**Batch: B - 1**

**Roll No.: 16014022050**

**Experiment No.: 5**

|  |
| --- |
| Title: To perform forecasting using time series analysis. |

**Aim:**

To perform forecasting using time series analysis

# Course Outcome:

**CO4:** Perform Time series Analytics and forecasting.

# Books/Journals/Websites referred:

# Pre Lab / Prior Concepts:

Students should have a basic understanding of: Time series Analytics and forecasting.

# ---------------------------------------------------------------------------------------------------------------

# Procedure:

**Dataset Used –** Air Quality

**Step1: Select and load the dataset**

**Step2: Visualize the data**

**Step 3: Fit the model (ARIMA model is used)**

**Step 4: Forecast future values**

**Step 5: Create a Data Frame for the forecast**

**Step 6: Plot the results**

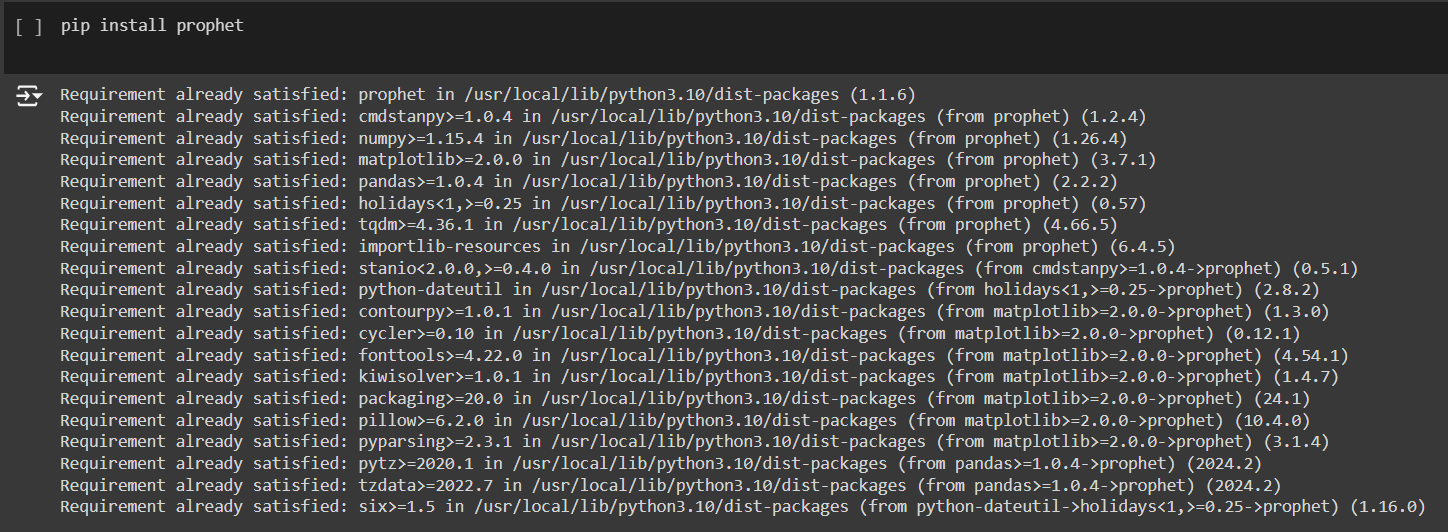
**Students have to perform all the tasks illustrated above by choosing any other time series related dataset.**

**Implementation details:**

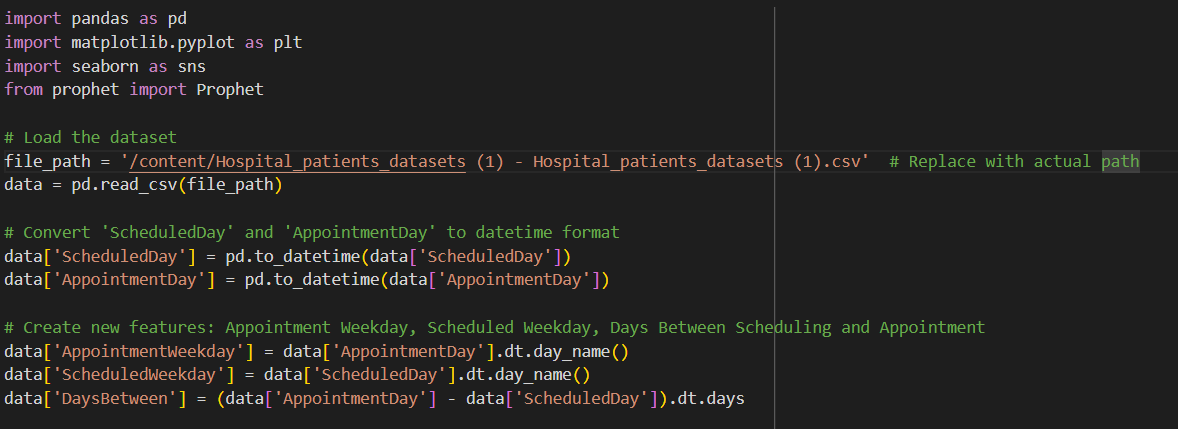
**Dataset Used** **–** Hospital Patients Datasets

**Output –**

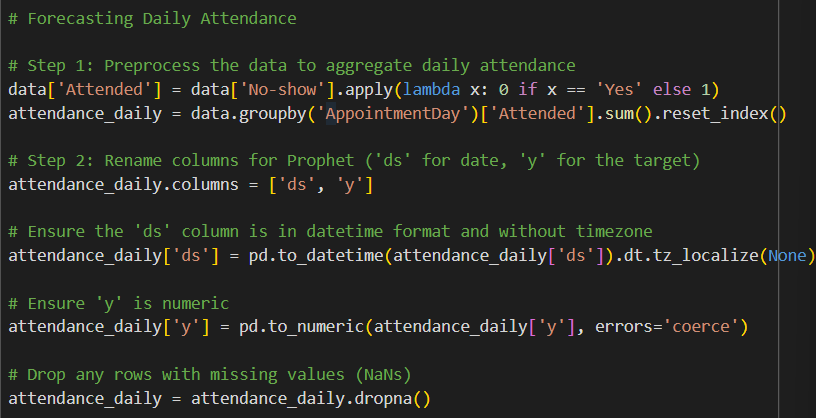
1. Installing dependencies:



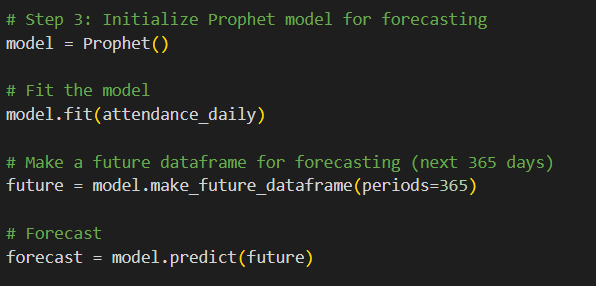
1. Loading dataset and converting data to appropriate data format:



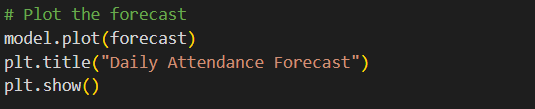
1. Preprocessing data and checking for any missing values:



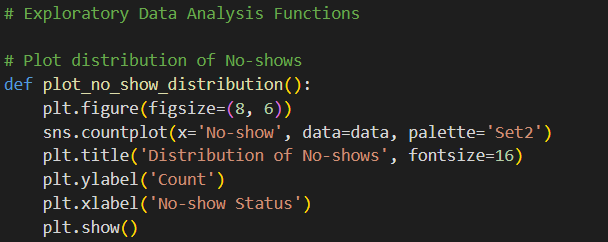
1. Initializing and fitting model:

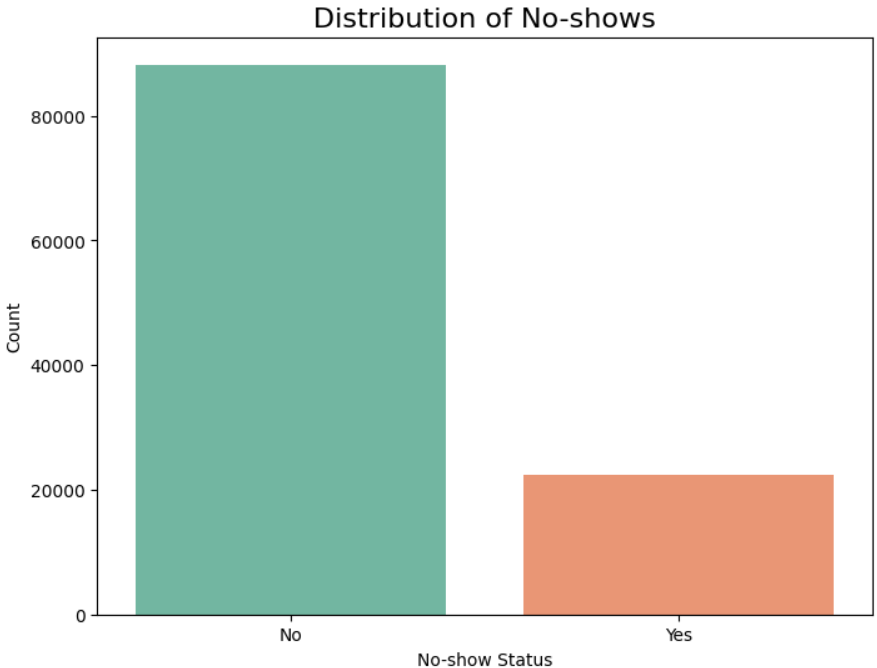


1. Plotting the forecast:

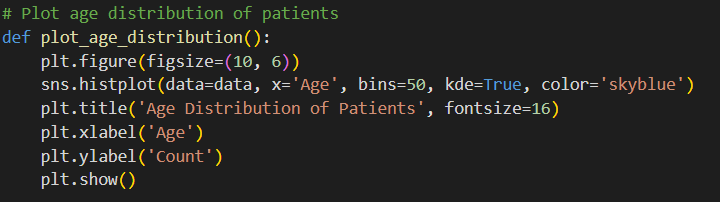


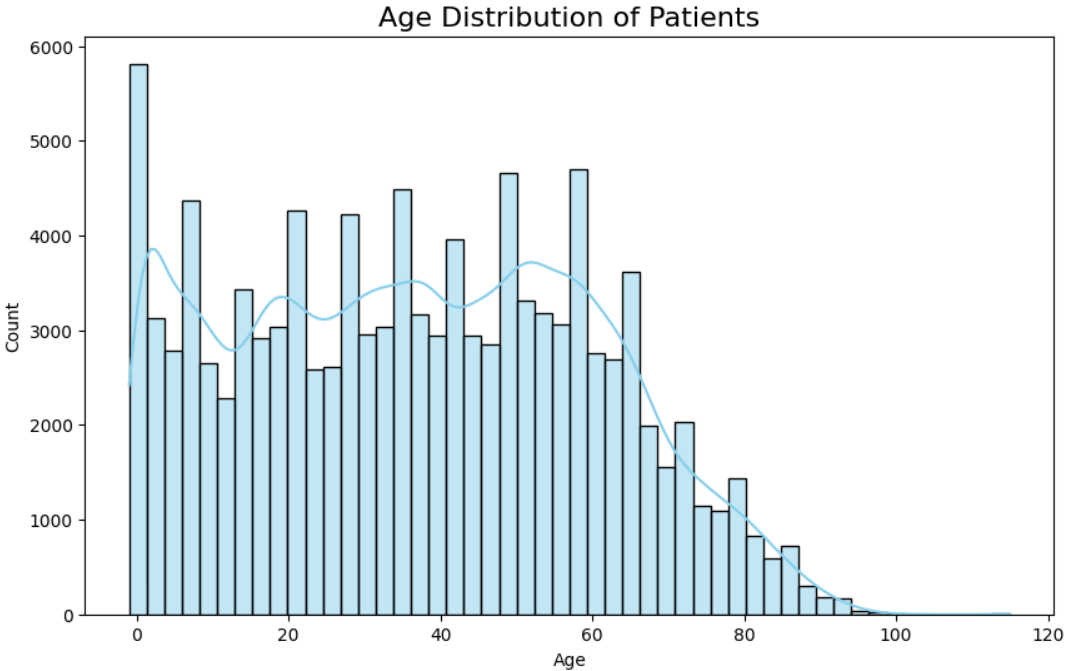
1. Visualizing data – Plot of Distribution of No-Shows:



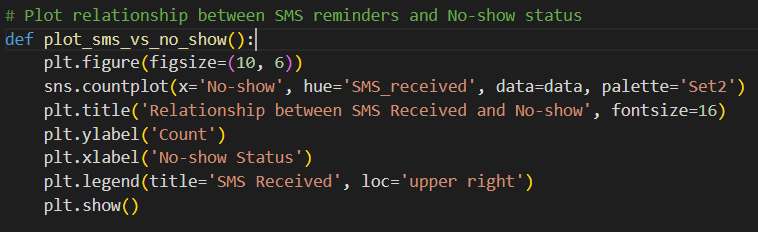


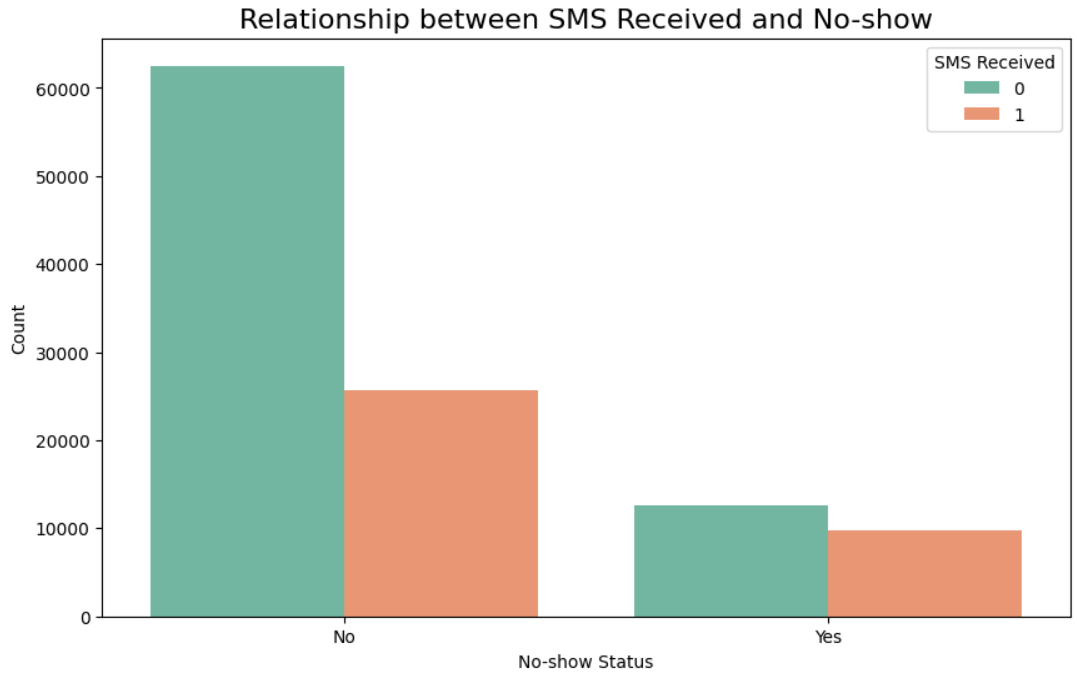
1. Visualizing data – Plot of Age Distribution of Patients:



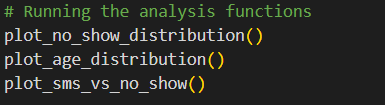


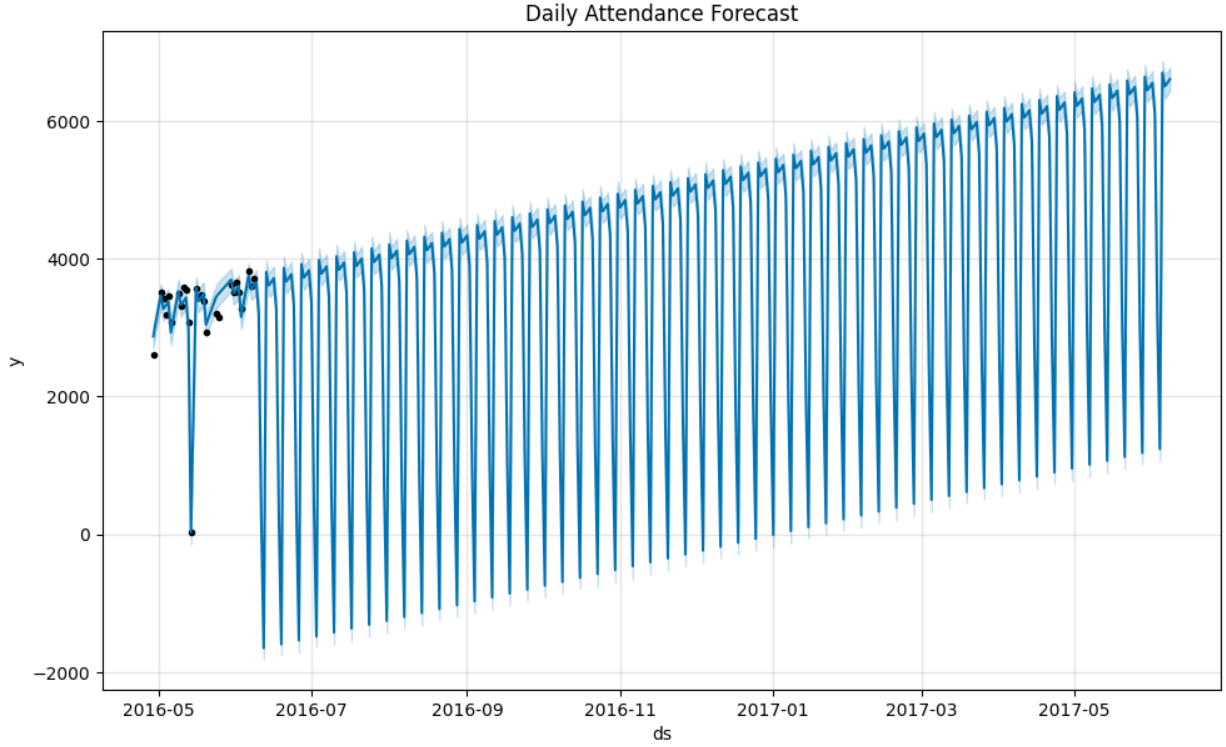
1. Visualizing data – Relationship between SMS reminders and No-Show Status:





1. Forecast Analysis:





**Date: \_\_\_\_\_\_\_\_\_\_\_\_\_ Signature of faculty in-charge**

**Post Lab Descriptive Questions**:

1. **What are the key components of a time series, and how do they affect the analysis?** 
   * Trend: The long-term movement or direction in the data over time. It reflects whether the values are increasing, decreasing, or remaining stable over a prolonged period. Trends help identify underlying growth or decline patterns.
   * Seasonality: The repeating short-term cyclical patterns in the data, often associated with a specific time of year, month, or week. Seasonal components are regular fluctuations that can be yearly, quarterly, monthly, etc.
   * Cyclic Component: Similar to seasonality but without a fixed periodicity. It reflects long-term oscillations due to factors like economic or market cycles.
   * Residual (Irregular): The random, unpredictable variations in the data that are not captured by the trend, seasonality, or cycles. These are often considered noise or error.

Effect on Analysis: Understanding and isolating these components is crucial for accurate forecasting, anomaly detection, and identifying patterns. Failure to account for seasonality or trends can lead to incorrect model fitting and poor predictions.

1. **What is the purpose of decomposing a time series into trend, seasonal, and residual components?**
   * Better Understanding: Decomposition helps to clearly observe the underlying structure of the time series by separating long-term trends, repeating seasonal patterns, and random noise.
   * Modeling: By decomposing, each component can be modeled and analyzed separately, leading to more accurate predictions. For example, trend and seasonality can be modeled, and then future values can be forecasted based on those patterns.
   * Forecasting: It allows for the improvement of forecasting models by focusing on predictable components (trend and seasonality) and ignoring noise (residuals).
   * Anomaly Detection: Residuals after decomposition can be examined to identify outliers or unusual patterns that deviate from normal behavior.
2. **Explain how the ARIMA model works and what the terms (p, d, q) represent.**

Autoregressive Integrated Moving Average (ARIMA) is a widely used model for analyzing and forecasting time series data. It combines three key components:

* + AR (Autoregressive): The AR part of the model represents the relationship between an observation and several lagged observations (previous time steps). The parameter p refers to the number of lag observations to be included in the model.
  + I (Integrated): This component represents the differencing of the observations to make the series stationary, which means removing trends or seasonal effects. The parameter d refers to the number of times the differencing is applied.
  + MA (Moving Average): The MA part models the dependency between an observation and a residual error from a moving average model applied to lagged observations. The parameter q refers to the size of the moving average window.

ARIMA (p, d, q):

* + p: The number of lag observations used in the autoregressive model.
  + d: The number of differences required to make the series stationary.
  + q: The size of the moving average window used to model the error term.

How ARIMA Works:

* + First, the series is differenced if needed to achieve stationarity (by applying d differences).
  + Then, the model is fitted using the specified lag order p for autoregressive terms and the specified q for moving average terms.
  + Once the model is trained, it can be used for forecasting future time steps based on past observations and errors.

ARIMA is effective for non-seasonal time series forecasting, while SARIMA (Seasonal ARIMA) is used for time series with seasonal components.