Process Improvement Project – Improve Health Condition

Process owner: Bing-Je Wu

Key Dates --->	Project Launch July 15, 2019	Define July 15, 2019	Measure July 15, 2019	Analyze August 13, 2019	Improve August 13, 2019	Control August 30, 2019

DEFINE

Problem Statement

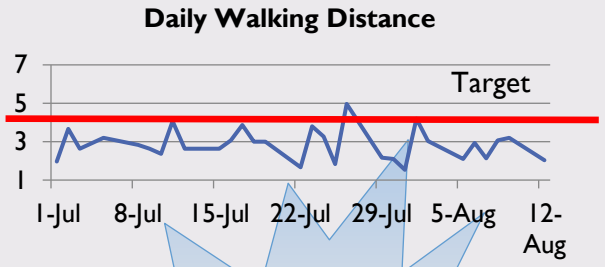
Health condition signal turns **red** !!

Business Impact

Payment on doctors and medicines for any sickness worth at least $300 per day.

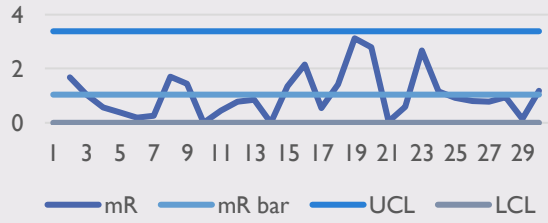
Defects appear many times in the process.
Current SQL = 0.2

MEASURE



Daily walking distances vary and are consistently below the target , 4 miles per day. The process is in control but not meeting specs.

IMR Chart before Improvement



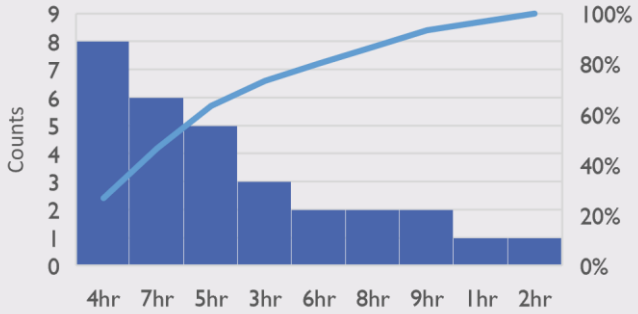
ANALYZE

n =	30
s =	0.81
$\bar{x}$ =	2.84
$Z(\frac{\alpha}{2})$ =	1.96
α =	0.05
Lower Bound	2.55
Upper Bound	3.13

95% confidence interval that the true walking distance is between 2.55 miles and 3.13 miles.

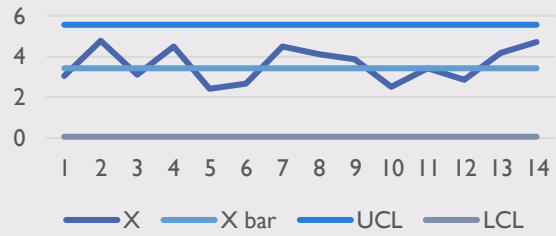
Pareto shows that 80% of times are at 3 to 4 active hour categories and some are at 7 active hour category.

Pareto Chart on Active Hour



IMPROVE

Individual Chart after Improvement



The process is in control. There are more spikes above the target (4 miles per day) and below the UCL.

Improved SQL = **1.3**
Average walking distance after improvement is **3.84** miles per day

CONTROL

- Stay active during the days.
- Keep the alarm reminders on in the weekends.
- Incorporate different exercises during the workout.

Project Team: Bing-Je Wu

DEFINE

To improve my health condition, I need to be more active, spending time on exercise. According to Centers for Disease Control and Prevention's recommendation, an adult gets at least 150 minutes of moderate-intensity aerobic physical activity each week. Walking 4 miles a day includes about 30 minutes of daily exercise. Therefore, My goal is to walk 4 miles per day. I will record my daily walking distance from my wearable device, Fitbit watch, every morning. The scope will be from Monday to Friday, except holidays. And the process starts from leaving at home to arriving at home. Two weeks of data will be used from history.

The variables that affect the goal are weathers, day of the week, how much water do I drink, going out for lunch, parking far, taking a lunch time walk, workout, and taking elevators, etc.

Success will be measured by whether I hit my goal of 4 miles per day or not. The business impact would be to reduce the risk on going to doctors and buying medicines for any sickness which worth at least $300 per day.

MEASURE: DATA COLLECTION METHOD

- I have been wearing the wearable devices for almost three years. I used the app report to see how active I am. The app report offers the information such as number of steps I took, how many floors I have climbed, distance of walking, active hours.
- Every morning, I will sync my watch to upload the data I have from yesterday. Once I have the report, I will manually log in the data into an excel sheet and record other variables, such as, weather, day of the week, parking, millimeter of water, workout, lunch walk, eat out etc.
- Weekend is not considered in the scope.
- There are 9 days of data are history data.

MEASURE: DATA MEASUREMENT PLAN

Data Stratification Tree

Questions About Process	Stratification factors X Variables	Measurements
Do I park the car far from building entrance?	Weather	Rain or not rain
What is the weather?	Day	Monday, Tuesday, Wednesday, Thursday, Friday
What day is today?	Water	Milliliter of water
Do I walk after lunch?	Lunch	Go out for lunch or not
How often do I go to the restroom?	Park far from entrance	Park far or not park far
How much water do I drink per day?	Walk	Walk after lunch or not
How often do I workout?	Steps	Number of steps
Do I order a lunch or bring it from home?	Workout	Workout or not
How many steps do I have per day?	Floors	Number of floors
Do I take stairs or elevator?	Hourly Activity	Number of hours being active
How long do I sit on my chair		

Output Y

Distance

MEASURE: VARIABLES AND SAMPLE SIZE

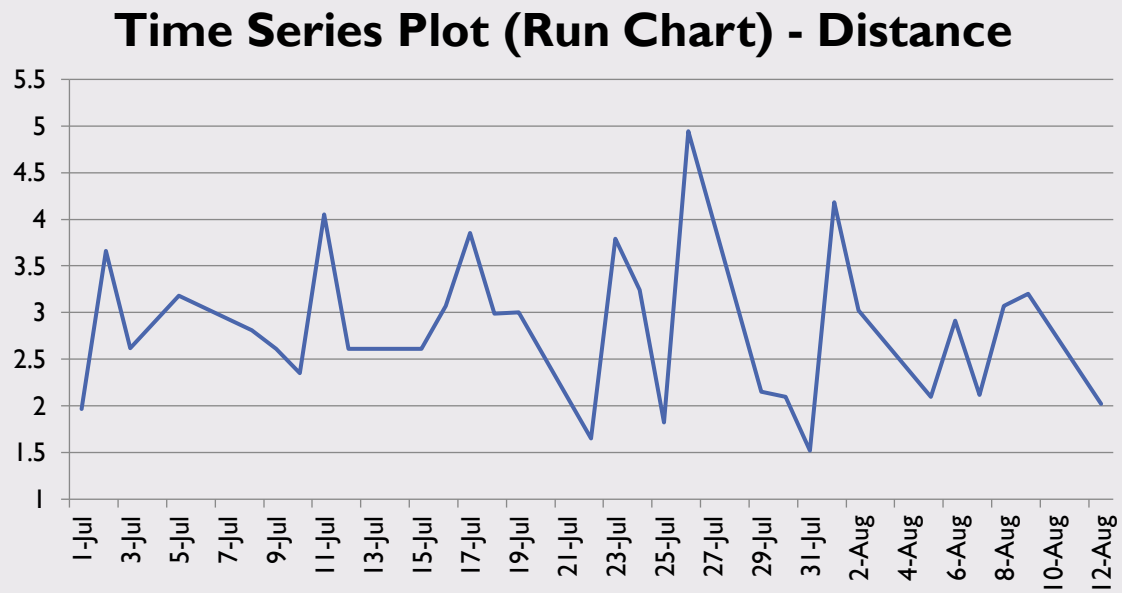
- My data contain discrete variables and continuous variables. The discrete variables are steps, floors, Rain or not., eat out or not, park far or not, lunch walk or not, workout or not, number of active hours. The continuous variables are the distance, milliliter of water, .
- I used some of existing data and I collected my own data.
- I collected 30 days of data in total.
- Sample mean = 2.84 ; sample standard deviation = 0.81
- I set my acceptable margin of error at 0.3 miles. By using the sample size equation:
- $n = \left(\frac{Z(\frac{\alpha}{2})(\sigma)}{E}\right)^2 = \left(\frac{1.96(0.81)}{0.3}\right)^2 = 28.01$
- The risk occurred if I collected too few samples and it will cause a larger margin of error.
- The actual margin of error with 30 samples, a standard deviation of 0.81 and an alpha level of 0.05:
- $E = \frac{Z(\frac{\alpha}{2})(\sigma)}{\sqrt{n}} = \frac{1.96(0.81)}{\sqrt{30}} = 0.29.$
- The actual margin of error is lower than my acceptable margin of error.

MEASURE: MEASUREMENT ERROR

- I may have measurement error in defining the day is rain day or not, depending on the day I logged in the data based on my memories on how severe the rain was. The risk of the measurement error is that the rain variable may not able to provide sufficient evidence that the performance of the output y, distance, is effected by the independent variable, rain. To reduce the risk of the measurement, I should have logged in the data for the rain variable in that day instead of relying on memories.

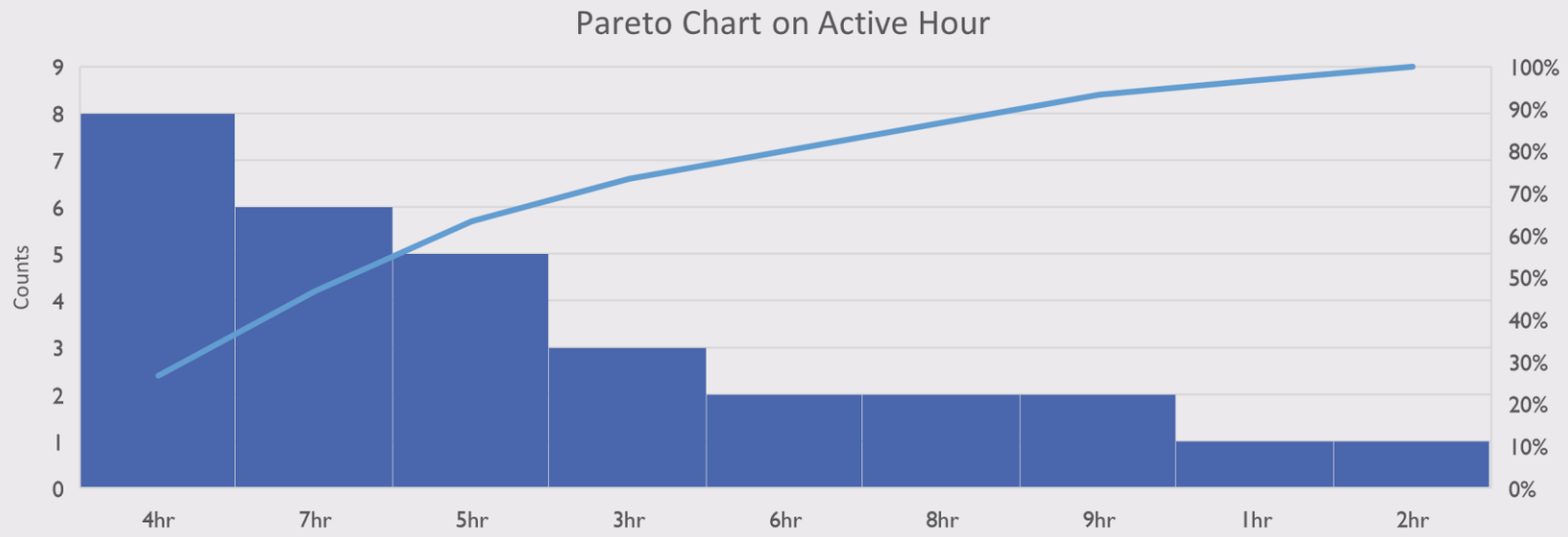
ANALYZE: SQL

Calculate SQL	Value
Defect opportunities per unit =	1
Units produced per timeframe	30
Total possible defects per timeframe	30
Actual Defects	27
Defects Per Opportunity =	0.9
Defects Per Million Opportunities (DPMO)	900,000
SQL value (from SQL table)	0.2



With the above table and chart, we can observe that my old process is severely broken, and an improvement is required immediately. I have been missing my target goal 90% of the time, and the target achieve rate is merely 10%. From the run chart, it is easy to see that there are clear spikes in time when I hit the goal (4 miles) and there are spikes below the goal, which provide further evidence of a poor process.

ANALYZE : PARETO CHART



I tool a look at the active hours on each day to see how active I am in majority of time. From the Pareto Chart above, I can see that 20% of time I stay active as 4 hours per day. This could be the case that I failed to hit my target goal (4 miles) during the past 30 days. In order to hit the target, I should be more active. Maybe take a walk around the office every hour and refill water at the same time can improve my output, walking distance.

ANALYZE: ONE SAMPLE RIGHT-TAILED HYPOTHESIS TEST

The average daily walking distance in 30 days before improvement was 2.84 miles per day. Sample standard deviation was 0.8 miles.

Ho: $\mu \leq 4$ miles

Ha: $\mu > 4$ miles

I used a z-test to calculate p because $n = 30 \geq 30$.

P(z < -7.84)
= NORM.S.DIST(-7.84,TRUE)
= 0
P (z > -7.84)
= NORM.S.DIST(7.84,TRUE)
= 1

n	30
s =	0.81
$\bar{x} =$	2.84
μ_0(*target distance*) =	4
z =	-7.84
p =	1
α =	0.05

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{2.84 - 4}{0.81/\sqrt{30}} = -7.84$$

P(z > - 7.84) = 1 – P (z < -7.84) = 1

P-value > 0.05. Fail to reject the null hypothesis. There is no enough evidence to prove that my average daily walking distance is more than 4 miles. It also supports the poor SQL value I have previously to indicate my process needs to be improved.

ANALYZE: CONFIDENCE INTERVAL OF DAILY WALKING DISTANCE

Lower Bound = $\bar{x}$ - E

$= \bar{x} - Z(\frac{\alpha}{2})\frac{(\sigma)}{\sqrt{n}}$

$= 2.84 - 1.96\left(\frac{0.81}{\sqrt{30}}\right)$

$= 2.84 - 0.29$

Upper Bound = $\bar{x}$ + E

$= \bar{x} + Z(\frac{\alpha}{2})\frac{(\sigma)}{\sqrt{n}}$

$= 17.05 + 1.96\left(\frac{0.81}{\sqrt{30}}\right)$

$= 2.84 + 0.29$

n =	30
s =	0.81
$\bar{x}$ =	2.84
$Z(\frac{\alpha}{2})$ =	1.96
α =	0.05
Lower Bound	2.55
Upper Bound	3.13

My population mean with confidence level of 95% on daily walking distance is between 2.55 miles and 3.13 miles. There is a lot of room for improvement to hit my goal of 4 miles per day.

ANALYZE: MULTIPLE LINEAR REGRESSION

Variable	Variable Name	Coefficients	P-value
	Intercept	1.061387586	0.009161
x_1	Number of Floors	0.079404852	0.000086
x_2	Rain (1) or Not Rain (0)	0.352552437	0.119534
x_3	Millimeter of Water drinking	-0.000054981	0.809613
x_4	Park Far (1) or Not Park Far (0)	0.176751267	0.411428
x_5	Lunch Walk (1) or No Lunch Walk (0)	0.558974290	0.031852
x_6	Workout (1) or Not Workout (0)	0.713794444	0.003920
x_7	Number of active hours	0.106955050	0.022795

Regression Statistics	
Multiple R	0.875505153
R Square	0.766509274
Adjusted R Square	0.692216771

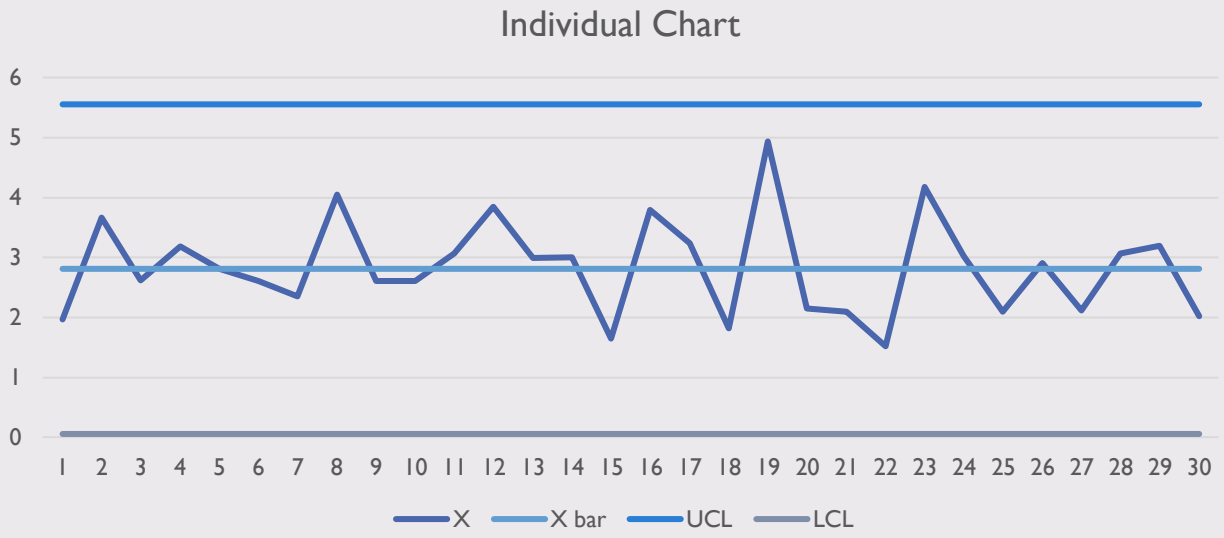
Regression Equation: $\hat{y} = 1.061 + 0.079x_2 + 0.352x_2 - 0.00005x_3 + 0.176x_4 + 0.558x_5 + 0.713x_6 + 0.106x_7$

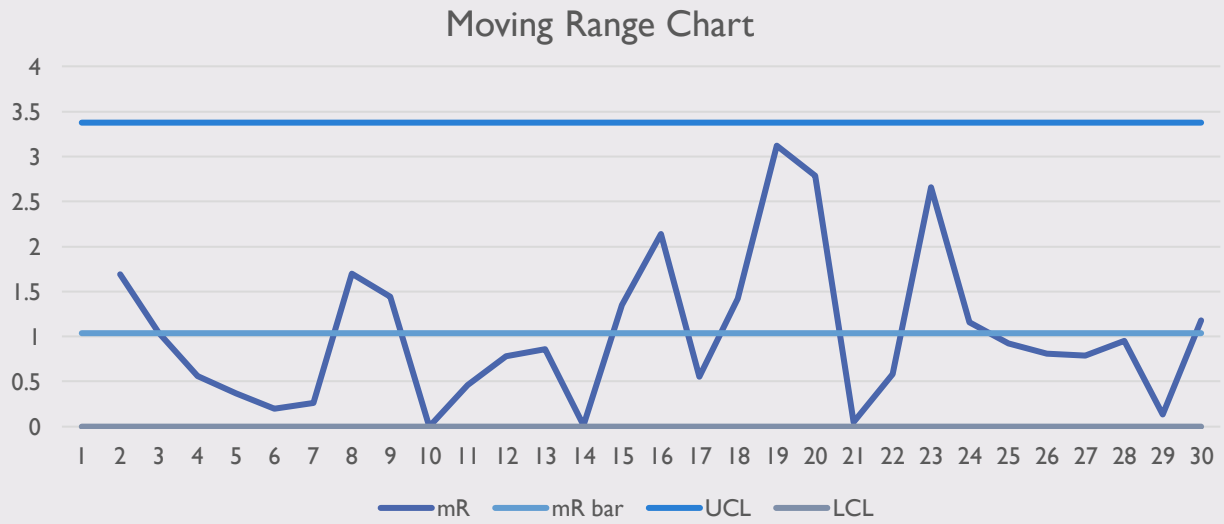
A Multiple Linear Regression has been run with the variables that I believe it has the largest impact in output, daily walking distance.

As we can see the p-value values on each variable. Three variables have p-value greater than 0.05. One is Rain variable; another is Milliliter of Water variable and the other is Parking variable. Those variables are not significant enough to have impact on my output, daily walking distance. Additionally, the Multiple Regression Model has adjusted R square at 0.69. It means 69% of variation in dependent variable can be explained by the independent variables. The model is not very convincing since it does not perfectly fit the data and has other factors that are not considered into account.

Although the Water variable shows no statistical significance, I will still use this information in my Improvement.

ANALYZE: CONTROL CHARTS





By looking at my control charts, I noticed that my process does not have out-of-control signals. The Individual Chart shows no seven or more consecutive points above or below the center line as a sawtooth pattern. However, the Moving Range Chart has many dramatic shifts. It indicates that I may need to review and try to figure out why.

IMPROVE

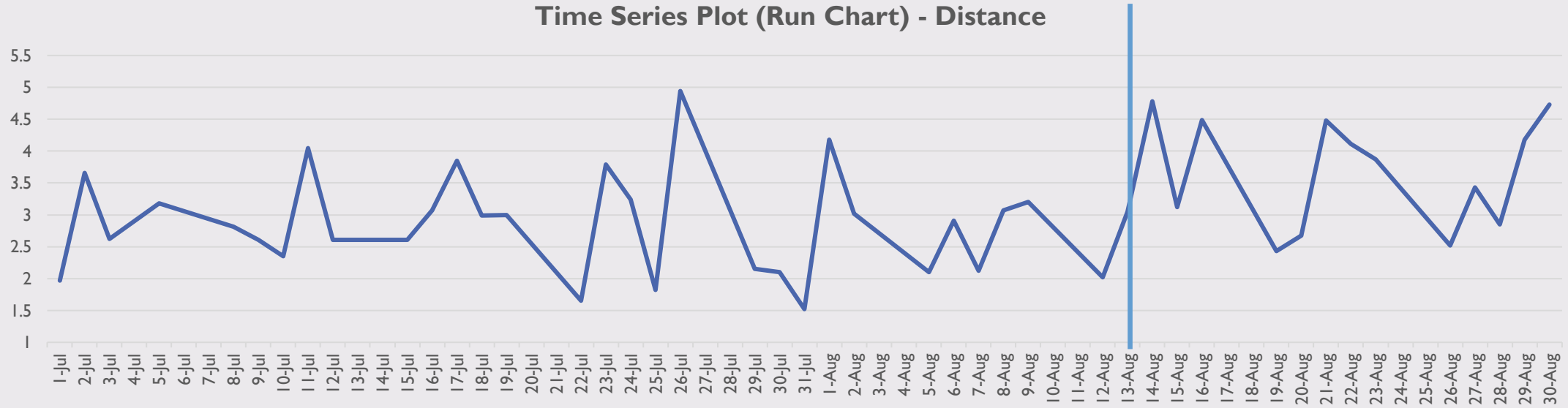
The improvements I made to my process are as follows:

- Although the Multiple Linear Regression shows no statistically significant on Water variable, I decide to increased the amount of water I drink everyday to increase the change to go to restroom and be more active.
- Set up a reminder each hour to remind me to take a walk around our office.
- Taking stairs and not taking elevators no matter what.
- Found a workout buddy to go to the gym together three times a week.
- Increased the exercise duration time from 45 minutes to an hour.

IMPROVED SQL

My SQL improved from 0.2 to 1.3. There is still much improvement to be made. Looking run chart below, we can see that after the improvement on August 13th, the chart has more spikes over the 4 miles threshold and spikes are more intense than measurement phase. This is a huge improvement on my process.

Calculate SQL	Value
Defect opportunities per unit	1
Units produced per timeframe	14
Total possible defects per timeframe	14
Actual Defects	8
Defects Per Opportunity	0.571428
Defects Per Million Opportunities (DPMO)	571,428
SQL value (from SQL table)	1.3



IMPROVE: RIGHT TAILED HYPOTHESIS

The average daily walking distance in 14 days after improvement was 3.84 miles per day. Sample standard deviation was 0.85 miles.

Ho: $\mu \leq 4$ miles

Ha: $\mu > 4$ miles

A *t* test was used because $n = 14 < 30$

P(t > -0.704)
= T.DIST.RT(-0.704, 13)
= 0.753

$\bar{x} =$	3.84
$\mu_0 =$	4
$s =$	0.85
$n =$	14
$df =$	13
$t =$	-0.704
$\alpha =$	0.05
$p =$	0.753

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{3.84 - 4}{0.85/\sqrt{14}} = -0.704$$

P(t > - 7.94) = 0.753

P-value = 0.753 > 0.05. Fail to reject the null hypothesis. There is no evidence to suggest that my daily walking distance is greater than 4 miles. The P-value is smaller than the one before the improvement. We can tell from that the process is improving and it echoes in the improved SQL score of 1.3 from 0.2.

IMPROVE: CONFIDENCE INTERVAL OF WALKING DISTANCE

Lower Bound = $\bar{x}$ - E

$= \bar{x} - Z(\frac{\alpha}{2})\frac{(\sigma)}{\sqrt{n}}$

$= 3.84 - 1.96\left(\frac{0.85}{\sqrt{14}}\right)$

$= 3.84 - 0.45$

Upper Bound = $\bar{x}$ + E

$= \bar{x} + Z(\frac{\alpha}{2})\frac{(\sigma)}{\sqrt{n}}$

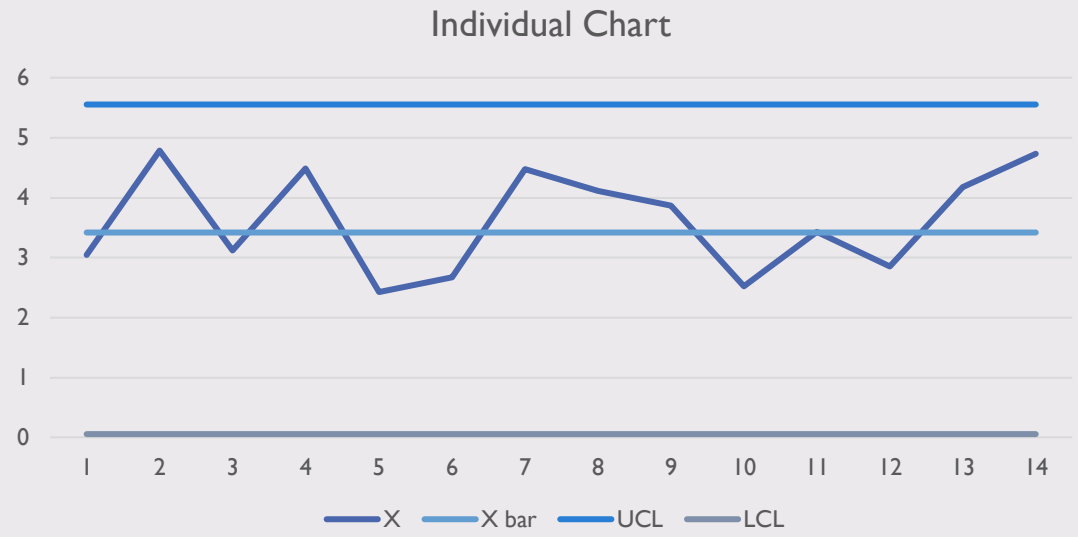
$= 3.84 + 1.96\left(\frac{0.85}{\sqrt{14}}\right)$

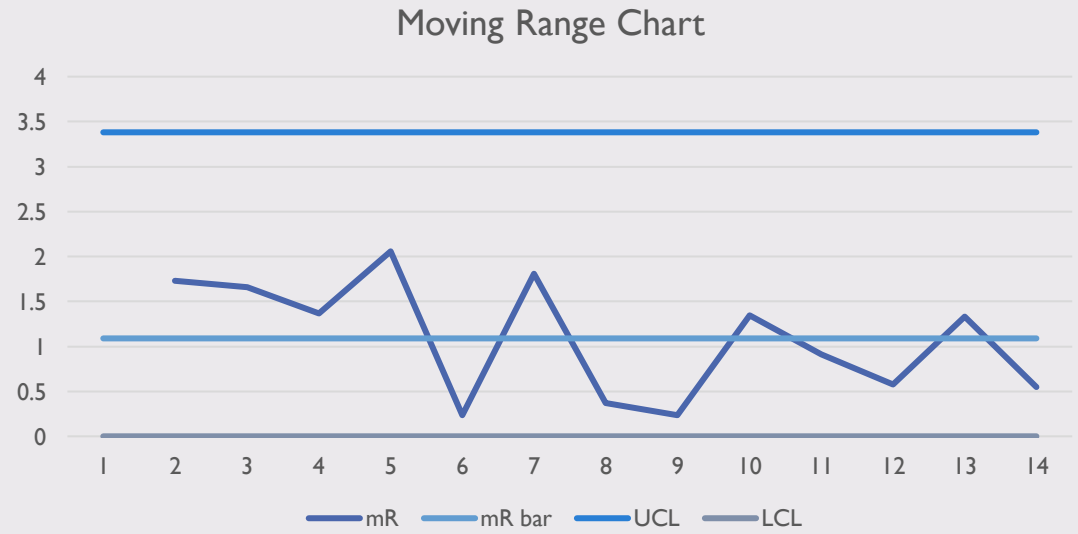
$= 3.84 + 0.45$

n =	14
s =	0.85
$\bar{x}$ =	3.84
$Z(\frac{\alpha}{2})$ =	1.96
α =	0.05
Lower Bound	3.39
Upper Bound	4.29

My population mean with confidence level of 95% on daily walking distance is between 3.39 miles and 4.29 miles. There is still a room for improvement to hit my goal of 4 miles per day. However, my target goal is well within the confidence interval which is a vast improvement from the start of the project.

IMPROVE: CONTROL CHART RESULTS





After implementing the improvements, the process is in control. Comparing with the chart before improvement, the spikes on the Individual Chart are wider. The shifts on Moving Range Chart are moderated. Both Individual Chart and Moving Range Chart are within the upper and lower control limits.

CONTROL

I have improved my process by using many statistical tools and techniques. My average walking distance has been improved to 3.84 miles per day which is close to my target goal 4 miles per day.

In order to keep meeting my goal and increase the average of daily walking distance, I will utilize the information from the analysis to continue stay active during the days. I will also leave the alarm reminders on in the weekends. Then, my body will get used to the time to move.

Incorporate different exercises instead of running at treadmill along. In this case, it not only can increase the duration of workout time but also train the muscles on other part of body.