# Forecasting algorithm for Airlines data with data driven approach

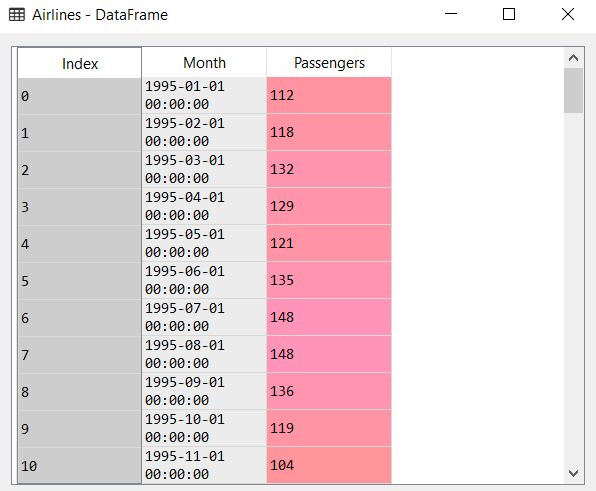
Problem Statement:

Forecast Airlines Passengers data set. Prepare a document for each model explaining

how many dummy variables you have created and RMSE value for each model. Finally which model you will use for

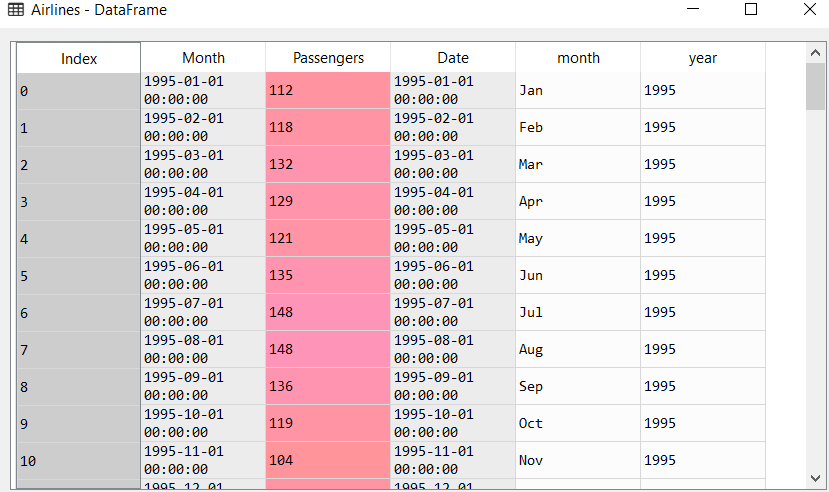
Forecasting.

Considering the airline dataset as follows:

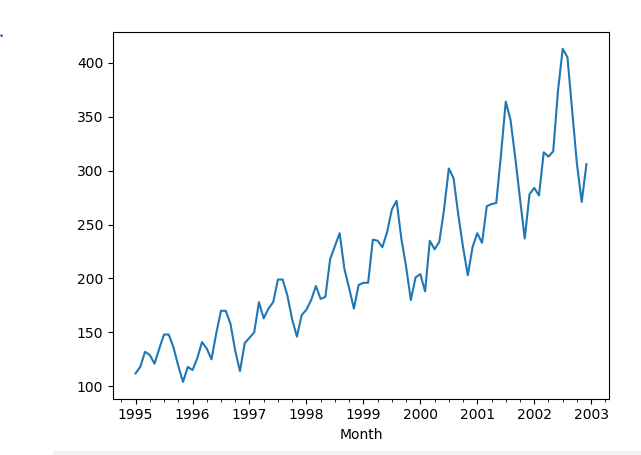


From the above dataset. The date containing the month and the year is extracted:

We get the following dataset:

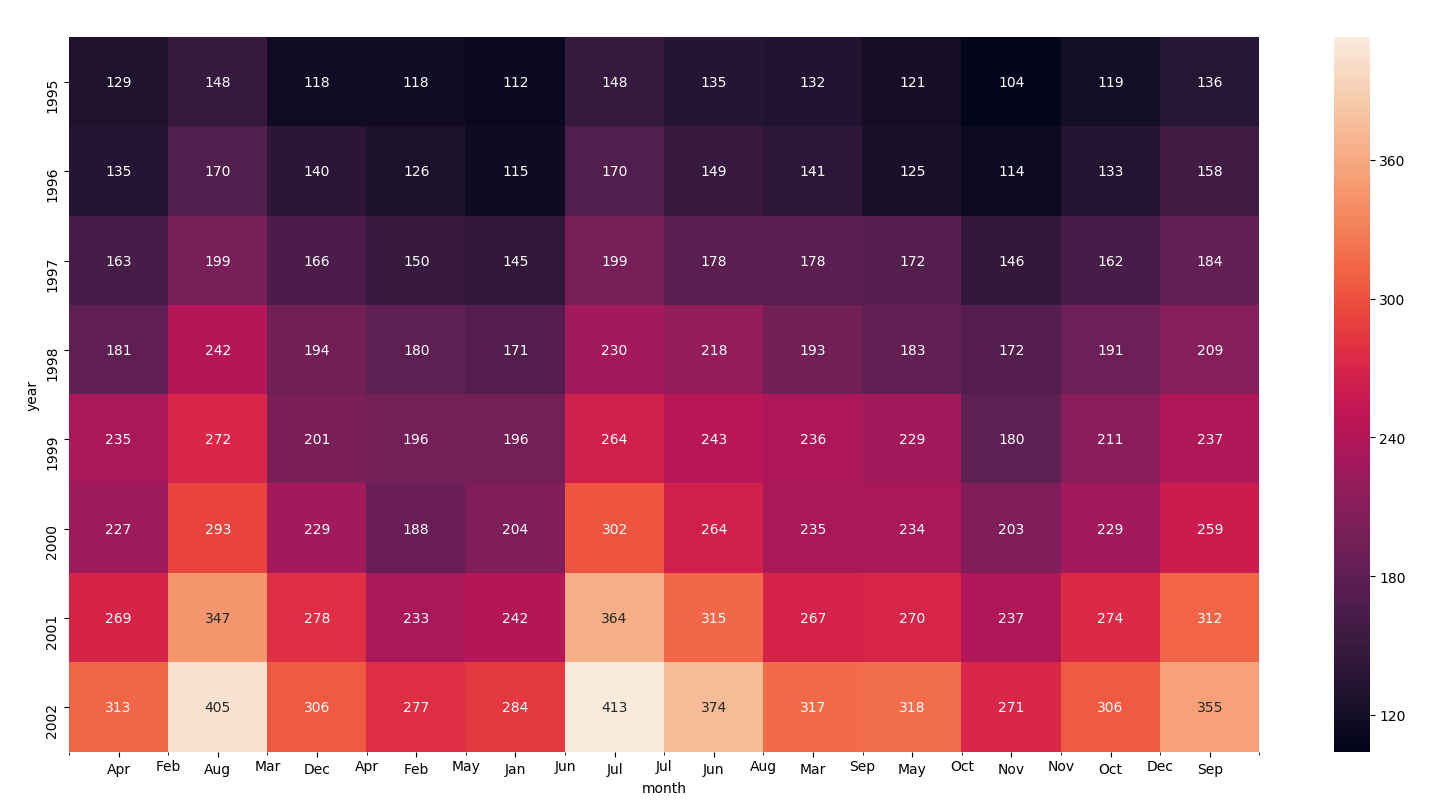


The following is the time series plot:



From the above plot, we can say that the trend is an upward trend with multiplicative seasonality.

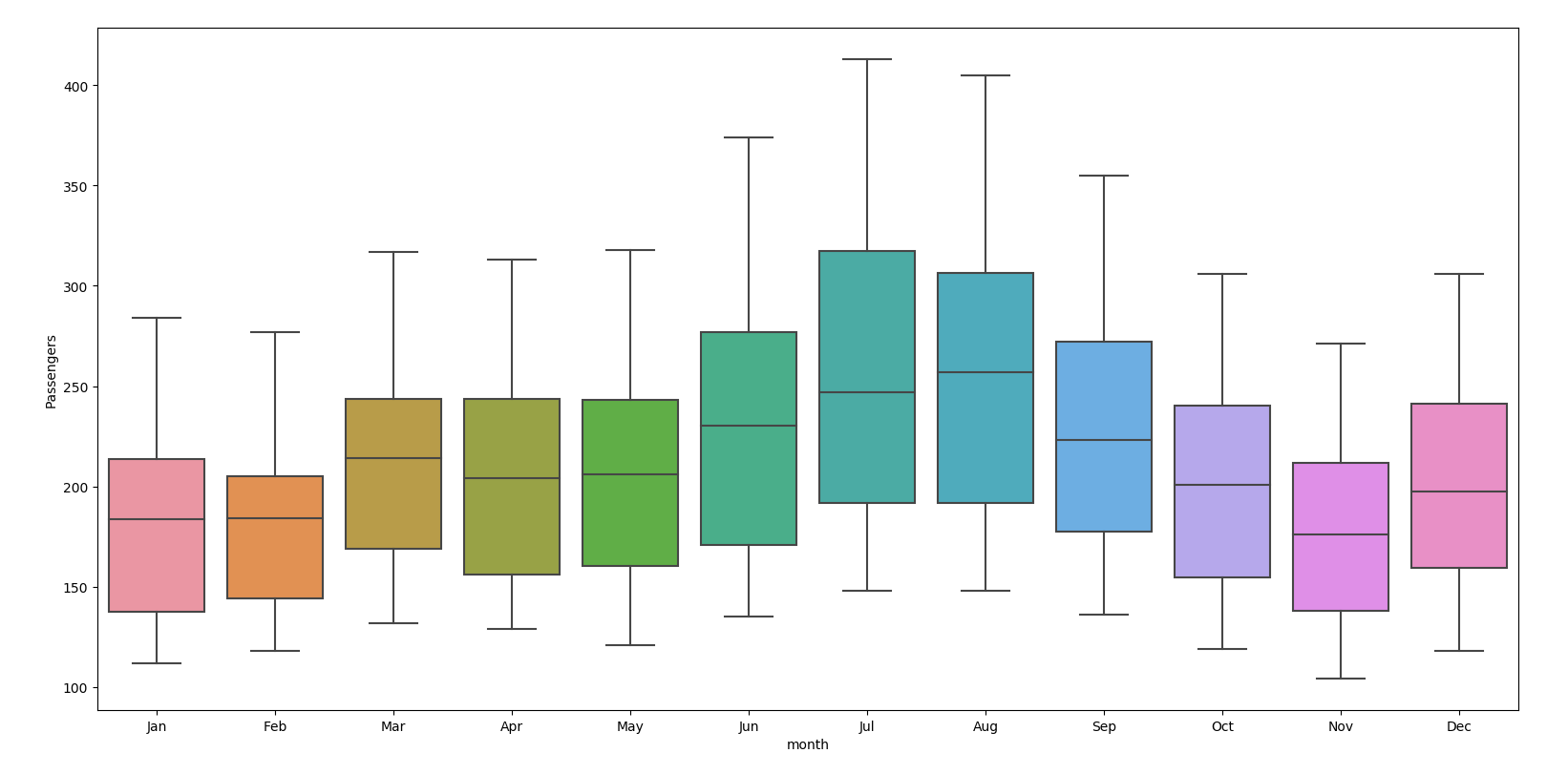
The following is the heatmap of the airline passengers trend data



From the above trend we can identify that:

1. The number of passengers travelling via airlines are almost exponentially increasing by the years
2. The number of passengers travelling during a year are usually more during the month of Aug, July and Sept.

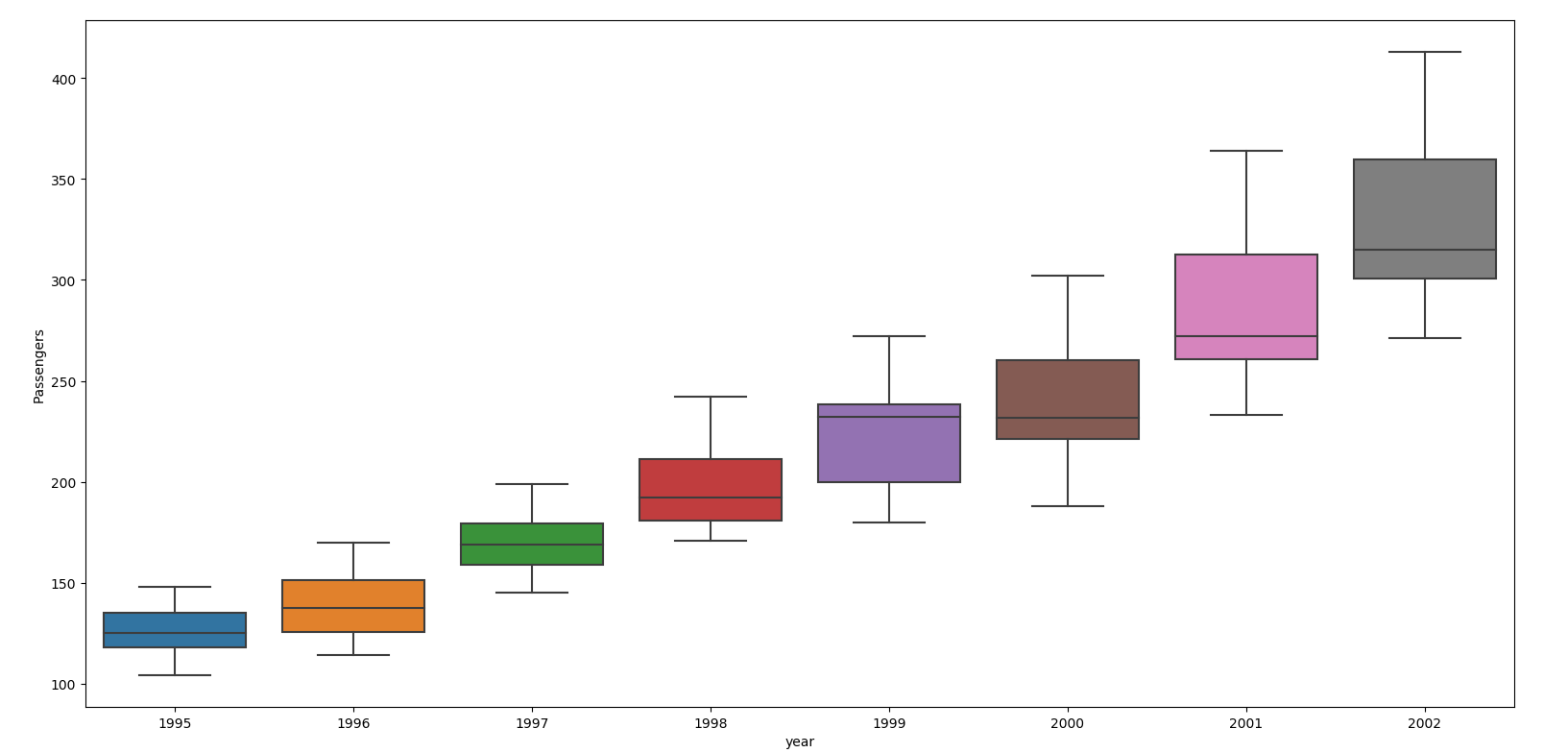
The following are the box plot for the month data:



From the above boxplots we get the following insights:

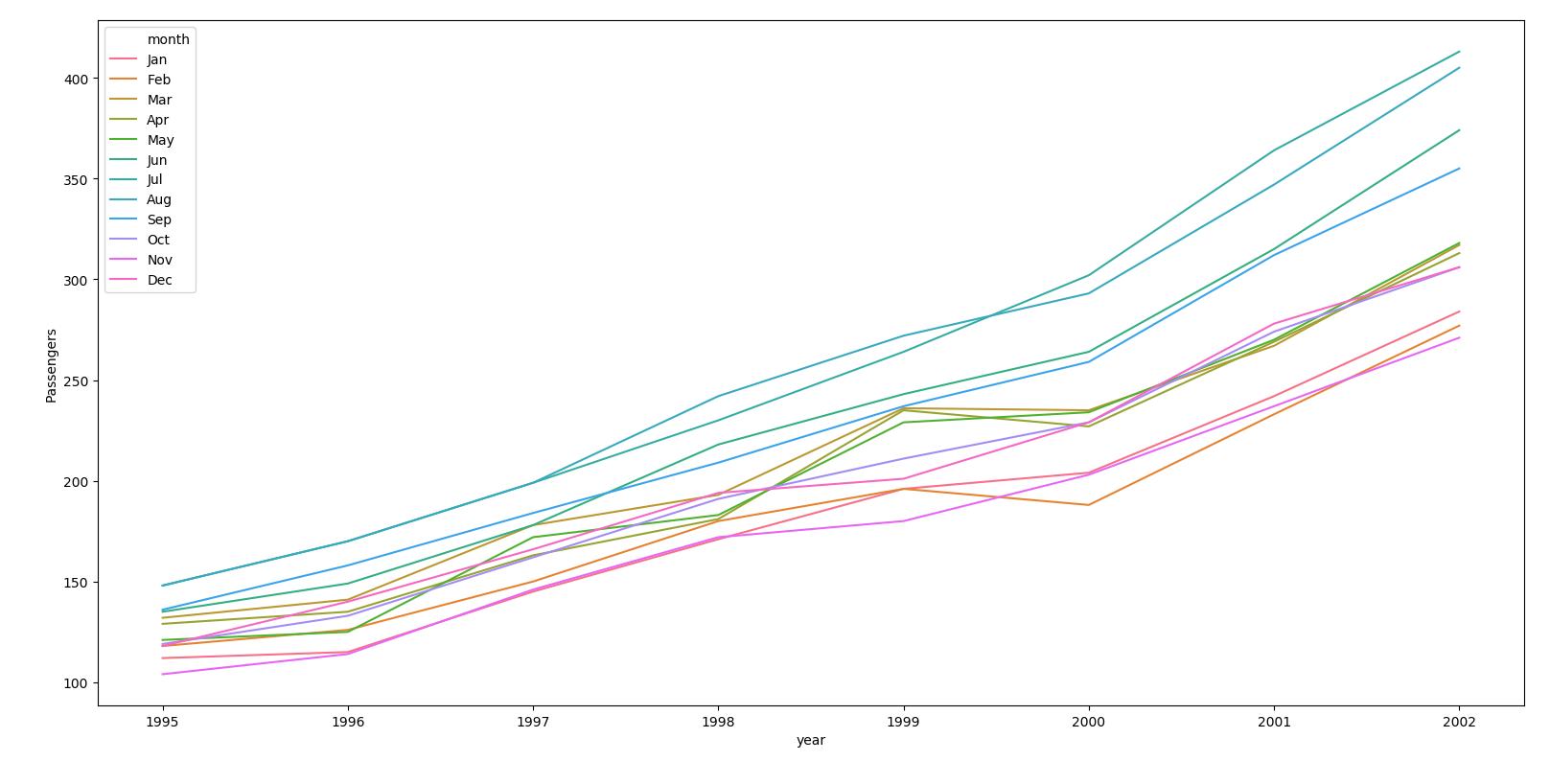
1. The number of passengers are higher in the month of Aug, July and Sept as backed by the heatmap diagram
2. There are no outliers in the above boxplots that means that there is no sudden increase or decrease in the number of passengers flying with an airline

The following are the boxplots for yearly data:

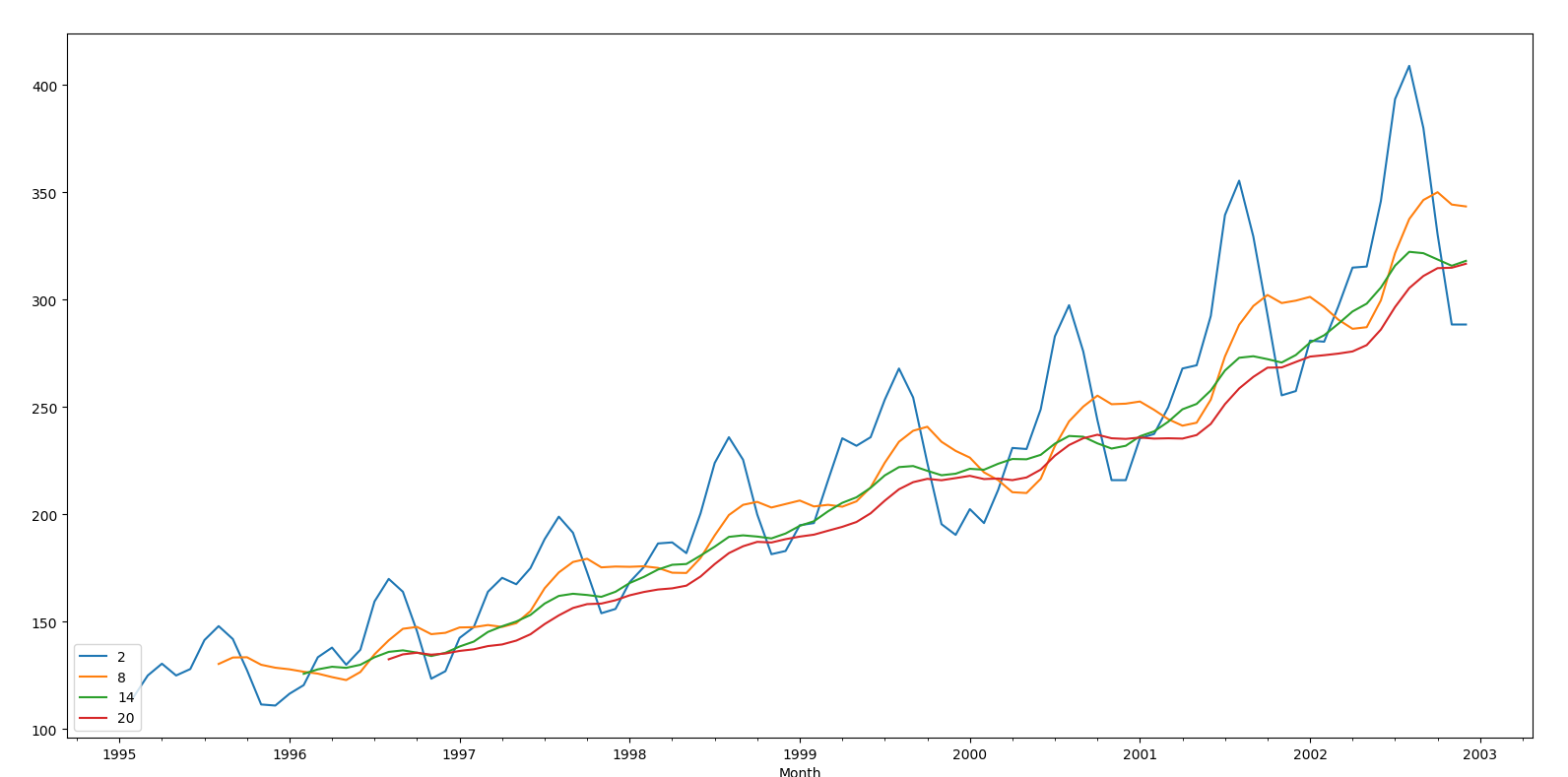


From the above, we estimate that there is a steady and almost an exponential rise in the number of passengers year on year.

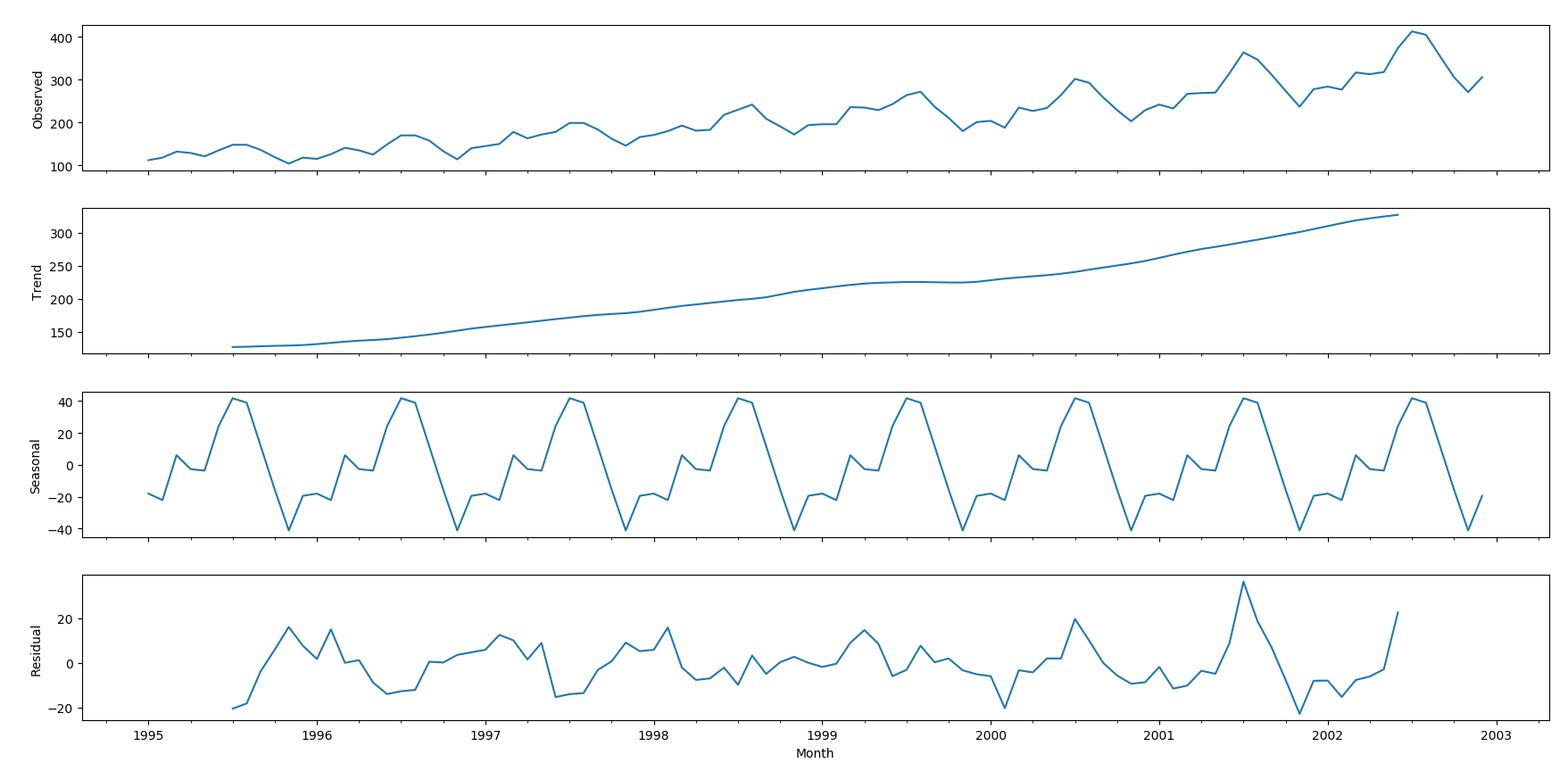
The following is the line plot reflecting the assumption of the year on year rise in number of passengers



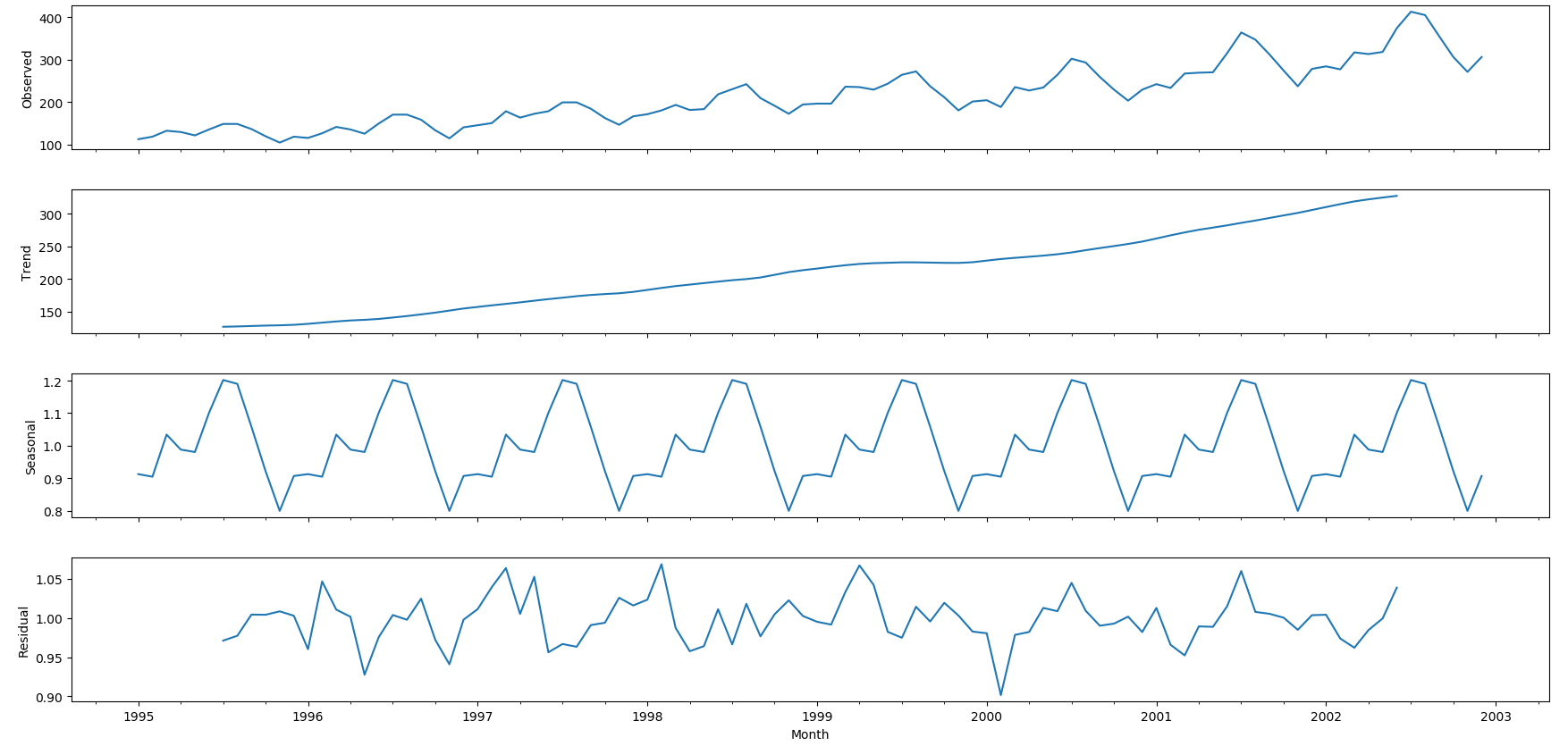
Moving average for the time series to understand better about the trend character in Airlines



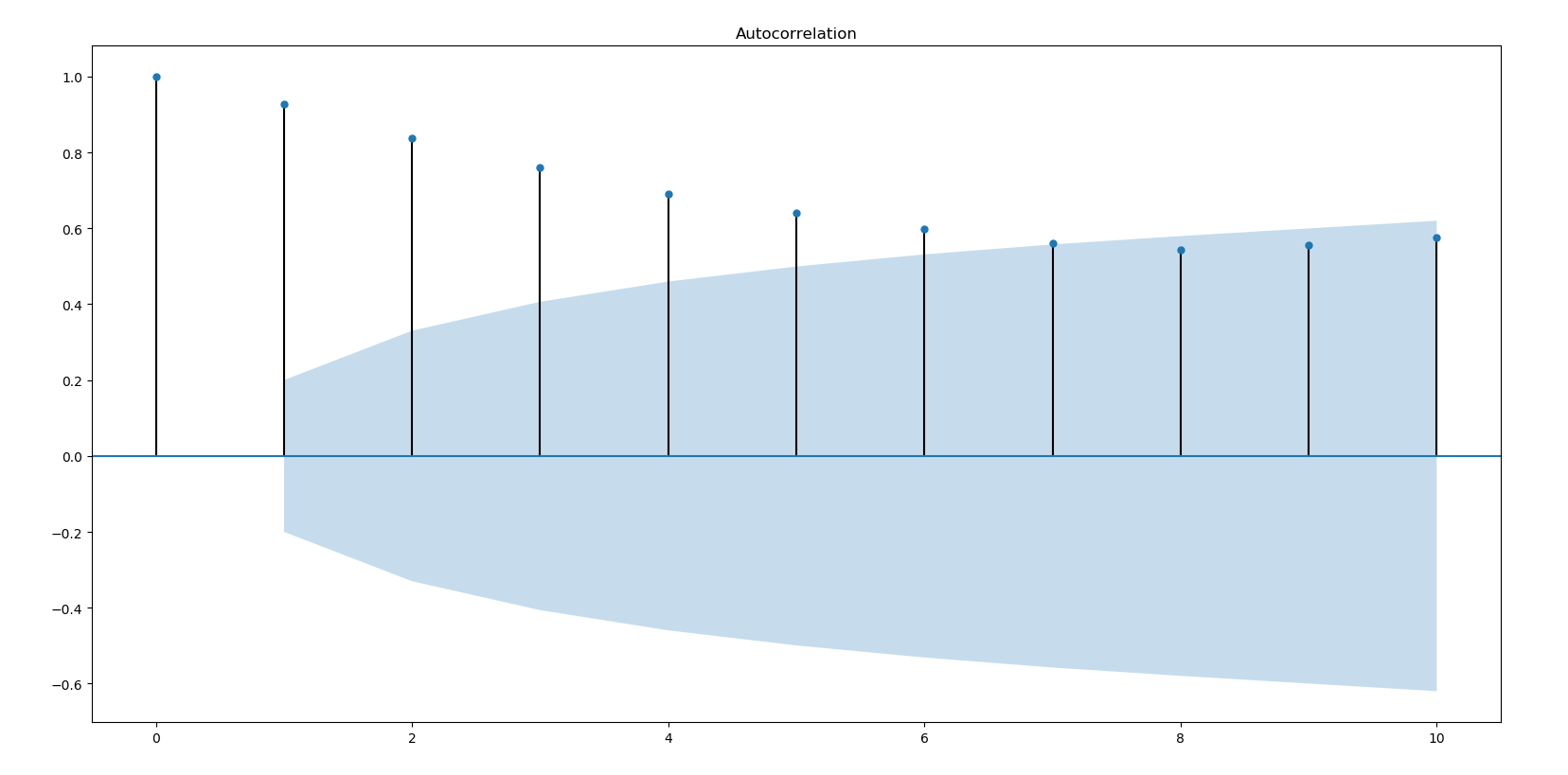
Time series decomposition plot for additive decomposition



Time series decomposition plot for multiplicative decomposition

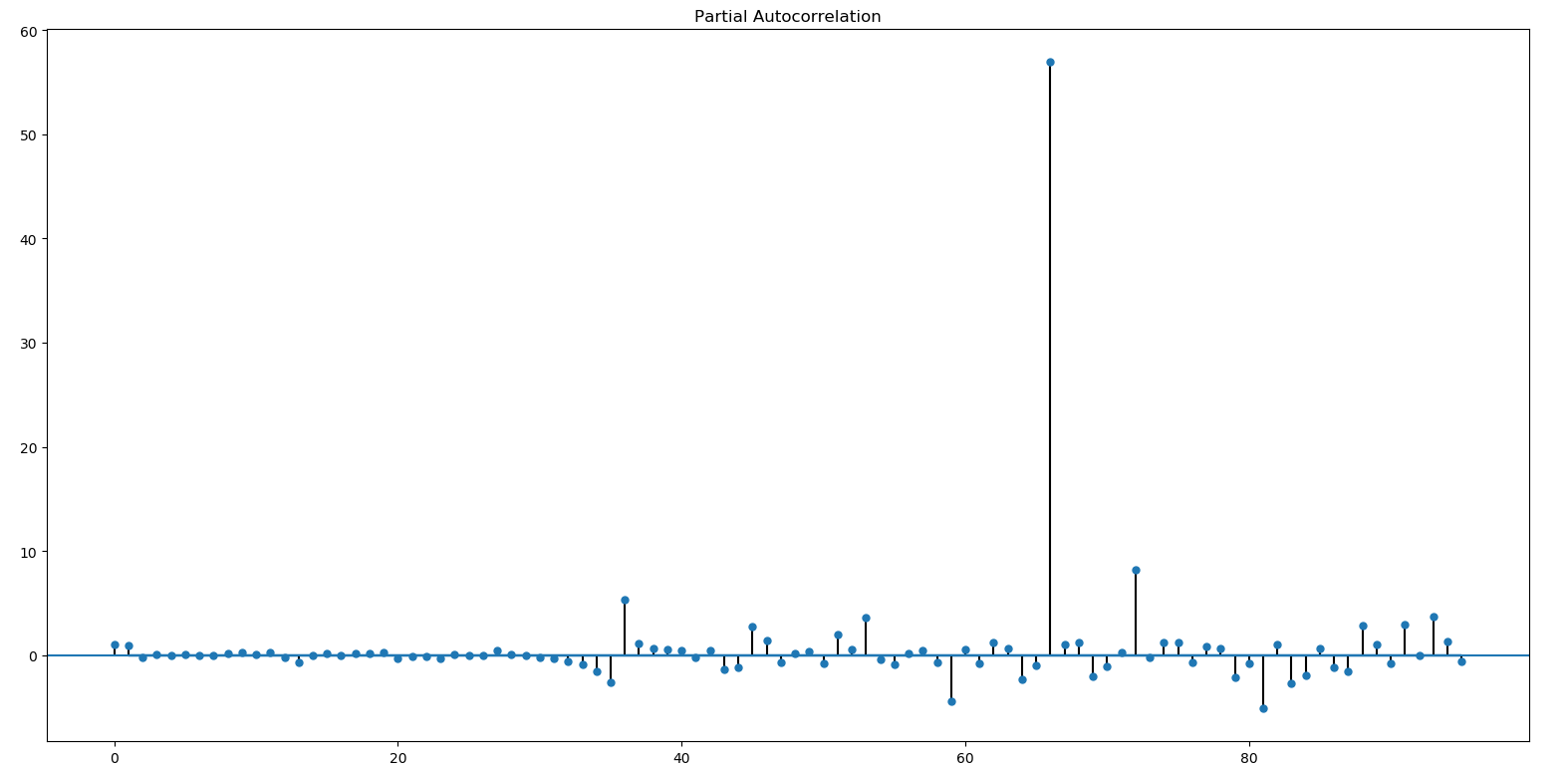


The following is the autocorrection correction factor ACF diagram:



In the above plots we estimate that the autocorrection between points 0 to 7 have correction factor with lagged data beyond the confidence bands

The following is the partial auto correlation factor diagram:

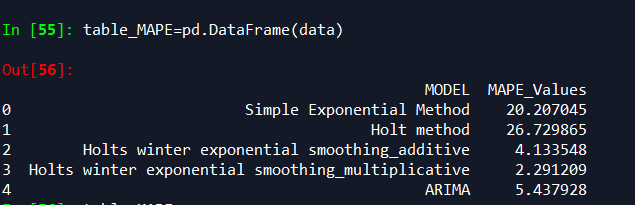


We can notice that there is an extreme value between 60 and 80

After splitting the data in 80:20 ratio, we will train and test models to develop prediction models

Using Simple Exponential method, Holt method, Holts winter exponential smoothing with additive seasonality & additive trend, Holts winter exponential smoothing with multiplicative seasonality & additive trend and finally with ARIMA

Using the above models we compare the results of each of the models using Mean Absolute Percentage Error. The following is the table:



From the above table with Holts winter exponential smoothing with multiplicative seasonality and additive trend model we get the least MAPE value of 2.289. Which provides comparatively accurate prediction results as compared to other models.

Holts method has the highest MAPE value of 26.729 whereas Holts winter with additive and ARIMA provide a MAPE value of 4.13 and 5.43 respectively.