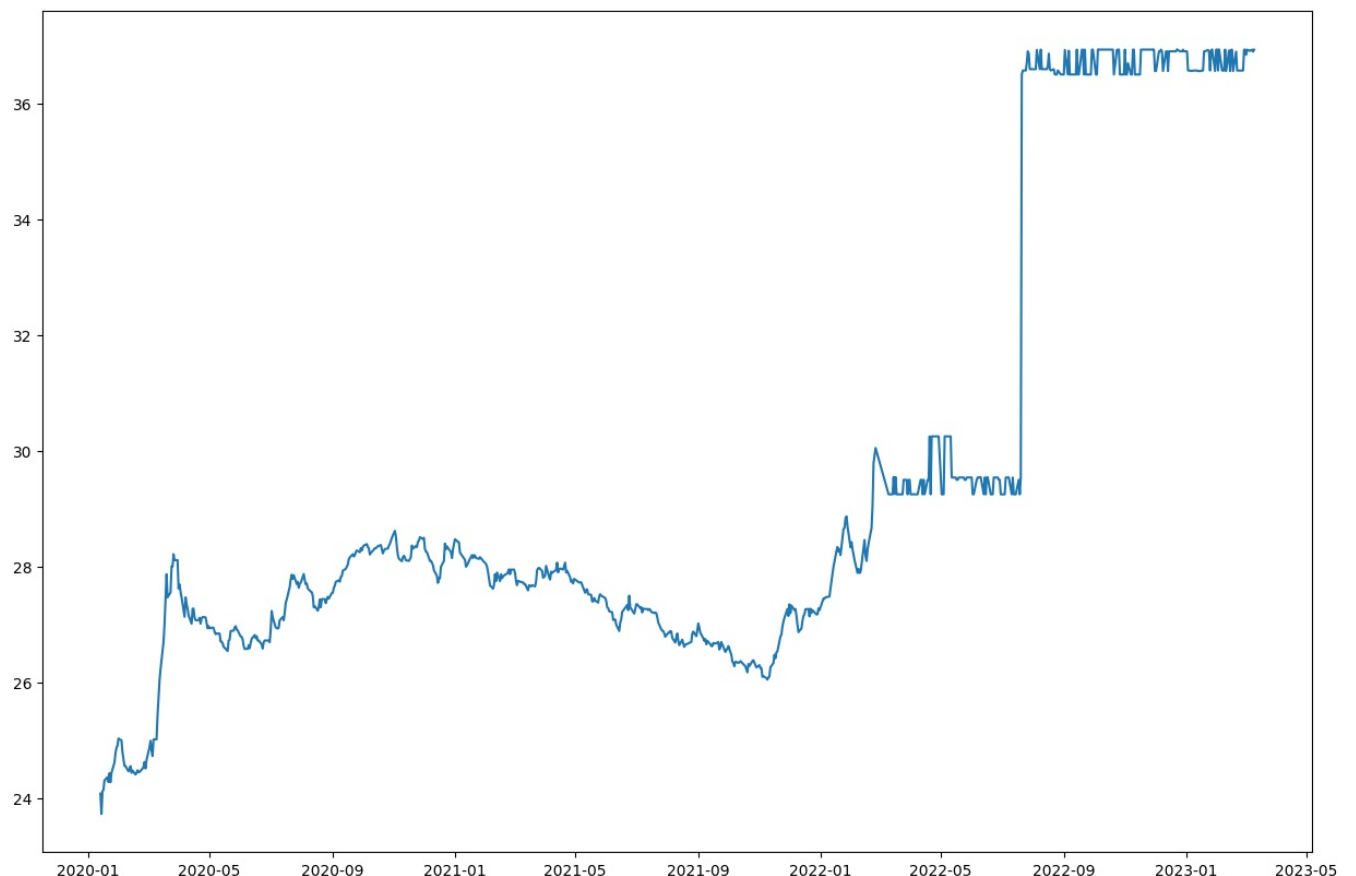
2023-03-11 22:17:23.738023: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38] TF-TRT Warning: Cannot dlopen some TensorRT libraries. If you would like to use Nvidia GPU with TensorRT, please make sure the missing libraries mentioned above are installed properly.

```
Out[10]: Date
2020-01-13    24.0813
2020-01-14    23.7300
2020-01-15    24.1175
2020-01-16    24.1450
2020-01-17    24.3084
...
2023-03-06    36.9070
2023-03-07    36.9200
2023-03-08    36.9290
2023-03-09    36.8950
2023-03-10    36.9290
Name: Price, Length: 798, dtype: float64

Rate Statistic
```

```
In [11]: fig, axis = plt.subplots(figsize=(15, 10))
axis.plot(df)
```

```
Out[11]: [<matplotlib.lines.Line2D at 0x7fdeed1e17e0>]
```



```
In [29]: dataset = df.values.reshape(-1, 1)
```

Training Data Length

```
In [13]: training_data_len = math.ceil(dataset.shape[0] * 0.8)
training_data_len
```

```
Out[13]: 639
```

Scale Data

```
In [30]: scaler = MinMaxScaler(feature_range=(0, 1))
scaled_data = scaler.fit_transform(dataset)
```

Training Data

```
In [15]: train_data = scaled_data[0:training_data_len, :]
train_period = 60
x_train, y_train = [], []
for i in range(train_period, len(train_data)):
    x_train.append(train_data[i-train_period:i, 0])
    y_train.append(train_data[i, 0])
x_train, y_train = np.array(x_train), np.array(y_train)
```

Reshape

```
In [16]: x_train = np.reshape(x_train, (x_train.shape[0], x_train.shape[1], 1))
x_train.shape
```

```
Out[16]: (579, 60, 1)
```

LSTM Model

```
In [17]: model = Sequential()
model.add(LSTM(50, return_sequences=True, input_shape=(x_train.shape[1], 1)))
model.add(LSTM(50, return_sequences=False))
model.add(Dense(25))
model.add(Dense(1))
```

```
2023-03-11 22:17:25.248382: W tensorflow/compiler/xla/stream_executor/platform/default/dso_loader.cc:64] Could
not load dynamic library 'libcuda.so.1'; dLError: libcuda.so.1: cannot open shared object file: No such file or
directory
2023-03-11 22:17:25.248416: W tensorflow/compiler/xla/stream_executor/cuda/cuda_driver.cc:265] failed call to c
uInit: UNKNOWN ERROR (303)
2023-03-11 22:17:25.248448: I tensorflow/compiler/xla/stream_executor/cuda/cuda_diagnostics.cc:156] kernel driv
er does not appear to be running on this host (localhost): /proc/driver/nvidia/version does not exist
2023-03-11 22:17:25.248843: I tensorflow/core/platform/cpu_feature_guard.cc:193] This TensorFlow binary is opti
mized with oneAPI Deep Neural Network Library (oneDNN) to use the following CPU instructions in performance-cri
tical operations: AVX2 FMA
To enable them in other operations, rebuild TensorFlow with the appropriate compiler flags.
```

Compile model

```
In [18]: model.compile(optimizer='adam', loss='mean_squared_error')
```

Train Model

```
In [19]: model.fit(x_train, y_train, batch_size=1, epochs=1)
```

```
579/579 [=====] - 18s 27ms/step - loss: 0.0051
```

```
Out[19]: <keras.callbacks.History at 0x7fdeecd136d0>
```

Testing Dataset

```
In [20]: test_data = scaled_data[training_data_len - train_period:, :]
x_test = []
y_test = dataset[training_data_len:, :]
for i in range(train_period, len(test_data)):
    x_test.append(test_data[i-train_period:i, 0])
x_test = np.array(x_test)
x_test = np.reshape(x_test, (x_test.shape[0], x_test.shape[1], 1))
```

Predictions

```
In [21]: predictions = model.predict(x_test)
predictions = scaler.inverse_transform(predictions)
```

```
5/5 [=====] - 1s 12ms/step
```


Get the root mean squared value

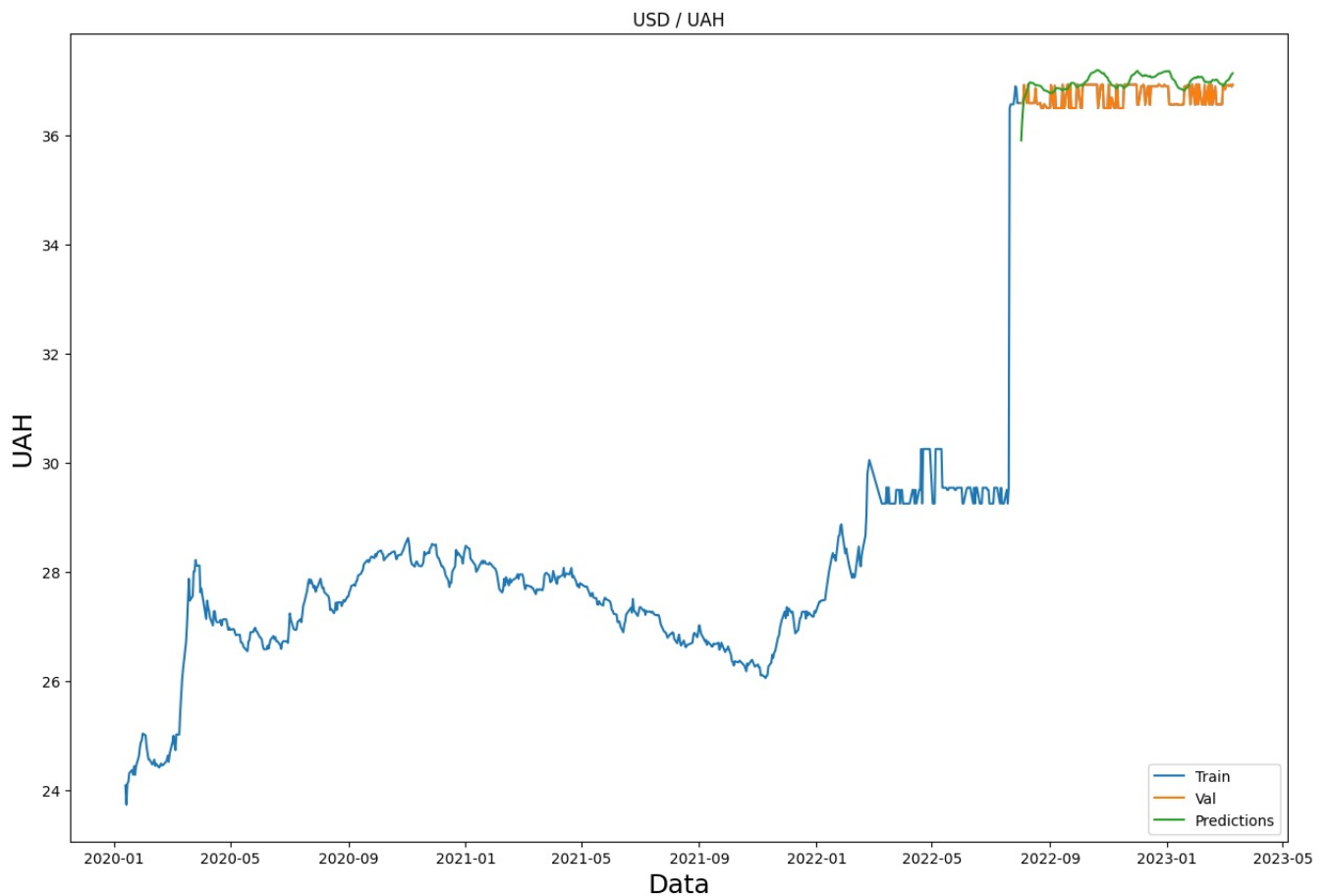
```
In [22]: rmse = np.sqrt(np.mean(predictions - y_test)**2)
rmse
```

```
Out[22]: 0.23806488436333012
```

Plot Data

```
In [23]: train = df[:training_data_len]
valid = df[training_data_len:].to_frame()
valid.columns = ['Valid']
valid['Predictions'] = predictions
plt.figure(figsize=(15, 10))
plt.title('USD / UAH')
plt.xlabel('Data', fontsize=18)
plt.ylabel('UAH', fontsize=18)
plt.plot(df)
plt.plot(valid[['Valid', 'Predictions']])
plt.legend(['Train', 'Val', 'Predictions'], loc='lower right')
```

```
Out[23]: <matplotlib.legend.Legend at 0x7fdea4496d40>
```



Additional Assignment

```
In [24]: from my_predictions import get_predictions
import datetime as dt
path = 'data/seattleWeather_1948-2017 copy.csv'
df = pd.read_csv(path, index_col=['DATE'], parse_dates=['DATE'])
def plot_predictions(df: pd.DataFrame, predictions: np.array, column_train: str):
    train = df[column_train]
    days = predictions.shape[0]
    end_date: dt.date = df.tail(1).index.item().to_pydatetime().date()
    index = [end_date + dt.timedelta(days=i) for i in range(days)]
    valid = pd.Series(predictions, index=index)
    plt.figure(figsize=(15, 10))
    plt.title(column_train)
    plt.xlabel('Date', fontsize=18)
    plt.ylabel(column_train, fontsize=18)
    plt.plot(train)
    plt.plot(valid)
    plt.legend(['Train', 'Val'], loc='lower right')
df
```

Out[24]:

	PRCP	TMAX	TMIN	RAIN
DATE				
1948-01-01	0.47	51	42	True
1948-01-02	0.59	45	36	True
1948-01-03	0.42	45	35	True
1948-01-04	0.31	45	34	True
1948-01-05	0.17	45	32	True
...
1953-03-27	0.47	52	39	True
1953-03-28	0.51	49	38	True
1953-03-29	0.00	52	38	False
1953-03-30	0.18	49	34	True
1953-03-31	0.04	48	32	True

1917 rows × 4 columns

In [27]:

```
predictions = get_predictions(df, 'TMAX', 365, 100, 40)
plot_predictions(df, predictions, 'TMAX')

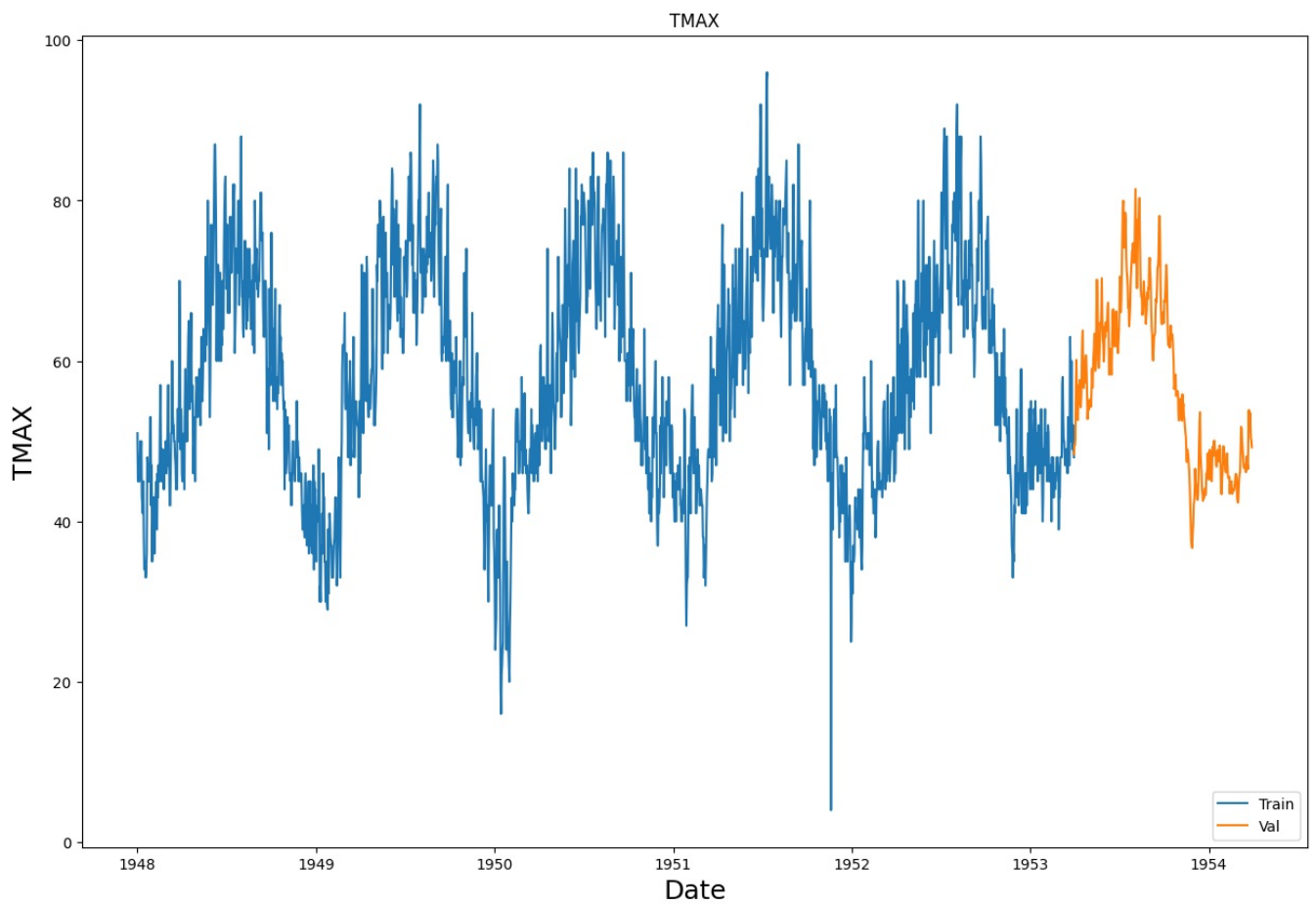
1552/1552 [=====] - 135s 85ms/step - loss: 0.0082
1/1 [=====] - 0s 496ms/step
1/1 [=====] - 0s 49ms/step
1/1 [=====] - 0s 41ms/step
1/1 [=====] - 0s 46ms/step
1/1 [=====] - 0s 39ms/step
1/1 [=====] - 0s 40ms/step
1/1 [=====] - 0s 41ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 42ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 40ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 39ms/step
1/1 [=====] - 0s 41ms/step
1/1 [=====] - 0s 41ms/step
1/1 [=====] - 0s 51ms/step
1/1 [=====] - 0s 39ms/step
1/1 [=====] - 0s 41ms/step
1/1 [=====] - 0s 39ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 39ms/step
1/1 [=====] - 0s 36ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 39ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 40ms/step
1/1 [=====] - 0s 44ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 42ms/step
1/1 [=====] - 0s 39ms/step
1/1 [=====] - 0s 40ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 43ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 36ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 40ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 36ms/step
1/1 [=====] - 0s 35ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 36ms/step
1/1 [=====] - 0s 43ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 36ms/step
1/1 [=====] - 0s 46ms/step
```

[illegible]

[illegible]

[illegible]

1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 39ms/step
1/1 [=====] - 0s 41ms/step
1/1 [=====] - 0s 45ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 46ms/step
1/1 [=====] - 0s 39ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 36ms/step
1/1 [=====] - 0s 36ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 39ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 40ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 58ms/step
1/1 [=====] - 0s 45ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 47ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 42ms/step
1/1 [=====] - 0s 40ms/step
1/1 [=====] - 0s 42ms/step
1/1 [=====] - 0s 39ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 39ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 36ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 40ms/step
1/1 [=====] - 0s 39ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 40ms/step
1/1 [=====] - 0s 39ms/step
1/1 [=====] - 0s 40ms/step
1/1 [=====] - 0s 39ms/step
1/1 [=====] - 0s 45ms/step
1/1 [=====] - 0s 41ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 36ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 39ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 41ms/step
1/1 [=====] - 0s 39ms/step



```
In [28]: predictions = get_predictions(df, 'TMIN', 365, 100, 40)
plot_predictions(df, predictions, 'TMIN')
```

```
1552/1552 [=====] - 138s 87ms/step - loss: 0.0102
1/1 [=====] - 1s 507ms/step
1/1 [=====] - 0s 36ms/step
1/1 [=====] - 0s 36ms/step
1/1 [=====] - 0s 35ms/step
1/1 [=====] - 0s 44ms/step
1/1 [=====] - 0s 36ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 35ms/step
1/1 [=====] - 0s 35ms/step
1/1 [=====] - 0s 35ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 40ms/step
1/1 [=====] - 0s 41ms/step
1/1 [=====] - 0s 36ms/step
1/1 [=====] - 0s 36ms/step
1/1 [=====] - 0s 35ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 40ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 44ms/step
```

[illegible]

[illegible]

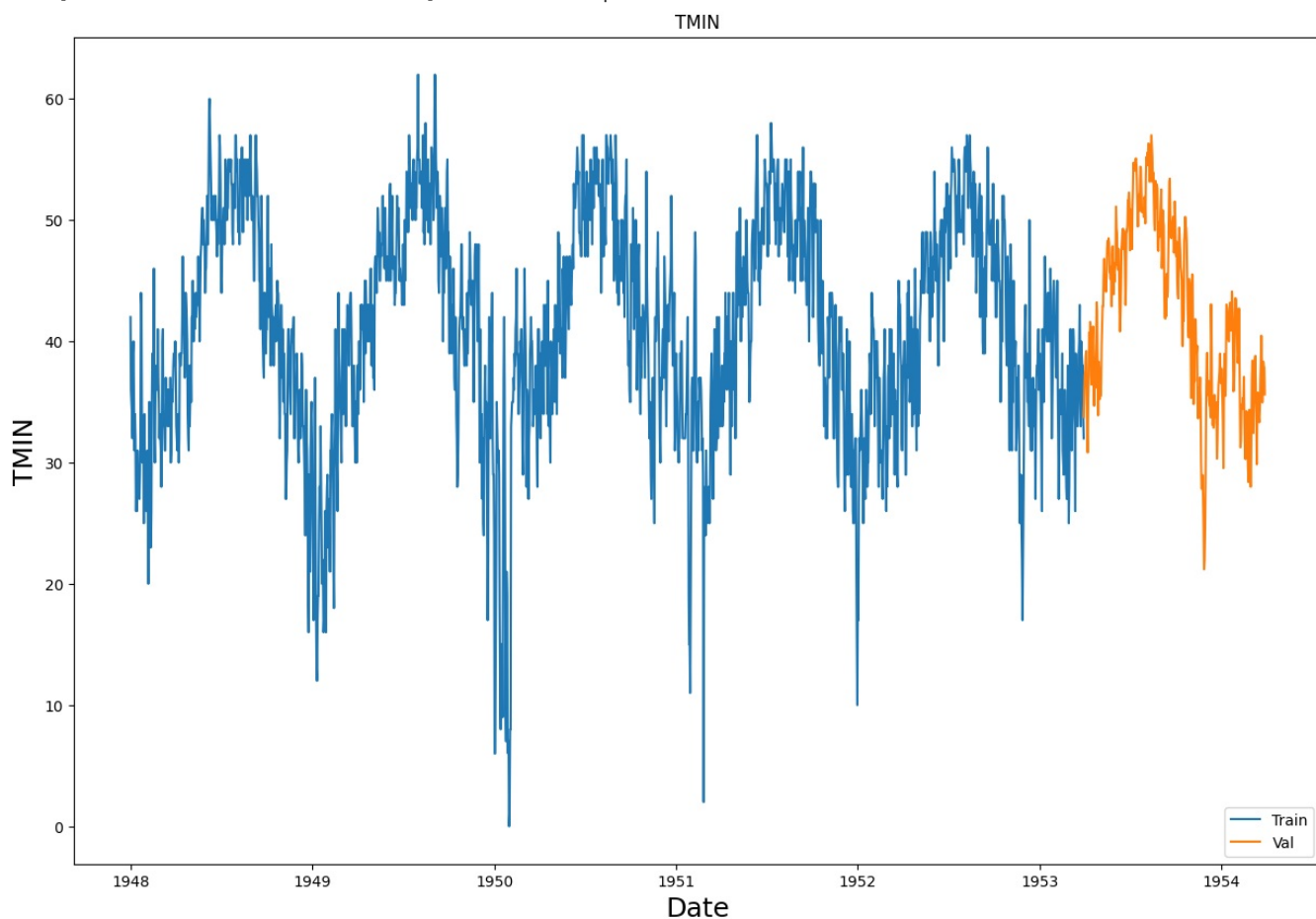
[illegible]

[illegible]

```

1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 40ms/step
1/1 [=====] - 0s 36ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 43ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 39ms/step
1/1 [=====] - 0s 39ms/step
1/1 [=====] - 0s 39ms/step
1/1 [=====] - 0s 38ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 37ms/step
1/1 [=====] - 0s 37ms/step

```



Results

Під час лабораторної роботи було виконано тренування моделей для прогнозування опадів, курсу гривні до долара. Курс гривні до долара показує, гарні тестові результати з невеликою середньоквадратичною похибкою. Під час тренування моделі для опадів бачимо графіки, прогнозовані значення на яких відповідають періодам опадів минулих років; можна сказати, що модель працює добре. Також була проаналізована статистика поширення COVID-19, побудовано графіки сезонних декомпозицій.