Weather Forecasting

With SARIMA & SARIMAX

Table of contents

01 **Opening**
Background and purpose about this project
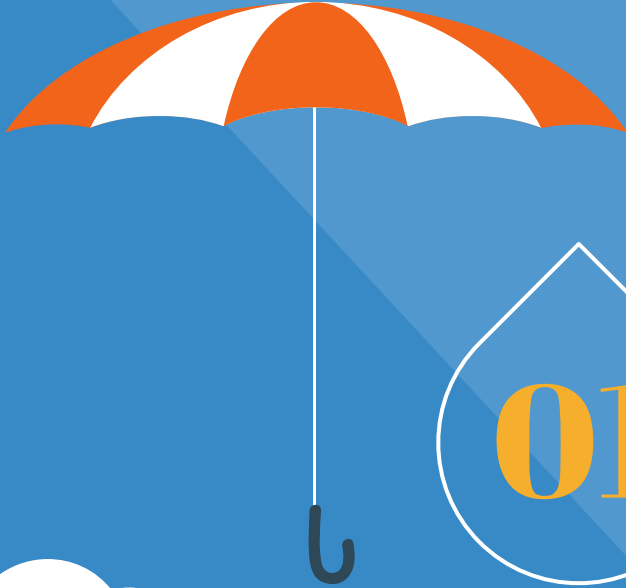
02 **Forecasting**
Model implementation to data

03 **Conclusion**
Summary about this project

Opening

Background and purpose
of this project

01

Purpose

To compare between the use of SARIMA and SARIMAX model in weather forecasting using Python.

Data Overview

Weather data have 2913 rows and time from 2015 to 2022 *red blocked are the selected variables*			
NAME	Station Name	**WDSP**	Windspeed (m/s)
Latitude	Latitude of Station	**PRCP**	Precipitation (Inch)
Longitude	Longitude of station	**DEWP**	Dewpoint (Farenheit)
DATE	Date the data recorded	**SNDP**	Snowdepth
SLP	Slope	**VISIB**	Visibility
TEMP	Temperature (Farenheit)	**Etc.**	Many more …

Data source: https://www.ncei.noaa.gov/

Data Preprocessing

Handle Missing Value
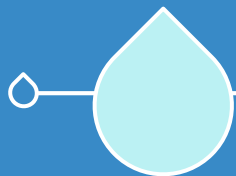
Feature Engineering

Handle Outlier

Model Implementation

With IQR

Change unit to our local unit

It's the biggest planet of them all
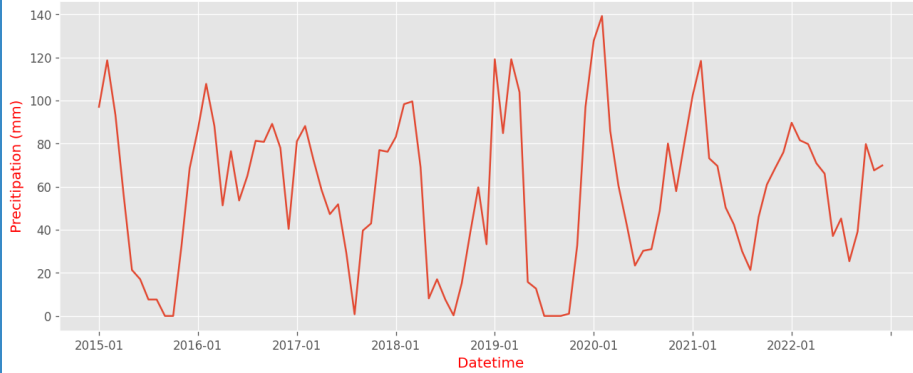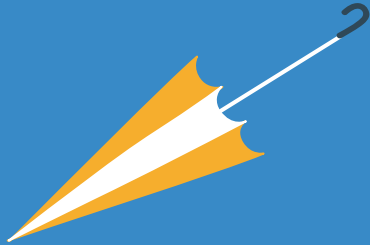
Time Series

The precipitation is **high at the beginning and final of year**, but **low at the middle of the year**. The season is 12 months, so **s = 12**.
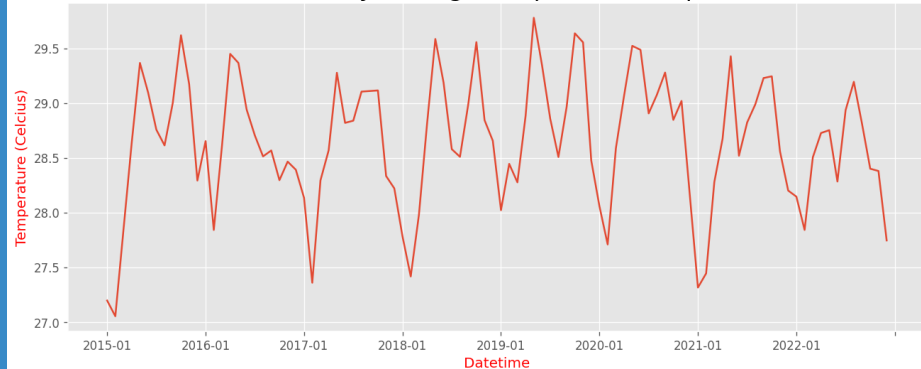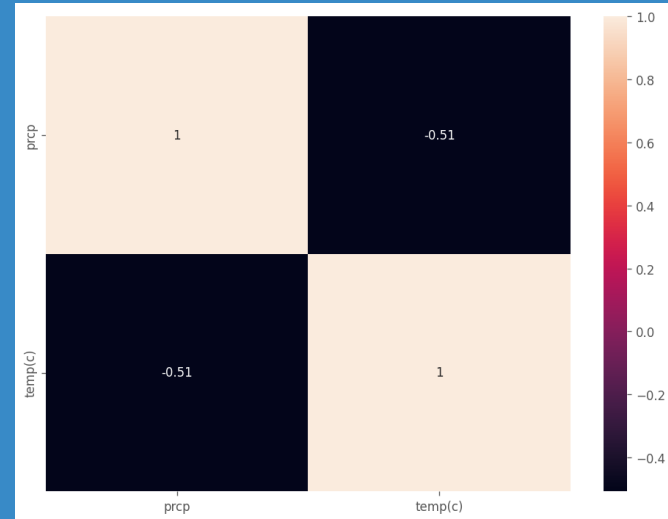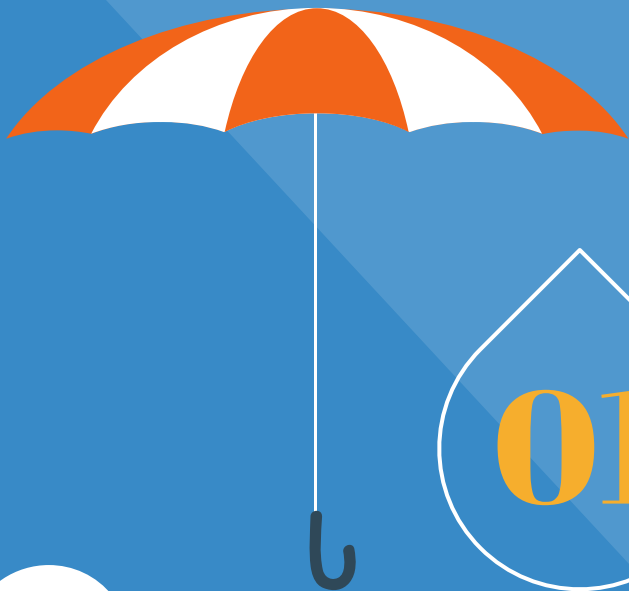
Correlation

The correation between precipitation and temperature is high but negative (-0.51). That makes the relation between them is inverse. So, if the **precipitation is high, then the temperature is low**, vice versa. These 2 variables selected because they have high correlation.

01

Forecasting

Model Implementation to data
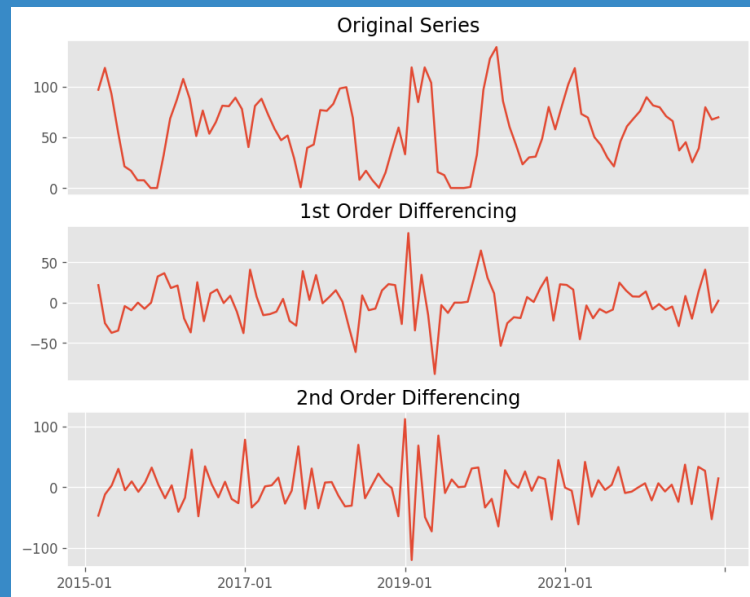
Need to Differentiate?

Time series chart beside is showed data after first and second differencing, but **the differencing don't need to be done** because the data is already stationer (p-value < 0.05)

Original Series
1st Order Differencing
2nd Order Differencing

```
ADF Statistic: -6.549938
p-value: 0.000000
Critical Values:
        1%: -3.505
        5%: -2.894
        10%: -2.584
```

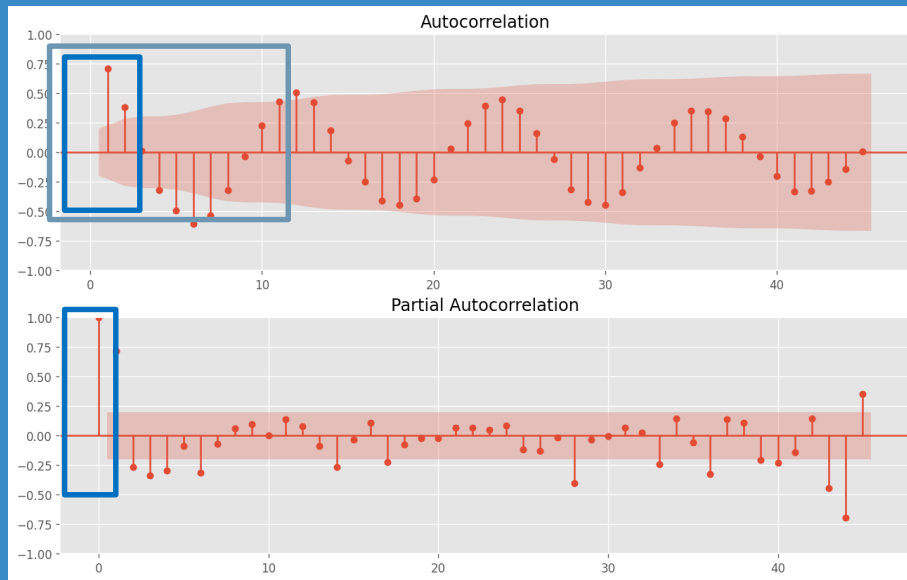# Parameter?

From ACF (Autocorrelation) and PACF (Partial Autocorrelation) Chart, selected:

- p = 1 or 2
- q = 1 or 2
- d = 0 (not differentiate)
- s = 12

I use SARIMAX (1,0,1)x(1,0,1,12) then, because this is the best model after tried many models.

Parameter?

After that, I train the data and I select:
- p = 1
- q = 1

Because the model have the lowest p-value and AIC value.

SARIMAX Results

Dep. Variable:	prcp		No. Observations:	84	
Model:	SARIMAX(1, 0, 1)x(1, 0, 1, 12)		Log Likelihood	-384.588	
Date:	Tue, 11 Apr 2023		AIC	779.177	
Time:	00:04:31		BIC	791.331	
Sample:	01-01-2015		HQIC	784.062	
	- 12-01-2021				

Covariance Type: opg

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.7935	0.080	9.888	0.000	0.636	0.951
ma.L1	-0.1159	0.150	-0.775	0.438	-0.409	0.177
ar.S.L12	0.9989	0.050	19.902	0.000	0.901	1.097
ma.S.L12	-0.9523	1.082	-0.880	0.379	-3.072	1.168
sigma2	422.0939	425.396	0.992	0.321	-411.666	1255.854

Ljung-Box (L1) (Q):	0.20	Jarque-Bera (JB):	2.89
Prob(Q):	0.65	Prob(JB):	0.24
Heteroskedasticity (H):	0.95	Skew:	-0.03
Prob(H) (two-sided):	0.88	Kurtosis:	3.91

Testing Data
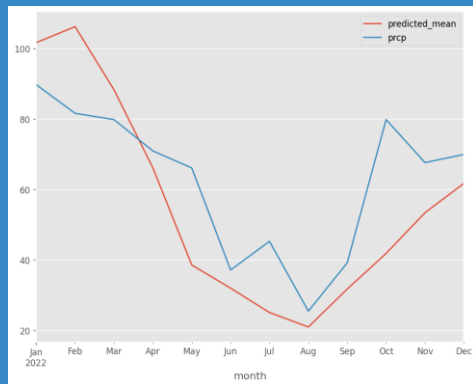
Without exogenuos variable (temperature) (SARIMA)
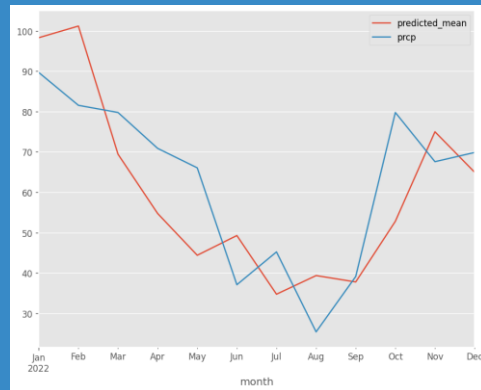
With exogenuos variable (temperature) (SARIMAX)

Forecasting

Prediction of Monthly Precipitation Graph
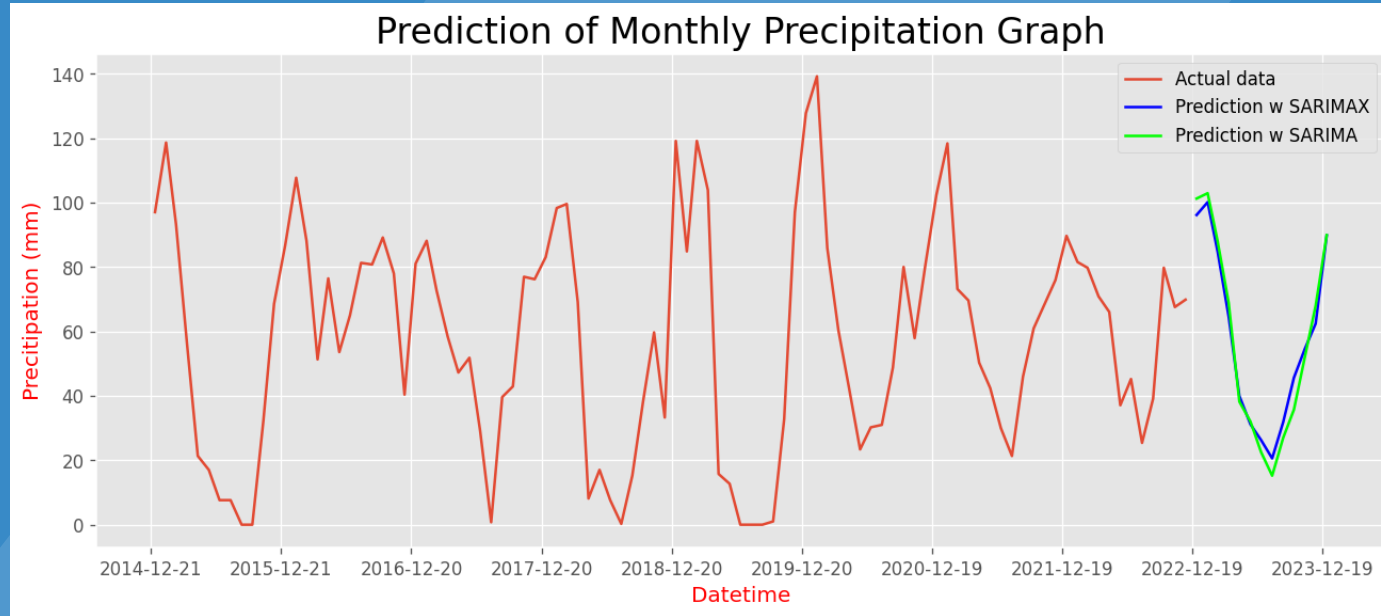
Forecasting with SARIMA & SARIMAX models

Validation

	Root Mean Error (RMSE)	Mean Absolute Error (MAE)	Mean Absolute Percent Error (MAPE)
SARIMA	17.85	14.58	23.18
SARIMAX	14.60	12.80	22.37

Conclusion

03

Summary about this project
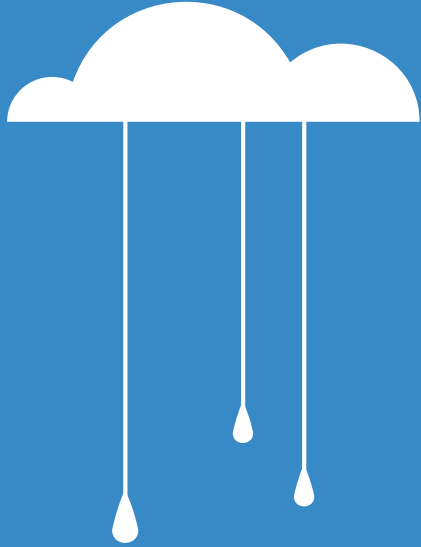
- From testing data, we know that the result of SARIMAX model can follow the line chart more than SARIMA model.
- From forecasting result, the chart made by SARIMAX model is a bit more notched than SARIMA.
- From validation, SARIMAX model is better than SARIMA model.

Thanks

Link to this code:
https://github.com/ahyaramdha/precip-forecast