

Answers to DELOS Aqua Questions

Tools used is python notebook with package below :

1. Pandas for data wrangling
2. Sklearn linear model for simple estimator
3. Pyplot and seaborn for plotting

Answers :

1. How can you generate a random number between 1 – 7 with only 1 dice?

With only 1 dice, if each dice roll observation is independent then you can only generate a random number between 1 through 6 but not 7.

2. 1. regarding dataset :

- Upon data inspection, there are 3 missing Afternoon Temperature and Evening Temperature on the same row/record and 85 missing Feed/Day.
- To impute the missing Temperature record, I sum the rounded average temperature changes between morning to afternoon(+2) and afternoon to evening(-1) to each temperature measurement. Reason for imputing by offsetting after the changes is to accurately represent the temp changes throughout the day, instead of just taking the typical temperature over the course of all measurements.

Below are before and after imputing.

	Days of Production	Pond ID	Morning Temp	Afternoon Temp	Evening Temp
76	77	A1	27	NaN	NaN
153	77	A2	28	29.0	29.0
247	77	A3	28	NaN	NaN
324	77	A4	27	29.0	29.0
418	77	A5	28	29.0	29.0
512	77	A6	27	29.0	29.0
606	77	A7	27	29.0	29.0
700	77	A8	27	NaN	NaN
777	77	A9	28	29.0	29.0
871	77	AA10	27	29.0	29.0

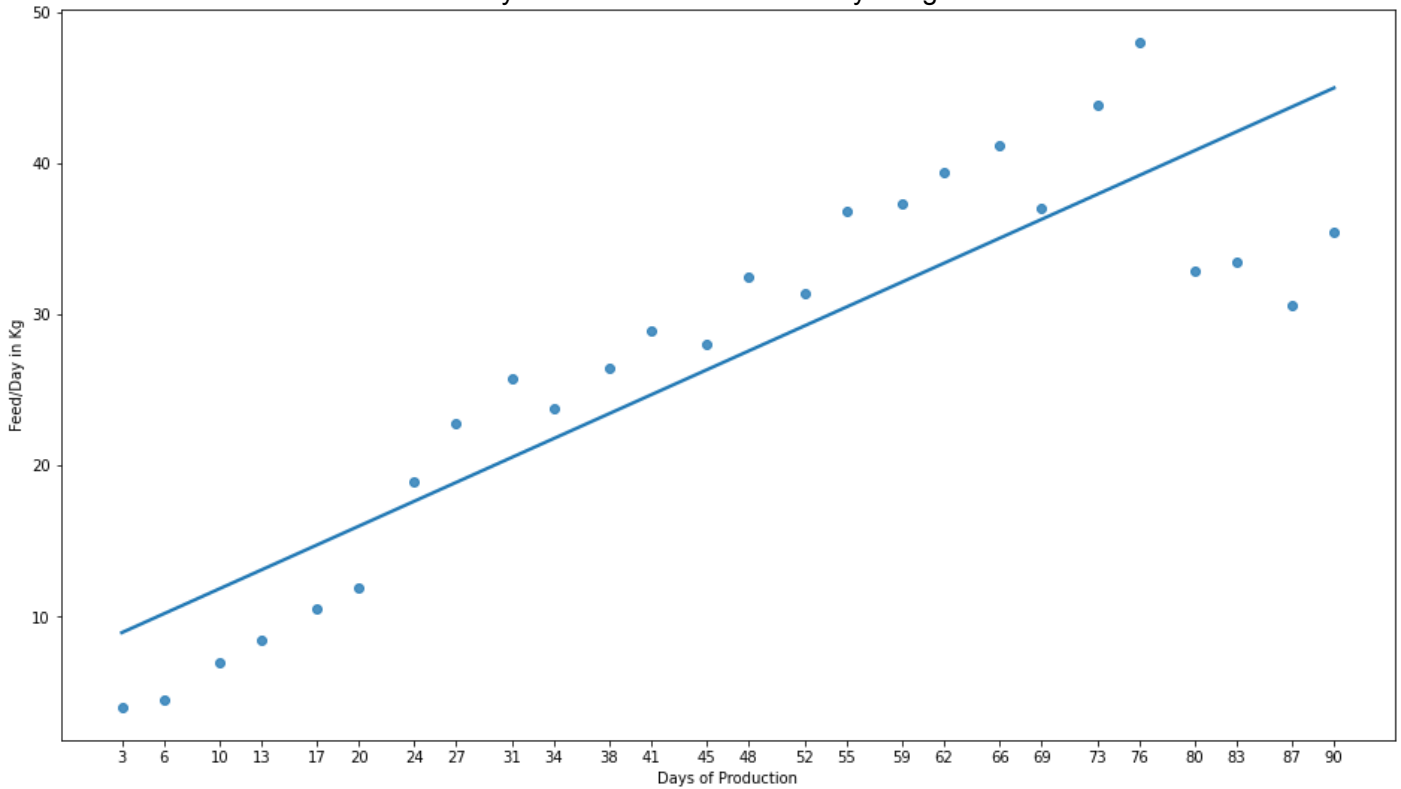
	Days of Production	Pond ID	Morning Temp	Afternoon Temp	Evening Temp
76	77	A1	27	29	28
153	77	A2	28	29	29
247	77	A3	28	30	29
324	77	A4	27	29	29
418	77	A5	28	29	29
512	77	A6	27	29	29
606	77	A7	27	29	29
700	77	A8	27	29	28
777	77	A9	28	29	29
871	77	AA10	27	29	29

- To impute the missing amount of Feed/Day, based on information given, I estimated Feed/Day by running a regression estimator on 'Days of Production' and Average Temperature on any given day. Reason for using Average Temperature instead of each Temperature measurement individually is because I assume the feed/day in the data represents the total accumulation feed given over multiple feeding sessions.

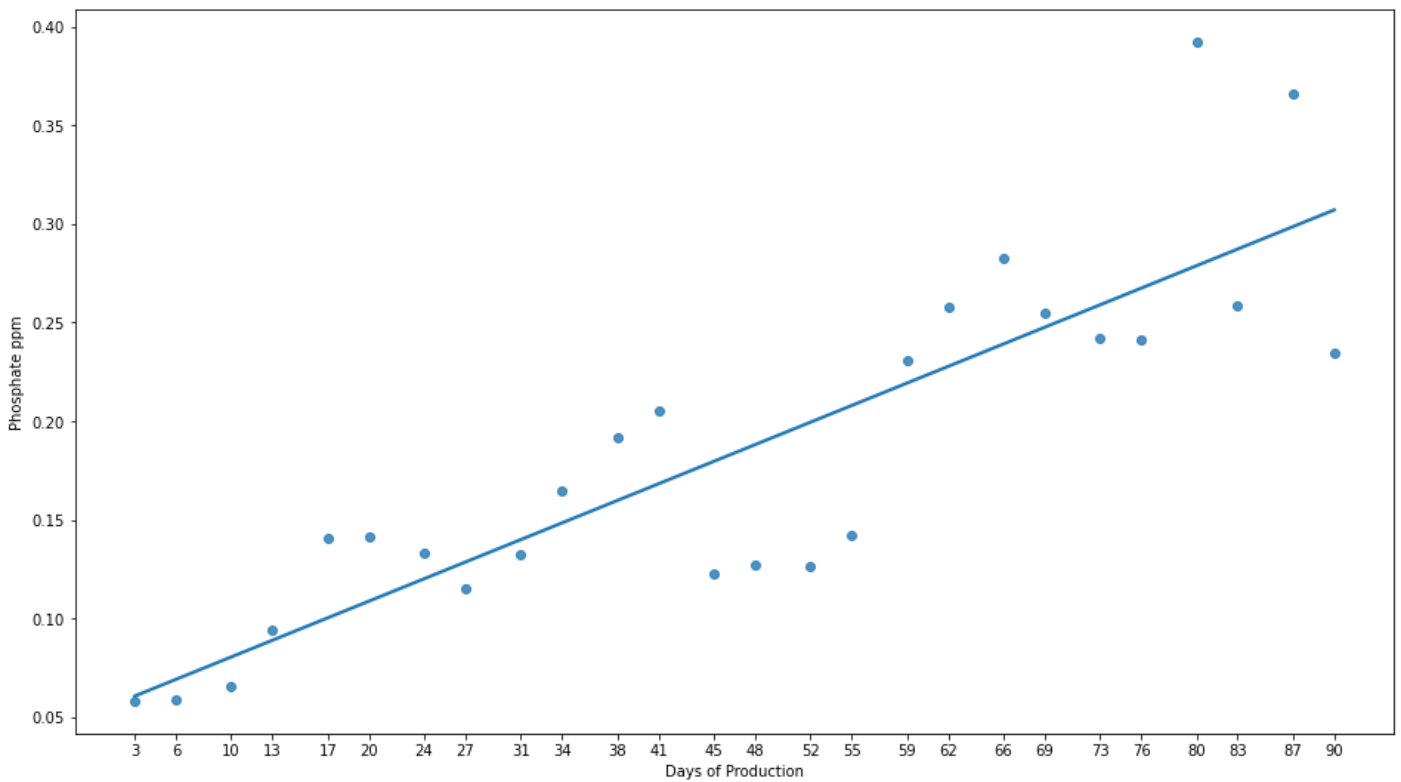
2. After plotting each inputs against the estimator here is what i can infer :

1. Over the course of Days of production, Feed/Day, Phosphate ppm and Phytoplankton cells/ml generally increases.

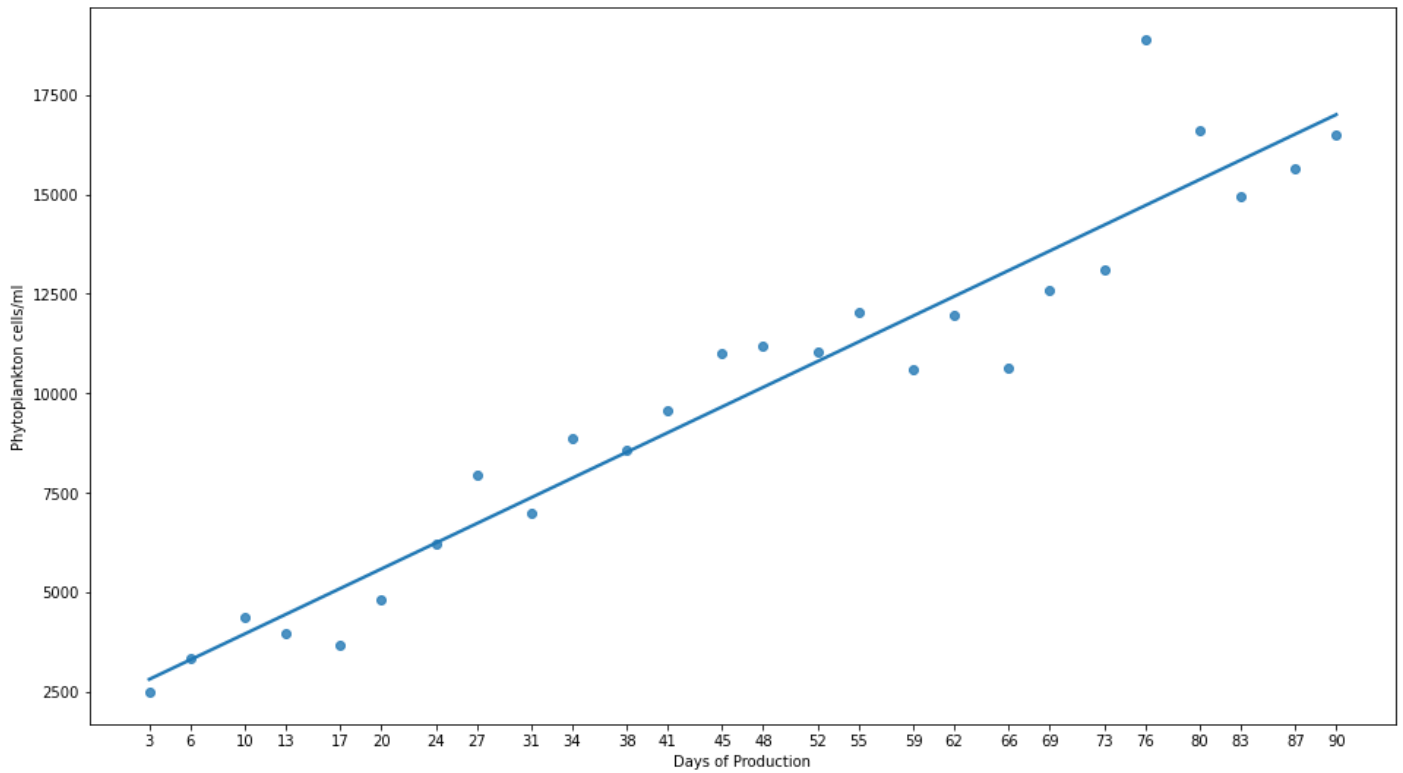
Days of Production vs Feed/Day in Kg



Days of Production vs Phosphate ppm

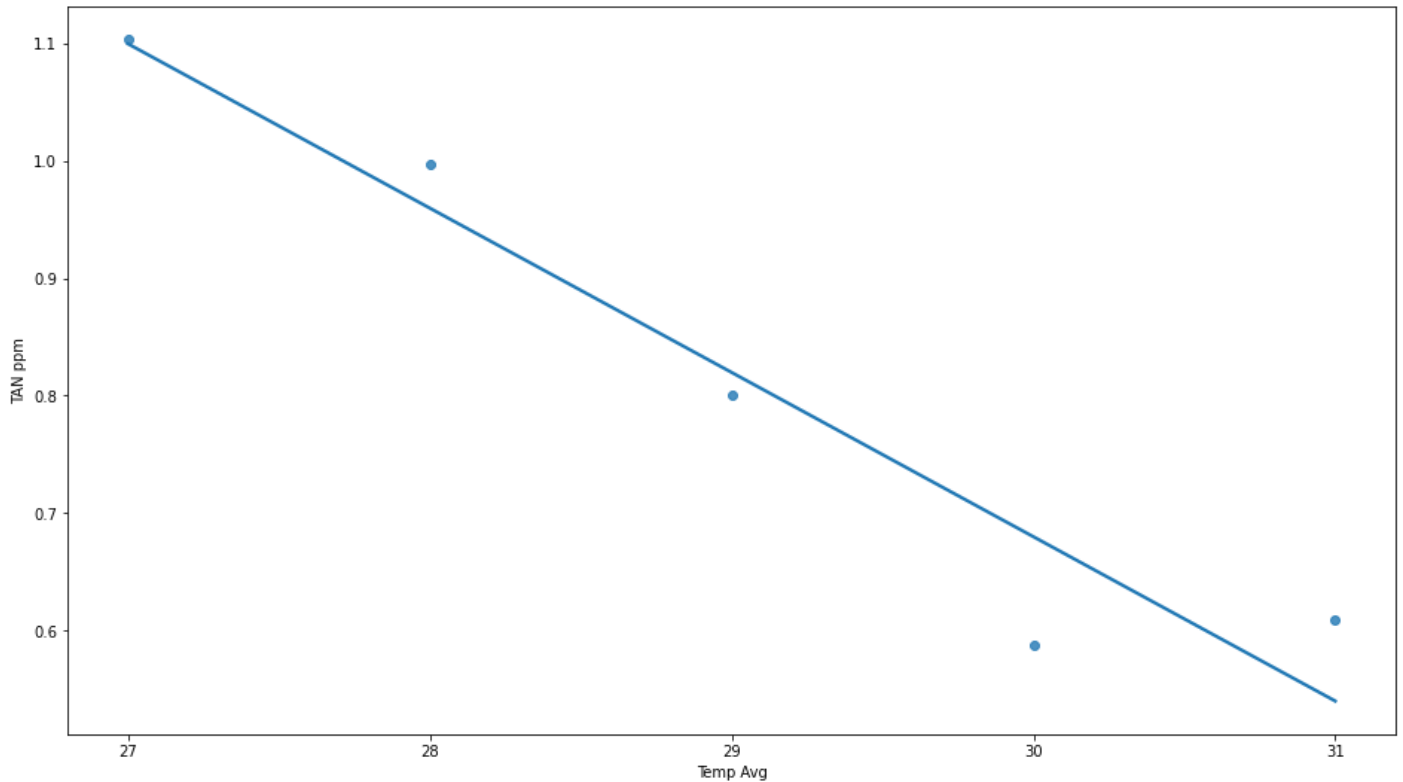


Days of Production vs Phytoplankton cells/ml



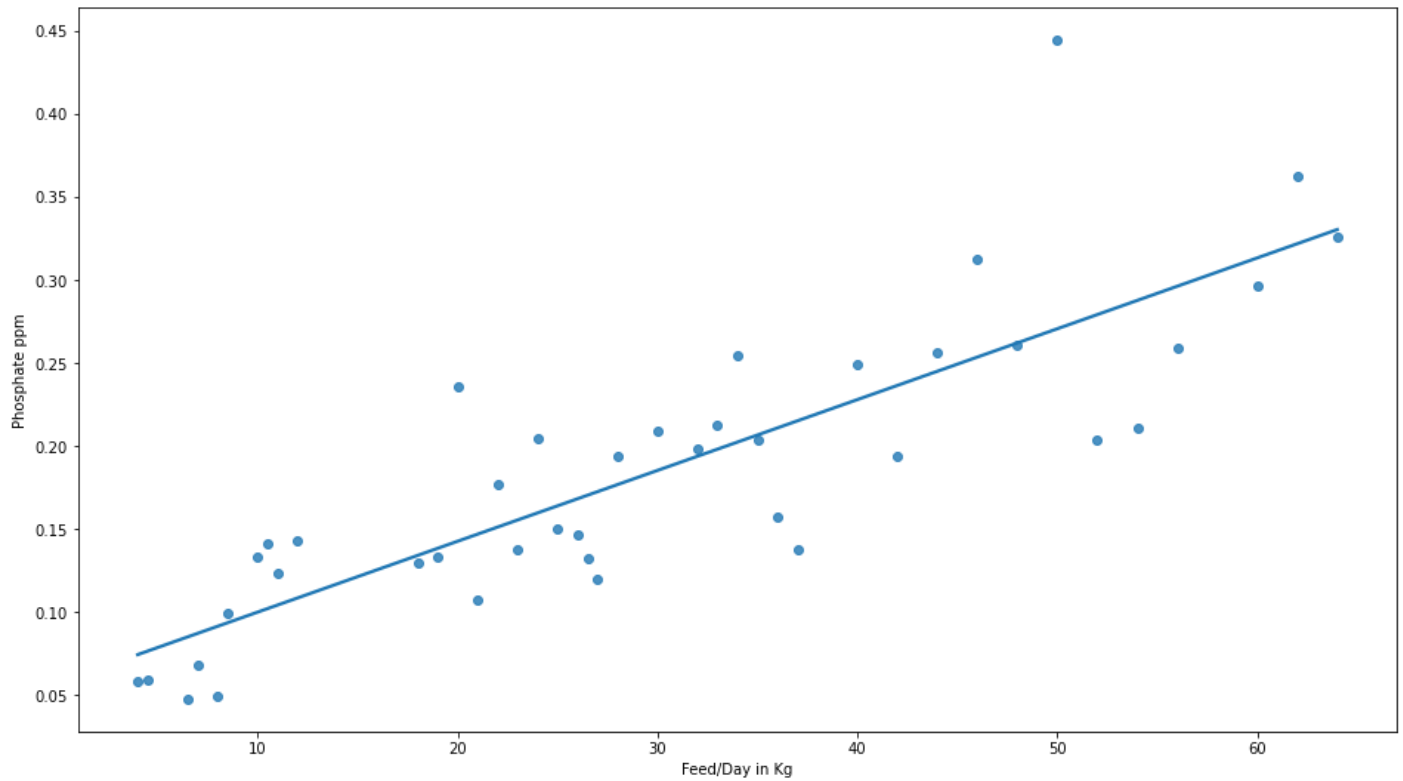
2. As temperature rises, 'TAN ppm' decreases.

Daily Temperature Average vs TAN ppm

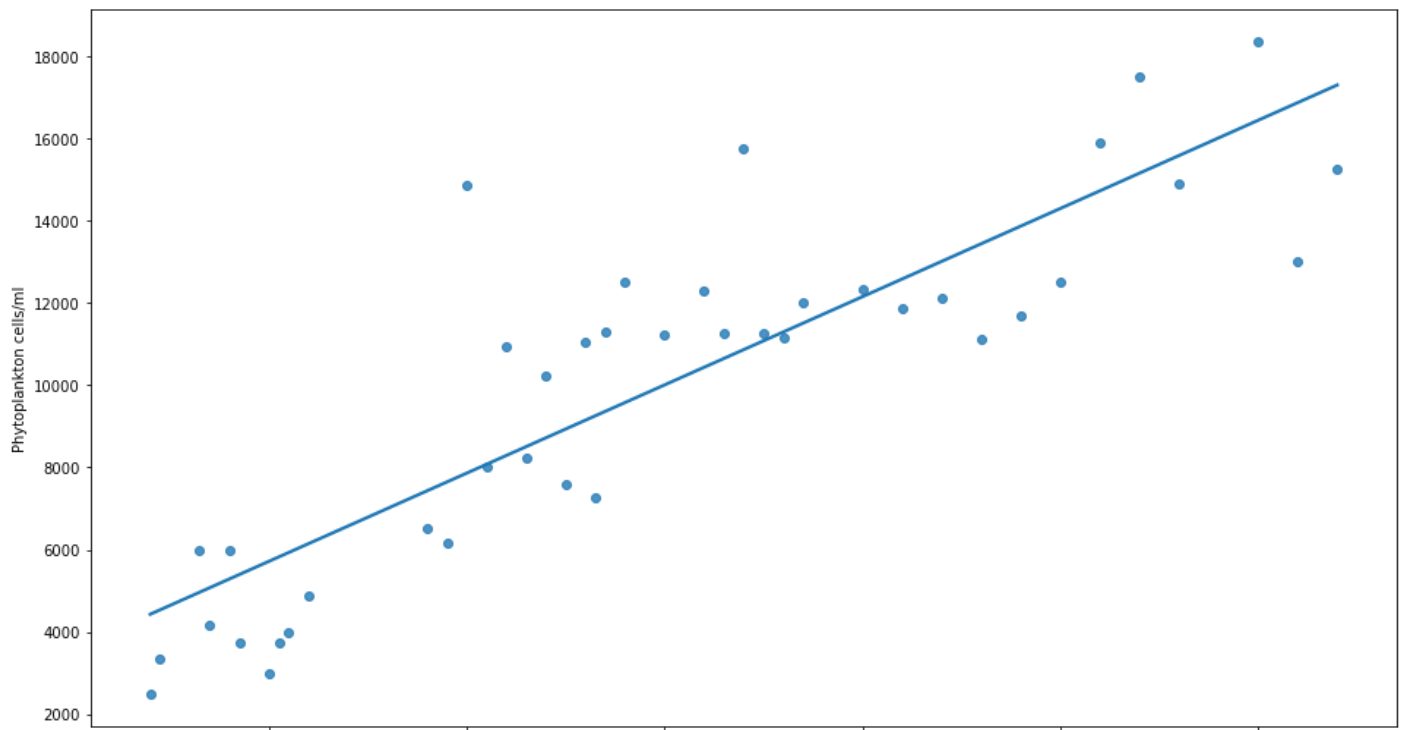


3. As the daily feed amount increases, 'Phosphate ppm' and 'Phytoplankton cells/ml' also generally increases.

Feed/Day in Kg vs 'Phosphate ppm'



Feed/Day in Kg vs 'Phytoplankton cells/ml'



Regarding 'Phosphate ppm' and 'Phytoplankton cells/ml', because Feed/Day increases as the production days goes, we can't be sure if the increase in Phosphate ppm and Phytoplankton cells/ml is caused due to either the production days or the feed amount or both.

More detail regarding analysis in python notebook.

3. If possible, record feeding amounts on each feeding session rather than taking the accumulations each day, also take parameters measurement on each temperature section (Morning, Afternoon, Evening) to see if changes in temperature throughout the day affect the parameters.

With Pond ID presented, i suspect that each pond have properties/conditions/treatment that might affect the parameters, if such data is recorded, further analysis on each pond can be done to further understand factors of changes in parameters.