

[Open in app](#)[Sign up](#)[Sign in](#)**Medium**

Search



Write



# Time Series forecasting using SARIMA in Python

Tirtha Mutha · [Follow](#)

5 min read · Nov 10, 2023

138

2



A time series is a series of data points ordered in time. In a time series, time is often the independent variable, and the goal is usually to make forecast for the future. To understand time series models and how to analyze them, it helps to know their three main characteristics: autocorrelation, seasonality and stationarity. Modeling these characteristics is necessary to get an accurate forecast.

There are many types of time series models, but the main ones include moving average, exponential smoothing, autoregressive moving average(ARMA), autoregressive integrated moving average(ARIMA) and seasonal autoregressive integrated moving average (SARIMA).

I will be using seasonal autoregressive integrated moving average (SARIMA) for forecasting my data.

## WHAT IS SARIMA ?

A seasonal autoregressive integrated moving average, or SARIMA, is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends.

A seasonal autoregressive integrated moving average mode is a form of regression analysis that gauges the strength of one dependent variable relative to other changing variables. A SARIMA model can be understood by outlining each of its components as follows:

- Seasonal (S): refers to a model with seasonal component.
- Autoregression (AR): refers to a model that shows a changing variable that regresses on its own lagged, or prior, values.
- Integrated (I): represents the differencing of raw observations to allow the time series to become stationary (i.e., data values are replaced by the difference between the data values and the previous values).
- Moving average (MA): incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.

## ABOUT THE DATASET

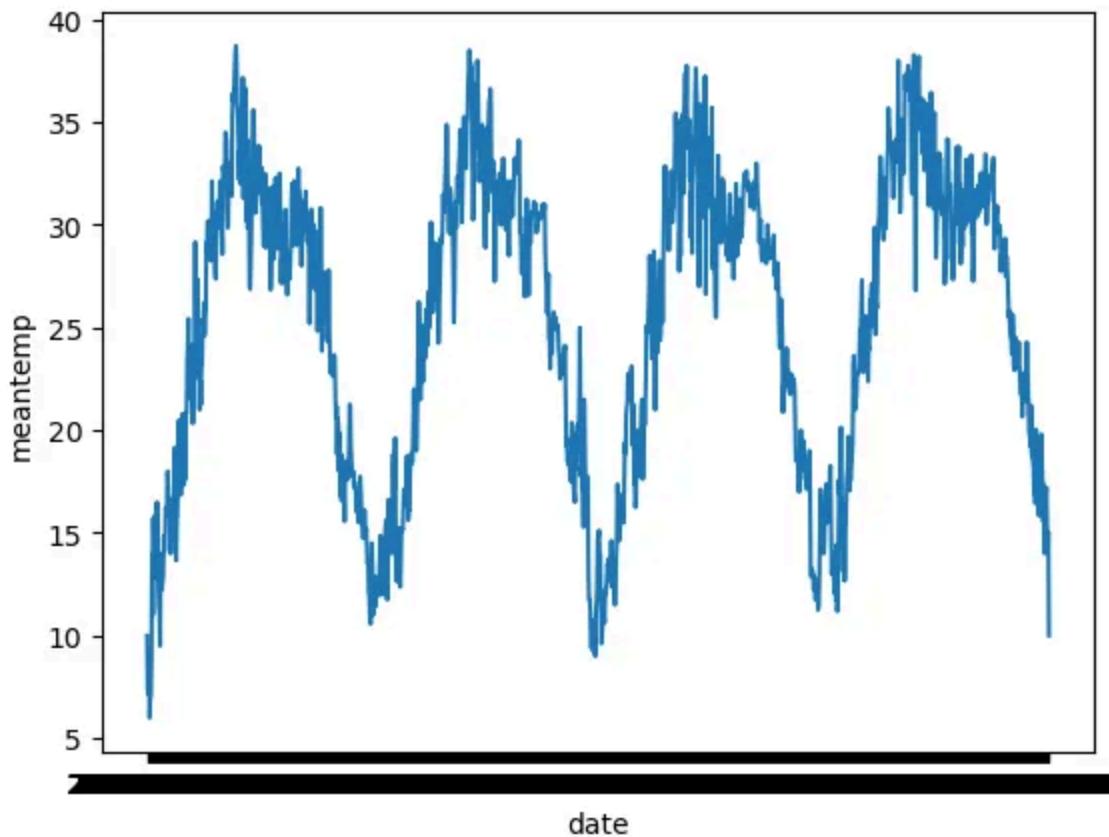
This dataset I have chosen contains weather data collected in the city of Delhi from the period of 4 years (1st January 2013 to 24th April 2017). It can be found [here](#). The four parameters are Mean temperature, humidity, windspeed, mean pressure.

I will be forecasting the mean temperature in Delhi.

	date	meantemp	humidity	wind_speed	meanpressure
0	2013-01-01	10.000000	84.500000	0.000000	1015.666667
1	2013-01-02	7.400000	92.000000	2.980000	1017.800000
2	2013-01-03	7.166667	87.000000	4.633333	1018.666667
3	2013-01-04	8.666667	71.333333	1.233333	1017.166667
4	2013-01-05	6.000000	86.833333	3.700000	1016.500000

These are the first five rows of the dataset

```
sns.lineplot(x='date',y='meantemp',data=data)
```



Here I can see that the data has seasonal variations hence I have used SARIMA

## How to use SARIMA in Python?

The SARIMA time series forecasting method is supported in Python via the statsmodel library.

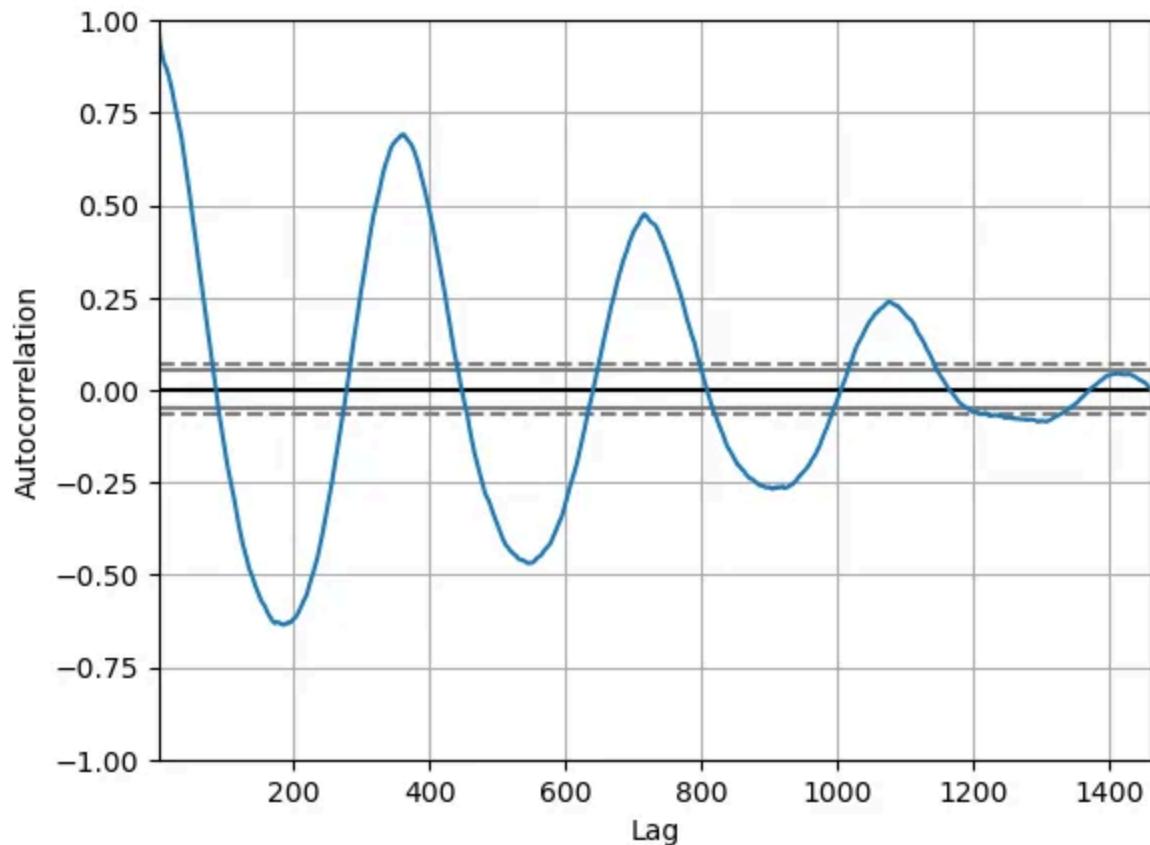
To use SARIMA there are three steps, they are:

1. Define the model.
2. Fit the defined model.
3. Make a prediction with the fit model.

## Autocorrelation, Seasonality and Stationarity

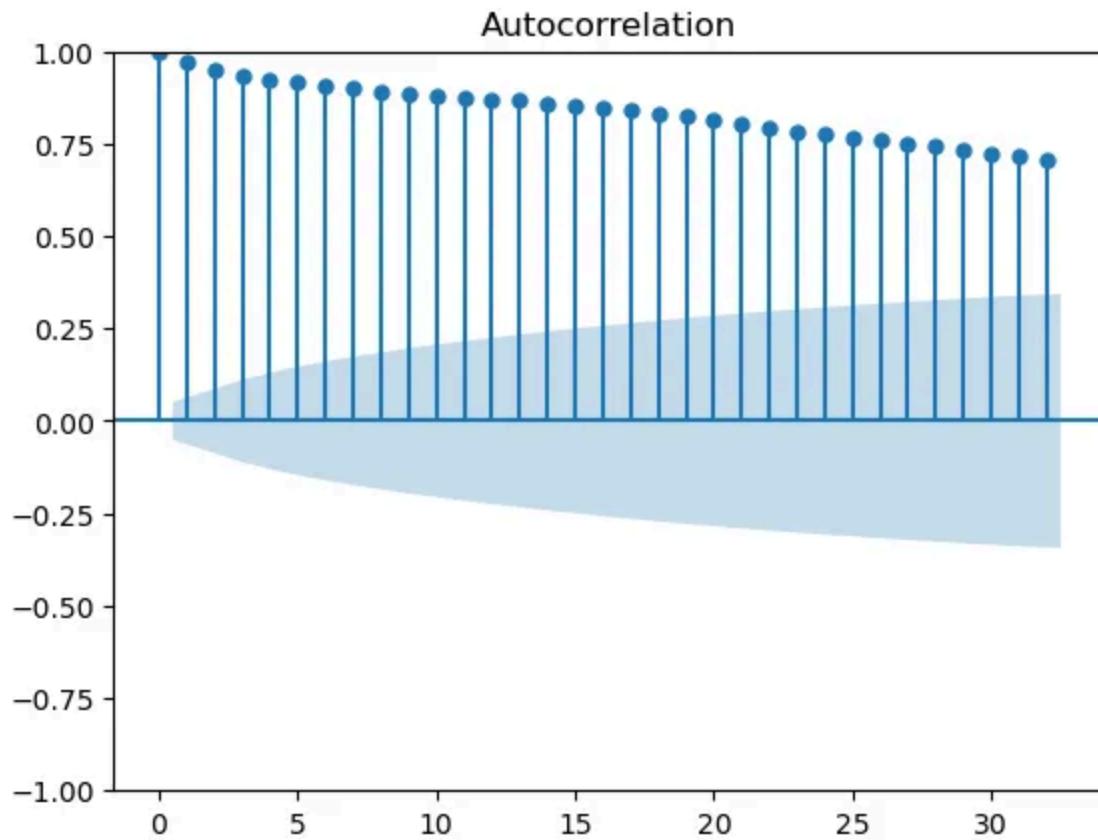
Autocorrelation refers to the degree of correlation of the same variables between two successive time intervals. It measures how the lagged version of the value of a variable is related to the original version of it in a time series.

```
from pandas.plotting import autocorrelation_plot  
autocorrelation_plot(data['meantemp'])
```



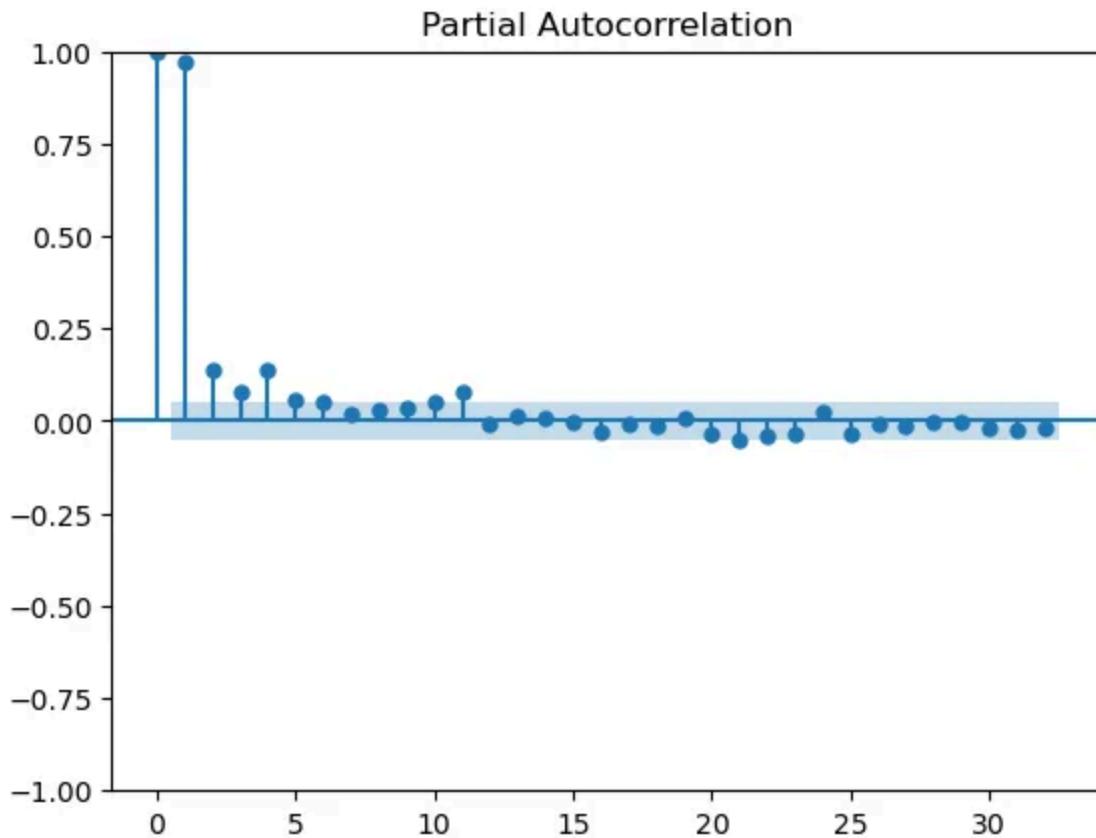
From the above graphs it is demonstrated that there is seasonality in the dataset. A **damped sinusoidal wave** is seen with a trough around every 350–400 lags.

```
from statsmodels.graphics.tsaplots import plot_acf  
  
plot_acf(data['meantemp'])  
plt.show()
```



ACF plots show the correlation between a time series and a lagged version of itself. The ACF graph is decreasing exponentially.

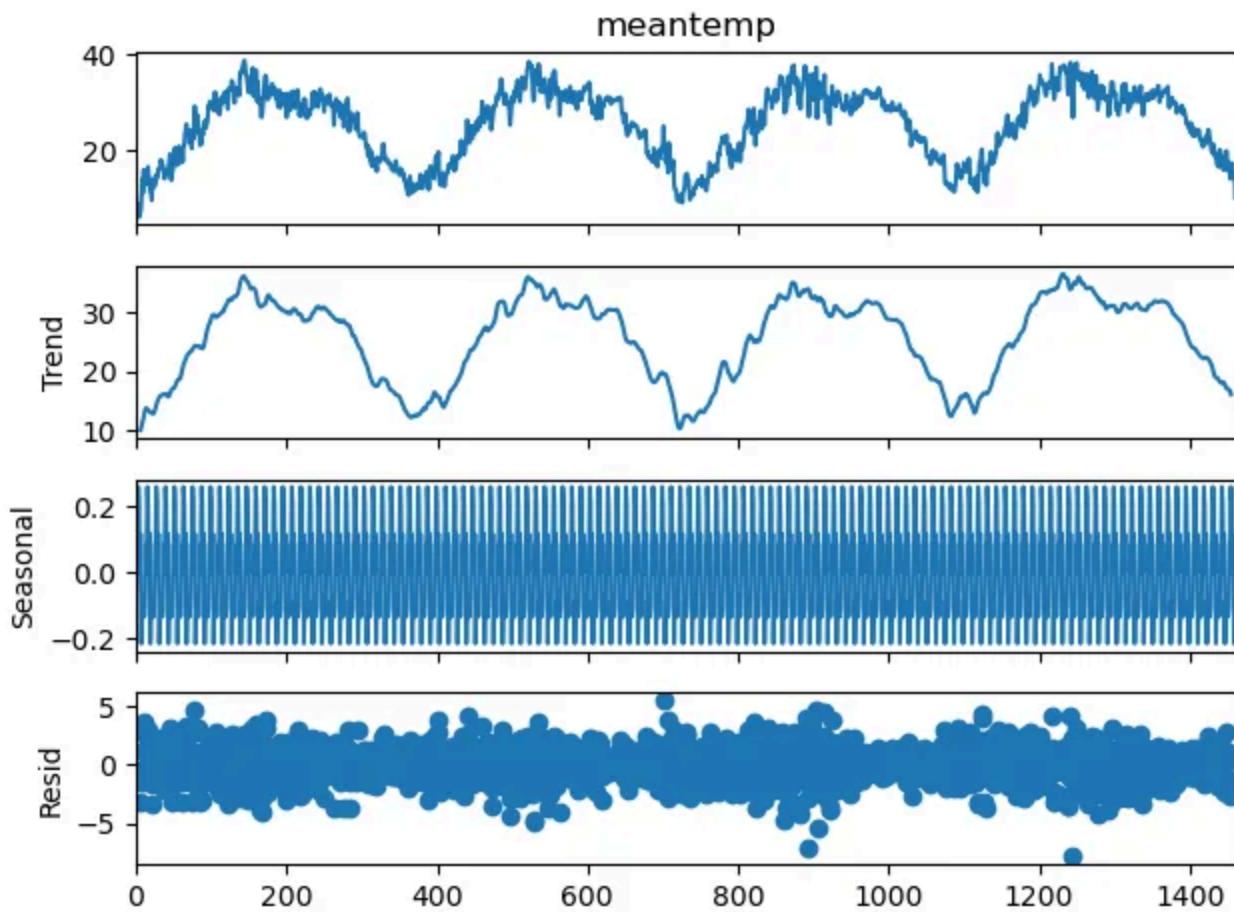
```
from statsmodels.graphics.tsaplots import plot_pacf  
  
plot_pacf(data['meantemp'])  
plt.show()
```



PACF plots show the correlation between a time series and a lagged version of itself after removing the effects of intervening observations. The PACF graph is abrupt.

Time series decomposition is a statistical technique that involves breaking down a time series into its underlying components to better understand its patterns and improve forecasting accuracy. The most common components of a time series are trend, seasonality, and remainder.

```
decomposition=seasonal_decompose(data['meantemp'],model='additive',period=12)
decomposition.plot()
plt.show()
```



## Augmented Dickey-Fuller Test

ADF (Augmented Dickey-Fuller) test is a statistical significance test which means the test will give results in hypothesis tests with null and alternative hypotheses. As a result, we will have a p-value from which we will need to make inferences about the time series, whether it is stationary or not.

```
dftest = adfuller(data.meantemp, autolag = 'AIC')
print("1. ADF : ", dftest[0])
print("2. P-Value : ", dftest[1])
print("3. Num Of Lags : ", dftest[2])
print("4. Num Of Observations Used For ADF Regression and Critical Values Calcul
print("5. Critical Values :")
for key, val in dftest[4].items():
    print("\t",key, ":", val)
```

P-Value of our time series is 0.2774121372301601.

the p-value obtained should be less than the significance level ( 0.05) in order to reject the null hypothesis. Thereby, inferring that the series is non-stationary.

If you wish to use ARIMA instead of SARIMA you need to detrend the time series data there are certain transformation techniques used and they are listed as follows.

- Differencing the data
- Log transforming of the data
- Taking the square root of the data
- Taking the cube root
- Proportional change

```
data_diff=data['meantemp'].diff(periods=350)

dftest = adfuller(data_diff.dropna(), autolag = 'AIC')
print("1. ADF : ",dftest[0])
print("2. P-Value : ", dftest[1])
print("3. Num Of Lags : ", dftest[2])
print("4. Num Of Observations Used For ADF Regression and Critical Values Calcul
print("5. Critical Values :")
for key, val in dftest[4].items():
    print("\t",key, ":", val)
```

p-value is 0.00138 so data is **stationary** and the new data can be used for forecasting using ARIMA

## Performing the Seasonal ARIMA

In order to use SARIMA we will need to choose p,d,q values for the ARIMA, and P,D,Q values for the Seasonal component. There are many ways to choose these values statistically, such as looking at auto-correlation plots, correlation plots, domain experience, etc.

In Auto ARIMA, the model itself will generate the optimal p, d, and q values which would be suitable for the data set to provide better forecasting.

```
import pmldarima as pmd

model=pmd.auto_arima(data['meantemp'],start_p=1,start_q=1,test='adf',m=12,season
```

Best model: ARIMA(1,1,1)(1,0,1)[12]

I will use the above p,d,q values for setting up my SARIMA model.

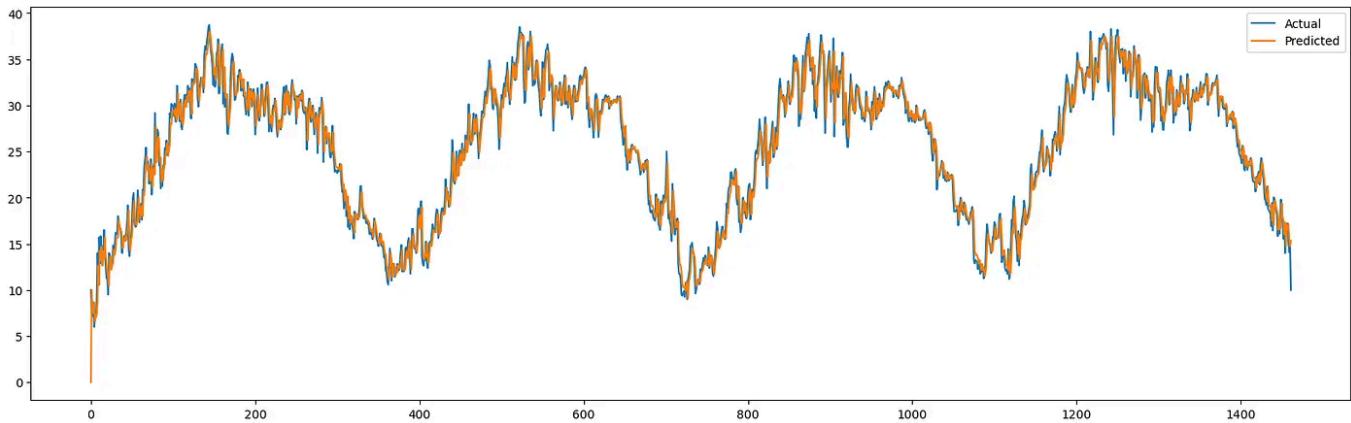
```
sarima=SARIMAX(data['meantemp'],order=(1,1,1),seasonal_order=(1,0,1,12))
predicted=sarima.fit().predict();predicted
```

0	0.000000
1	9.999995
2	7.841684
3	7.555284
4	8.665322
	...
1457	16.702556
1458	17.159179
1459	15.521542
1460	14.752933

1461 15.341090

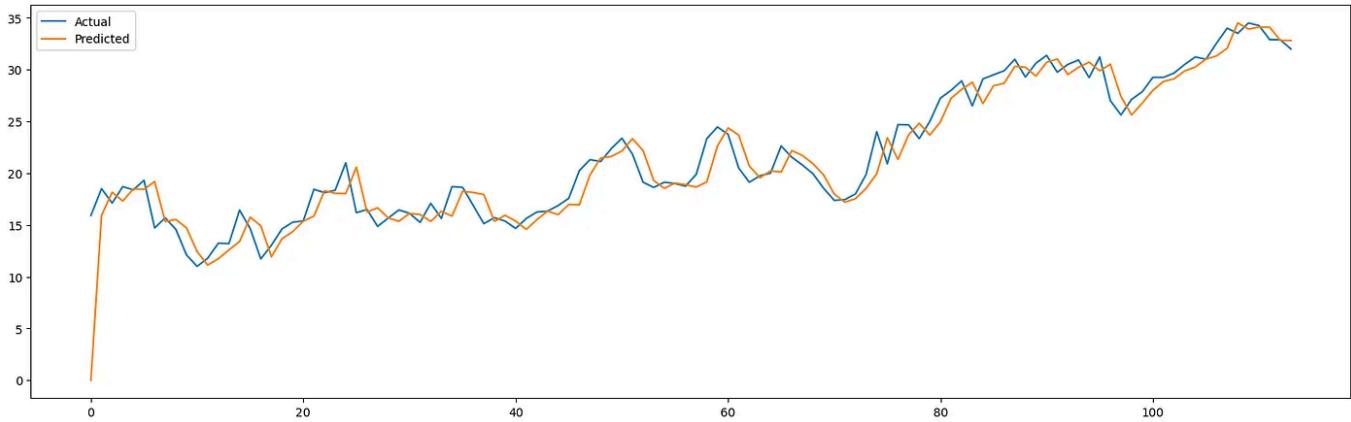
Name: predicted\_mean, Length: 1462, dtype: float64

The above data points are the predicted mean temperatures for the next 5 days.



Actual vs. Predicted graph of meantemp (training data)

Now that I have evaluated our data on the test set and am satisfied with the performance, the next step would be to refit our model to test data set and then forecast into the real future.



Actual vs. Predicted graph of meantemp (test data)

You can check out my GitHub repository containing all the code discussed in detail [here](#).

Thankyou:)

Data

Data Science

Forecasting

Time Series Analysis

Sarima



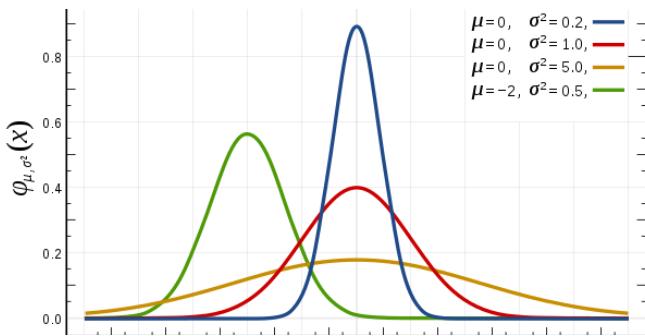
**Written by Tirtha Mutha**

26 Followers · 6 Following

Follow

I knew it was time to start a stats blog the day I misread passion as poisson.

## More from Tirtha Mutha



**INSPIRED**

HOW TO  
CREATE  
TECH  
PRODUCTS



Tirtha Mutha

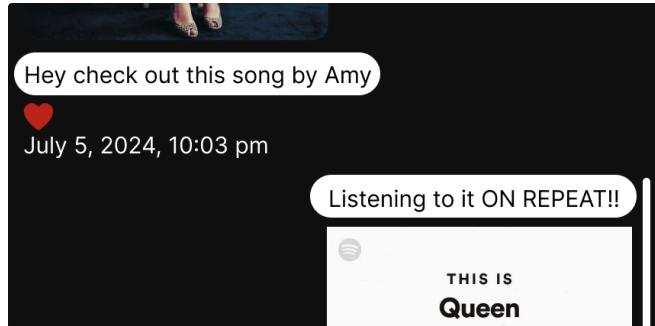


Tirtha Mutha

## Understanding Different Types of Probability Distributions

Probability distributions play a vital role in the field of statistics and data analysis. They...

Apr 23 1



In Bootcamp by Tirtha Mutha

## Designing Spotify's Messaging System: Making Music Social Again

Spotify has been my go-to music platform for years, yet something has always felt missing....

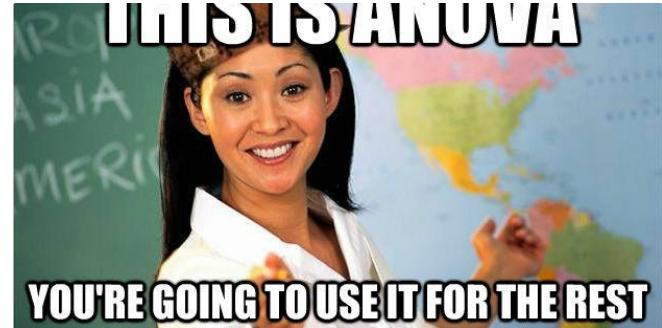
Aug 28 1



## Building Great Tech Products: Lessons from Marty Cagan's...

Creating products that customers truly love requires a balance of strategic alignment,...

Nov 24 22



Tirtha Mutha

## A Brief Theoretical Approach to Understanding One-Way ANOVA...

One-way ANOVA (Analysis of Variance) is a statistical technique that allows researchers...

May 12 9



[See all from Tirtha Mutha](#)

## Recommended from Medium



Irina (Xinli) Yu, Ph.D.

## Mastering Time Series Forecasting with ARIMA Models

Time series forecasting is a critical component in many domains, including...

Jun 6 · 2



L Data PR

## ARIMA Models in R part-1

Hello Guys,

6d ago · 10



## Lists



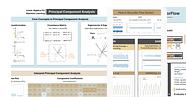
### Predictive Modeling w/ Python

20 stories · 1699 saves



### data science and AI

40 stories · 296 saves



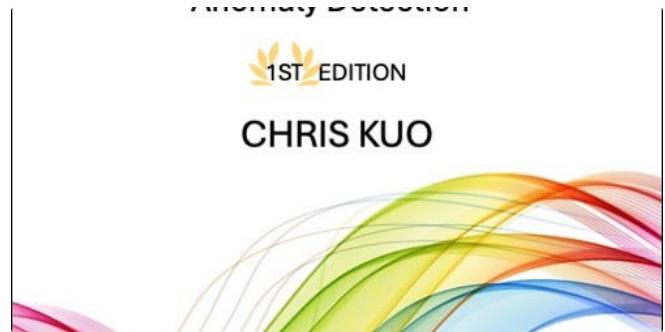
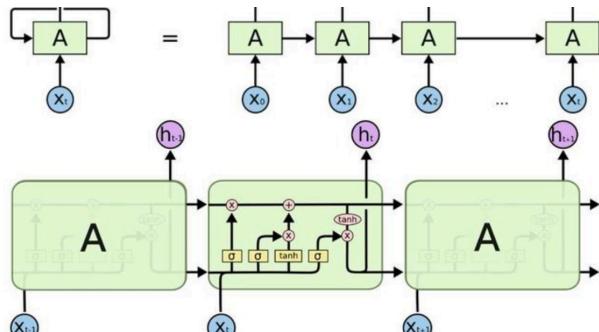
### Practical Guides to Machine Learning

10 stories · 2066 saves



### Coding & Development

11 stories · 926 saves



Pelin Okutan

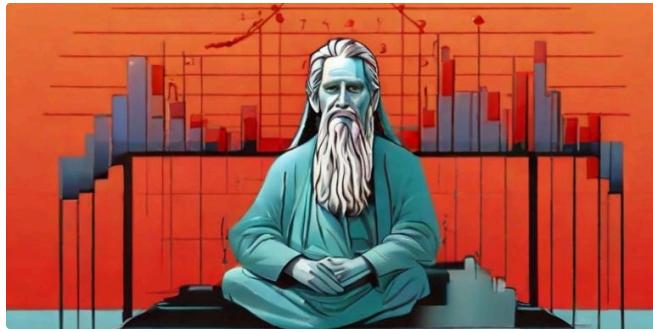


In Dataman in AI by Chris Kuo/Dr. Dataman

## Time Series Forecasting with Python: Practical Implementation...

In my previous article, Time Series Forecasting: A Comparative Analysis of...

⭐ Sep 4 ⚡ 109 🗣 1



**tds** In Towards Data Science by Jonas Dieckmann

## Getting Started Predicting Time Series Data with Facebook Prophet

This article aims to take away the entry barriers to get started with time series...

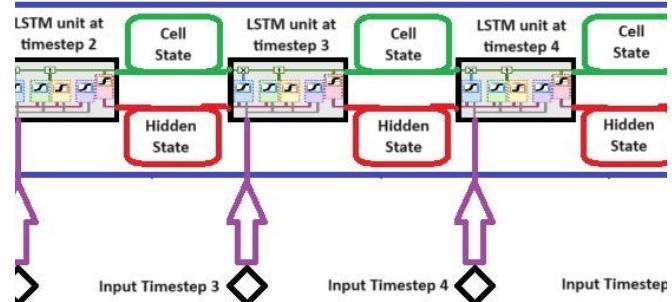
Jan 30 ⚡ 349 🗣 3



## Temporal Fusion Transformer for Interpretable Time Series...

Sample eBook chapters (free):  
<https://github.com/dataman-git/modern-...>

⭐ Apr 18 ⚡ 381 🗣 3



**GoPenAI** In GoPenAI by Seyed Mousavi

## Temporal Data in LSTM Time Series Forecasting

If you noticed in our previous articles, when building LSTM models, we either specified t...

⭐ Jun 7 ⚡ 24



See more recommendations