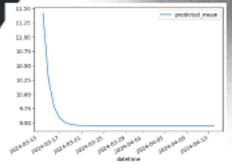
Chicago Climate Forecasting

Priyadarshini Rajendran | A20476470 | Kyung Jin Kwak | A20497336 |
Shwedha Srinivasan | A20514543 |

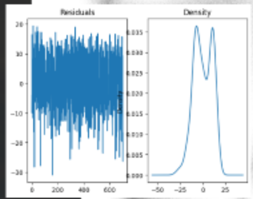
MATH 546-01: INTRODUCTION TO TIME SERIES:

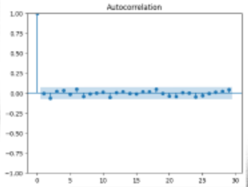
ARIMA

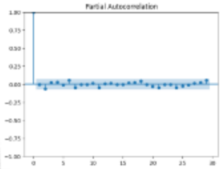


1. This class of ARIMA models has a general term ARIMA(p, d, q) that includes three components (or terms): autoregressive (AR), integrated (or differencing), and moving average (MA) terms with the corresponding order of p, d, and q. The magnitude of temporal correlation exhibited in the time series will determine the AR and MA terms while the differencing term can transform a nonstationary series to be stationary

2. p = 1| d = 1 | q = 1 from the ACF and PACF plots as well as the AIC and BIC values based on Python model fit.







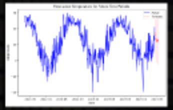
3. Residuals should be white noise and make sure not to display seasonality.

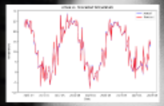
SARIMAX

1. SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables) is an extension of the ARIMA model that incorporates additional variables, known as exogenous variables or regressors, into the model.

2. The (P,D,Q,M) Order refers to the seasonal component of the model for the Auto Regressive parameters, differences, Moving Average parameters, and periodicity:

- D indicates the integration order of the seasonal process (the number of transformation needed to make stationary the time series)
- P indicates the Auto Regressive order for the seasonal component
- Q indicated the Moving Average order for the seasonal component
- M indicates the periodicity, i.e. the number of periods in season, such as 12 for monthly data.

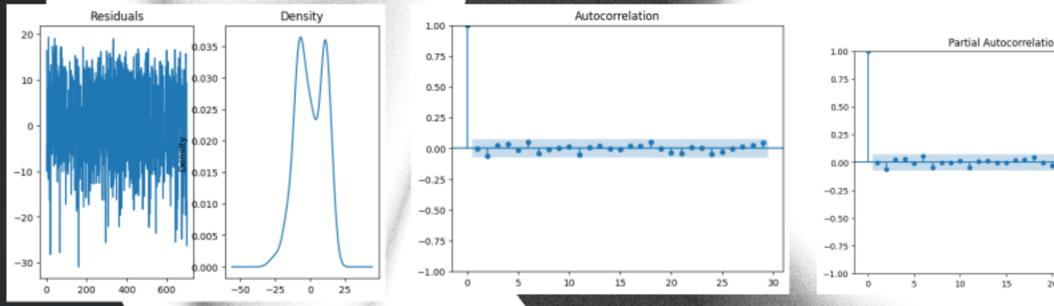




FUTURE WORK

- A wider array of data is to be included for more accurate predictions.
- Add categorical features in the model.
- Integrate machine learning model.

MATH 546-01: INTRODUCTION TO TIME SERIES:



Chicago Climate Forecasting

Priyadarshini Rajendran | A20476470 | Kyung Jin Kwak | A20497336 |
Shwedha Srinivasan | A20514543 |

INTRODUCTION

1.

Our team developed a system to predict Chicago's weather, focusing on extreme temperature changes due to global warming.

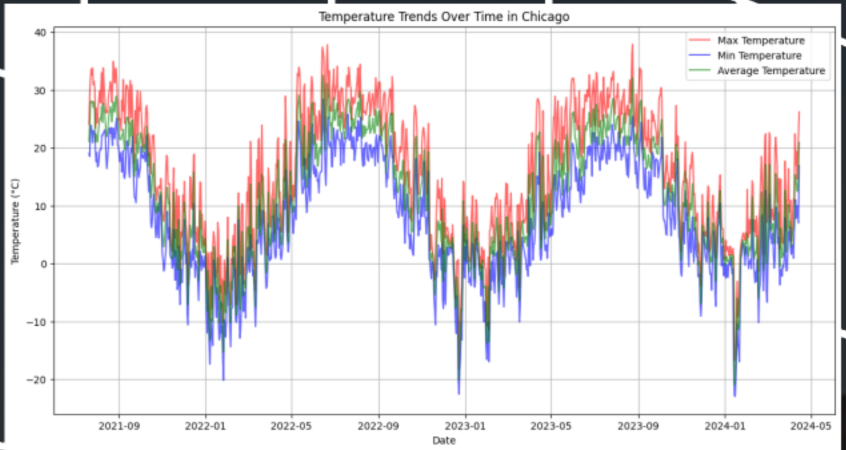
2.

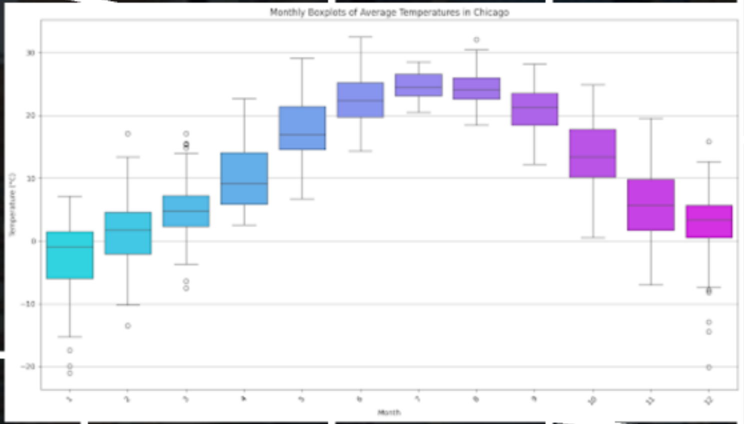
In-depth Data Analysis: We used three years of historical and real-time data from Chicago, combining traditional and advanced methods to forecast upcoming weather conditions.
climate change.

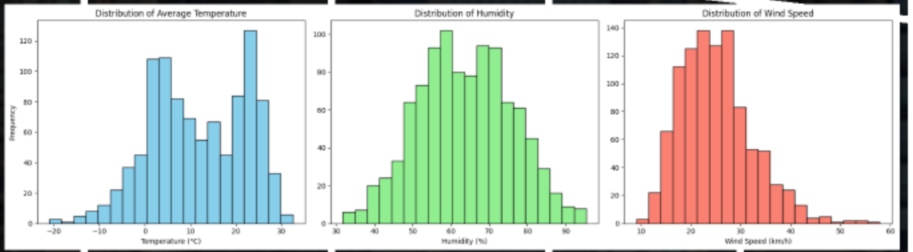
3.

Our team developed a system to predict Chicago's weather, focusing on extreme temperature changes due to global warming.

Exploratory Data Analysis

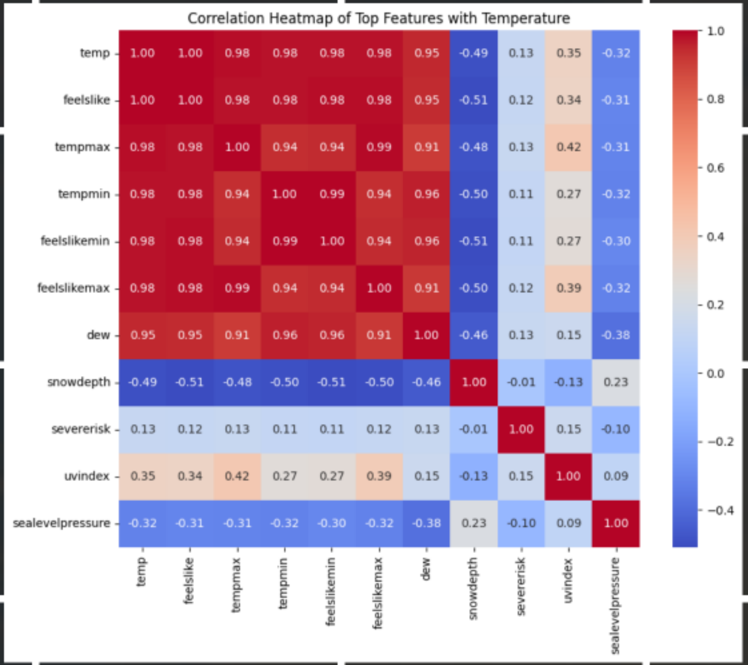







```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 999 entries, 0 to 998
Data columns (total 33 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   name              999 non-null    object
 1   datetime          999 non-null    datetime64[ns]
 2   tempmax           999 non-null    float64
 3   tempmin           999 non-null    float64
 4   temp              999 non-null    float64
 5   feelslikemax      999 non-null    float64
 6   feelslikemin      999 non-null    float64
 7   feelslike         999 non-null    float64
 8   dew               999 non-null    float64
 9   humidity          999 non-null    float64
 10  precip            999 non-null    float64
 11  precipprob        999 non-null    int64
 12  precipcover       999 non-null    float64
 13  preciptype        558 non-null    object
 14  snow              999 non-null    float64
 15  snowdepth         999 non-null    float64
 16  windgust          999 non-null    float64
 17  windspeed         999 non-null    float64
 18  winddir           999 non-null    float64
 19  sealevelpressure  999 non-null    float64
 20  cloudcover        999 non-null    float64
 21  visibility        999 non-null    float64
 22  solarradiation    999 non-null    float64
 23  solarenergy       999 non-null    float64
 24  uvindex           999 non-null    int64
 25  severerisk        826 non-null    float64
 26  sunrise           999 non-null    datetime64[ns]
 27  sunset            999 non-null    datetime64[ns]
 28  moonphase         999 non-null    float64
 29  conditions        999 non-null    object
 30  description       999 non-null    object
 31  icon              999 non-null    object
 32  stations          999 non-null    object
dtypes: datetime64[ns](3), float64(22), int64(2), object(6)
memory usage: 257.7+ KB
```

Numberical Features:



Categorical Features:

Chi-Square Test

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

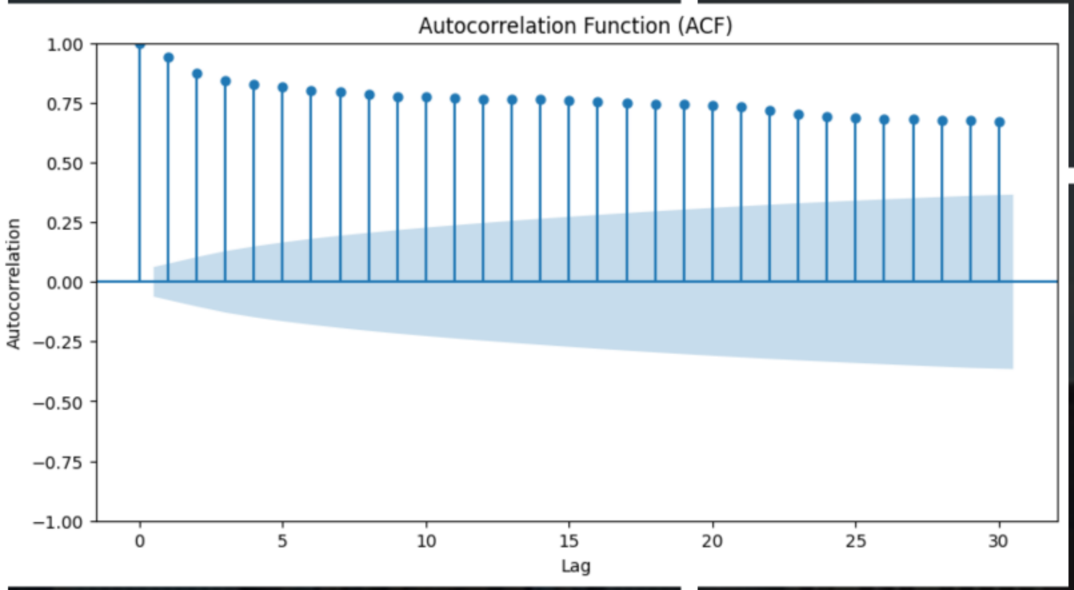
χ^2 = the test statistic $\quad \sum$ = the sum of

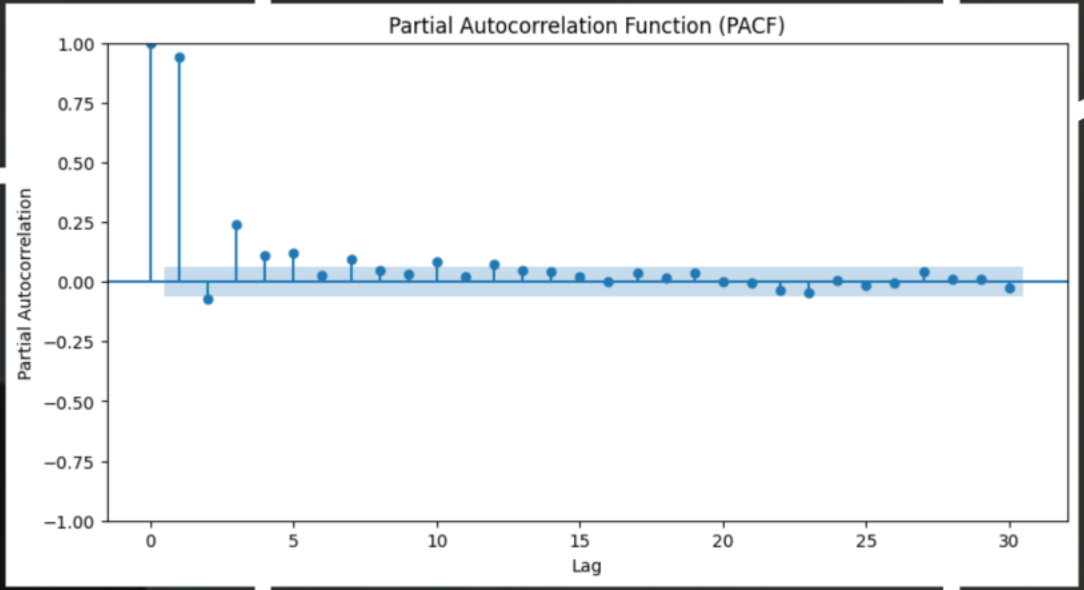
O = Observed frequencies E = Expected frequencies

Cramer's V

Highly Related Variable Pairs:
'precipprob – conditions'
'precipprob – description'
'precipprob – icon'
'preciptype – conditions'
'conditions – description'
'conditions – icon'
'description – icon'

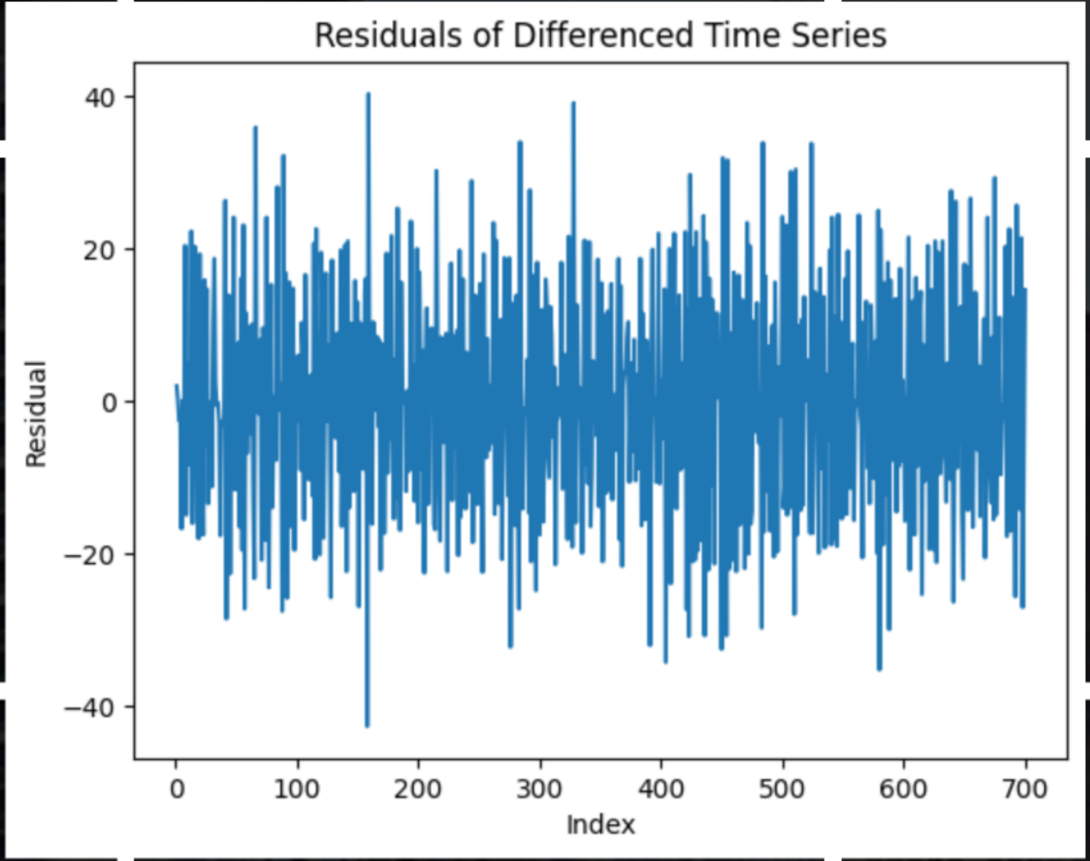
ACF AND PACF

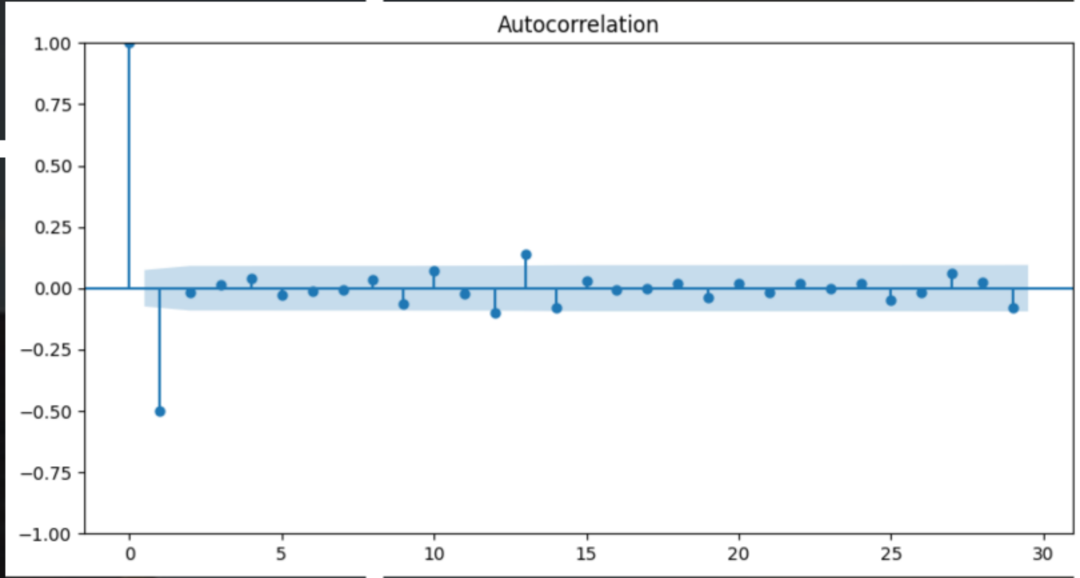


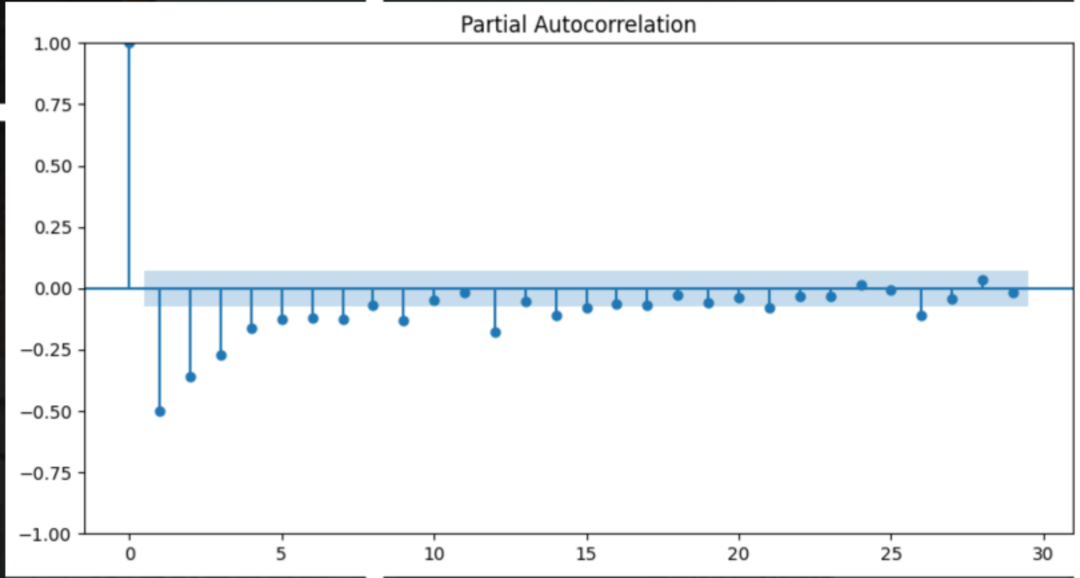


Augmented Dickey-Fuller(ADF) method → p-value: 0.19

After Differentiation



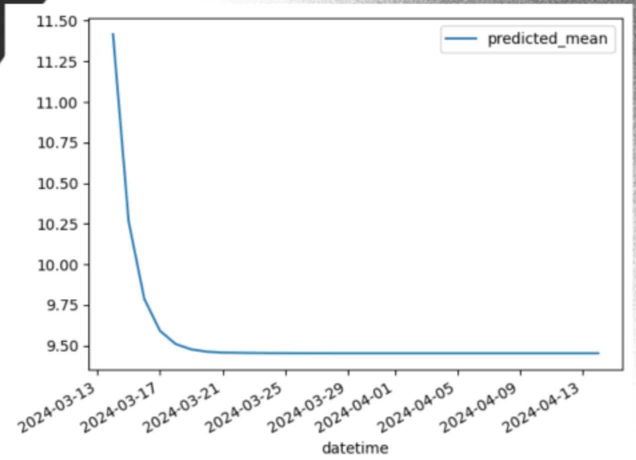




p-value: 2.480128867451355e-21

ARIMA

2024-04-14	18.966004
2024-04-15	15.731907
2024-04-16	14.354709
2024-04-17	13.847867
2024-04-18	13.667512
2024-04-19	13.603889
2024-04-20	13.581496
2024-04-21	13.573619
2024-04-22	13.570849
2024-04-23	13.569875
2024-04-24	13.569533
2024-04-25	13.569412
2024-04-26	13.569370
2024-04-27	13.569355
2024-04-28	13.569350
2024-04-29	13.569348
2024-04-30	13.569347
2024-05-01	13.569347
2024-05-02	13.569347
2024-05-03	13.569347
2024-05-04	13.569347
2024-05-05	13.569347
2024-05-06	13.569347
2024-05-07	13.569347
2024-05-08	13.569347
2024-05-09	13.569347
2024-05-10	13.569347
2024-05-11	13.569347
2024-05-12	13.569347
2024-05-13	13.569347
2024-05-14	13.569347



1.

This class of ARIMA models has a general term ARIMA(p, d, q) that includes three components (or terms): autoregressive (AR), integrated (or differencing), and moving average (MA) terms with the corresponding order of p, d, and q. The magnitude of temporal correlation exhibited in the time series will determine the AR and MA terms while the differencing term can transform a nonstationary series to be stationary

2.

p = 1| d = 1 | q = 1 from the ACF and PACF plots as well as the AIC and BIC values based on Python model fit.

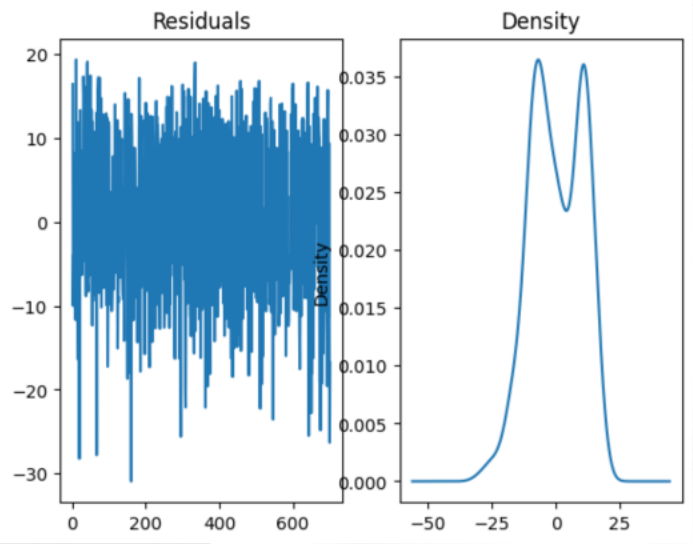
```python
from statsmodels.tsa.arima.model import ARIMA
model = ARIMA(train['temp'],order=(1,1,2))
model = model.fit()
model.summary()
```

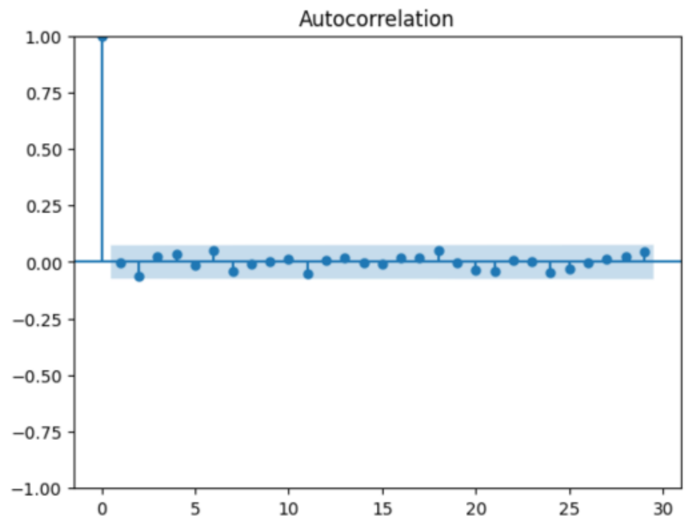
SARIMAX Results

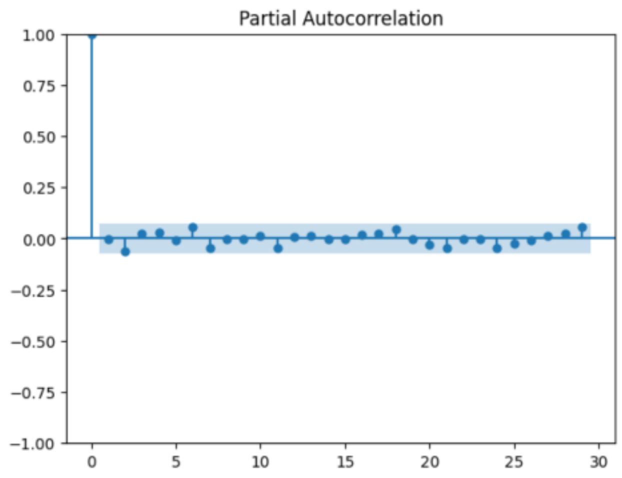
Dep. Variable:	temp	No. Observations:	846
Model:	ARIMA(1, 1, 2)	Log Likelihood	-2258.480
Date:	Fri, 26 Apr 2024	AIC	4524.960
Time:	00:01:31	BIC	4543.918
Sample:	0	HQIC	4532.224
	- 846		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.4129	0.061	6.723	0.000	0.293	0.533
ma.L1	-0.4801	0.061	-7.880	0.000	-0.600	-0.361
ma.L2	-0.3403	0.041	-8.373	0.000	-0.420	-0.261
sigma2	12.2675	0.397	30.938	0.000	11.490	13.045

Ljung-Box (L1) (Q):	0.00	Jarque-Bera (JB):	306.58
Prob(Q):	0.96	Prob(JB):	0.00
Heteroskedasticity (H):	0.79	Skew:	-0.55
Prob(H) (two-sided):	0.05	Kurtosis:	5.74







3.

Residuals should be white noise and make sure not to display seasonality.

SARIMAX

1.

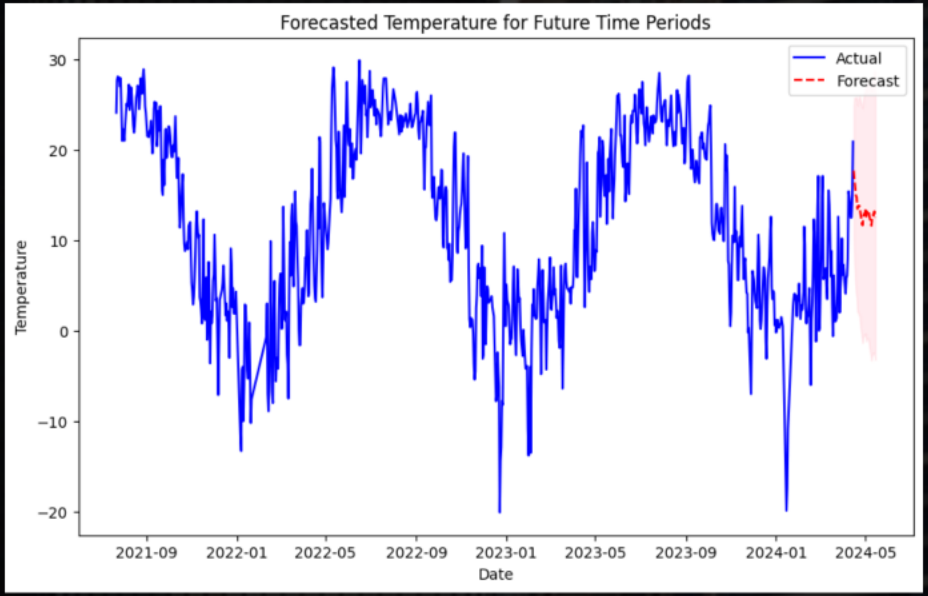
SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables) is an extension of the ARIMA model that incorporates additional variables, known as exogenous variables or regressors, into the model.

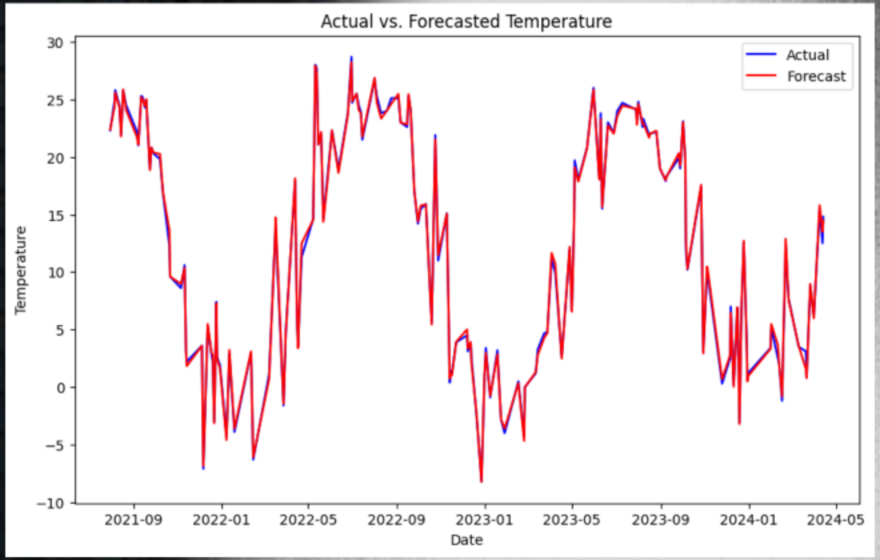
2.

The (P,D,Q,M) Order refers to the seasonal component of the model for the Auto Regressive parameters, differences, Moving Average parameters, and periodicity:

- D indicates the integration order of the seasonal process (the number of transformation needed to make stationary the time series)
- P indicates the Auto Regressive order for the seasonal component
- Q indicated the Moving Average order for the seasonal component
- M indicates the periodicity, i.e. the number of periods in season, such as 12 for monthly data.

	Date	Temperature
876	2024-04-15	17.679708
877	2024-04-16	16.735517
878	2024-04-17	15.844386
879	2024-04-18	14.930744
880	2024-04-19	14.791846
881	2024-04-20	13.477907
882	2024-04-21	13.800308
883	2024-04-22	13.720334
884	2024-04-23	13.772164
885	2024-04-24	13.069728
886	2024-04-25	12.550053
887	2024-04-26	12.055914
888	2024-04-27	11.650195
889	2024-04-28	12.501227
890	2024-04-29	12.911662
891	2024-04-30	12.850900
892	2024-05-01	13.318255
893	2024-05-02	12.631241
894	2024-05-03	13.168204
895	2024-05-04	13.225209
896	2024-05-05	13.341076
897	2024-05-06	12.731912
898	2024-05-07	12.345577
899	2024-05-08	12.082945
900	2024-05-09	11.613497
901	2024-05-10	12.424509
902	2024-05-11	12.805104
903	2024-05-12	12.724904
904	2024-05-13	13.178164
905	2024-05-14	12.475019





FUTURE WORK

1. A wider array of data is to be included for more accurate predictions.
2. Add categorical features in the model.
3. Integrate machine learning model.

THANK YOU AND HAVE A GREAT DAY, HOPEFULLY THE WEATHER IS RIGHT!!!