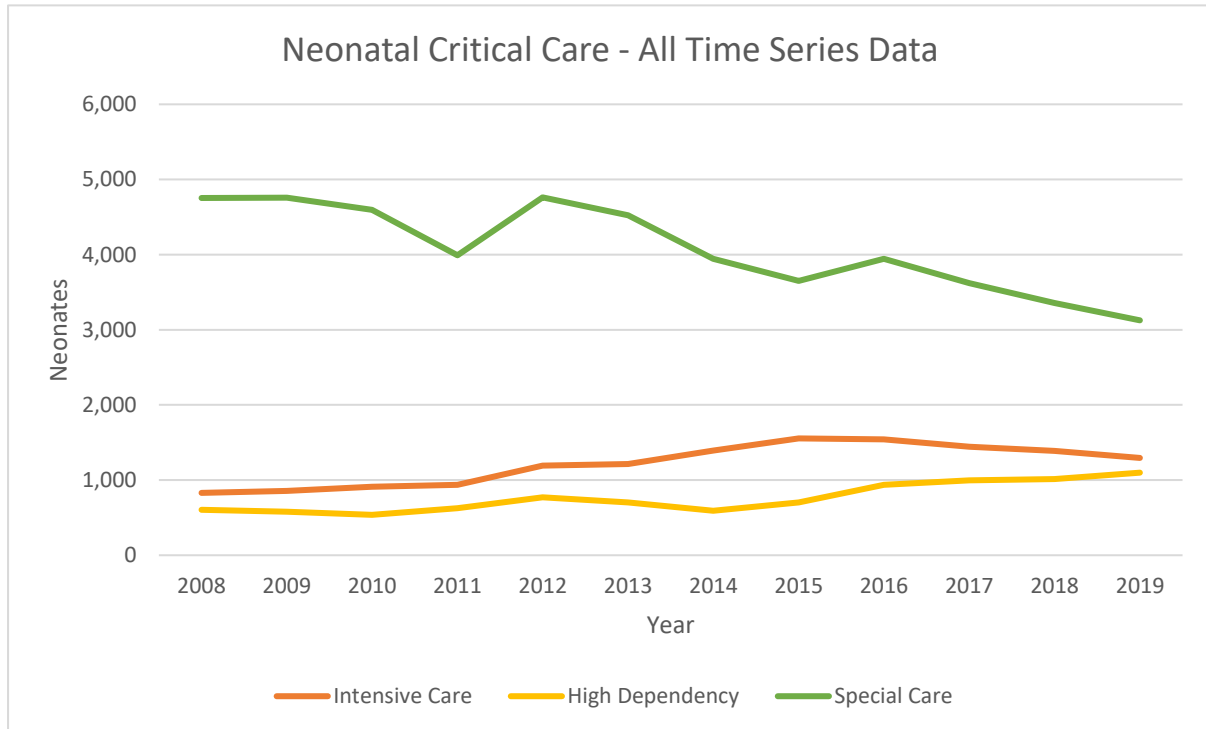


5BUIS002W - Business Analytics - Coursework

TOMASZ WASOWSKI – W1684891

PART A: FORECASTING

QUESTION 1



The above chart displays the three levels of neonatal critical care. From it, we can see that the number of neonates in Special Care seems to be much higher than that in Intensive or High Dependency care. If we take a closer look at the individual time series data, we can see some clear patterns emerging across the years:

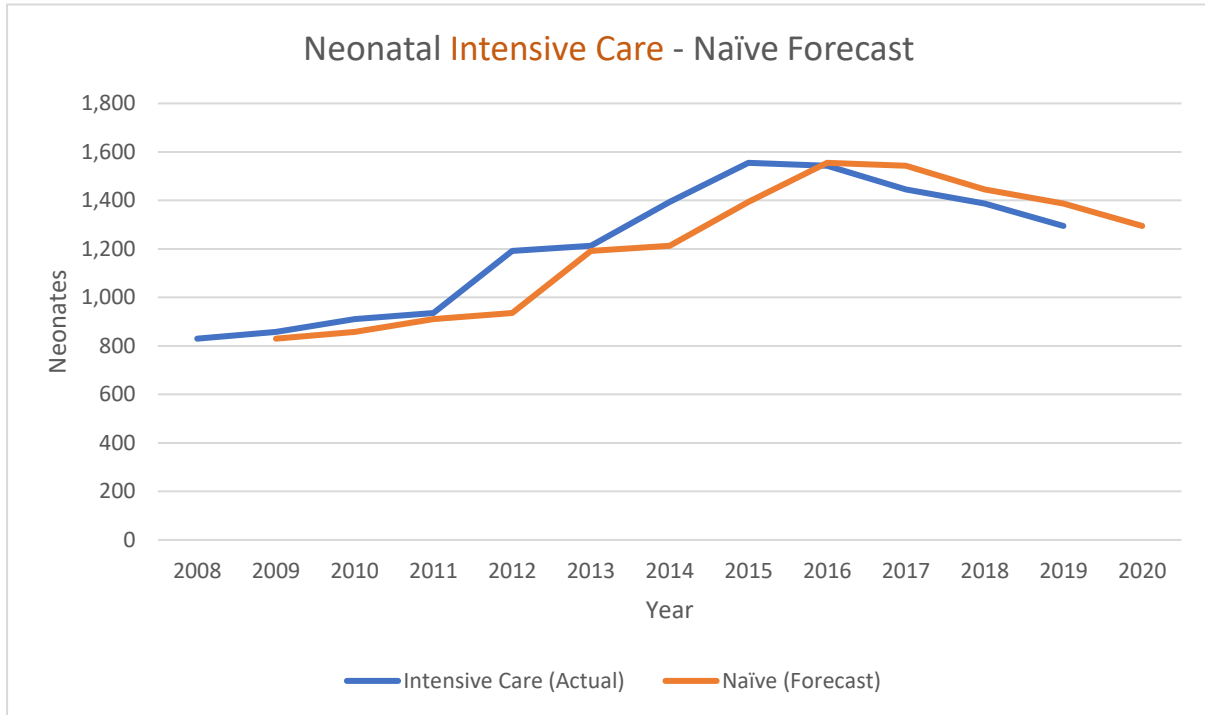
- **Intensive Care** – It appears that this data is following an up-to-down trend; raising steadily up from 2008 to 2015, at which point it seems to gently diminish.
- **High Dependency Care** – It seems that, across the timeline, this data is displaying a slight upward trend with a potential cyclical pattern.
- **Special Care** – This data appears to be following a downward trend, displaying what could be a seasonal pattern that seems to repeat every leap year.

All of these conclusions have been deducted not only from the above graph, but also after displaying the three critical levels, as a whole and individually, both in the forms of bar and line charts. I found these different chart types, as well as individual graphing, to be very helpful in determining the possible patterns within this data time series data.

To save space on this report, I will not include all of the produced graphs, but please refer to the 'Task 1 – Graph Data' sheet in the 'Coursework – Part 1 – Forecasting' workbook in the attached zip file to have more insight into my rationale.

QUESTION 2

A. NAÏVE METHOD

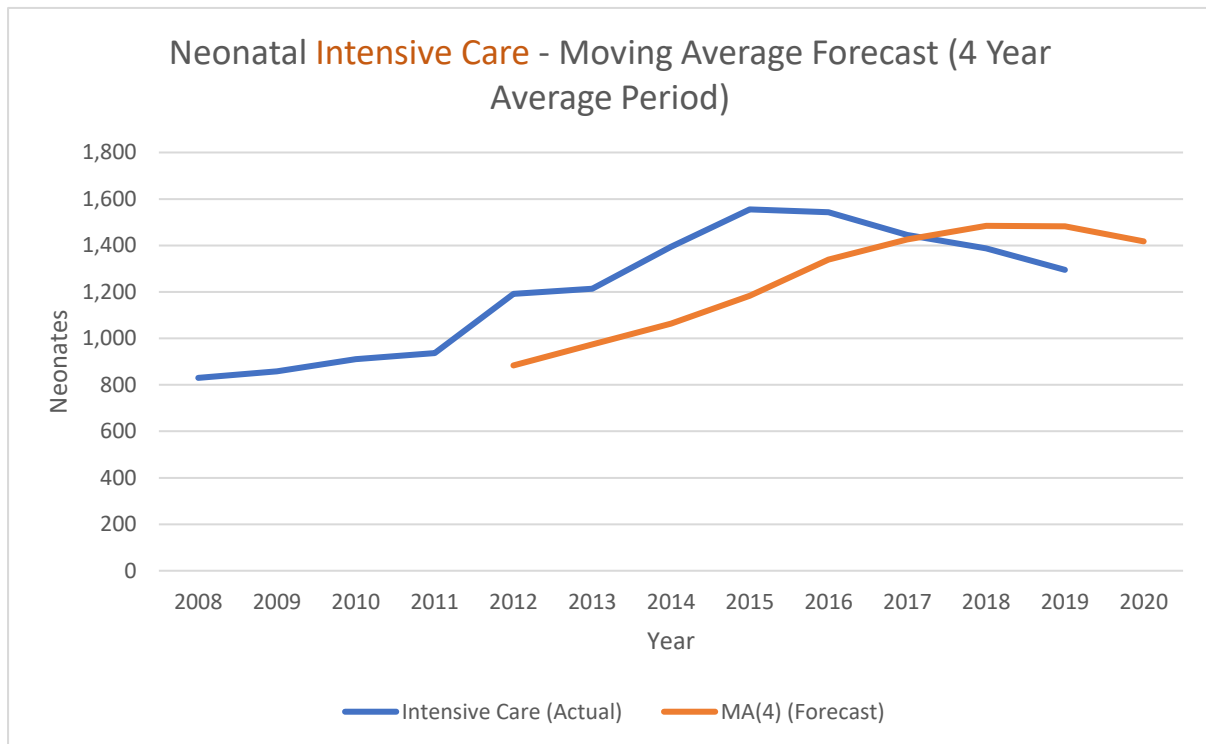


Year	Intensive Care (Actual)	Naïve (Forecast)	Forecast Error	Squared Forecast Error	Forecast Error	Forecast Error / Actual
2008	830					
2009	858	830	-28	784	28	3.26%
2010	910	858	-52	2,704	52	5.71%
2011	936	910	-26	676	26	2.78%
2012	1,192	936	-256	65,536	256	21.48%
2013	1,213	1,192	-21	441	21	1.73%
2014	1,394	1,213	-181	32,761	181	12.98%
2015	1,555	1,394	-161	25,921	161	10.35%
2016	1,543	1,555	12	144	12	0.78%
2017	1,445	1,543	98	9,604	98	6.78%
2018	1,387	1,445	58	3,364	58	4.18%
2019	1,295	1,387	92	8,464	92	7.10%
2020		1,295		13,673	90	7.01%
				MSE	MAD	MAPE

In the table above you can see the forecast number of neonates in intensive critical care for 2020 highlighted in yellow. This forecast was determined using the Naïve method, which is comprised of simply using the previous actual value for the future forecast.

Algebraic formula used for forecast: $F_t = A_{t-1}$

B. MOVING AVERAGE (4) METHOD

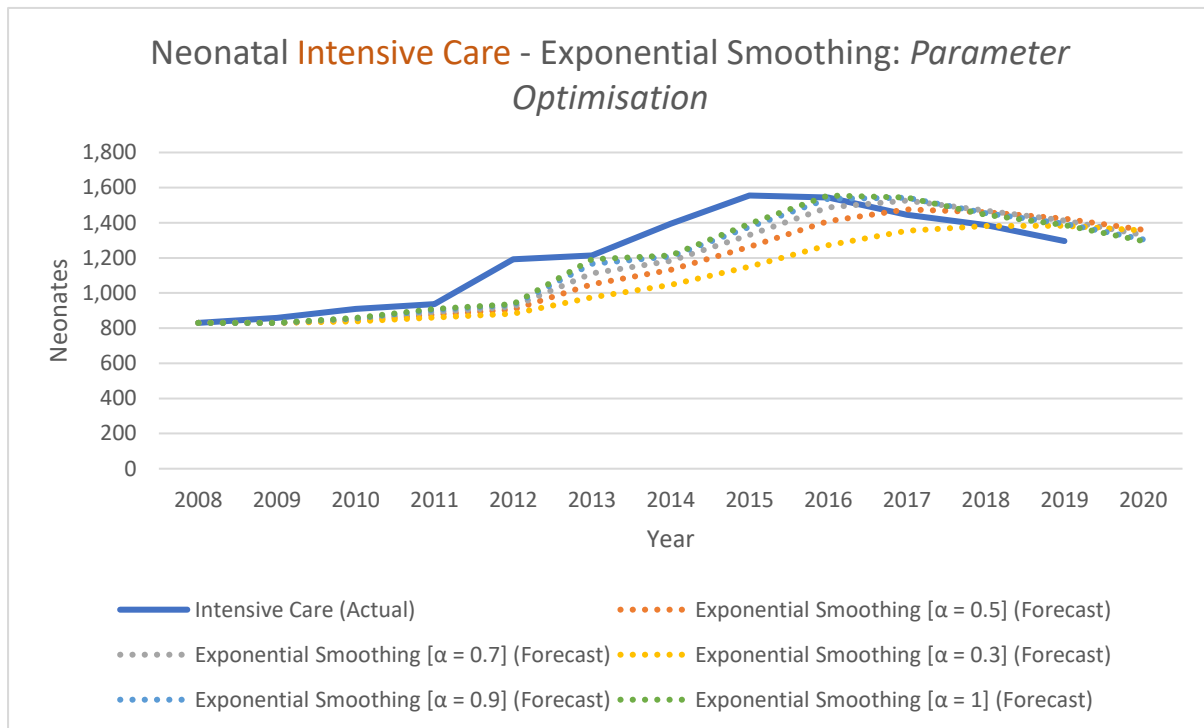


Year	Intensive Care (Actual)	MA(4) (Forecast)	Forecast Error	Squared Forecast Error	Forecast Error	Forecast Error / Actual
2008	830					
2009	858					
2010	910					
2011	936					
2012	1,192	884	-309	95,172	309	25.88%
2013	1,213	974	-239	57,121	239	19.70%
2014	1,394	1,063	-331	109,727	331	23.76%
2015	1,555	1,184	-371	137,827	371	23.87%
2016	1,543	1,339	-205	41,820	205	13.25%
2017	1,445	1,426	-19	352	19	1.30%
2018	1,387	1,484	97	9,458	97	7.01%
2019	1,295	1,483	188	35,156	188	14.48%
2020		1,418		60,829	220	16.16%
				MSE	MAD	MAPE

As before, the table above shows the forecast of the number of neonates in intensive critical care, this time concluded with the use of the 4-period moving average method. This method takes the average of the 4 most recent actual values to generate a forecast.

Algebraic formula used for forecast: $F_t = (A_{t-1} + A_{t-2} + A_{t-3} + A_{t-4})/4$

C. EXPONENTIAL SMOOTHING & PARAMETER OPTIMISATION



In the above chart, we can see the result of using different values of the smoothing constant to develop exponential smoothing models to forecast the number of neonates in intensive critical care in 2020. I have tried a variety of different constant values in order to minimise the MAPE. It seems that, due to the inconsistent trend in the data, higher values of the smoothing constant showed better results, as they adjusted to changes faster. I will list the values that I have compared and their following forecasts and MAPEs below:

- $\alpha = 0.3$ | Forecast = 1356 | MAPE = 13.34%
- $\alpha = 0.5$ | Forecast = 1359 | MAPE = 10.73%
- $\alpha = 0.7$ | Forecast = 1330 | MAPE = 9.03%
- $\alpha = 0.9$ | Forecast = 1305 | MAPE = 7.52%
- $\alpha = 1$ | Forecast = 1295 | MAPE = 7.01%

Full tables and charts for these figures are available on the 'Task 2 – IC Forecast' sheet of the 'Coursework – Part 1 – Forecasting' workbook in the included zip file.

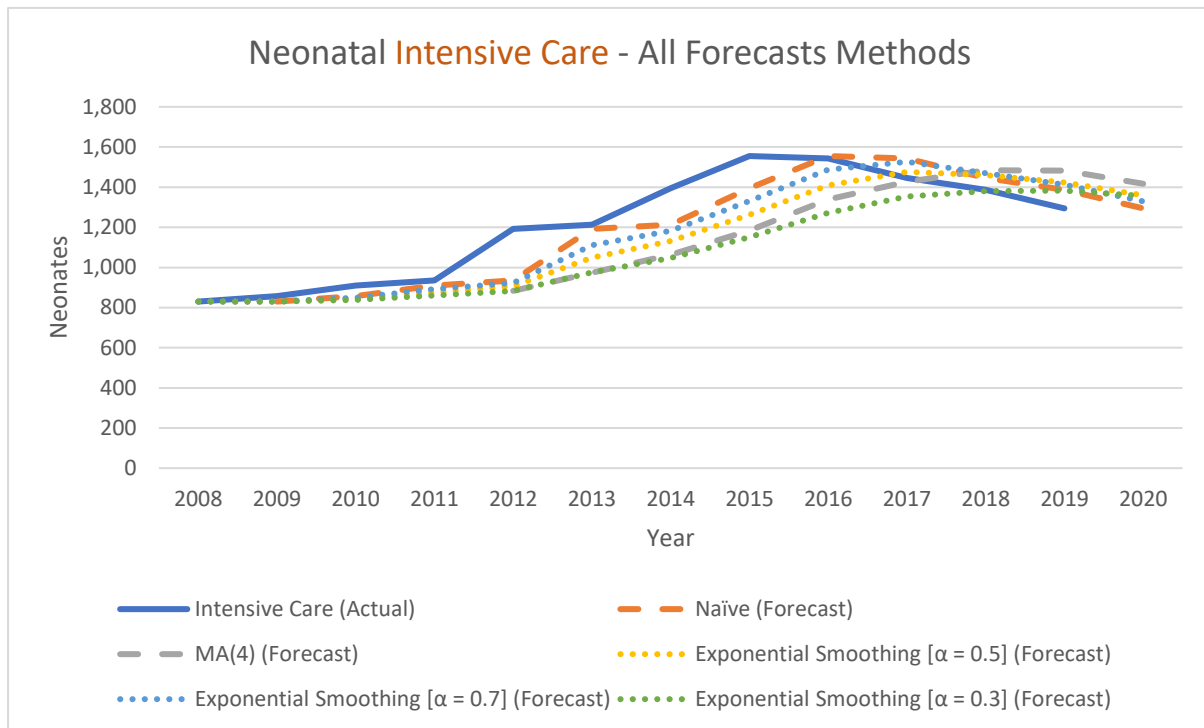
As can be seen from the above smoothing constants and their relative MAPEs, the higher the smoothing constant value the lower the MAPE, and hence, the more accurate the forecast. It seems that the best smoothing constant for this data set is an extreme 1, making the exponential smoothing model very reactive to changes across the timeline.

Year	Intensive Care (Actual)	Exponential Smoothing [$\alpha = 1$] (Forecast)	Forecast Error	Squared Forecast Error	Forecast Error	Forecast Error / Actual
2020		1,295		19,050	113	7.01%
				MSE	MAD	MAPE

The MAPE for all of these exponential smoothing models was calculated using the average of all the percentage error values excluding the initial forecast, as it is assumed to be equal to the actual value. As a result, it produces a percentage error of 0.00% which, if counted, would lower the overall average.

Algebraic formula used for forecast: $F_t = \alpha A_{t-1} + (1 - \alpha)F_{t-1}$

D. COMPARISON & CONCLUSION



The chart above displays the each of the forecast methods together, including an exponential smoothing with 3 different values of the smoothing constant. I decided not to include exponential smoothing with a smoothing constant of 1, as it is equivalent to the Naïve method, and the two would simply overlap on the graph.

The **Naïve method** does a relatively good job of forecasting the number of neonates expected in intensive critical care in 2020, as denoted by its MAPE of only 7.01%. This method manages to follow the trend of the data, despite not taking it into account, as the forecast is only ever one step behind the actual value. That said, it is still not perfect, as the maximum percentage error experienced was 21.48% in 2012.

The **4-period moving average** method provided a relatively poor forecast when compared to the simpler Naïve method. Given the MAPE of 16.16% and a maximum percentage error of 25.88% in 2012, we can see that the forecasts provided by this method have struggled to keep up with the trend of this data set. Perhaps a lower value of moving average would make this method more viable, as it would be more responsive to change. That said, even at a 1-year moving average, this method will only ever be as accurate as the Naïve method.

The **exponential smoothing** models gave a variety of results, seemingly increasing in accuracy with an increase in smoothing constant. That said, even with the extreme, maximal smoothing constant of 1, the exponential smoothing method was only as good as the much simpler Naïve method, with its resulting MAPE also at 7.01% and a maximum percentage error of 21.48% in 2012. Lower values of the constant were much less reactive to the changes in the data and were not able to adjust fast enough to follow the shifting trends.

To conclude; given its simplicity and highest accuracy, I would say that the best of these three methods for the forecast of this data is the Naïve method. Without any smoothing, it mirrors the actual values with a single time period offset and does a relatively good job at following the pattern of this data set. That said, I believe that all of these three techniques are best used when the time series is relatively stable and are not as suitable for this case, where the data set clearly exhibits a shifting trend pattern.

QUESTION 3

At first, I wanted to exclude any variations of forecasting techniques that we have exercised in the previous question. As such, I have performed a forecast using the following methods, and got the following results:

- Simple Mean | **Forecast** = 1213 | **MAPE** = 17.88%
- Moving Average (2) | **Forecast** = 1341 | **MAPE** = 10.63%
- Moving Average (3) | **Forecast** = 1376 | **MAPE** = 13.62%
- Moving Average (5) | **Forecast** = 1445 | **MAPE** = 16.49%
- Moving Average (6) | **Forecast** = 1437 | **MAPE** = 16.88%
- Moving Average (7) | **Forecast** = 1405 | **MAPE** = 16.67%

As can be seen, the results are not great when compared to the Naïve method, which forecast 1295 with a MAPE of 7.01%.

Full tables and charts for these figures are available on the 'Task 3 – More Methods' sheet of the 'Coursework – Part 1 – Forecasting' workbook in the included zip file. Those charts also include the MSE and MAD accuracy measures for each forecast.

After doing some research on the topic, I have found several sources which discussed forecasting time series data that might contain trend, seasonal or cyclical patterns. I used one article to verify that I am indeed dealing with an “Up-to-Down Trend” (Silva, 2019). I have also found an analytics blog (Singh, 2018) which discussed in greater details the various methods for forecasting data that displays a trend.

Using this information, I decided to try the methods that were not only recommended, but ones that I was also already familiar with to see if I could find a technique which minimised the MSE, MAD and MAPE. These methods involved the Weighted Moving Average, Naïve with Trend and Least Squares. The forecasts and MAPEs of these methods are listed below:

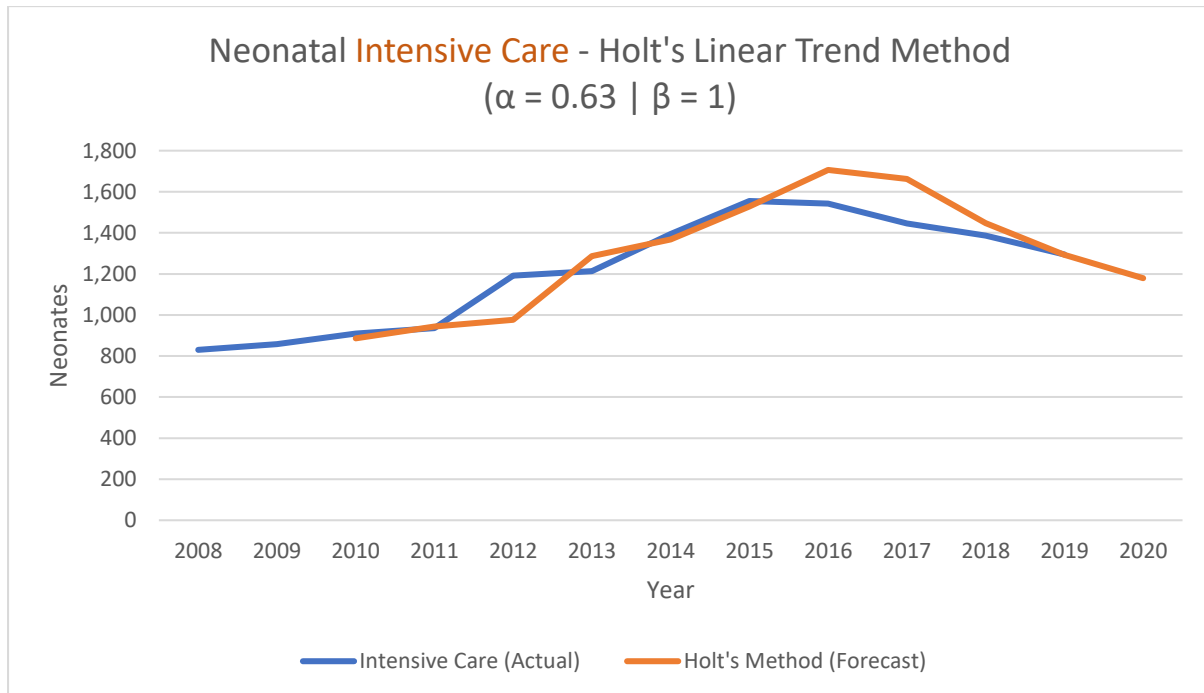
- Weighted Moving Average (2) (0.65, 0.35) | **Forecast** = 1327 | **MAPE** = 9.61%
- Weighted Moving Average (3) (0.55, 0.33, 0.12) | **Forecast** = 1343 | **MAPE** = 11.15%
- Weighted Moving Average (4) (0.50, 0.25, 0.15, 0.10) | **Forecast** = 1365 | **MAPE** = 12.82%
- Naïve with Trend | **Forecast** = 1203 | **MAPE** = 7.95%
- Least Squares | **Forecast** = 1620 | **MAPE** = 9.65%

Though much more promising in their results, especially the Naïve with Trend and the Least Squares methods, these techniques still did not manage to beat the basic Naïve method when it comes to their MAPE. It seems that due to the trend variable of the Naïve with Trend method, the forecast was overcompensating with every variation in the data. As for the Least Squares method, the fact that the line of best fit was linear mean that the method did not account for the trend shifting around 2015 from an upward trend to a downward one.

Before trying different techniques, I decided to perform the Naïve with Trend and Least Squares methods on only a fraction of the data, from 2015 onwards, as this would focus on forecasting solely with the second, downward trend in consideration. The results and accuracy measures of this approach are listed below:

- Naïve with Trend (2015 Onwards) | **Forecast** = 1203 | **MAPE** = 3.82%
- Least Squares (2015 Onwards) | **Forecast** = 1242 | **MAPE** = 1.09%

This resulted in forecasts with fantastic MAPE results, and if only considering the trend from the previous 5 years was an option, I would say that the Least Squares method is probably the most accurate. That said, if we want to take the entire data set into consideration, the analytics blog (Singh, 2018) suggests that there are two more techniques that should be considered: Holt's Linear Trend Method and ARIMA (Autoregressive Integrated Moving Average) Method. After researching both Holt's Method (OTexts™, no date) and ARIMA (Brownlee, 2017), I decided to try to use Holt's method to achieve a more accurate forecast.



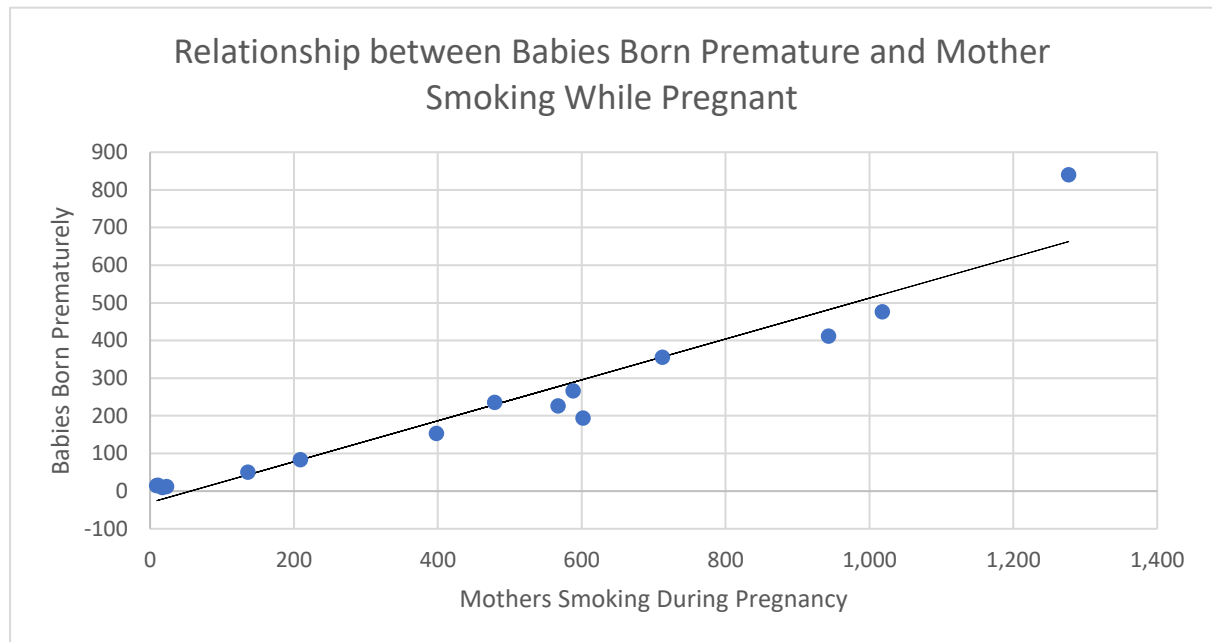
Year	Intensive Care (Actual)	Level	Trend	Holt's Method (Forecast)	Forecast Error	Squared Forecast Error	Forecast Error	Forecast Error / Actual
2008	830							
2009	858	858	28					
2010	910	901.1183	43.11831	886	-24	576	24	2.64%
2011	936	939.0481	37.92982	944	8	68	8	0.88%
2012	1,192	1112.427	173.3786	977	-215	46,234	215	18.04%
2013	1,213	1239.943	127.5164	1,286	73	5,301	73	6.00%
2014	1,394	1384.178	144.235	1,367	-27	704	27	1.90%
2015	1,555	1545.161	160.9829	1,528	-27	707	27	1.71%
2016	1,543	1603.375	58.21371	1,706	163	26,616	163	10.57%
2017	1,445	1525.153	-78.2218	1,662	217	46,911	217	14.99%
2018	1,387	1409.179	-115.974	1,447	60	3,592	60	4.32%
2019	1,295	1294.336	-114.843	1,293	-2	3	2	0.14%
2020				1,179		13,071	81	6.12%
Alpha:	0.63	Beta:	1			MSE	MAD	MAPE

With the aid of a great source that explained the implementation of this technique (Major, 2018), I was able to achieve a forecast, which is more accurate than the previously unbeaten Naïve method. This method appears to work better than the alternatives as it applies smoothing with the use of two smoothing constants as opposed to just one. To optimise both of these constants, I have used an Excel data analysis plug in called Solver. Using this solver, I was able to set constraints and automatically optimise the constants for minimising the MAPE. As such, I am confident to recommend this technique and a next year forecast of 1179.

Algebraic formula used for forecast: $F_t = L_{t-1} + T_{t-1}$

Where: $L_{t-1} = \alpha A_{t-1} + (1 - \alpha)(L_{t-2} + T_{t-2})$ and $T_{t-1} = \beta(L_{t-1} - L_{t-2}) + (1 - \beta)T_{t-2}$

QUESTION 4



<i>NHS Board of Residence (i)</i>	<i>Mothers Smoking (x_i)</i>	<i>Premature Babies (y_i)</i>	<i>x_i y_i</i>	<i>x_i²</i>	<i>Regression Line</i>
1	567	226	128142	321,489	277.4514
2	136	50	6800	18,496	43.53679
3	209	83	17347	43,681	83.15575
4	602	194	116788	362,404	296.4468
5	479	235	112565	229,441	229.6916
6	712	355	252760	506,944	356.1466
7	1,277	840	1072680	1,630,729	662.7865
8	398	153	60894	158,404	185.7309
9	943	411	387573	889,249	481.5162
10	1,018	476	484568	1,036,324	522.2206
11	11	15	165	121	-24.3039
12	17	9	153	289	-21.0475
13	588	266	156408	345,744	288.8487
14	23	12	276	529	-17.7912
15	9	14	126	81	-25.3893
Totals:	6,989	3,339	2,797,245	5,543,925	
Averages:	466	223	186,483	369,595	
n =	15	b1 =	0.543	b0 =	-30.274

In the chart above we can see the line of best fit drawn using the Least Squares method. This line shows a clear correlation between mothers smoking while pregnant and babies being born prematurely.

Algebraic formula used for regression line: $\hat{y} = b_0 + b_1x$

Where: $b_0 = \bar{y} - b_1\bar{x}$ and $b_1 = ((\sum x_i y_i) - ((\sum x_i \sum y_i) / n)) / ((\sum x_i^2) - ((\sum x_i)^2 / n))$

In order to measure the goodness-of-fit of the least squares line to the data, I decided to calculate the coefficient of determination. It the percentage of variation of the dependant variable, in our case that is babies born prematurely, explained by regression.

<i>NHS Board of Residence (i)</i>	<i>Mothers Smoking (x_i)</i>	<i>Gestational age < 37 weeks (y_i)</i>	<i>Estimate Gestational Age < 37 weeks (ŷ)</i>	<i>(y_i - ŷ)²</i>	<i>(y_i - ȳ)²</i>
1	567	226	277.4514496	2,647	11.56
2	136	50	43.53679142	42	29,790.76
3	209	83	83.15574745	0	19,488.16
4	602	194	296.4468395	10,495	817.96
5	479	235	229.6916122	28	153.76
6	712	355	356.1466363	1	17,529.76
7	1,277	840	662.7865014	31,405	381,182.76
8	398	153	185.7308528	1,071	4,844.16
9	943	411	481.5162095	4,973	35,494.56
10	1,018	476	522.2206163	2,136	64,211.56
11	11	15	-24.30388671	1,545	43,097.76
12	17	9	-21.04753416	903	45,624.96
13	588	266	288.8486836	522	1,883.56
14	23	12	-17.79118161	888	44,352.36
15	9	14	-25.38933756	1,552	43,513.96
Average:		223	Totals:	58,207	731,997.60

Coefficient of determination (R²)	0.9205	#Good fit to the data!
---	---------------	------------------------

As we can see from the above table, the coefficient of determination, which would normally range from 0 to 1, is at 0.9205. The higher the coefficient, the better the fit of the least squares line. As such, with our coefficient, our line seems to be a very good fit to the data, as over 90% of it fits the regression model.

Algebraic formula used to calculate the coefficient of determination: $R^2 = 1 - (\sum(y_i - \hat{y}_i)^2 / (\sum(y_i - \bar{y})^2)$

PART B: SIMULATION

QUESTION 1

Below are the appropriate intervals for random numbers to be used to generate time between customer calls from the provided discrete probability distribution:

<i>Time between calls (min)</i>	<i>Probability</i>	<i>Cumulative</i>	<i>Min</i>	<i>Max</i>
5	0.10	0.10	0.00	0.10
8	0.20	0.30	0.10	0.30
10	0.40	0.70	0.30	0.70
12	0.30	1.00	0.70	1.00

The intervals are between 'Min' and 'Max' and should reflect the probability of each time between calls.

QUESTION 2

Below is the simulated time between calls for the first 10 customers, using the provided random numbers:

<i>Call No.</i>	<i>Random Number</i>	<i>Time Between Calls (min)</i>
1	0.53933	10
2	0.71344	12
3	0.83459	12
4	0.37075	10
5	0.00997	5
6	0.31444	10
7	0.48636	10
8	0.12296	8
9	0.09736	5
10	0.28647	8

<i>Min</i>	<i>Time between calls (min)</i>
0.00	5
0.10	8
0.30	10
0.70	12

For this simulation, I have used the VLOOKUP() function in Excel to match the random numbers to their appropriate intervals and to retrieve the corresponding time between calls. The average time between calls from the above 10 iterations is 9.

QUESTION 3

Below is the simulation of scenario 1; a spreadsheet model simulating the first 10 customers in the system:

<i>Custo mer</i>	<i>Interarrival Time</i>	<i>Arrival Time</i>	<i>Service Start Time</i>	<i>Waiting Time</i>	<i>Service Time</i>	<i>Completion Time</i>	<i>Time In System</i>
1	10	10	10	0	10	20	10
2	12	22	22	0	10	32	10
3	12	34	34	0	10	44	10
4	10	44	44	0	10	54	10
5	5	49	54	5	10	64	15
6	10	59	64	5	10	74	15
7	10	69	74	5	10	84	15
8	8	77	84	7	10	94	17
9	5	82	94	12	10	104	22
10	8	90	104	14	10	114	24

From the above simulation, we can deduct that:

- Average 'Time In System' spent per caller: **14.8 minutes**
- Average 'Waiting Time' spent per caller: **4.8 minutes**

We can see that, after a short warmup period, the queue begins to fill up and that customer waiting times (as well as total time in the system) get progressively longer. This is most likely due to the fact that the average service time is 10 minutes, however the average call interarrival time we have calculated is 9 minutes. As such, given enough time and iterations, the queue will continue to grow larger causing longer waiting times.

QUESTION 4

A. SCENARIO 2 (ONE REP)

Below is the simulation for scenario 2; a spreadsheet model simulating the first 100 customers with an exponential probability distribution for interarrival time (mean of 9 minutes) and a discrete probability distribution for service time.

<i>Service Time (min)</i>	<i>Probability</i>	<i>Cumulative</i>	<i>Min</i>	<i>Max</i>	<i>Min</i>	<i>Service Time (min)</i>
5	0.24	0.24	0.00	0.24	0.00	5
6	0.20	0.44	0.24	0.44	0.24	6
7	0.15	0.59	0.44	0.59	0.44	7
8	0.14	0.73	0.59	0.73	0.59	8
9	0.12	0.85	0.73	0.85	0.73	9
10	0.08	0.93	0.85	0.93	0.85	10
11	0.05	0.98	0.93	0.98	0.93	11
12	0.02	1.00	0.98	1.00	0.98	12

Time between Calls (Exponential Distribution) - Mean:	9
---	---

Custo mer	Interarrival Time	Arrival Time	Service Start Time	Waiting Time	Service Time	Completion Time	Time In System
1	1.31	1.31	1.31	0.00	5	6.31	5.00
2	20.40	21.71	21.71	0.00	10	31.71	10.00
3	34.70	56.41	56.41	0.00	5	61.41	5.00
4	11.64	68.05	68.05	0.00	7	75.05	7.00
5	12.72	80.77	80.77	0.00	5	85.77	5.00
96	14.48	753.40	795.94	42.54	6	801.94	48.54
97	5.41	758.81	801.94	43.13	8	809.94	51.13
98	0.87	759.68	809.94	50.26	5	814.94	55.26
99	3.95	763.64	814.94	51.31	10	824.94	61.31
100	1.07	764.70	824.94	60.24	8	832.94	68.24

The table above displays the first 5 and the last 5 customers in the simulation, though the entire 100 are available in the supporting 'Task 4 – Report Scenario 2' sheet of the 'Coursework – Part 2 – Simulation' workbook. Please note that formulas are found on the 'Task 4 – Scenario 2 (One Rep)' sheet.

As you can see from the above data, waiting time starts off at a comfortable 0, but by the end of the simulation it has reached up to an hour in the last case.

The method of simulating this is by using a variety of excel formulas to calculate the different times. This is actually the reason that the first and second column are highlighted in blue. Since there are no calls in the queue at $t=0$, the first row does not use formulas to calculate its values. The second row, on the other hand, introduces all the formulas that are then repeated in every row of the simulation.

The most notable formula is the one that calculates the service start time. It uses a conditional IF() function to calculate whether the customer has to wait in the queue, or whether they can be serviced directly at their arrival time.

A. SCENARIO 3 (TWO REPS)

Below is the simulation for scenario 3; a spreadsheet model simulating the first 100 customers using the same probability distributions as scenario 2 but simulating the effects of having two customer representatives.

Custo mer	Interarrival Time	Arrival Time	Service Start Time	Waiting Time	Service Time	Completion Time	Time In System
1	1.31	1.31	1.31	0.00	5	6.31	5.00
2	1.11	2.42	2.42	0.00	5	7.42	5.00
3	0.24	2.66	6.31	3.65	10	16.31	13.65
4	3.00	5.66	7.42	1.75	6	13.42	7.75
5	7.67	13.34	13.42	0.08	6	19.42	6.08
96	4.80	726.40	726.40	0.00	6	732.40	6.00
97	30.15	756.56	756.56	0.00	5	761.56	5.00
98	3.09	759.64	759.64	0.00	5	764.64	5.00
99	4.26	763.90	763.90	0.00	8	771.90	8.00
100	5.51	769.41	769.41	0.00	5	774.41	5.00

Once again, the table above displays the first 5 and the last 5 customers in the simulation, though the entire 100 are available in the supporting 'Task 4 – Report Scenario 3' sheet of the 'Coursework – Part 2 – Simulation' workbook. Please note that formulas are found on the 'Task 4 – Scenario 3 (Two Reps)' sheet.

As you can see from the above data, the scenario with two customer service representatives performs much better, as the waiting times stay significantly lower throughout the entire simulation.

The method of simulating this scenario is almost identical to the previous one, the difference being that this time there are two customer representatives. As such, the first two rows do not use any formulas for the arrival and service start time, as they are picked up directly by the first and second representative, respectively.

The key difference in this simulation is within the service start time formula, as now it has to calculate which of the two representatives is free to pick up the next call. This is calculated with the help of the Excel MIN() function within the original conditional IF() function. The MIN() determines which rep has had the earliest completion time and uses the lowest value to compare to the customers arrival time and is potentially set as the customers service start time if the customer is required to wait.

QUESTION 5

A. SUMMARY STATISTICS (SCENARIO 2 & 3)

Below are the summary statistics calculated from the simulation of scenario 2 (with one customer service representative):

Scenario 2 (One Rep) Summary Statistics	
Total Average Time in System	38.87
Average Waiting Time	31.58
Maximum Time in System	77.11
Maximum Waiting Time	69.11
Number Waiting > 3 Mins	81
Probability of Waiting > 3 Mins	81.00%

As can be seen from these statistics, the average waiting time has reached 31 minutes, which is significantly higher than the company's guideline requirement of 3 minutes. We can see that 81 out of the 100 customers had to wait over 3 minutes, with the maximum waiting time reaching a staggering 69 minutes. Clearly, a single representative is not enough to handle the demand of this service.

Having run this simulation multiple times, it appears that these statistics are a little on the harsher side, though certainly not at the extreme either. During the vast majority of simulations, it seems that the number of customers waiting over the guideline 3 minutes is at least 50%.

Next are the summary statistics calculated from the simulation of scenario 3 (with two customer service representatives):

Scenario 3 (Two Reps) Summary Statistics	
Total Average Time in System	7.54
Average Waiting Time	0.50
Maximum Time in System	15.06
Maximum Waiting Time	8.02
Number Waiting > 3 Mins	7
Probability of Waiting > 3 Mins	7.00%

In improvement from the previous statistics, the average waiting time in this case is 0.5 minutes, which is significantly better, and far below the company's guideline requirement of 3 minutes. We can see that 7 out of the 100 customers had to wait over 3 minutes, with the maximum waiting time of 8 minutes. We can see that two customer representatives offer a significant increase in performance within the system.

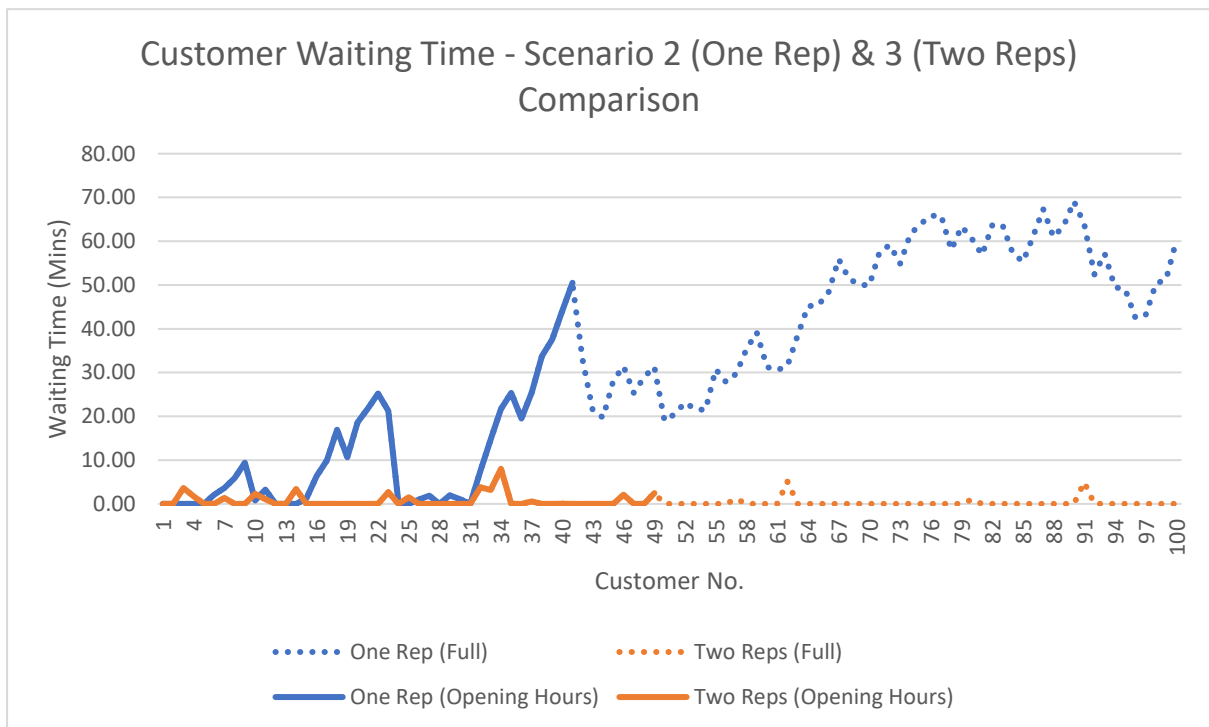
To conclude the above, I have decided to calculate all statistics within the opening hours of the customer service. The service is open from 8:00 to 14:00, which is 6 hours or 360 minutes. As such, I have created a cut of point at customer arrival time < 360. The statistics now only include the first n customers serviced during the services operating hours:

Scenario 2 (One Rep) Summary Statistics (Within Opening Hours)	
Total Average Time in System	18.17
Average Waiting Time	10.81
Maximum Time in System	56.50
Maximum Waiting Time	50.50
Number Waiting > 3 Mins	22
Probability of Waiting > 3 Mins	53.66%
Customers Served	41

Scenario 3 (Two Reps) Summary Statistics (Within Opening Hours)	
Total Average Time in System	8.16
Average Waiting Time	0.77
Maximum Time in System	15.02
Maximum Waiting Time	8.02
Number Waiting > 3 Mins	5
Probability of Waiting > 3 Mins	10.20%
Customers Served	49

As can be seen from the above, the statistics for scenario 2 have improved significantly, as the queue does not have as much time to build up.

B. COMPARISON OF SCENARIO 2 AND SCRNARIO 3



There are some start differences between the two scenarios. For the sake of this comparison, I am going to only consider the **statistics from the operating hours** of both scenarios.

As can be seen in the graph and statistics above, the first difference is in the number of customers that have been served during work hours. Scenario 2, with one customer representative only managed to serve 41 people, whereas in Scenario 3, the two representatives managed to serve 49.

The main reason for this is waiting times. As can be seen in the graph, it is very easy for the customer queue to build up when only one representative is available to serve the customers. As a matter of fact, the average

waiting time (within operating hours) with one rep is almost 11 minutes: significantly above the sub-one-minute waiting times with two customer representatives.

Another significant difference is around the extremes. In Scenario 2, both the maximum time in system, as well as the maximum waiting time are over 50 minutes. Having two reps reduces these maximums to 15 minutes in the system, and 8 minutes waiting.

Not only that, but the probability of the customers waiting over 3 minutes is significantly lower in Scenario 3, with customers facing only a 10.20% percent likelihood compared to the 53.66% probability in Scenario 2.

As such, it is very clear that Scenario 3, with an additional customer representatives, shows a significant increase in performance and is the better of the two.

C. RECOMMENDATION REGARDING BEST STAFFING PLAN

After running all of the aforementioned simulations, I think it would be most reasonable to recommend that the company invests in more than one representatives.

That said, after looking at the scenario statistics and considering the company guideline requirements of a sub-three-minute average wait time, I think it would also be reasonable to suggest that two customer representatives working full time during the operating hours might be somewhat overkill.

To help reduce redundancy from customer service representatives waiting around for calls, I would recommend having **one rep working full time**, and considering **hiring another rep part time**; particularly for the latter half of the day, during which the number of customers in the queue tends to increase. An example of this might be: **One rep for the first 3 hours of operating, and two reps for the last 3 hours of operating.**

Alternatively, we know from our simulations that with a single rep, the customer wait time slowly increases over time. With two reps, however, the wait time tends to stay close to 0 during operating hours. As such, it may be beneficial for the company to look for ways that minimise the period of time that the customer service is operated by only one representative, to prevent the queue from building up. An example of this might be: **One rep for the first hour, two reps for the second hour, one rep for the third, two for the fourth, one for the fifth and two for the sixth.**

In both of these examples the number of work hours for the reps are the same, but in the second example, the customer queue only ever has one hour to build up before a second rep comes to spread the workload and bring the average wait time closer to 0 again.

In conclusion, if feasible, I would recommend that the company hires one full time rep and one part time rep, while optimising their work hours to minimise the maximum time that the customer queue is given to build up with a single rep.

REFERENCES

- Silva, T. (2019). Fundamentals of Time Series Data and Forecasting. BetterProgramming. Available from <https://betterprogramming.pub/fundamentals-of-time-series-data-and-forecasting-15e9490b2618> [Accessed 03 March 2021].
- Singh, G. (2018). 7 Methods to Perform Time Series Forecasting (with Python Codes). Analytics Vidhya. Available from <https://www.analyticsvidhya.com/blog/2018/02/time-series-forecasting-methods/> [Accessed 03 March 2021].
- OTexts™ (no date). 7.2 Trend Methods. OTexts. Available from <https://otexts.com/fpp2/holt.html> [Accessed 03 March 2021].
- Brownlee, J. (2017). How to Create an ARIMA Model for Time Series Forecasting in Python. Machine Learning Mastery. Available from <https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/> [Accessed 03 March 2021].
- Major, L. (2018). How to do Holt's Method in Excel. YouTube. Available from https://www.youtube.com/watch?v=x-hxg4pE-Ns&ab_channel=LeslieMajor [Accessed 03 March 2021].