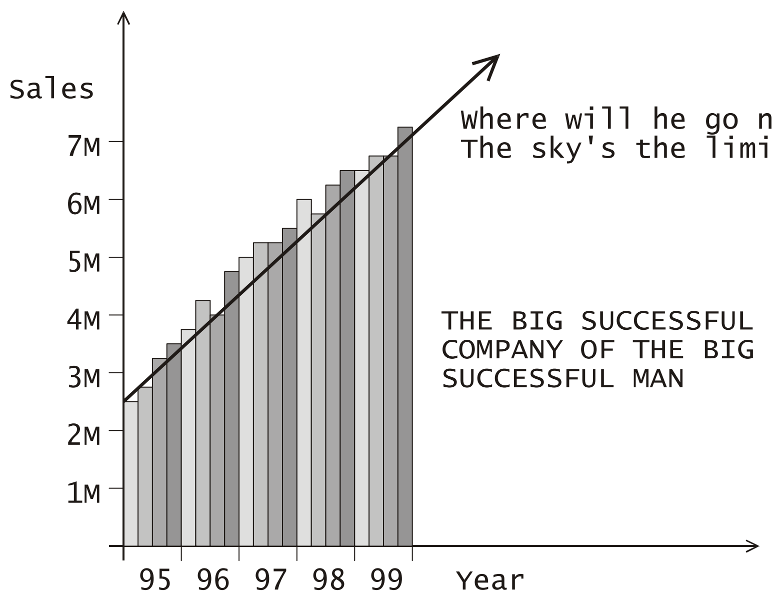
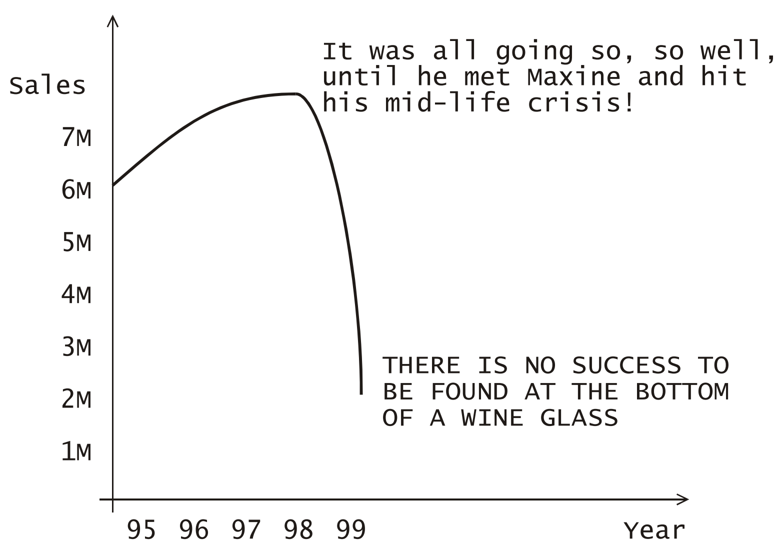
Time Series Analysis

Introduction

Time series analysis is a method of forecasting. It works on the assumption that trends in such things as sales or measurements of rainfall, will continue. This is the picture familiar from certain films and comedy series, where in the background of an office you can see a picture like this



The trend seems upwards. But we are also familiar with the limitations of this idea



Despite the limitations of the idea of predicting the future on the basis of past trends, we still use this approach, because it is not unreasonable to suppose that a trend in the past will continue into the future, and because we have to base our decisions about future expectations on something.

In fact, trends come in more than one shape or size. The above sketches illustrate the idea of an underlying trend – in the case of the first graph, an underlying trend of increasing sales. Not all the points fitted the graph, some there was also some random fluctuation about the trend line. However, in some businesses, we witness other kinds of patterns – such as seasonal fluctuations in sales. There is more demand for toys in the Christmas period than during the winter period immediately after it. There is a greater demand for beachwear during the summer. There can also be other cycles that affect such things as sales and profits – business cycles in terms of demand for products in the economy as a whole.

Additionally, the idea of spotting trends can obviously be extended to non-business contexts. There might be a trend in the erosion of a beach, for instance, or a trend in the growth of a tree.

In order to find the trend we need to be able to iron out the fluctuations due to random forces. For example, the trend in sales at an island holiday resort that is losing touch with the market and is getting to be boring and old-fashioned, may be downwards; however, if there is boycott of petrol stations in other countries, that holiday resort may be fully-booked, but only for a week. This is an example of a random fluctuation, and its effect needs to be ironed out.

The obvious way to do this is by taking moving averages. What this means, is that we average the data (say sales figures) over a given time period. The average is "moving" because as we progress in time, we drop one piece of data out of our average and include another. But all of this is best illustrated by worked example.

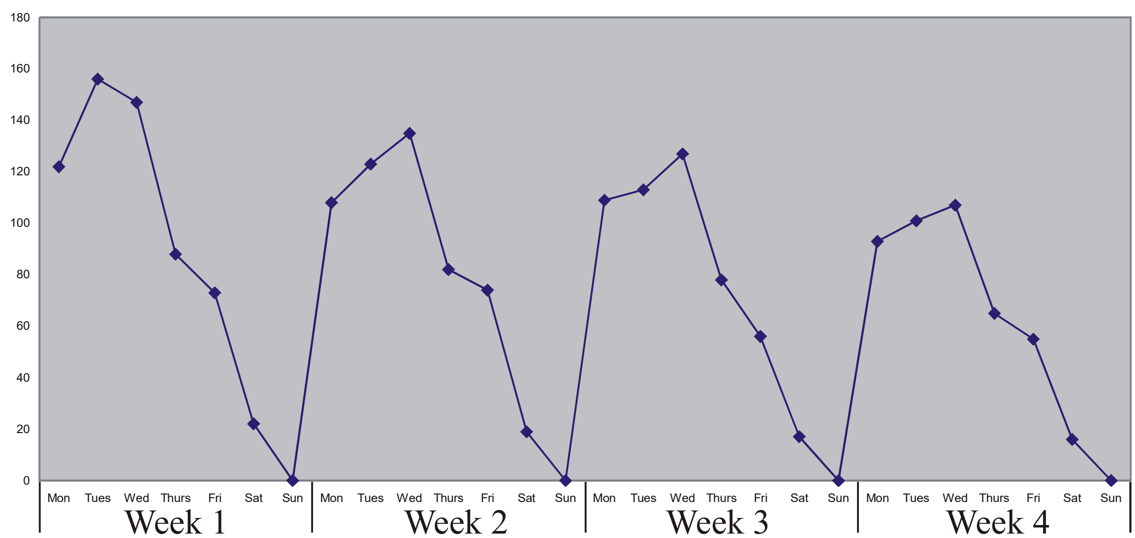
Example (1)

The following data gives the number of students attending Philosophy seminars and lectures during term time at a certain college. Draw a graph of the data. Construct a seven day moving average. Analyse the data to determine whether there is a seven day periodic cycle. Predict the attendance during Week 8 – the last week of the term.

Week	Day	Attendance
1	Mon	122
	Tues	156
	Wed	147
	Thurs	88
	Fri	73
	Sat	22

	Sun	0
2	Mon	108
	Tues	123
	Wed	135
	Thurs	82
	Fri	74
	Sat	19
	Sun	0
3	Mon	109
	Tues	113
	Wed	127
	Thurs	78
	Fri	56
	Sat	17
	Sun	0
4	Mon	93
	Tues	101
	Wed	107
	Thurs	65
	Fri	55
	Sat	16
	Sun	0

Solution



The graph shows us that there is a clear trend – downwards – and also a strong period fluctuation over a week.

We have been asked to construct a seven day moving average. The first thing to do is to sum the first seven terms.

Week	Day	Attendance
1	Mon	122
	Tues	156
	Wed	147
	Thurs	88
	Fri	73
	Sat	22
	Sun	0
2	Mon	108
	Tues	123
	Wed	135

Sum of first 7 terms =
122 + 156 + 147 + 88 + 73 + 22 + 0 = 608

We then take the average of these 7 terms. The average is inserted at the mid-point of the sequence – that is, next to Thursday – in a new column

$$\text{The average is} = \frac{608}{7} = 86.9\left(1.D.P.\right)$$

Which gives the table

Week	Day	Attendance	Moving average
1	Mon	122	
	Tues	156	
	Wed	147	
	Thurs	88	86.9
	Fri	73	
	Sat	22	
	Sun	0	
2	Mon	108	
	Tues	123	
	Wed	135	

We now sum the next 7 attendances – that is, we drop Monday of Week 1 out of the sum, and add in Monday of Week 2.

Week	Day	Attendance
1	Mon	122
	Tues	156
	Wed	147
	Thurs	88
	Fri	73
	Sat	22
	Sun	0
2	Mon	108
	Tues	123
	Wed	135

Sum of next 7 terms =
156 + 147 + 88 + 73 + 22 + 0 + 108 = 594

(Just a word of advice. If you are doing this "by hand", that is without the help of a computer programme, you will be tempted to use the method of subtracting on number from the running total and adding the next one on. If you make a mistake with this method, your error will accumulate throughout the process. Although it sounds tedious – and this topic is a little tedious – you are advised to sum the data afresh each time – it cuts down on error in the long run, unless you have very considerable powers of concentration!)

We take the average of this sum, and place it in the appropriate mid point, which is next to Friday of Week 1.

$$\text{The average is} = \frac{594}{7} = 84.9(1.D.P.)$$

This gives the table

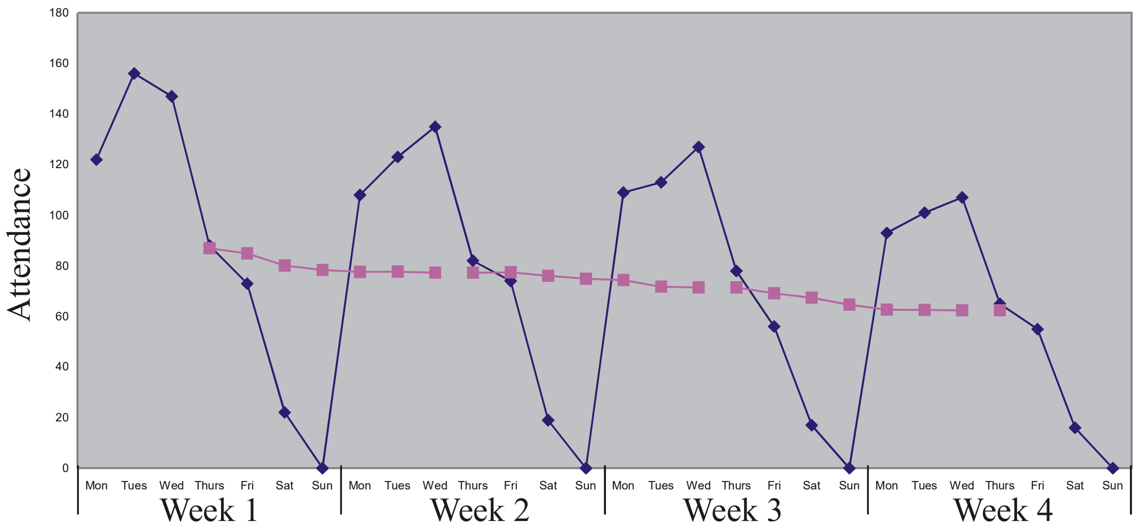
Week	Day	Attendance	Moving average
1	Mon	122	
	Tues	156	
	Wed	147	
	Thurs	88	86.9
	Fri	73	84.9
	Sat	22	
	Sun	0	
2	Mon	108	
	Tues	123	
	Wed	135	

We now repeat the process for the entire set of data. Note, that it will not be possible to find a moving average for some of these days at the beginning and end of the time period, because there will not be enough entries to add up and average.

Week	Day	Attendance	Moving	Moving

			total	average
1	Mon	122		
	Tues	156		
	Wed	147		
	Thurs	88	608	86.9
	Fri	73	594	84.9
	Sat	22	561	80.1
	Sun	0	549	78.4
2	Mon	108	543	77.6
	Tues	123	544	77.7
	Wed	135	541	77.3
	Thurs	82	541	77.3
	Fri	74	542	77.4
	Sat	19	532	76.0
	Sun	0	524	74.9
3	Mon	109	520	74.3
	Tues	113	502	71.7
	Wed	127	500	71.4
	Thurs	78	500	71.4
	Fri	56	484	69.1
	Sat	17	472	67.4
	Sun	0	452	64.6
4	Mon	93	439	62.7
	Tues	101	438	62.6
	Wed	107	437	62.4
	Thurs	65	437	62.4
	Fri	55		
	Sat	16		
	Sun	0		

We are now in a position to analyse the weekly trends. Before we do so, let us redraw the graph, showing the moving average as a trend line.



What the graph clearly shows is that the weekly fluctuations take the form of some days involving an attendance above the moving average, and some days involving an attendance below the moving average. So we represent the fluctuation as a + or – from the average.

Do to this, we subtract the value for the moving average from the actual value for that day.

Week	Day	Attendance	Moving total	Moving average
1	Mon	122		
	Tues	156		
	Wed	147		
	Thurs	88	608	86.9
	Fri	73	594	84.9
	Sat	22	561	80.1
	Sun	0	549	78.4
2	Mon	108	543	77.6
	Tues	123	544	77.7

Daily variation = 88 – 86.9 = +1.1

These daily fluctuations are included in a further column of the table.

	Day	Attendance	Moving total	Moving average	Daily variation
1	Mon	122			
	Tues	156			
	Wed	147			
	Thurs	88	608	86.9	+1.1
	Fri	73	594	84.9	–11.9
	Sat	22	561	80.1	–58.1

	Sun	0	549	78.4	–78.4
2	Mon	108	543	77.6	+30.4
	Tues	123	544	77.7	+45.3
	Wed	135	541	77.3	+57.7
	Thurs	82	541	77.3	+4.7
	Fri	74	542	77.4	–3.4
	Sat	19	532	76.0	–57.0
	Sun	0	524	74.9	–74.9
3	Mon	109	520	74.3	+34.7
	Tues	113	502	71.7	+41.3
	Wed	127	500	71.4	+55.6
	Thurs	78	500	71.4	+6.6
	Fri	56	484	69.1	–13.1
	Sat	17	472	67.4	–50.4
	Sun	0	452	64.6	–64.6
4	Mon	93	439	62.7	+30.3
	Tues	101	438	62.6	+38.4
	Wed	107	437	62.4	+44.6
	Thurs	65	437	62.4	+2.6
	Fri	55			
	Sat	16			
	Sun	0			

We have been also asked to predict the attendance during Week 8. To do this, we will make the prediction for each day on the basis of

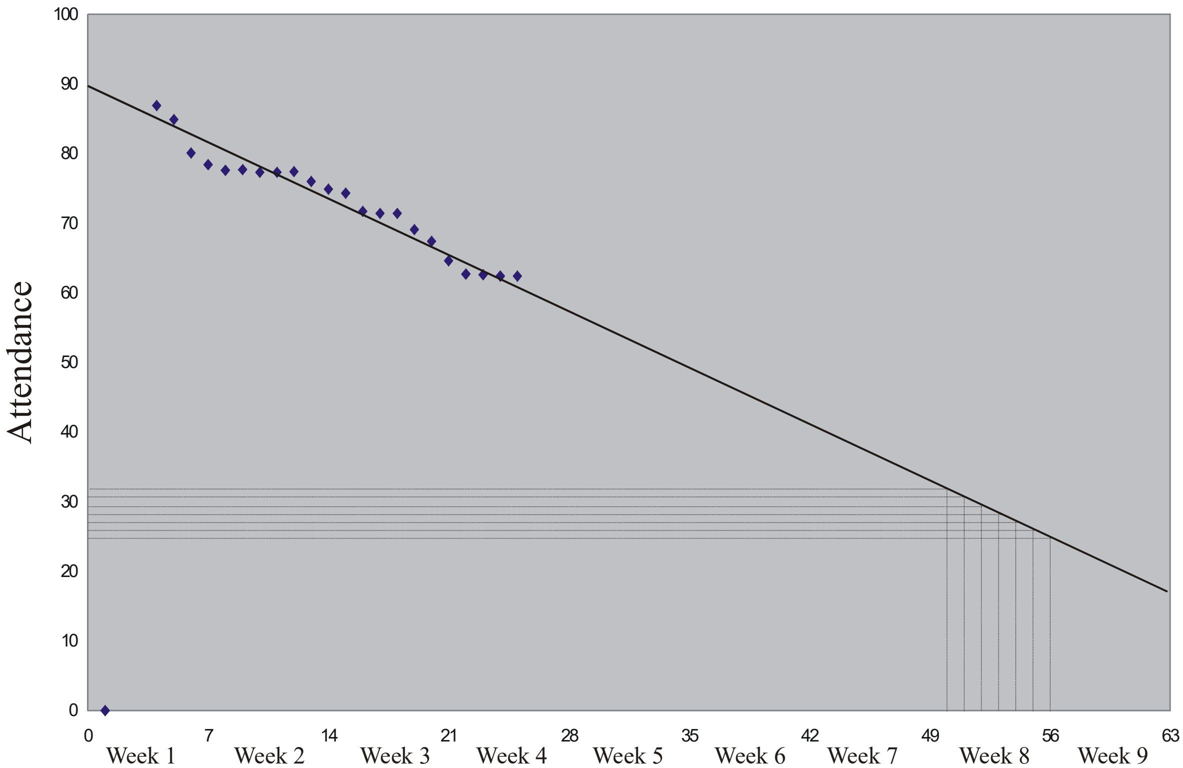
Prediction = Trend + Daily variation

For this, we will have to decide what "daily variation" we will add (or subtract). One approach here could be to take the average of the daily variations for the data we do have.

Day	Week 1	Week 2	Week 3	Week 4	Total	Average
Mon		+30.4	+34.7	+30.3	+95.4	+31.8
Tues		+45.3	+41.3	+38.4	+125	+41.7
Wed		+57.7	+55.6	+44.6	+157.9	+52.6
Thurs	+1.1	+4.7	+6.6	+2.6	+15.0	+3.75
Fri	-11.9	–3.4	–13.1		–28.4	–9.5
Sat	–58.1	–57.0	–50.4		–165.5	–55.2
Sun	–78.4	–74.9	–64.6		–217.9	–72.6

However, when making our prediction for Sunday of Week 8, we will set that attendance to 0, since that is the clear trend from our existing figures.

Now we have to find the trend values. To do this accurately we have to find the equation of the straight line that fits the data the best. There is a mathematical technique for this known as the method of least squares regression, which is dealt with in detail in units in the course called regression lines, and, at a higher level, least squares regression. Here, we will use the simpler graphical technique of drawing the line of best fit, and extrapolation.



To the nearest whole number this method of extrapolation gives the following values for average attendance during week 8:

Day	Average attendance
Mon	32
Tues	31
Wed	29
Thurs	28
Fri	27
Sat	26
Sun	25

These are only the values corresponding to the trend line, so we should add the daily variation, which we have just calculated. However, we will only add them to the nearest number, since the prediction could not be even that accurate.

Day	Average attendance	Daily variation	Predicted attendance
Mon	32	+32	64
Tues	31	+42	73
Wed	29	+53	82
Thurs	28	+4	32
Fri	27	−10	17
Sat	26	−55	0
Sun	25	−73	0

Note that attendance on Saturdays and Sundays cannot fall below 0, so that is our prediction for these days.

This illustrates the basic technique of time series analysis. However, in this above example we were taking a moving average over an odd number of time intervals – over seven days in fact. We need a small modification when dealing with moving averages of even numbers of time intervals. We proceed to illustrate this idea.

Example (2)

The following data derives from the bed occupancy at Mythical Island, which is a holiday resort renown for offering increasingly old-fashioned accommodation and entertainment at a high price.

Year	Quarter	Bed occupancy
1985	Q1	978
	Q2	1245
	Q3	1687
	Q4	1011
1986	Q1	865
	Q2	1132
	Q3	1303
	Q4	877
1987	Q1	843
	Q2	1034
	Q3	1271
	Q4	790
1988	Q1	723

	Q2	961
	Q3	1023
	Q4	695

Construct a yearly moving average, and show the quarterly variation. Display the results of your analysis graphically.

Solution

Here the problem is that in averaging 4 items we cannot strictly place the average alongside any one quarter. What we do first is calculate the moving total and place it between two quarters, in the middle of the 4 items that are used to construct it.

Year	Quarter	Bed occupancy	Yearly moving total
1985	Q1	978	
	Q2	1245	
			4921
	Q3	1687	
	Q4	1011	
1986	Q1	865	
	Q2	1132	
	Q3	1303	
	Q4	877	
1987	Q1	843	
	Q2	1034	
	Q3	1271	
	Q4	790	
1988	Q1	723	
	Q2	961	
	Q3	1023	
	Q4	695	

$$978 + 1245 + 1687 + 1011 = 4921$$

We work down the entire column in this way, filling in all the yearly totals.

Year	Quarter	Bed occupancy	Yearly moving total
1985	Q1	978	
	Q2	1245	
	Q3	1687	4921
	Q4	1011	3808
1986	Q1	865	4695
	Q2	1132	4311
	Q3	1303	4177
	Q4	877	4155
1987	Q1	843	4057
	Q2	1034	4025
	Q3	1271	3938
	Q4	790	3818
1988	Q1	723	3745
	Q2	961	3479
	Q3	1023	3672
	Q4	695	

In order to realign the values of the totals with the quarters, we now add the yearly moving totals in pairs to produce a two yearly moving total.

Year	Quarter	Bed occupancy	Yearly moving total	Two yearly moving total	Moving average	Quarterly variation
1985	Q1	978				
	Q2	1245				
	Q3	1687	4921	8729	1091	+596
	Q4	1011	3808			
1986	Q1	865	4695			
	Q2	1132	4311			
	Q3	1303	4177			
	Q4	877	4155			
1987	Q1	843	4057			
	Q2	1034	4025			
	Q3	1271	3938			
	Q4	790	3818			
1988	Q1	723	3745			
	Q2	961	3479			
	Q3	1023	3672			
	Q4	695				

The first of these two-yearly totals is the sum of the first two year totals.

$4921 + 3808 = 8729$

The average is then found by dividing by the number of quarters in the two-yearly total – that is 8.

$8729 \div 8 = 1091$

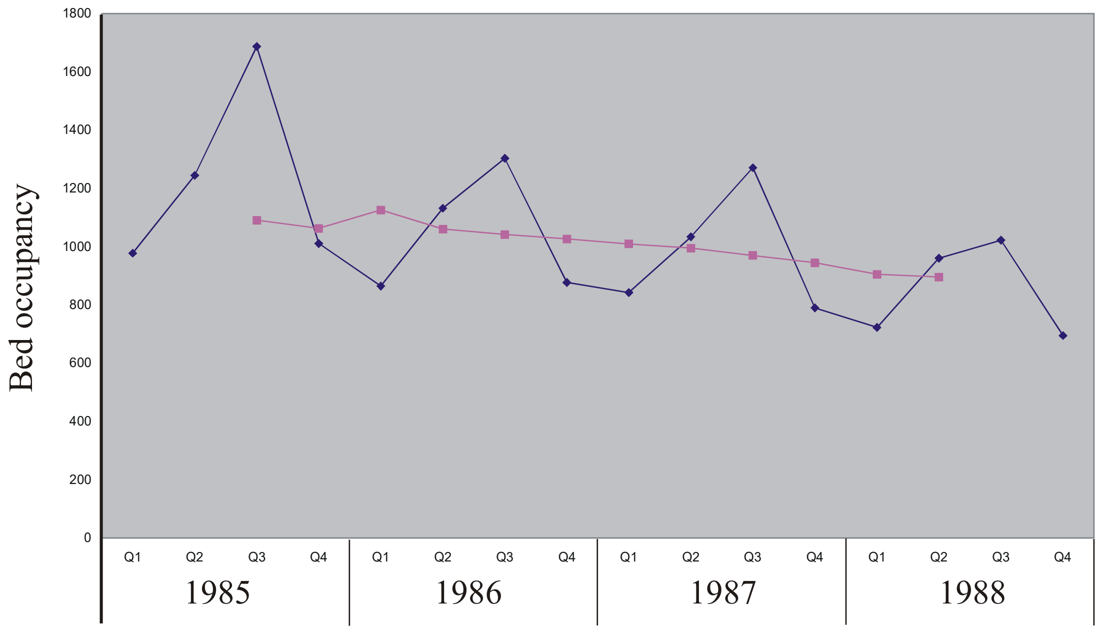
This is to the nearest bed. The quarterly variation is the bed occupancy – the moving average

$1687 - 1091 = 596$

The full solution is

Year	Quarter	Bed occupancy	Yearly moving total	Two yearly moving total	Moving average	Quarterly variation
1985	Q1	978				
	Q2	1245				
	Q3	1687	4921	8729	1091	+596
	Q4	1011	3808	8503	1063	–52
1986	Q1	865	4695	9006	1126	–261
	Q2	1132	4311	8488	1061	+71
	Q3	1303	4177	8332	1042	+261
	Q4	877	4155	8234	1027	–150
1987	Q1	843	4057	8082	1010	–167
	Q2	1034	4025	7963	995	+39
	Q3	1271	3938	7756	970	+301
	Q4	790	3818	7563	945	–155
1988	Q1	723	3745	7242	905	–182
	Q2	961	3479	7169	896	+65
	Q3	1023	3672			
	Q4	695				

A graphical presentation of these results is as follows


Bed occupancy
1800
1600
1400
1200
1000
800
600
400
200
0
Q1
Q2
Q3
Q4
1985
1986
1987
1988