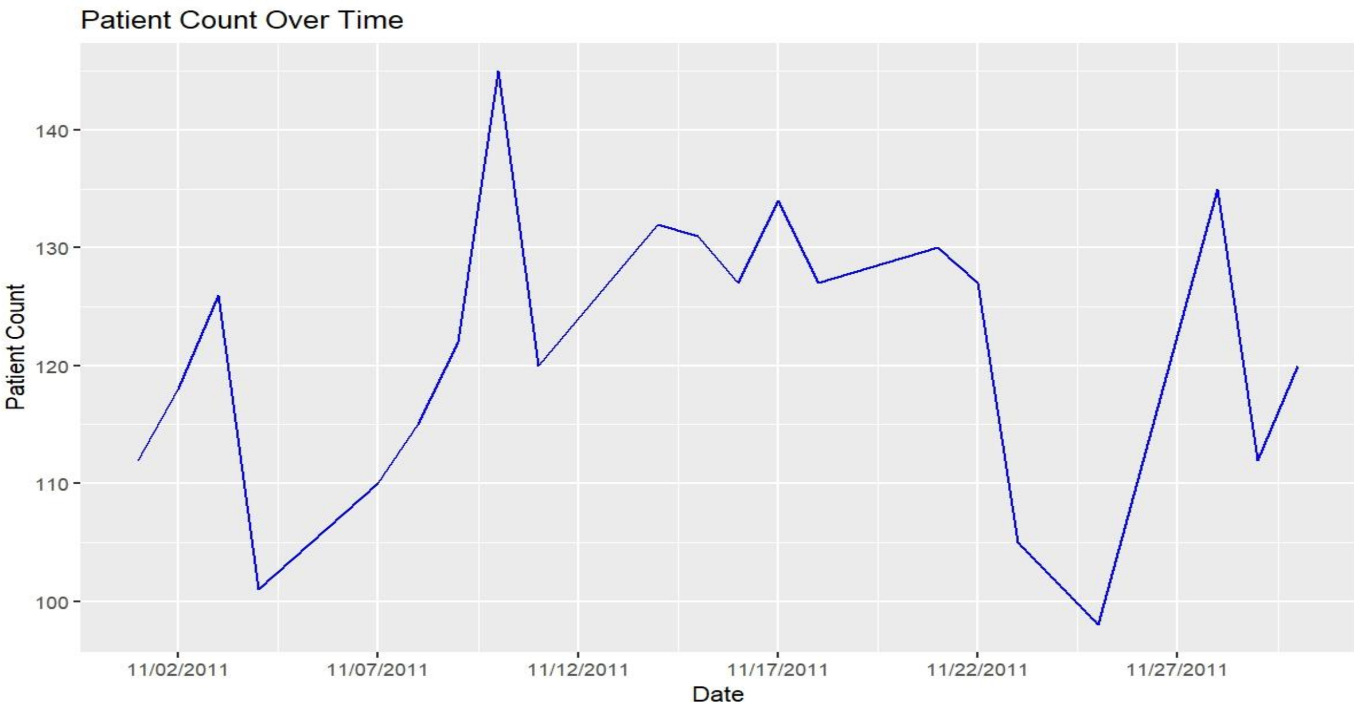
1. Plot the historic data and analyze the trend.



Analysis:

In the above line trend the x-axis represents the Date and y axis represents the patient count. We can notice that the patient count data for November 2011 displays an initial rise, reaching its peak at 145 patients around 11/10/2011, but rest of the month sees a decline in patient numbers, with occasional fluctuations but an overall decrease including significant drops on November 23rd which is 105 patients and November 25th which is 98 patients. These reductions could potentially be attributed due to holiday observances or other external factors.

2. Descriptive statistics:

Descriptive statistics	Values
Mean	**121.28**
Median	**122.00**
Absolute Deviation	**9.70**
Minimum	**98.00**
Maximum	**145.00**
1st Quartile	**112**
3rd Quartile	**130**

Skewness	**-0.196**

Analysis:

From the descriptive statistics we notice that minimum patient count is 98 and maximum patient count is 145, with a mean of 121.28 and a median of 122. The interquartile ranges from the patient count 112 to 130, with a slight negative skewness. Variance and standard deviation are 144.51 and approximately 12.02, respectively. The absolute deviation is 9.70.

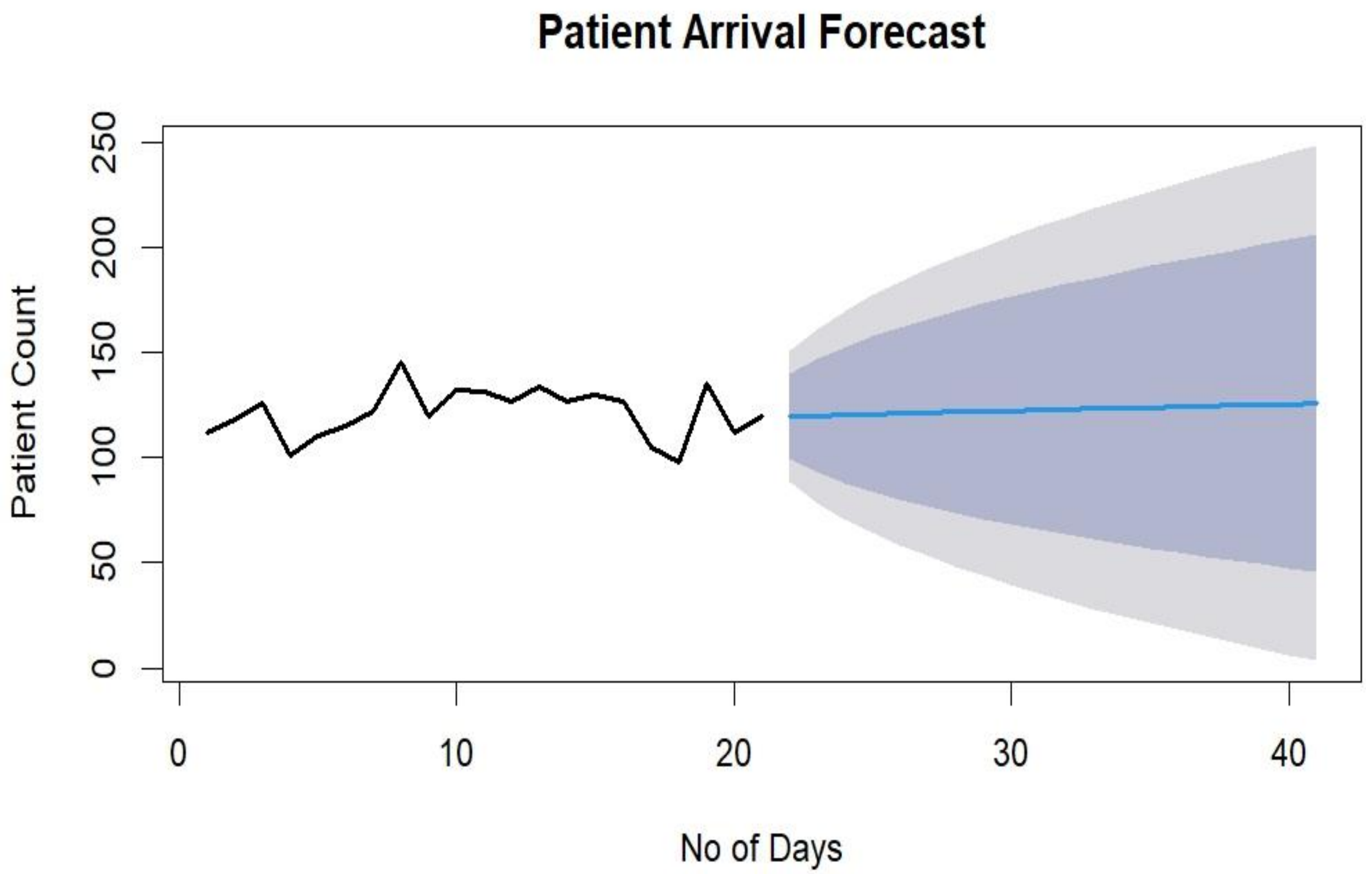
3. Apply exponential smoothing to calculate the forecast for the next month (calendar days - 20 d ays). You need to find the optimal alpha value.

Best alpha: 0.87.

MAE: 9.53

Days	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
22	119.6418	99.2697	140.014	88.48529	150.7984
23	119.9652	92.9609	146.969	78.665707	161.2646
24	120.2885	87.9851	152.592	70.884757	169.6922
25	120.6118	83.7627	157.461	64.25597	176.9676
26	120.9351	80.0416	161.829	58.393931	183.4763
27	121.2584	76.6852	165.832	53.089606	189.4273
28	121.5818	73.6095	169.554	48.214543	194.949
29	121.9051	70.7585	173.052	43.683114	200.1271
30	122.2284	68.0926	176.364	39.434846	205.022
31	122.5517	65.5827	179.521	35.425157	209.6783
32	122.8751	63.2066	182.544	31.620053	214.1301
33	123.1984	60.9469	185.45	27.992917	218.4039
34	123.5217	58.7896	188.254	24.522442	222.521
35	123.845	56.7233	190.967	21.191257	226.4988
36	124.1684	54.7388	193.598	17.984979	230.3518
37	124.4917	52.828	196.155	14.891539	234.0918
38	124.815	50.9843	198.646	11.900694	237.7293
39	125.1383	49.202	201.075	9.00366	241.273
40	125.4617	47.476	203.447	6.192838	244.7305
41	125.785	45.802	205.768	3.4616	248.1084

4. Plot the historic and forecast values and analyze the results of the confidence values.



Analysis:

The x axis represents the no of days and y axis represents the patient count. For example, consider the day 30 which predicts a point estimate of 122.2284 with an 80% confidence interval ranging from 68.09261 to 176.3642 and a 95% confidence interval ranging from 39.434846 to 205.022, we can observe a gradual increase in the forecasted values over the days. The point forecast indicates a steady upward trend, suggesting potential growth or accumulation in the variable being forecasted. Overall, while the point forecast indicates a positive trend, the widening confidence intervals highlight the importance of monitoring and so we need to keep checking and adjusting the predictions as we get more information.

5. Provide and analyze the forecast error measures.

The prediction model's accuracy is analyzed using four metrics: Mean Absolute Deviation (MAD), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE), Mean Squared Error. These metrics all indicate lower values are better. The MAD of 9.53 shows the average difference between predicted and actual values. The MAPE of 8.17% reveals the average percentage difference.

With a mean squared error (MSE) of 144.0521, the squared discrepancies between the actual and predicted values are, on average, 144.0521 units. Higher values indicate greater variability and dispersion of errors. Finally, the RMSE of 12.00 reflects the average variability of predictions from actual values. While these results provide a baseline, improvement is possible to reduce errors.

R code

```
#for attaching the file to the Rstudio
attach(Historic_demand_ER_2024)



# Loading  necessary libraries
library(forecast)
library(Metrics)
library(ggplot2)
library(DescTools)
library(fBasics)






# Read the Excel file with correct date format
Date1 <- as.Date(Date)

# plot line graph to represent the patient count over time
ggplot(Historic_demand_ER_2024, aes(x = Date1, y = Patients)) +
  geom_line(color="blue") +
  labs(x = "Date", y = "Patient Count") +
  ggtitle("Patient Count Over Time") +
  scale_x_date(date_breaks = "5 day", date_labels = "%m/%d/%Y")
```

```
#Descriptive statistics mean, median, absolute deviation, min, max


#To Determine the Detail summary of the demand
basicStats(data.frame(Patients))
#To Determine the abs_deviation
abs_deviation<-mean(abs(Patients-mean(Patients)))
cat("Absolute Deviation: ", abs_deviation, "\n")



#To find the optimal alpha value
# Choose the appropriate method for your data: "ZZZ", "AAN", "ANA", "AAA",
#Generate a time series.
my_ts <- ts(c(Patients), frequency = 1)


#count of the validation set's observations
validation_length <- 20
#initialize variables with the minimal possible error and best alpha
best_alpha <- NULL
min_error <- Inf
for(alpha in seq(0.1, 0.9, by = 0.11)){
  ets_model <- ets(my_ts, model = "AAN", alpha = alpha)
  forecast_values <- forecast(ets_model, h = validation_length)
  error <- mae(forecast_values$mean, my_ts[(length(my_ts) - validation_length + 1):
length(my_ts)])
  if(error < min_error){
    min_error <- error
    best_alpha <- alpha
  }
```

```
}


# Print the forecasted values
print(paste("best alpha:", best_alpha))
print(paste("mae:", min_error))
print(forecast_values)


# Plot the historic and forecast values
plot(forecast_values,
    main = "Patient Arrival Forecast",
    xlab = "No.of Days",
    ylab = "Patient Count",lwd = 2)

# Calculate forecast error measures
actual_values <- my_ts[(length(my_ts) - validation_length + 1): length(my_ts)]
forecasted_values <- forecast_values$mean


# Mean Absolute Percentage Error (MAPE)
mape_error <- mean(abs((actual_values - forecasted_values) / actual_values)) * 100


# Mean Absolute Deviation (MAD)
mad_error <- mean(abs(actual_values - forecasted_values))

# Mean Squared Error (MSE)
```

```
mse_error <- mean((actual_values - forecasted_values)^2)

# Root Mean Squared Error (RMSE)
rmse_error <- sqrt(mean((actual_values - forecasted_values)^2))

# Print the error measures
cat("Mean Absolute Percentage Error (MAPE): ", mape_error, "%\n")
cat("Mean Absolute Deviation (MAD): ", mad_error, "\n")
cat("Mean Squared Error (MSE): ", mse_error, "\n")
cat("Root Mean Squared Error (RMSE): ", rmse_error, "\n")
```