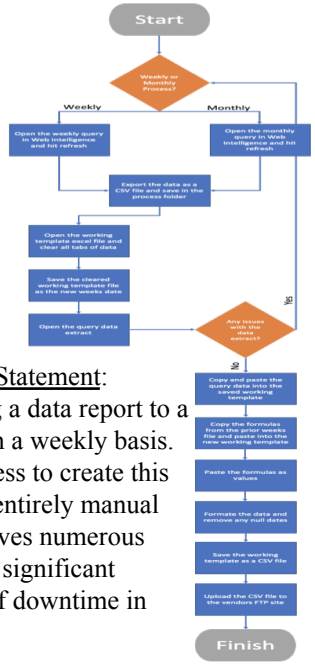
Pricetek Data Report

Process owner: Courtney Zimmer

Key Dates --->	Team Launch: April 2019	Define: April 2019	Measure: April - May 2019	Analyze: May 2019	Improve: May 2019	Control: Ongoing

DEFINE



Problem Statement: Providing a data report to a vendor on a weekly basis. The process to create this report is entirely manual and involves numerous steps and significant amount of downtime in between.

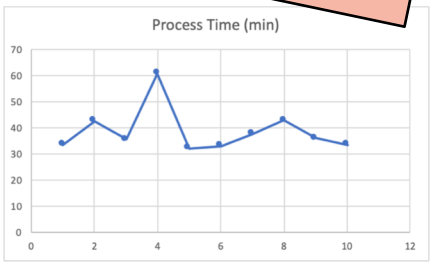
Business Impact: Streamlining this process would allow us to offboard this process to another team. Freeing up hours to focus on new reporting of the organization and over savings of $1,500/yr

MEASURE

SQL = 2.0

Mean	39
Median	36
Range	29
Std dev	8.7
Variance	74.4

Large Variance



ANALYZE

Alpha = .2

	M/Tu	W/Th/F	Total
Courtney	4	2	6
Andrew	1	3	4
Total	5	5	10

	f Observed	F Expected	(f-F)^2/F
Courtney + M/Tu	4	3	0.333
Andrew + M/Tu	1	2	0.500
Courtney + W/Th/F	2	3	0.333
Andrew + W/Th/F	3	2	0.500
		Chi-Square Value>>	1.667

df = (rows-1)*(columns-1) 0.197 << p value
df = 1

Using Chi-square we were able to prove that there is a relationship between report producer and time of the week the report was run.

$H_0 = \mu_1 < \mu_2$
$H_a = \mu_1 > \mu_2$

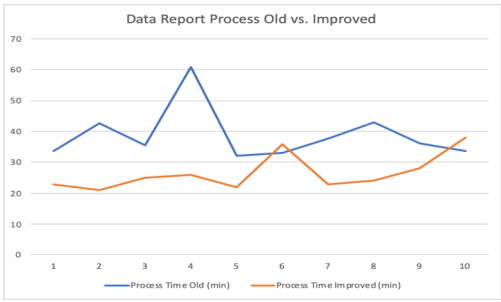
$$t_{data} = \frac{39-27}{\sqrt{\frac{8.7^2}{10}+\frac{5.9^2}{10}}} = 3.63$$

SQL = 2.75

df = 10+10-2 = 18
Alpha = 0.10
$t_{crit} = 1.33$

$3.63 \geq 1.33$ H_0 is rejected

IMPROVE

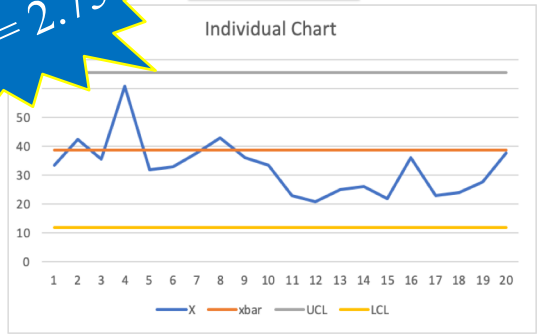


$x_{bar} = 27$	alpha = 0.1
s = 5.85	df = 19
n = 10	t = 1.729

$$Upper\ Bound = 27 + 1.729\left(\frac{5.85}{\sqrt{10}}\right) = 30.19$$

$$Lower\ Bound = 27 - 1.729\left(\frac{5.85}{\sqrt{10}}\right) = 23.8$$

CONTROL

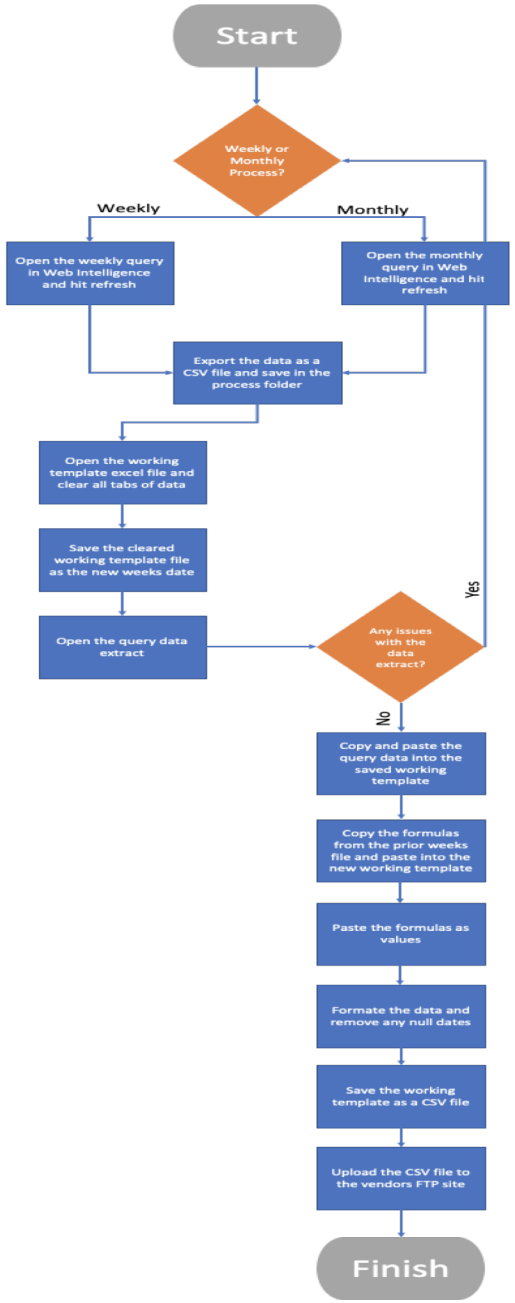


TEAM MEMBERS

Andrew (Business Analyst), Business Intelligence Team, & Anna (Supervisor)

Define

- My team is responsible for running a weekly data report for our vendor, that has not been evaluated or improved upon since it was initiated. As you can see from the process map, there are numerous manual steps involved and a significant amount of down time in between.
- My goal is to streamline the process of producing the weekly data report for our vendor, Pricetek.
- I will measure the success of my process improvement but a reduction in the overall amount of time needed to produce this report. This includes reducing the amount of downtime and working to eliminate small instances where the report produces an error and the report must be recreated.
- Because this is an existing process, the input and output must remain the same, but the process to get from raw data to report can be improved upon.



Measure

- Data Collected
  - Process time in min (continuous)
  - Who produced the report (discrete)
  - Day of the week the report was run (discrete)
  - Did you have to start the report over (discrete)
  - Total amount of downtime in between steps in min (continuous)
  - Number of steps (continuous)
  - Number of defects
- The data was collected by both myself and my co-worker. We maintained a spreadsheet with the various columns and the continuous data was collected using an online timer.
- In order to have a larger sample size, we ran the report twice a week for 5 weeks and obtained a sample size of 10.
- With a 80% confidence level and a 3 min margin of error, I calculated an ideal sample size of 14

$$n = \left(\frac{1.28 * 8.7}{3}\right)^2 = 14$$

- We used timers to measure both our process time and the total amount of downtime in between steps, but since we recorded the time ourselves, there is opportunity for measurement error. One way to measure measurement error could be for myself and my co-worker to separately measure running the same process and compare times. We could also use two separate timers and measure the variance between the two timers.

Measure

Data Stratification Tree

Question About the Process	Output	X Variables
Does the report process time vary by day of the week?	Y = Weekly Data Report	Report Producer
Does the process time change depending on who produces the report?		Day of the week
Does skill level have an impact?		Did the report have to be redone?
Is there an unnecessary amount of downtime in between steps?		Total amount of downtime
Did the process have an error?		Number of steps
Did the process have to be redone?		
Which step is the longest?		
Which step is the shortest?		
Number of training hours per month		

Analyze

I used chi-square to see if there was a relationship between who produced the report and when the report was run during the week. The calculated p value is 0.197, which is lower then our alpha of 0.2. This means that we reject the H_0 and report producer and time in the week the report is run are not independent. One issue with this test, is that with such a small sample size, I was not able to have more then 5 instances in any of my columns. In the future I would like to continue to gather sample data to confirm this relationship.

Alpha = .2

	M/Tu	W/Th/F	**Total**
Courtney	4	2	**6**
Andrew	**1**	3	**4**
Total	**5**	**5**	**10**

	f Observed	F Expected	(f-F)^2/F
Courtney + M/Tu	4	3	0.333
Andrew + M/Tu	1	2	0.500
Courtney + W/Th/F	2	3	0.333
Andrew + W/Th/F	3	2	0.500
		Chi-Square Value>>	1.667

df = (rows-1)*(columns-1)

df = 1

0.197 << p value

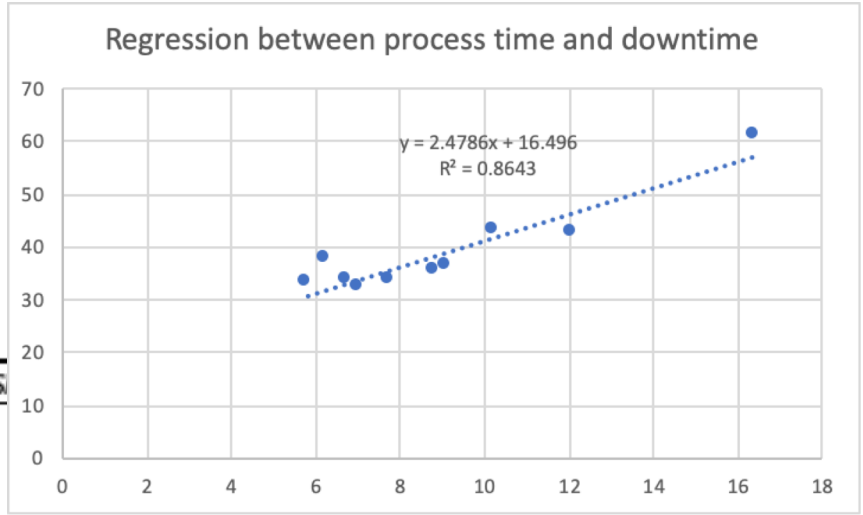
Analyze

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.92969525
R Square	0.86433326
Adjusted R Square	0.84737492
Standard Error	3.37985358
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*S*
Regression	1	582.228678	582.228678	50.9680269	
Residual	8	91.3872816	11.4234102		
Total	9	673.61596			



	Coefficients	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	16.4955026	3.3102142	4.98321306	0.00107516	8.86213498	24.1288702	8.86213498	24.1288702
X Variable 1	2.47855689	0.34717623	7.13918951	9.8129E-05	1.67796707	3.27914671	1.67796707	3.27914671

Next, I looked at the simple linear regression between my y (process time) and the amount of downtime recorded during the production of each report. First I graphed the two variables, in which you can see there does appear to be a correlation between them. The regression summary output, also tells me what the correlation coefficient, 0.929 is extremely close to 1 which tells me that there is a strong correlation between process time and downtime between steps. I can also look at the p-value of my x variable, just to be sure that this is an important variable to my process. What I took from this is that if I can reduce the overall number of steps in my process, it would help decrease the amount of downtime and help to improve my process overall.

Analyze

$H_0 = \mu_1 < \mu_2$ $\quad t = \frac{x_{bar1} - xb_{ar2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

$H_a = \mu_1 > \mu_2$

$$t_{data} = \frac{39-27}{\sqrt{\frac{8.7^2}{10} + \frac{5.9^2}{10}}} = 3.63$$

df = 10+10-2 = 18
Alpha = 0.10a
t_{crit} = 1.33

One of the tools I used was the Hypothesis Test. I used the mean of the process time from the original data collected and then from my improved process. Because my sample size was so small, I chose to use the t test and an upper-right tail hypothesis test. The result of my data the H_0 is rejected because $3.63 \geq 1.33$. This gives us the confidence that we really did make improvements to our process and that it not just a fluke in the data.

$x_{bar} = 27$	alpha = 0.1
s = 5.85	df = 19
n = 10	t = 1.729

$$Upper\ Bound = 27 + 1.729\left(\frac{5.85}{\sqrt{10}}\right) = 30.19$$

$$Lower\ Bound = 27 - 1.729\left(\frac{5.85}{\sqrt{10}}\right) = 23.8$$

Using the confidence interval formula, we calculated with 90% confidence that the population mean would be in between the Upper and Lower boundaries

$$23.8 \leq \mu \leq 30.19.$$

Analyze

Some takeaways from the various test I ran on my data. I was fearful that I wouldn't have a large enough sample size, to really be able to analyze my data. Being able to calculate the sample size needed for 80% confidence and a small margin of error and having that sample size be very close to what I was able to gather, gave me the confidence in my sample. The regression analysis told me a lot about the correlation between process time and downtime. Looking at my process map, I already knew that their was too many steps in my data, but I know strongly believe that in reducing the number of steps, I can reduce the amount of downtime and that would allow me to improve my process time.

Sigma Quality level for the original process

$$\frac{6}{(2 * 10)} = 0.3 * 1{,}000000 = 300{,}000$$

SQL = 2.0

Sigma Quality level for the improved process

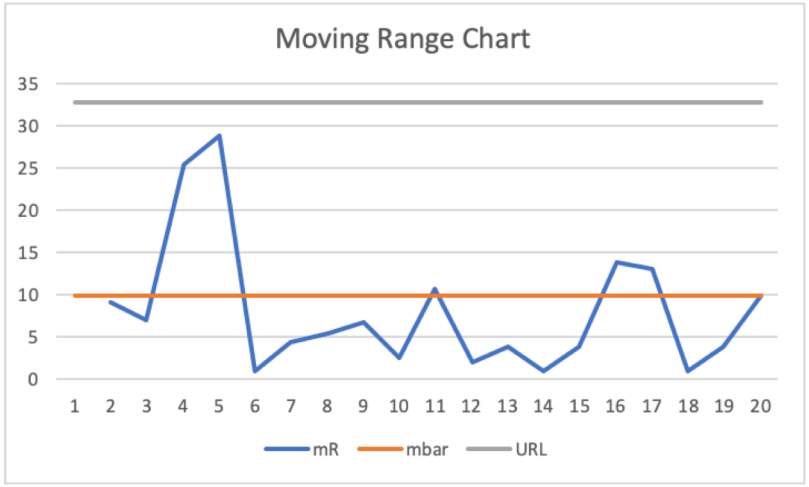
$$\frac{2}{(2 * 10)} = 0.1 * 1{,}000000 = 100{,}000$$

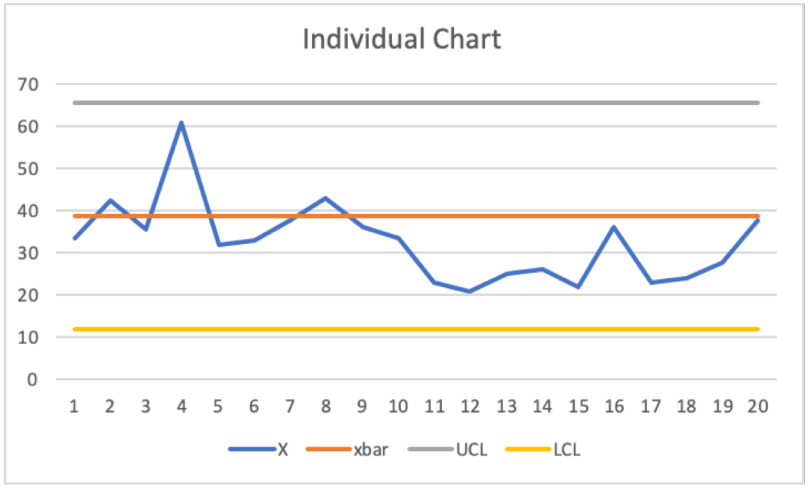
SQL = 2.75

Improve

1. The first improves I made were to the number of steps involved in producing the report. After querying the raw data, we manually format the data within excel, which I felt was an unnecessary step. Working with our business intelligence team, they were able to make some changes to the query that allowed the date to be changed into the correct format. They also created some custom variables, that removed default null values and replaced them with blanks, this was done to 3 data fields. This automation allowed us to remove 4 steps from our process.
2. One of the steps in our process is downloading the data and putting it into a template file. This is where a large part of the downtime is recorded, because it takes a significant amount of time to export our data and transfer it to the template file. Working with my co-worker, Andrew, we worked on rearranging the query so that the data would be exported in the same format at the template.
3. Another improvement we made, was more robust training. After updating the query and eliminating some formatting steps, we wrote a new process document that documented the best way to perform each step to maximum efficiency. We also made the suggestion to only run the report on Monday, as earlier in the week tend to have the fastest process time and prevented the producer from having unnecessary data in the report.
4. One improvement that we as a team still need to work on, is eliminating the potential for errors in the data export that requires the entire process to be redone. We are still unsure of what causes this error, which is a duplication of the data in the export, so there are still spikes in our data processing time due to this.

Control





Looking at my moving range chart, I can see that my process was and still is in control. There are no data points outside of my control limit lines. Looking at the individual chart, I can clearly see that my process time is trending downwards after completing my process improvements. This downward trend is exactly what I wanted to see, and I'm happy with these results. There are still some improvements to be made to diminish the larger spikes in my process time. However, the improvements and training that we've put in place have made noticeable improvements and I will continue to measure the control of my process.