**4. Forecasting and Seasonality Analysis**

To obtain an accurate information of the demand levels, the number of patients that arrive to the ED over specific periods of time is explored. This allows us to have an actual state simulation that is very close to the reality and propose solutions that don’t compromise service levels and capacity, even in the worst-case scenario.

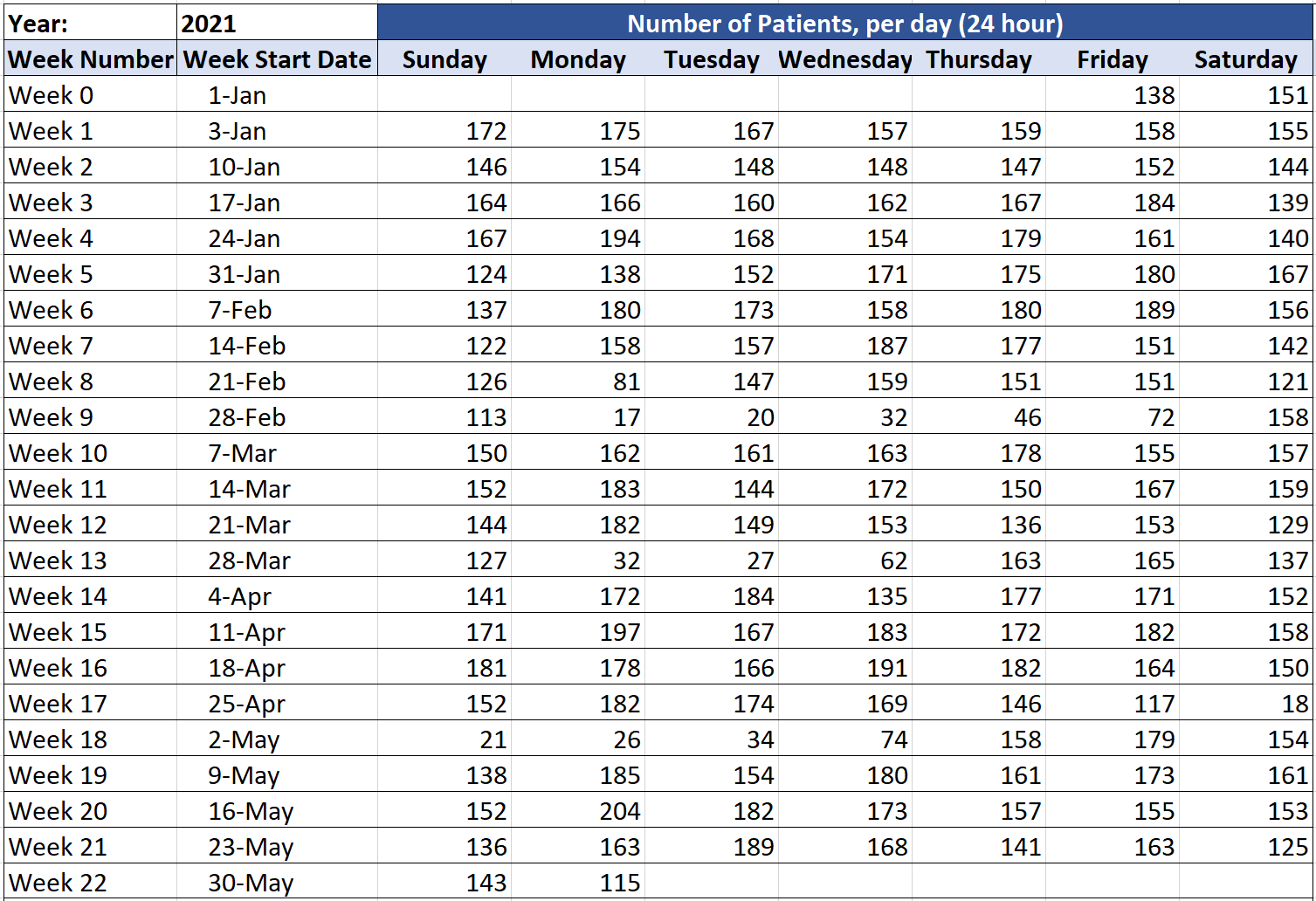
**4.1 Forecasting Patient Arrival**

As most of the resources needed to run the ED are planned per shift or per hour within a day, the forecast is done per day of the week to obtain at least 1 additional week of patient arrival data.

The approach taken to ensure a reliable and accurate forecast is outlined below.

* + 1. **Obtain the data**

The Bixler ED patient data provided was used, which covers from January 1st to May 31st of 2021. The arrival date and time information were filtered to obtain the count of patients that enter to the ED per day, and it was classified per day of the week to also explore seasonality trends within days of the week. Below is shown the table of all data observations.



* + 1. **Establish the forecast methods**

To explore the data and obtain the most accurate predicted values 2 different forecast methods were explored: The Moving Average and the Exponential Smoothing methods.

The **Moving Average (MA)** forecast method consist in obtaining the next period values based on the arithmetic mean on the most recent observations of a specified timeframe. The formula used is the following:

Where:

is the forecasted value

are the data observations values.

is the total of periods that have the observations

is the identifier number of the observation in the data ()

is the number of data observations that are considered to be averaged.

The **Exponential Smoothing (ES)** method consist in a weighted moving average forecast that is done based on the previous observation and an optimal factor. The formula used to calculate the forecasted values is:

Where:

is the new forecasted value

Dt is the value of observation in the data

is the previous forecasted value

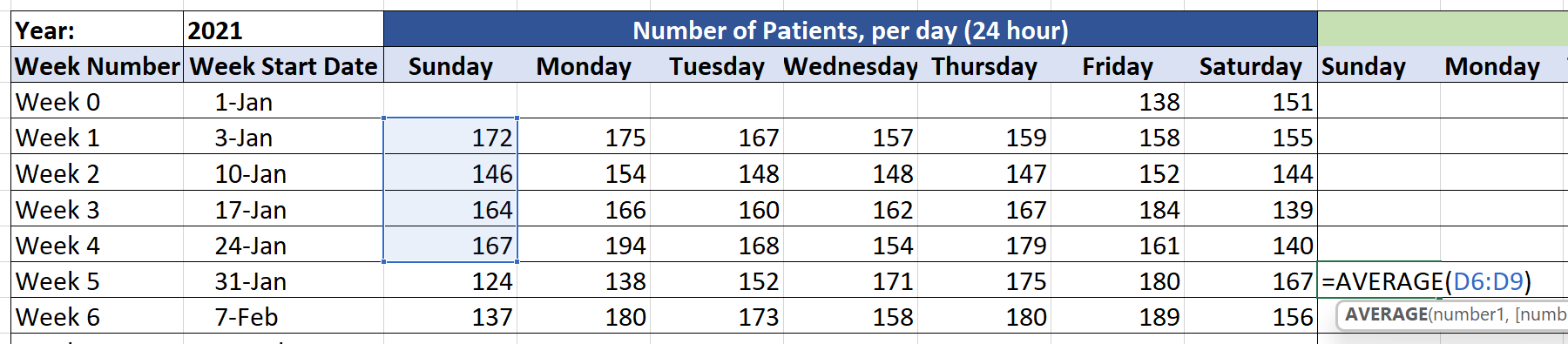
is the smoothing constant, which is 0 < α < 1, but generally is between 0.1 and 0.2 for stability of the forecast

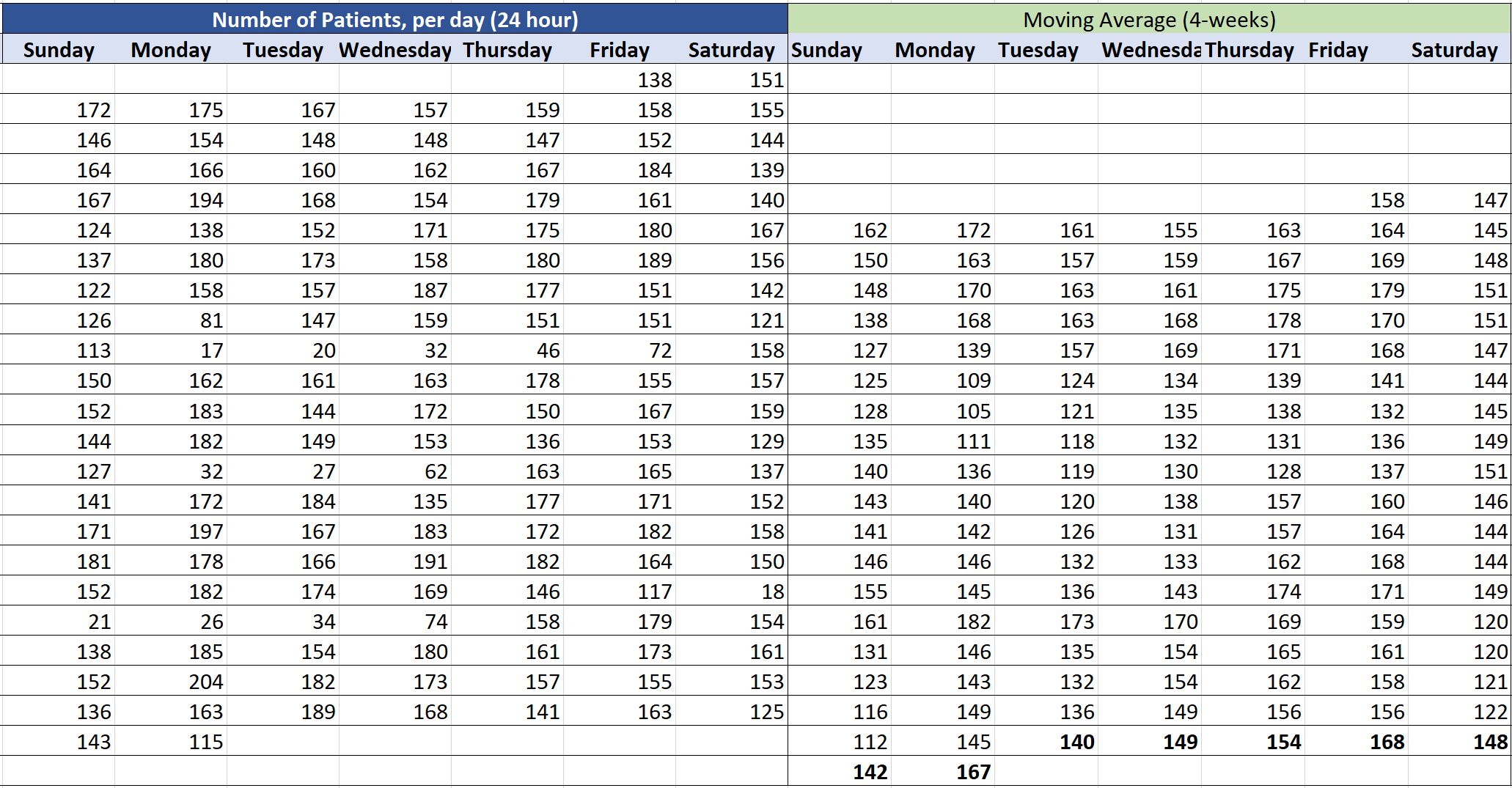
* + 1. **Forecast current demand with both methods**

The forecasts of the demand levels in the ED were computed using both methods described above, for the periods covered by the original data (Jan-1 to May-31).

**Moving Average Forecast**

This forecast was computed obtaining the number of patients per day, initially based on the average of 1 month of previous data (N = 4). Below is shown how the values were computed in the spreadsheet, and the results obtained.



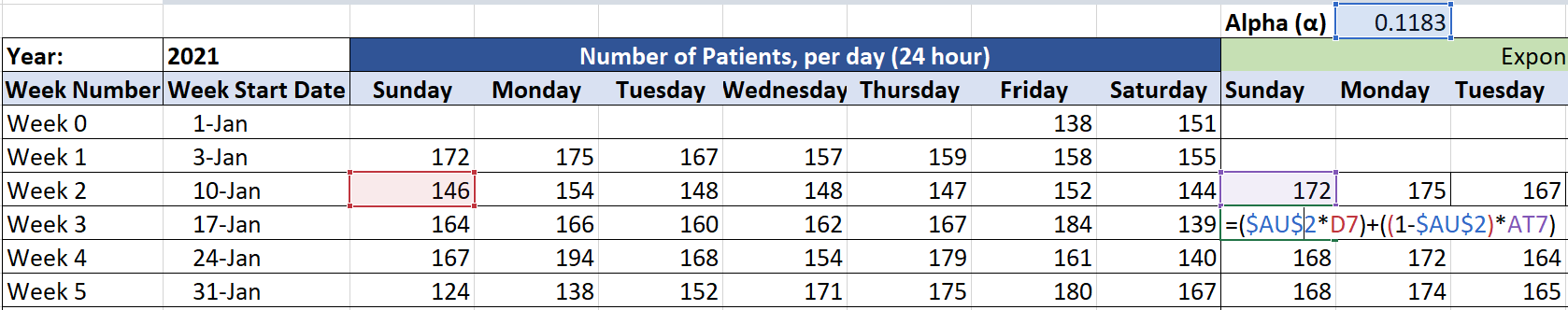


For a 4-week based average, the new forecasted values are:

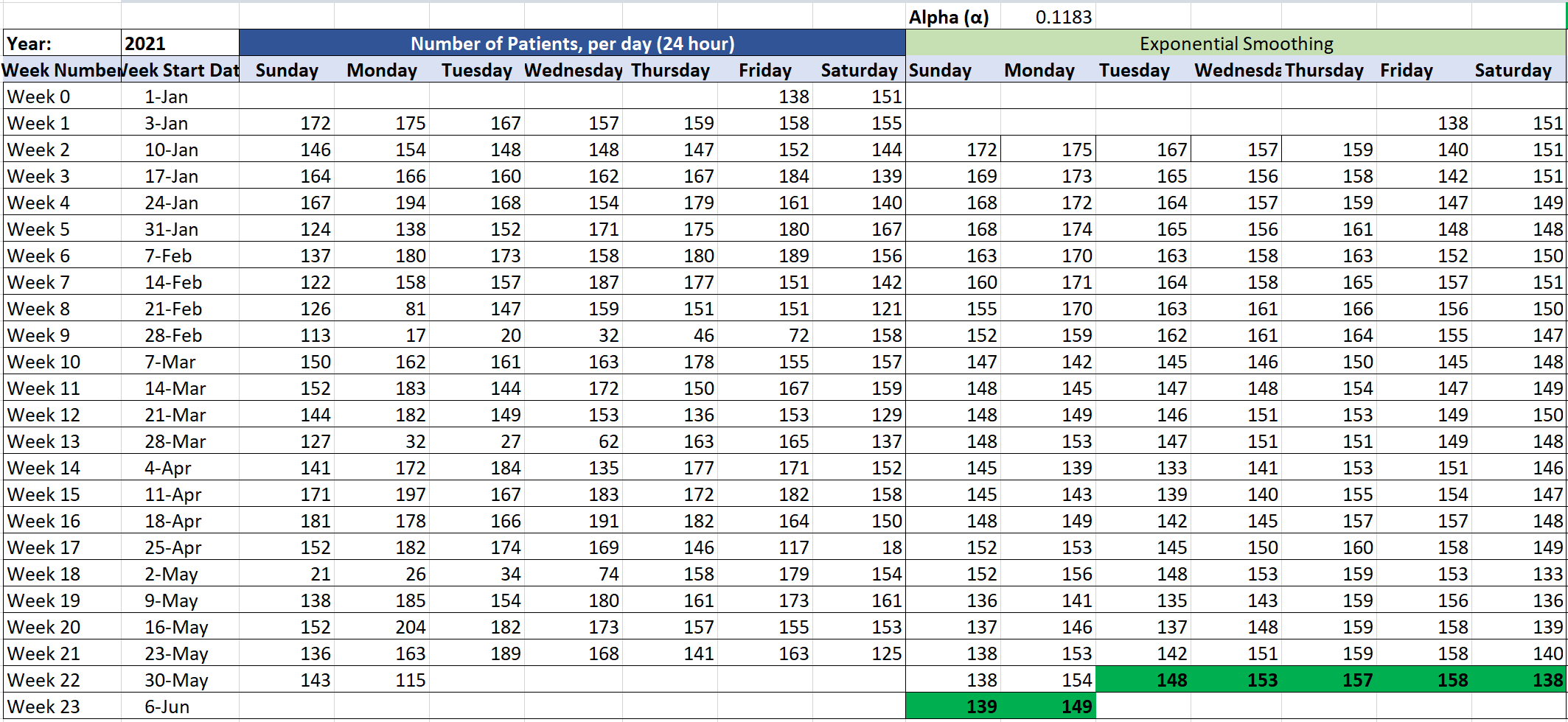
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Moving Average Forecast, (N=4) | | | | | | |
| **Sunday** | **Monday** | **Tuesday** | **Wednesday** | **Thursday** | **Friday** | **Saturday** |
| 142 | 167 | 140 | 149 | 154 | 168 | 148 |

**Exponential Smoothing Forecast**

The formula described for the Exponential Smoothing Forecast was computed in the Microsoft Excel spreadsheet to use it with the Bixler ED data, as it is shown in the following table.



The results obtained for all the data are shown in the table below.



* + 1. **Measure each forecast error and choose the most accurate forecast method**

To evaluate the forecast methods and select the most accurate, the error for each was computed using 3 methods, which are described below.

The first method that was considered is the **Mean Absolute Deviation (MAD)**, which is the average of the differences between the forecasted value and the original observation. The formula used to compute this is the following:

Where:

|ei| is the absolute value of the error

Di is the value of the data of the observation i

Fi is the forecasted value of the observation i

n is the total number of observations that are evaluated

The second method used is the **Mean Square Error (MSE)**, which measures the average of the squared difference between the forecasted value and the original observation. This error evaluation method is more sensitive, as it emphasizes any significant deviation the forecast may have. The formula of this method is the following:

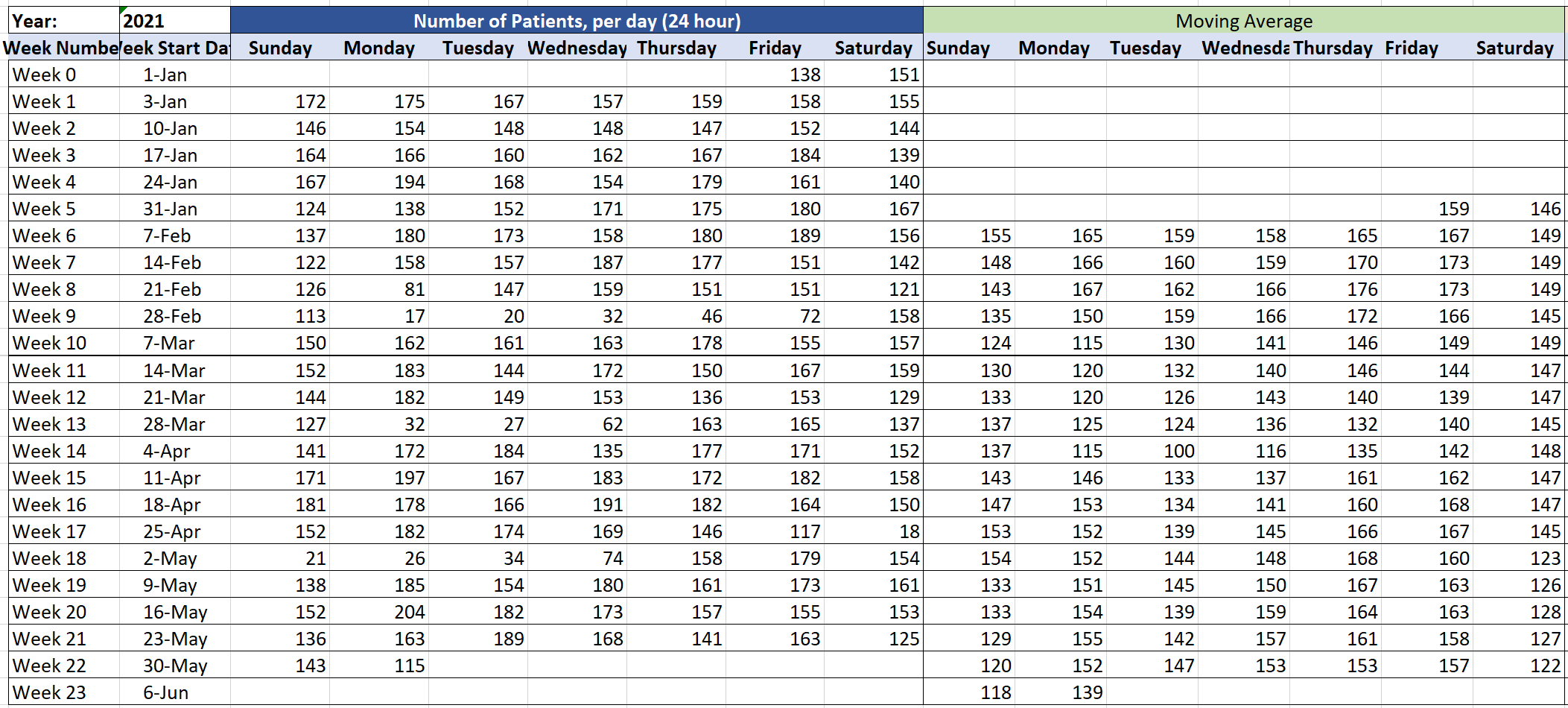
The third evaluation method is the **Mean Absolute Percentage Error (MAPE)** which consists in the average of the sum of each absolute errors (the difference between the forecasted and the original value), divided by the demand, to measure a general percentage error. The formula is the following:

Before comparing and selecting the best forecast method, an error evaluation for the parameters used in each forecast was performed, in order to get the most accurate results in each method.

In the Moving Average method, the accuracy of the forecast estimation relies on the number of observations considered for the average (how much is the N). In that way, many N values were tested, and the errors were measured using the 3 evaluators. The table below shows the results of the evaluations.

|  |  |  |  |
| --- | --- | --- | --- |
| **Optimal N value for Moving Average Forecast** | | | |
| **N** | **MAD** | **MSE** | **MAPE** |
| 2 | 36.27 | 2866.94 | 0.5772 |
| 3 | 36.05 | 2693.91 | 0.6056 |
| 4 | 33.56 | 2323.79 | 0.5871 |
| 5 | 32.30 | 2172.68 | 0.5863 |
| 6 | 33.55 | 2370.66 | 0.6254 |
| 7 | 35.06 | 2560.98 | 0.6668 |
| 8 | 35.52 | 2633.39 | 0.6998 |

For this data, the most accurate moving average forecast was obtained based on 5 previous observations (N = 5).



For a 5-week based average, the most accurate values using this forecast method are:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Moving Average Forecast, (N = 5)** | | | | | | |
| **Sunday** | **Monday** | **Tuesday** | **Wednesday** | **Thursday** | **Friday** | **Saturday** |
| 118 | 139 | 147 | 153 | 153 | 157 | 122 |

For the Exponential Smoothing forecast, another approach was taken. As the forecast estimation highly relies on the smoothing factor (α), the optimal value was obtained running a Generalized Reduced Gradient Non-linear program, using the Microsoft Excel solver. This value must satisfy the minimum error level condition and is constrained by its magnitude, which must be between 0 and 1.

Based on each error method, the optimal smoothing factor obtained were the following:

|  |  |  |
| --- | --- | --- |
| **Optimal α value for ES Forecast** | | |
| **Error Eval. Method** | **Error Value** | **Optimal** **α** |
| Mean Absolute Deviation (MAD) | 26.89 | 0.0438 |
| Mean Square Error (MSE) | 1787.24 | 0.0860 |
| Mean Absolute Percentage Error (MAPE) | 0.4964 | 0.1183 |

With each optimal solution three different optimal smoothing factors are obtained. For the purposes of this forecast, the value considered was α = 0.1183 computed based on the minimum MAPE as small numbers between 0.1 and 0.2 give stable predicted results. Nevertheless, the other error indicators were considered to compare the forecast methods. The forecasted value obtained using this method were:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Exponential Smoothing Forecast** | | | | | | |
| **Sunday** | **Monday** | **Tuesday** | **Wednesday** | **Thursday** | **Friday** | **Saturday** |
| 139 | 149 | 148 | 153 | 157 | 158 | 138 |

After evaluating the error of the best forecasts obtained with both methods, each error indicator was computed to select the one with less error. To ensure an equitable comparison, the factors of each forecast method were converted to have the same distribution of forecast error. This was done adjusting the smoothing factor (α) to the optimal N of the moving average method, through the formula:

Where for this case:

The results of the error evaluation are shown below:

|  |  |  |
| --- | --- | --- |
| **Forecast Error Comparison** | | |
| **Error Indicator** | **Forecasting Method** | |
| **MA (N=5)** | **ES (α=0.3333)** |
| MAD | 32.30 | 30.51 |
| MSE | 2172.68 | 2027.14 |
| MAPE | 0.5863 | 0.5114 |

The evaluation results show the Exponential Smoothing has less error levels, according to all methods, therefore is the best forecast estimator for this case. If the same evaluation is applied with the optimal smoothing factor obtain above, the comparison would look like the following:

|  |  |  |
| --- | --- | --- |
| **Forecast Error Comparison** | | |
| **Error Indicator** | **Forecasting Method** | |
| **MA (N=5)** | **ES (α=0.118)** |
| MAD | 32.30 | 27.51 |
| MSE | 2172.68 | 1797.14 |
| MAPE | 0.5863 | 0.4964 |

The results of the forecast obtained using this smoothing factor are shown in the bar chart below.

There can be seen that in the following weeks more patients will arrive from Wednesday and increasing until Friday, and less patients will come in from Saturday to Sunday.

**4.2 Seasonality Analysis**

The other important factor about capacity and resources planning is the knowledge of variation during specific periods of time. A seasonality analysis was performed to the patient’s arrival data, to ensure proper capacity levels during peak or low times. For this case, the multiplicative model was used to determine the percentage of variations of demand levels, based on the overall average of patient’s arrival to ED per day. This analysis was done to days of the week and monthly periods, but due to limitations of the data provided, only five months were analyzed.

To apply the multiplicative model, the steps described below were taken.

**4.2.1 Obtain the data**

To have the data for the day of week seasonality, the data was gathered and organized as described for the forecast analysis in section 4.1.1. For the monthly seasonality, the number of patients arriving per day was placed in a table with the days as rows and the months as columns, as shown below.

Table

Description automatically generated

**4.2.2 Obtain periodic and overall averages**

After organizing the data, the average number of patients per periods and the overall average were obtained. The overall average is the average of each observation of the data, which is numbers of patients per day, from January 1st to May 31st. The average per day of the week is computed with the mean of all the observations of each day. The formula for the overall average is the following:

Where:

is the average of all observations

is the observation

is the identified number of the observation in the data

is the total number of observations in the data

The formula used for the average per period is the following:

Where:

is the average of the observations of the period

is the observation of the period

is the identified number of the observation in the data

is the number of the period in the data (example: 1 = Sunday, 2 = Monday, 3 = Tuesday)

is the total number of observations in the period (example: 20 entries for Monday)

Below is shown the results and the bar charts.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Average Number of Patients, per day of the week | | | | | | |
| **Sunday** | **Monday** | **Tuesday** | **Wednesday** | **Thursday** | **Friday** | **Saturday** |
| 140 | 147 | 144 | 150 | 157 | 158 | 142 |

Overall Average: 148.38

The day-of-week data is showing important seasonality information here. On average, the ED have most of the patients on Thursday and Friday, and less patients on Saturday and Sunday, per week. This is aligned with the forecast analysis results.

The monthly averages were obtained by computing each month mean. The results are shown in the table and chart below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Average Number of Patients, per day of the week | | | | |
| **January** | **February** | **March** | **April** | **May** |
| 158 | 154 | 126 | 166 | 140 |

Overall Average: 148.38

From January to May, the month that have more patients on average is April, and the one with less patients is March.

**4.2.3 Obtain Seasonality Index**

Based on the averages obtained, the seasonality index is computed per period. The seasonality index is the percentage of variation over the overall average in a period. The formula used was:

Where:

is the Seasonality Index, per period (%)

is the average of observations per period (per day, day of the week or month)

is the average of all observations

For days of the week seasonality, the results were the following:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Seasonality Index (%)** | | | | | | | |
| **Weekday** | **Sunday** | **Monday** | **Tuesday** | **Wednesday** | **Thursday** | **Friday** | **Saturday** |
| **Average per weekday** | 140 | 147 | 144 | 150 | 157 | 158 | 142 |
| **Seasonality Index** | 94% | 99% | 97% | 101% | 106% | 107% | 96% |
| **Variation %** | -6% | -1% | -3% | 1% | 6% | 7% | -4% |
| **Overall Average** | 148.38 |  |  |  |  |  |  |

The results show that on average, more patients come from Wednesday to Friday and the number of patients drops from Saturday to Tuesday. For the purposes of capacity planning, the optimal setup would have 6% or 7% more resources available than the average days from Thursday and Friday, 6% of less resources on Saturday and Sunday and a normal amount of resources from Monday to Wednesday.

For the monthly seasonality, the results are:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Seasonality Index (%), per month** | | | | | |
| **Month** | **January** | **February** | **March** | **April** | **May** |
| **Average per Month** | 158 | 154 | 126 | 166 | 140 |
| **Seasonality Index** | 107% | 103% | 85% | 112% | 94% |
| **Variation** | 7% | 3% | -15% | 12% | -6% |
| **Overall Average** | 148.38 |  |  |  |  |

The results shows that there are 15% less patients than the average going to ED in March, and %12 more in April. As observations are limited and seasonality is not established based on a full year data, the capacity cannot be planned per month with accuracy. In general terms, the capacity can stay stable during the first five months of the years, with exceptions on a decrease in March and an increase in April.