Undergraduate Journal of Mathematical Modeling: One + Two

Volume 9 | 2018 Fall 2018 Issue 1 | Article 3

Trend Analysis and Completion Prediction of the Section Project

Ronald Jones
University of South Florida

Advisors:
Arcadii Grinshpan, Mathematics and Statistics
Andrea Link, Director of Engineering, Digarc

Problem Suggested By: Andrea Link

Recommended Citation
Jones, Ronald (2018) "Trend Analysis and Completion Prediction of the Section Project," *Undergraduate Journal of Mathematical Modeling: One + Two*: Vol. 9: Iss. 1, Article 3.
DOI: https://doi.org/10.5038/2326-3652.9.1.4896
Available at: https://scholarcommons.usf.edu/ujmm/vol9/iss1/3

Trend Analysis and Completion Prediction of the Section Project

Abstract

Creating an accurate prediction of the completion timeline of a software development project is complicated and error prone. Developers will gain a natural intuition as to how long a task should take then. However, this prediction can end up being wrong in many cases due to many factors. This paper will attempt to determine a formula which allows a more accurate prediction to be created.

Creating an accurate prediction of the completion timeline of a software development project is complicated and error prone. Developers will gain a natural intuition as to how long a task should take then. However, this prediction can end up being wrong in many cases due to many factors. This paper will attempt to determine a formula which allows a more accurate prediction to be created.

Keywords

software development, completion prediction, polynomial trendline, polynomial equation, polynomial roots

PROBLEM STATEMENT

Software development is a challenging field of engineering. One of the many challenges of accurate determining is when a specific project will be completed. This paper does a trend analysis on the number of tickets opened and completed during a period of 6 months for a specific product being developed by Digarc. With this analysis, the resulting trend should be able to predict when the project will be completed.

MOTIVATION

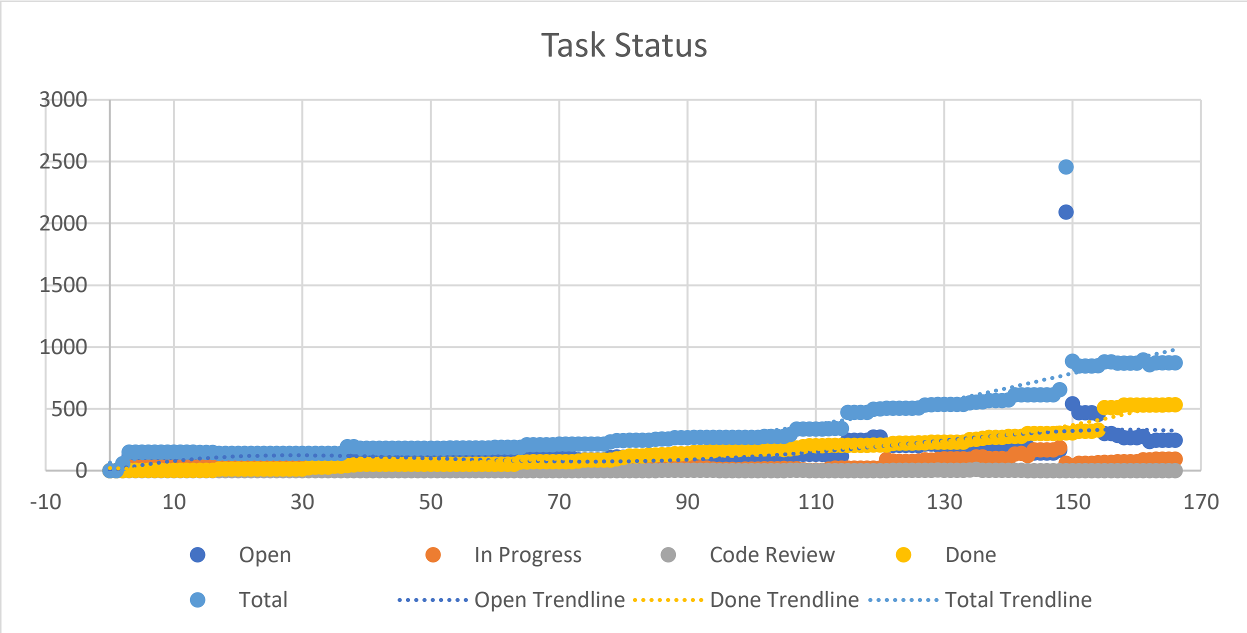
In the software development career field, accurate predictions for the completion of projects are important. Accurate prediction can lead towards better time and resource management of the development and quality assurance teams. It can also allow the sales and marketing teams to promise new features and software with a reasonable probability of deliverance. Many factors go into these predictions including accurate specifications for the new feature or product as well as understanding the difficulties in bringing a new feature or product to completion.

While it is the developers' job to prediction how long the engineering process of the new feature will take, this task is made difficult when accurate specifications are not given. On large projects as the one presented in this paper, not all the specifications can reasonably be provided before development begins. As specifications are created, the amount of tasks that the developers are required to do will grow. Tracking these tasks as they are created and completed are even more vital in this situation so a trend of the work can be determined and an informed prediction can be created.

MATHEMATICA DESCRIPTION AND SOLUTION APPROACH

Many programs are available for tracking the status of tasks needed to be done to complete a software engineering project. For the project looked at in this paper, Target Process, a product of Target Process Inc, is being used. Target Process tracks, among other things, the state of each task at any given time. We were able to export the number of tasks for the specific project looked at in this paper in each of the various states. While, the software tracks many states vital to the software development process, only the **Open**, **Done** and **Total** count of tasks for each day is looked at in this paper.

The values, which can be found in Appendix A, were imported in Excel and the following graph was generated.



Generating the trendline equations by Excel polynomial regression from this chart provide us with a fourth order polynomial equation for the **Open** and **Done** status as well as the **Total** number of tasks.

DOI: https://doi.org/10.5038/2326-3652.9.1.4896

The equation for the **Open** status is the following:

$$y = -8 * 10^{-6}x^4 + 0.0026x^3 - 0.2802x^2 + 10.363x. \quad (1)$$

The equation for the **Done** status is the following:

$$y = 4 * 10^{-6}x^4 - 0.0013x^3 + 0.1266x^2 - 3.0689x + 21.733. \quad (2)$$

The equation for the **Total** tasks is the following:

$$y = -5 * 10^{-6}x^4 + 0.0019x^3 - 0.1944x^2 + 7.9342x - 64.046. \quad (3)$$

We hope that it is possible to calculate the date when the project is likely to be completed using these trendlines. Since the most obvious definition of the project being completed is that all the tickets are in the **Done** state, we set the equations for **Done** (equation 2) and **Total** (equation 3) equal and solve for *x*:

$$4 * 10^{-6}x^4 - 0.0013x^3 + 0.1266x^2 - 3.0689x + 21.733$$

$$= -5 * 10^{-6}x^4 + 0.0019x^3 - 0.1944x^2 + 7.9342x - 64.046. \quad (4)$$

We start by moving both sides of equation (4) to the left side:

$$\frac{9x^4}{1000000} - \frac{2x^3}{625} + \frac{321x^2}{1000} - \frac{110031x}{10000} + \frac{85779}{1000} = 0 \,, \quad (5)$$

and then multiply (5) by 1000000 to remove the fraction which should give us the following:

$$9x^4 - 3200x^3 + 321000x^2 - 11003100x + 85779000 = 0. \quad (6)$$

However a one-digit misprint led us to a wrong equation:

$$9x^4 - 3200x^3 + 321000x^2 - 1003100x + 85779000 = 0. \quad (7)$$

We analyze both equations (6) and (7) to show the role of a one-digit misprint for the solution of the considered problem. We begin with the wrong equation (7).

One can use *WolframAlpha* to find the roots of polynomial equation (7):

$$x_1 \approx 0.2187 - 16.444i, \quad x_2 \approx 0.2187 + 16.444i,$$
$$x_3 \approx 177.559 - 60.94i, \quad x_4 \approx 177.559 + 60.94i. \tag{8}$$

Since none of the roots of (7) are real, no solution exists over the real numbers.

Alternatively, one can use a standard technique to come to the same conclusion about the roots of equation (7). Substitute *x* with *y* + 800/9 in (7) to eliminate the cubic term:

$$85779000 - 1003100\left(y + \frac{800}{9}\right) + 321000\left(y + \frac{800}{9}\right)^2 - 3200\left(y + \frac{800}{9}\right)^3 + 9\left(y + \frac{800}{9}\right)^4 = 0. \tag{9}$$

Write the depressed quartic polynomial equation (9) into the standard form:

$$y^4 - \frac{317000y^2}{27} + \frac{445148900y}{729} + \frac{205897337000}{2187} = 0. \tag{10}$$

Then reduce equation (10) to two quadratic equations, find their roots y_k $(k = 1, 2, 3, 4)$, and then define the roots of equation (7) by the formulas $x_k = y_k + \frac{800}{9}$ $(k = 1, 2, 3, 4)$.
Here are the basic steps based on Ferrari's method, https://en.wikipedia.org/wiki/Quartic_function. Complete the square in (10):

$$y^4 - \frac{317000y^2}{27} + \left(\frac{158500}{27}\right)^2 = -\frac{445148900y}{729} + \left(\frac{158500}{27}\right)^2 - \frac{205897337000}{2187}. \tag{11}$$

Express the left-hand side of (11) as a square:

$$\left(y^2 - \frac{158500}{27}\right)^2 = -\frac{445148900y}{729} + \left(\frac{158500}{27}\right)^2 - \frac{205897337000}{2187}. \tag{12}$$

Add $2\lambda\left(y^2 - \frac{158500}{27}\right) + \lambda^2$ to both sides of (12), where λ is a nonzero constant to be determined later:

$$\left(y^2 - \frac{158500}{27}\right)^2 + 2\lambda\left(y^2 - \frac{158500}{27}\right) + \lambda^2 = -\frac{445148900y}{729} + \left(\frac{158500}{27}\right)^2 - \frac{205897337000}{2187} + 2\lambda\left(y^2 - \frac{158500}{27}\right) + \lambda^2 . \tag{13}$$

Collect the right hand side of (13) in terms of *y*:

DOI: https://doi.org/10.5038/2326-3652.9.1.4896

$$\left(y^2-\frac{158500}{27}+\lambda\right)^2 = 2\lambda y^2-\frac{445148900y}{729}+\frac{1}{2\lambda}\left(\frac{222574450}{729}\right)^2+\left(\frac{158500}{27}\right)^2-\frac{205897337000}{2187} -\frac{317000}{27}\lambda+\lambda^2-\frac{1}{2\lambda}\left(\frac{222574450}{729}\right)^2. \tag{14}$$

As $\lambda \neq 0$ we complete the square on the right-hand side of (14):

$$\left(y^2-\frac{158500}{27}+\lambda\right)^2 = \left(\sqrt{2\lambda}\,y-\frac{1}{\sqrt{2\lambda}}\frac{222574450}{729}\right)^2-\frac{130530587000}{2187}-\frac{317000}{27}\lambda+\lambda^2- \frac{1}{2\lambda}\left(\frac{222574450}{729}\right)^2. \tag{15}$$

To express the right-hand side of (15) as a square, one can use any value of λ such that:

$$-\frac{130530587000}{2187}-\frac{317000}{27}\lambda+\lambda^2-\frac{1}{2\lambda}\left(\frac{222574450}{729}\right)^2=0. \tag{16}$$

It is easy to see that a nonzero root λ_w of equation (16) exists. We can use the depressed cubic "equivalent" of (16) generated by a linear substitution $\lambda = z - 317000/81$ such that the coefficient for z^2 is 0, applying Cardano's method and formula (see Appendix B) to find $\lambda_w \neq 0$ which satisfies equation (16). Afterward we use the value of $\lambda=\lambda_w$ to obtain the following equations from (15):

$$y^2-\frac{158500}{27}+\lambda_w = \pm\left(\sqrt{2\lambda_w}\,y-\frac{1}{\sqrt{2\lambda_w}}\frac{222574450}{729}\right). \tag{17}$$

Quadratic equations (17) lead to the four roots y_k $(k=1,2,3,4)$ of equation (10) and then to the four roots (8): $x_k = y_k+\frac{800}{9}$ $(k=1,2,3,4)$ of the wrong equation (7).

Now we use the similar steps to analyze the correct equation (6).

By *WolframAlpha* we find the roots of polynomial equation (6):

$$x_1 \approx 10.8934,\quad x_2 \approx 53.7847,\quad x_3 \approx 75.5456,\quad x_4 \approx 215.332\ . \tag{18}$$

Thus all the four roots of (6) are real. We can apply the above used standard technique to come to the same conclusion about the roots of equation (6). Substitute *x* with *y* + 800/9 in (6) to eliminate the cubic term:

$$85779000-11003100\left(y+\frac{800}{9}\right)+321000\left(y+\frac{800}{9}\right)^2-3200\left(y+\frac{800}{9}\right)^3+9\left(y+\frac{800}{9}\right)^4 = 0. \tag{19}$$

Write the depressed quartic polynomial equation (19) into the standard form:

$$y^4 - \frac{317000y^2}{27} - \frac{364851100y}{729} - \frac{10102663000}{2187} = 0. \tag{20}$$

Then reduce equation (20) to two quadratic equations, find their roots y_k $(k = 1, 2, 3, 4)$, and define the roots of equation (6) by the formulas $x_k = y_k + \frac{800}{9}$ $(k = 1, 2, 3, 4)$.

The basic steps based on Ferrari's method, https://en.wikipedia.org/wiki/Quartic_function are given below. Complete the square in (20):

$$y^4 - \frac{317000y^2}{27} + \left(\frac{158500}{27}\right)^2 = \frac{364851100y}{729} + \frac{10102663000}{2187} + \left(\frac{158500}{27}\right)^2. \tag{21}$$

Express the left-hand side of (21) as a square:

$$\left(y^2 - \frac{158500}{27}\right)^2 = \frac{364851100y}{729} + \frac{10102663000}{2187} + \left(\frac{158500}{27}\right)^2 . \tag{22}$$

Add $2\lambda\left(y^2 - \frac{158500}{27}\right) + \lambda^2$ to both sides in (22) where λ is a nonzero constant to be determined later:

$$\left(y^2 - \frac{158500}{27}\right)^2 + 2\lambda\left(y^2 - \frac{158500}{27}\right) + \lambda^2 = \frac{364851100y}{729} + \frac{10102663000}{2187} + \left(\frac{158500}{27}\right)^2 + 2\lambda\left(y^2 - \frac{158500}{27}\right) + \lambda^2. \tag{23}$$

Collect the right hand side (23) in terms of *y*:

$$\left(y^2 - \frac{158500}{27} + \lambda\right)^2 = 2\lambda y^2 + \frac{364851100y}{729} + \frac{1}{2\lambda}\left(\frac{1824250550}{729}\right)^2 + \frac{10102663000}{2187} + \left(\frac{158500}{27}\right)^2 - \lambda\frac{317000}{27} + \lambda^2 - \frac{1}{2\lambda}\left(\frac{1824250550}{729}\right)^2. \tag{24}$$

As $\lambda \neq 0$ we complete the square on the right-hand side of (24):

$$\left(y^2 - \frac{158500}{27} + \lambda\right)^2 = \left(\sqrt{2\lambda}\, y + \frac{1}{\sqrt{2\lambda}}\frac{1824250550}{729}\right)^2 + \frac{85469413000}{2187} - \lambda\frac{317000}{27} + \lambda^2 - \frac{1}{2\lambda}\left(\frac{1824250550}{729}\right)^2. \tag{25}$$

To express the right-hand side of (25) as a square, use any nonzero value of λ such that:

$$\frac{85469413000}{2187} - \lambda\frac{317000}{27} + \lambda^2 - \frac{1}{2\lambda}\left(\frac{1824250550}{729}\right)^2 = 0. \tag{26}$$

DOI: https://doi.org/10.5038/2326-3652.9.1.4896

It is easy to see that a nonzero root λ_c of equation (26) exists. To find it we use the depressed cubic "equivalent" of (26) generated by a linear substitution $\lambda = z - 317000/81$ such that the coefficient for z^2 is 0, and we apply Cardano's method and formula (see Appendix B). Then we use the value of $\lambda=\lambda_c$ to obtain the following equations from (25):

$$y^2 - \frac{158500}{27} + \lambda_c = \pm\left(\sqrt{2\lambda_c}\; y + \frac{1}{\sqrt{2\lambda_c}}\frac{1824250550}{729}\right).$$
(27)

Quadratic equations (27) lead to the four roots $y_k (k = 1, 2, 3, 4)$ of equation (20) and then to the four roots (18): $x_k = y_k + \frac{800}{9}$ $(k = 1, 2, 3, 4)$ of the correct equation (6).

DISCUSSION

When attempting to calculate the estimated completion date of a project using the fourth order polynomial equation derived from the number of tasks in each status, two polynomial equations were analyzed: the correct quartic equation and a wrong quartic equation with a one-digit misprint. With the wrong equation only one conclusion is found. That conclusion is that due to the nearly random nature of the data, the polynomials that are generated from it cannot be used to accurately predict a completion date. This result is not unexpected as the very nature of software development can be varied and difficult to quantify. The conclusion based on the correct equation is absolutely different: four real solutions are found. It is up to the company management to decide which one of them is more suitable in the considered case.

If this prediction were to be as easily quantified as was being attempted in that paper, some development prediction algorithm would probably have already been created and used by management to determine how long a project should take, and developers would not be needed to make these predictions.

Conclusions and Recommendations

As the results show, creating a formula to provide an accurate prediction for the completion of a software development project is not an easy task which can depend on a one-digit misprint. Experienced developers learn to anticipate the challenges and pitfalls of development, but the predictions are still a guess that can still be wrong on many reasons. Some of these reasons include but are not limited to not having full specifications for the project at the start, not understanding the requirements, detailed level of the code being created or modified, and unexpected side effects of modified code. It is the developers' job to learn to account for these and other factors, and try to give as accurate prediction as possible.

References

Section Value Delivery, Digarc, Lakeland, Florida, 2018.

Brian Albright, *Mathematical Modeling with Excel*, Jones & Bartlett, 2011.

Alicia Dickenstein, Ioannis Z. Emiris, *Solving Polynomial Equations: Foundations, Algorithms, and Applications*, Springer, 2006.

Quartic function, https://en.wikipedia.org/wiki/Quartic_function

Cubic function, Cardano's_method,

https://en.wikipedia.org/wiki/Cubic_function#Cardano's_method

DOI: https://doi.org/10.5038/2326-3652.9.1.4896

APPENDIX A

Task Status for Section project by date:

Date	Open	In Progress	Code Review	Done	Total
13-Nov-17	0	0	0	0	0
14-Nov-17	0	0	0	0	0
15-Nov-17	34	24	0	0	58
16-Nov-17	119	28	0	3	150
17-Nov-17	118	29	0	3	150
18-Nov-17	118	29	0	3	150
19-Nov-17	118	29	0	3	150
20-Nov-17	118	29	0	3	150
21-Nov-17	118	29	0	3	150
22-Nov-17	118	29	0	3	150
23-Nov-17	118	29	0	3	150
24-Nov-17	118	29	0	3	150
25-Nov-17	118	29	0	3	150
26-Nov-17	118	29	0	3	150
27-Nov-17	111	33	0	4	148
28-Nov-17	111	33	0	4	148
29-Nov-17	111	33	0	4	148
30-Nov-17	95	33	0	12	140
1-Dec-17	95	33	0	12	140
2-Dec-17	95	33	0	12	140
3-Dec-17	95	33	0	12	140
4-Dec-17	95	33	0	12	140
5-Dec-17	89	39	0	12	140
6-Dec-17	89	37	0	14	140
7-Dec-17	89	37	0	14	140
8-Dec-17	89	37	0	14	140
9-Dec-17	89	37	0	14	140
10-Dec-17	89	37	0	14	140
11-Dec-17	89	37	0	14	140
12-Dec-17	89	37	0	14	140
13-Dec-17	89	38	0	14	141
14-Dec-17	89	29	0	23	141
15-Dec-17	89	23	0	29	141
16-Dec-17	89	23	0	29	141
17-Dec-17	89	23	0	29	141
18-Dec-17	89	16	0	36	141

19-Dec-17	89	16	0	36	141
20-Dec-17	137	17	0	41	195
21-Dec-17	137	12	0	46	195
22-Dec-17	119	12	0	51	182
23-Dec-17	119	12	0	51	182
24-Dec-17	119	12	0	51	182
25-Dec-17	119	12	0	51	182
26-Dec-17	119	12	0	51	182
27-Dec-17	99	32	0	51	182
28-Dec-17	99	32	0	51	182
29-Dec-17	99	32	0	51	182
30-Dec-17	99	32	0	51	182
31-Dec-17	99	32	0	51	182
1-Jan-18	99	32	0	51	182
2-Jan-18	99	32	0	51	182
3-Jan-18	99	32	0	51	182
4-Jan-18	99	32	0	51	182
5-Jan-18	95	37	0	51	183
6-Jan-18	95	37	0	51	183
7-Jan-18	95	37	0	51	183
8-Jan-18	82	52	0	51	185
9-Jan-18	82	52	0	51	185
10-Jan-18	82	52	0	51	185
11-Jan-18	82	52	0	51	185
12-Jan-18	87	52	0	51	190
13-Jan-18	87	52	0	51	190
14-Jan-18	87	52	0	51	190
15-Jan-18	87	52	0	51	190
16-Jan-18	85	34	0	71	190
17-Jan-18	106	34	0	71	211
18-Jan-18	106	32	0	73	211
19-Jan-18	106	32	0	73	211
20-Jan-18	106	32	0	73	211
21-Jan-18	106	32	0	73	211
22-Jan-18	110	33	0	73	216
23-Jan-18	102	41	0	73	216
24-Jan-18	102	41	0	73	216
25-Jan-18	89	49	0	78	216
26-Jan-18	89	41	0	86	216
27-Jan-18	89	41	0	86	216
28-Jan-18	89	41	0	86	216
29-Jan-18	89	41	0	86	216
30-Jan-18	106	41	0	86	233

DOI: https://doi.org/10.5038/2326-3652.9.1.4896

31-Jan-18	112	35	0	97	244
1-Feb-18	110	29	0	105	244
2-Feb-18	110	17	0	119	246
3-Feb-18	110	17	0	119	246
4-Feb-18	110	17	0	119	246
5-Feb-18	98	20	0	128	246
6-Feb-18	95	31	0	128	254
7-Feb-18	91	35	1.5	128.5	256
8-Feb-18	91	35	1.5	128.5	256
9-Feb-18	91	35	2	138.5	266.5
10-Feb-18	91	35	2	138.5	266.5
11-Feb-18	91	35	2	138.5	266.5
12-Feb-18	95	27	2	146.5	270.5
13-Feb-18	95	24	2	149.5	270.5
14-Feb-18	94	21	2.5	152.5	270
15-Feb-18	94	13	2.5	160.5	270
16-Feb-18	94	10	0	166	270
17-Feb-18	94	10	0	166	270
18-Feb-18	94	10	0	166	270
19-Feb-18	93	7	0	170	270
20-Feb-18	87	13	0	170	270
21-Feb-18	87	13	0	170	270
22-Feb-18	84	12	0	174	270
23-Feb-18	85	15	1	175	276
24-Feb-18	85	15	1	175	276
25-Feb-18	85	15	1	175	276
26-Feb-18	85	18	1	172	276
27-Feb-18	98	15	4	176	293
28-Feb-18	133	15	3	184.5	335.5
1-Mar-18	133	8	2	193.5	336.5
2-Mar-18	133	2	0	202.5	337.5
3-Mar-18	133	2	0	202.5	337.5
4-Mar-18	133	2	0	202.5	337.5
5-Mar-18	122	12	0	204.5	338.5
6-Mar-18	117	21	0	204.5	342.5
7-Mar-18	117	20	0	205.5	342.5
8-Mar-18	245.5	21	0	205.5	472
9-Mar-18	245	20.5	1	205.5	472
10-Mar-18	245	20.5	1	205.5	472
11-Mar-18	245	20.5	1	205.5	472
12-Mar-18	271	20.5	0	206.5	498
13-Mar-18	271	21.5	0	206.5	499
14-Mar-18	208	90	0	207	505

15-Mar-18	205	79	2	219	505
16-Mar-18	205	75	2	223	505
17-Mar-18	205	75	2	223	505
18-Mar-18	205	75	2	223	505
19-Mar-18	201	78	2.5	226.5	508
20-Mar-18	214	89	2.5	226.5	532
21-Mar-18	214	86	2.5	230	532.5
22-Mar-18	208	92	5.5	230	535.5
23-Mar-18	205	95	4	231.5	535.5
24-Mar-18	205	95	4	231.5	535.5
25-Mar-18	205	95	4	231.5	535.5
26-Mar-18	205	95	4	231.5	535.5
27-Mar-18	199	94	6	250.5	549.5
28-Mar-18	180	111	10	253.5	554.5
29-Mar-18	180	111	3	262.5	556.5
30-Mar-18	197	98	2	270.5	567.5
31-Mar-18	197	98	2	270.5	567.5
1-Apr-18	197	98	2	270.5	567.5
2-Apr-18	197	96	3	276.5	572.5
3-Apr-18	197	136	3	276.5	612.5
4-Apr-18	194	139	3	276.5	612.5
5-Apr-18	192	121	0	299.5	612.5
6-Apr-18	143	169	1	301.5	614.5
7-Apr-18	143	169	1	301.5	614.5
8-Apr-18	143	169	1	301.5	614.5
9-Apr-18	143	169	1	301.5	614.5
10-Apr-18	165	186	0	304	655
11-Apr-18	2093	59	0	305	2457
12-Apr-18	542	39	0	305	886
13-Apr-18	468	60	1	318	847
14-Apr-18	468	60	1	318	847
15-Apr-18	468	60	1	318	847
16-Apr-18	455	64	0	330	849
17-Apr-18	300.5	67	1	511	879.5
18-Apr-18	300.5	67	1	511	879.5
19-Apr-18	285.5	72	1	511	869.5
20-Apr-18	265.5	72	0	532	869.5
21-Apr-18	265.5	72	0	532	869.5
22-Apr-18	265.5	72	0	532	869.5
23-Apr-18	276.5	88	0	532	896.5
24-Apr-18	235.5	89	0	532	856.5
25-Apr-18	246.5	94	0	532	872.5
26-Apr-18	246.5	94	0	532	872.5

DOI: https://doi.org/10.5038/2326-3652.9.1.4896

27-Apr-18	246.5	93	0	533	872.5
28-Apr-18	246.5	93	0	533	872.5

APPENDIX B

Cardano's formula for three roots of the depressed cubic equation $x^3 + ax + b = 0$ is

$$x_k = \varepsilon^k \sqrt[3]{\frac{-b}{2} + \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}} + \varepsilon^{2k} \sqrt[3]{\frac{-b}{2} - \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}},$$

where $\varepsilon = \frac{-1+i\sqrt{3}}{2}$ and k = 0, 1, 2.

Here the cube roots presented as radicals are defined to be any pair of cube roots whose product is $-\frac{a}{3}$, https://en.wikipedia.org/wiki/Cubic_function#Cardano's_method .