QN Immediate Fix Cycle Time Analysis

Raisa Huang

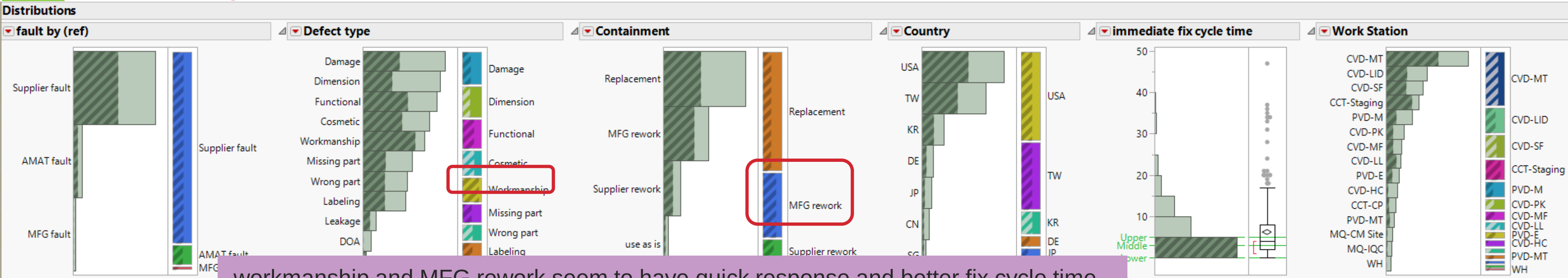
AGENDA

- Root Cause Analysis of QN Fix Cycle time
- Graphical Root Cause Analysis Summary
- Compare Fit Model, Partition, Neural Model
- Hybrid Text Mining & Data Mining Analysis
- Take Away Learnings

Histogram – 1st Layer of Root Cause Analysis of QN Fix Cycle time

- What scenarios impact on QN fix cycle time? The impact is endurable?
  - Criteria: within 5 days (In spec, success analysis); over 5 days (out of spec ,failure analysis)
  - **Use Histogram Conditional Mosaic Plot to conduct both Success Analysis and Failure Analysis**

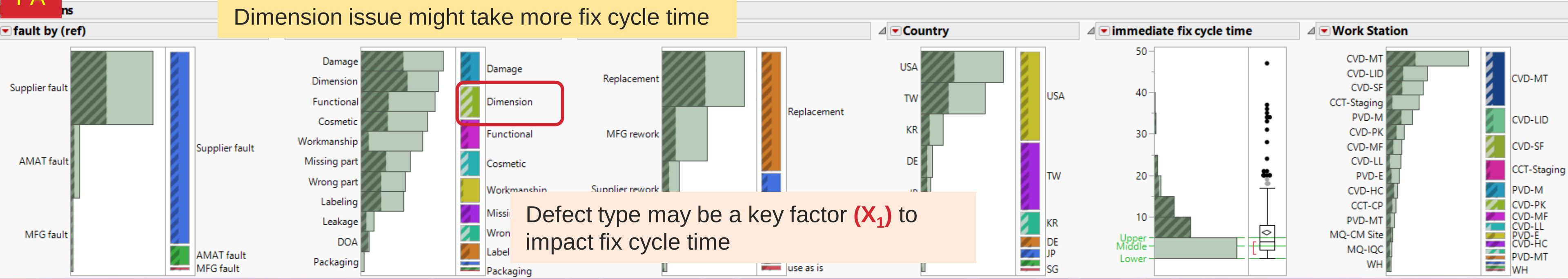
SA



workmanship and MFG rework seem to have quick response and better fix cycle time

FA

Dimension issue might take more fix cycle time



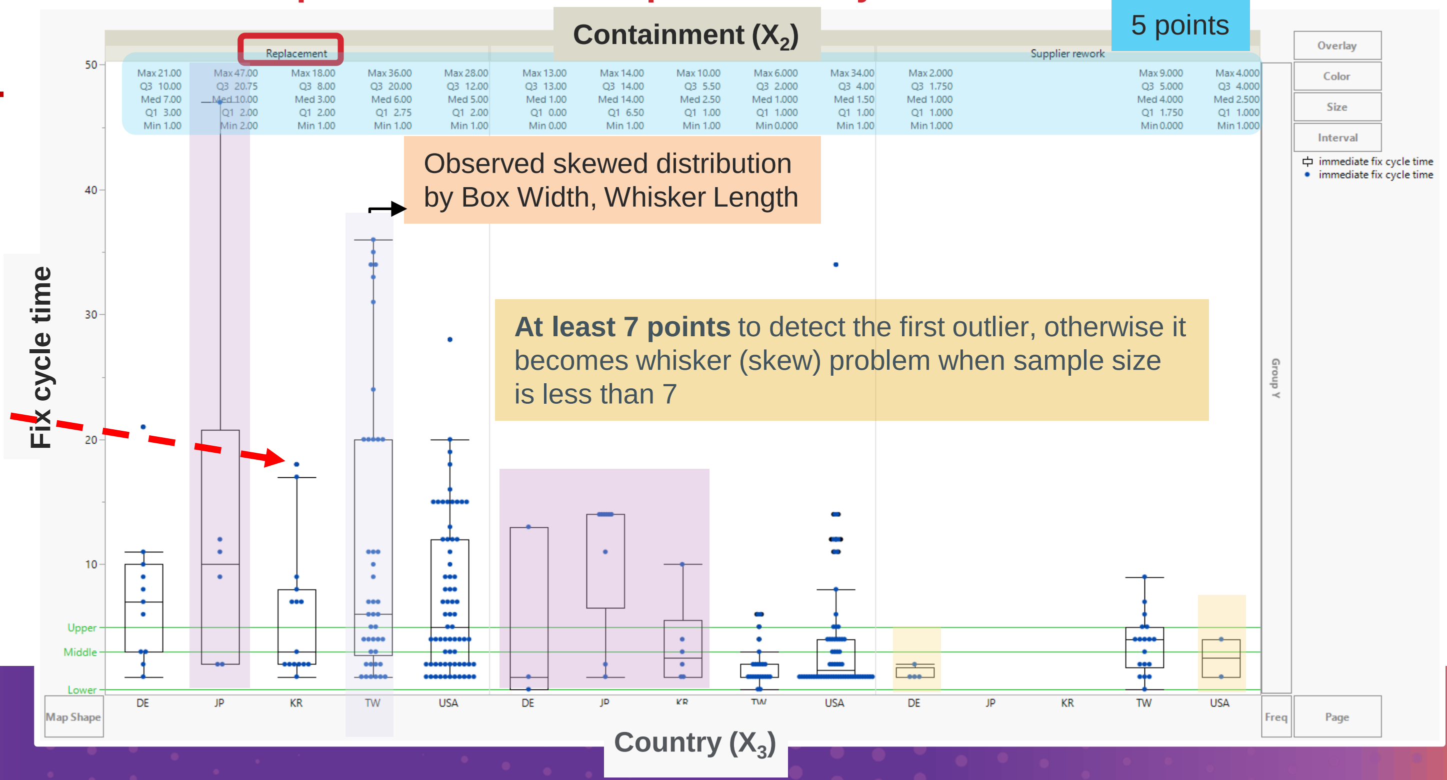
Defect type may be a key factor **(X_1)** to impact fix cycle time

Graph Builder Box Plot – 2^{nd} Layer of Root Cause Analysis

- Plot continuous fix cycle time vs. **nested structure**(categorical country X_3 under containment X_2)
  - » **The cycle time of Replacement is much longer than other containment actions**
  - » **Containment should be one of important factors to impact on fix cycle time**

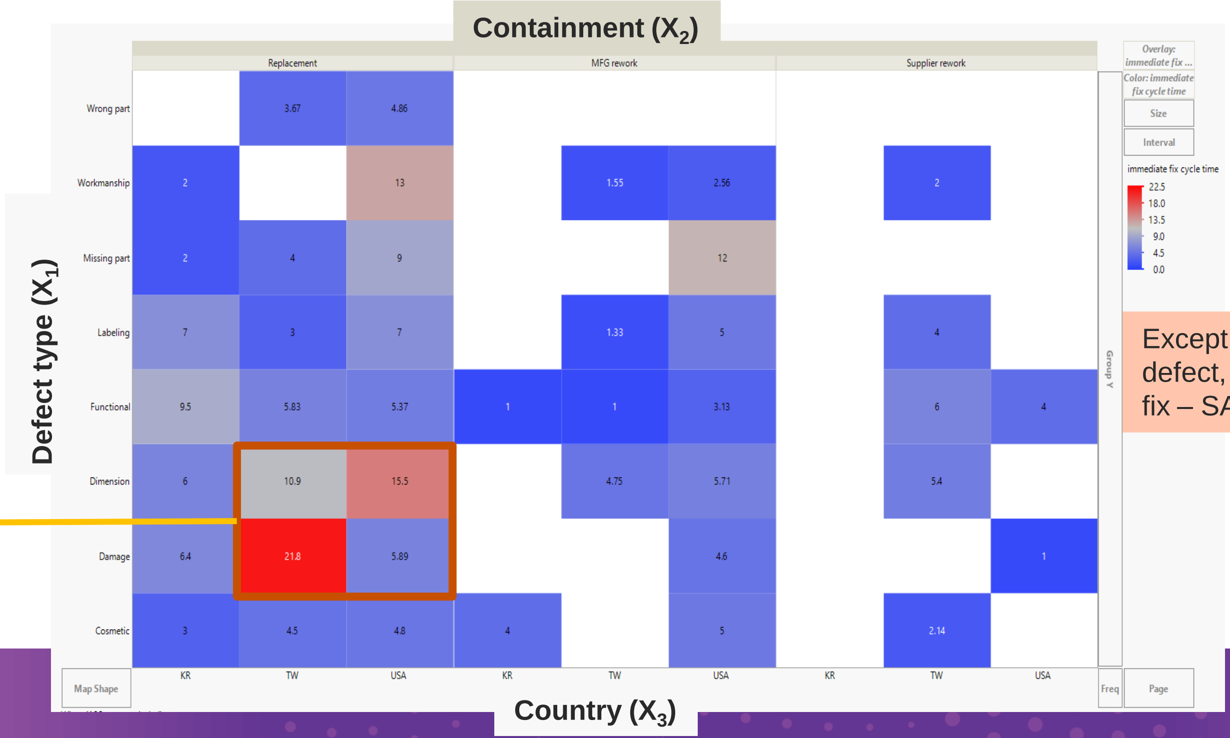
- Box plot is a **non-parametric** tool to use Median as central tendency

- How to handle **marginal outliers** which are within 2 σ GRR noise from the whisker?



Graph Builder Heatmap – 3rd Layer of Root Cause Analysis

- Add **back categorical** “**defect type** (X_1) ” on Y axis, color for fix cycle time
- Use 8 x 9 layout (**balanced**) to quickly catch out the max / min cycle time scenarios
  - **Replacing TW Damage** parts is the worst case for cycle time
  - **Replacing USA Dimension** issue parts is the 2nd worst scenario.



FA inference

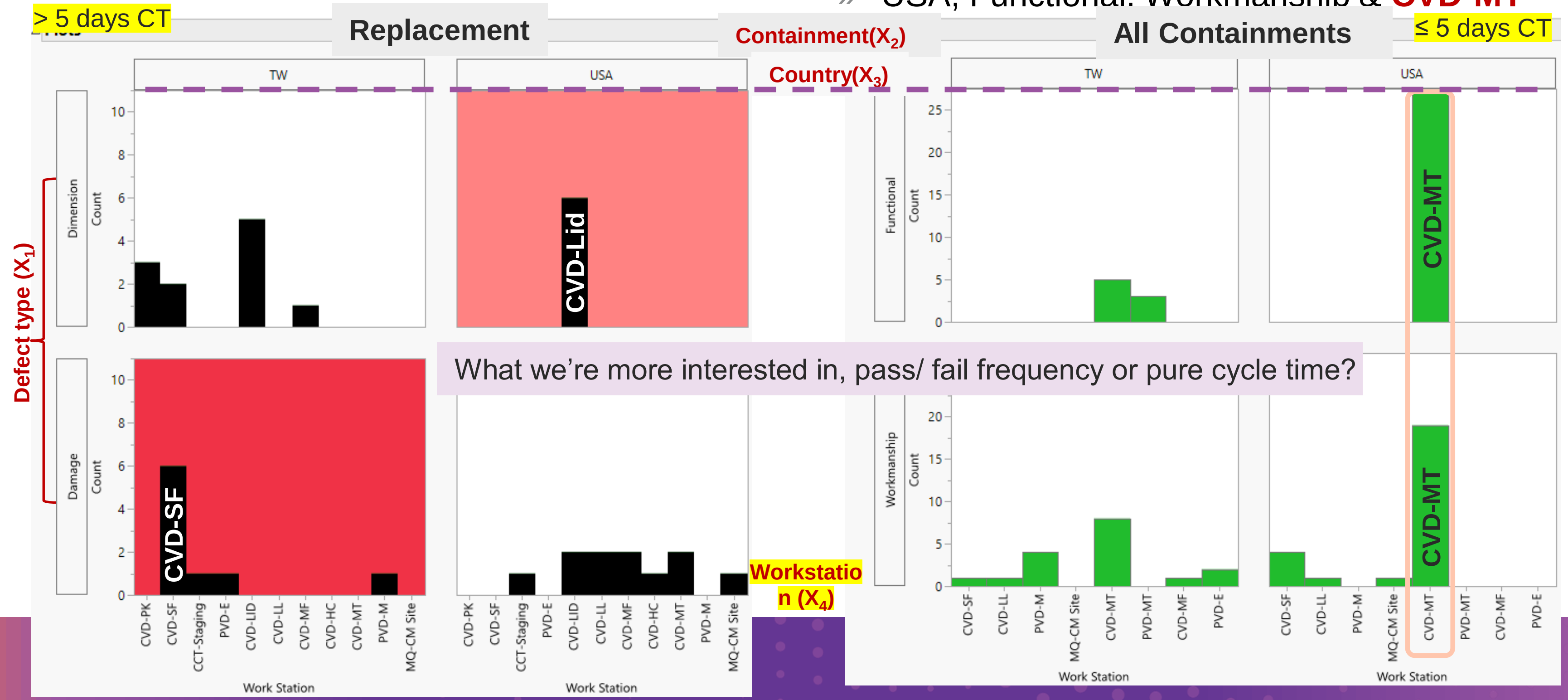
Except for Dimension and Damage defect, others are easy to quickly fix – SA inference

Pareto Chart – 4th Layer of Root Cause Analysis

- Add additional factor “**workstation** (X_4)” in Pareto Chart to visualize frequency event
  - Replacing TW damage & **CVD-SF**
  - Replacing USA Dimension & **CVD-Lid**

» Per previous inference : Except for Dimension and Damage, other defects are easy to quickly fix

» USA, Functional, Workmanship & **CVD-MT**



Tabulate – 5th Layer of Root Cause Analysis

- Show average and count on Tabulate table to do further comparison
- FA: CVD-SF Replacement TW damage issue
- SA: CVD-MT MFG rework USA Workmanship issue

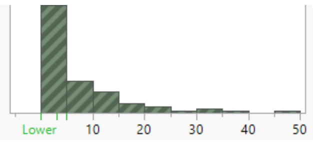
		immediate fix cycle time							
		Work Station							
		CVD-LID				CVD-SF			
		Defect type				Defect type			
		Dimension		Damage		Dimension		Damage	
Containment	Country	Mean	N	Mean	N	Mean	N	Mean	N
Replacement	TW	15	7	4	1	8	3	34	6
	USA	16	6	7	6	.	0	2	1
	All	15	13	7	7	8	3	29	7

		immediate fix cycle time			
		Work Station			
		CVD-MT			
		Defect type			
		Workmanship		Functional	
Containment	Country	Mean	N	Mean	N
MFG rework	USA	2	20	3	16
	TW	2	8	1	2
	All	2	28	3	18

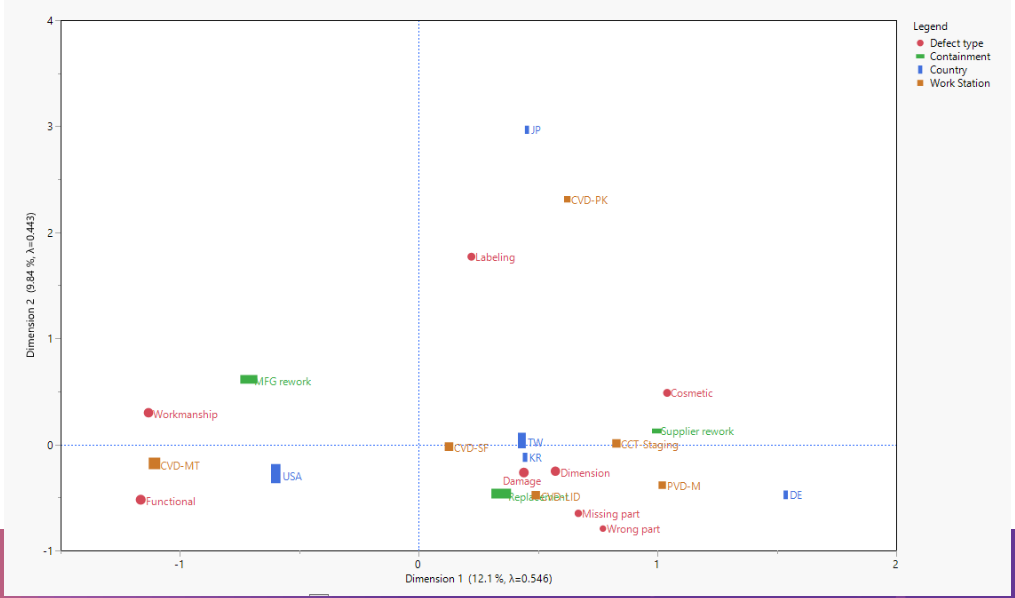
Root Cause Analysis Summary

- Use different **Graphical** JMP Platforms in **Engineering and Logical Sequence** to conduct **deeper** Root Cause Analysis
  - **1st Layer Histogram**: set Conditional Mosaic to investigate both SA and FA
  - **2nd Payer Box plot**: know how to investigate the process special variations (skewness, outliers)
  - **3rd Layer Heatmap plot**: narrow down the SA/FA root cause analysis scope to Defect Type X Country Square
  - **4th Layer Pareto Chart**: conduct 2-dimensional Pareto Chart from previous Heatmap results
  - **5th Layer Tabulate**: visualize the Pivot Table on integrating the previous layers of Root Cause Analysis
- **Identify the Potential inputs (X_s) to Predict the QN fix Cycle Time**
  - **1st Layer Histogram**: Defect type (X_1)
  - **2nd Layer Box plot**: Containment (X_2), Country (X_3)
  - **3rd Layer Heatmap**: Defect type (X_1), Containment (X_2), and Country (X_3)
  - **4th Layer Pareto Chart**: Defect type (X_1), Containment (X_2), and Country (X_3), Workstation (X_4)
  - **5th Layer Tabulate**: Narrow Down to Damage (Defect type X_1), Replacement (Containment X_2), TW (Country, X_3), CVD-SF (Workstation X_3)
- **Next Step: Build a model to predict the QN fix Cycle Time (Validation of Root Causes)**

Model Selection and Comparison



- The fit model challenge:
  - Skewed distribution: log transformation -> no help
  - All input variables are categorical type (filter out 60% of workstation category, R–square increase by 6%)
  - Dependency among categorical variables (low risk)



- Partition tree model:
  - » Distribution free model
  - » Split base on data available
  - » Little overfit concern
  - » Recursive split
  - » Random Forest Predictor Screening

Predictor Screening

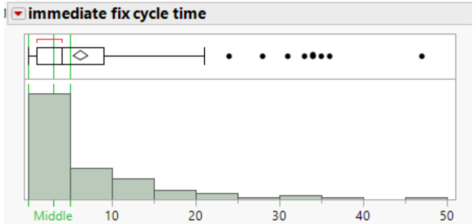
	immediate fix cycle time			
Predictor	Contribution	Portion		Rank
Defect type	1864.08	0.3384		1
Work Station	1780.23	0.3232		2
Country	1038.42	0.1885		3
Containment	702.34	0.1275		4
fault by (ref)	123.17	0.0224		5

Copy Selected

- Neural Network model:
  - » Strong transformation model
  - » Two steps (training & validation) model
  - » Significant overfit concern

(1) Fit Model – Main Effect Only

- **R square ~ 30%** is not adequate due to severe **right skewness**
- Observed significant **lack of fit** risk though **Max R-Square ~ 47%**
- **Log transformation** won't help model fit much



Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	25	30.23418	1.20937	1.7219
Pure Error	233	163.64544	0.70234	Prob > F
Total Error	258	193.87963		0.0208*
				Max RSq
				0.4745

Summary of Fit

RSquare	0.296308
RSquare Adj	0.239787
Root Mean Square Error	6.465428
Mean of Response	6.288889
Observations (or Sum Wgts)	270

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	20	4382.830	219.141	5.2424
Error	249	10408.637	41.802	Prob > F
C. Total	269	14791.467		<.0001*

Remembered Settings

Setting	Defect type	Containment	Country	Work Station	fault by (ref)	immediate fix cycle time	immediate fix cycle time Lower CI	immediate fix cycle time Upper CI	Desirability
Optimal	Workmanship	Supplier rework	KR	CCT-Staging	Supplier fault	-8.332262	-14.2495	-2.415021	0.999758

Setting	Defect type	Containment	Country	Work Station	fault by (ref)	immediate fix cycle time	immediate fix cycle time Lower CI	immediate fix cycle time Upper CI	Desirability
Optimal	Damage	Replacement	JP	CVD-SF	MFG fault	24.194545	17.305714	31.083375	0.743400

Pick Role Variables

Y	Log(immedi...cycle time 2)
	optional
Weight	optional numeric
Freq	optional numeric
By	optional

Construct Model Effects

Add	Defect type
Cross	Containment
Nest	Country
	Work Station
Macros ▾	fault by (ref)

Summary of Fit

RSquare	0.303036
RSquare Adj	0.247055
Root Mean Square Error	0.933638
Mean of Response	1.268714
Observations (or Sum Wgts)	270

Use **Log transformation** of the cycle time variable to transform the skewed cycle time distribution

Remembered Settings

Setting	Defect type	Containment	Country	Work Station	fault by (ref)	immediate fix cycle time 2	immediate fix cycle time 2 Lower CI	immediate fix cycle time 2 Upper CI	Desirability
Optimal	Wrong part	Supplier rework	KR	CCT-Staging	Supplier fault	0.5518602	0.2182356	1.3955084	0.895239

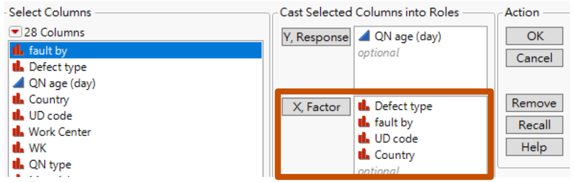
Setting	Defect type	Containment	Country	Station	fault by (ref)	cycle time 2	time 2 Lower CI	time 2 Upper CI	Desirability
Optimal	Dimension	Replacement	JP	CVD-MT	MFG fault	30.040394	9.7730486	92.338152	0.695992

(2) Partition Tree Model – Model Improvement & Comparison

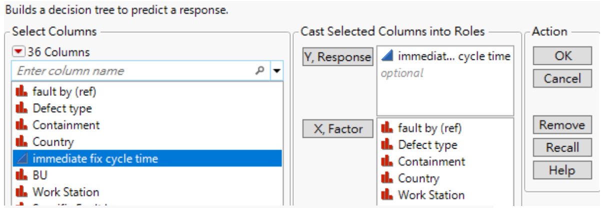
- **Baseline Model**
- Model is not adequate (R square 37.9%)
  - Originally **4** input factors

- **Model Augmentation**
  - Discuss with SME and select a **new output variable (Y)**
  - Add 5th input variable (**Workstation ,X**)
- R square has been improved around **20%**

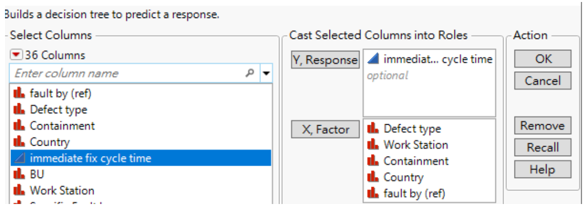
- **Model Simplification**
  - Utilize **Pareto principle** and data filter to screen any minor data category
  - Total sample size decrease to 270 from 426
- R square improve around **6%**



RSquare	RASE	N	Number of Splits	AICc
0.379	27.269101	426	37	4111.52



RSquare	RASE	N	Number of Splits	AICc
0.566	4.3510406	426	52	2585.74



RSquare	RASE	N	Number of Splits	AICc
0.623	4.5468444	270	35	1670.14

Model **Augmentation (R-square improved by 20%)**

- **0%** R-square Improvement

QN age (day)

Column Contributions

Term	Number of Splits	SS	Portion
Defect type	17	114415.148	0.5911
Country	14	59520.9863	0.3075
UD code	5	19591.3368	0.1012
fault by (ref)	1	50.4751834	0.0003

RSquare	RASE	N	Number of Splits	AICc
0.379	27.269101	426	37	4111.52

immediat... cycle time

Column Contributions

Term	Number of Splits	SS	Portion
Country	15	2430.52733	0.3555
Defect type	12	2407.45243	0.3522
UD code	8	1829.65564	0.2676
fault by (ref)	5	168.540489	0.0247

RSquare	RASE	N	Number of Splits	AICc
0.368	5.248367	426	40	2714.91

- **16%** R-square Improvement

- Add X factor: **MFG Workstation (the NO.2 ranking, around 28%)**
- **UD code less critical after adding workstation**

Y, Response immediat... cycle time

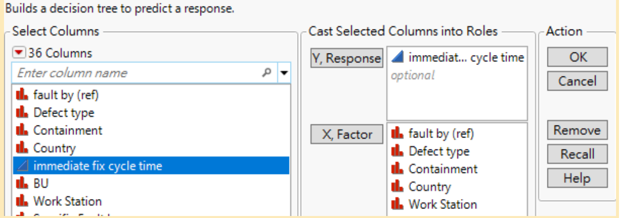
Column Contributions

Term	Number of Splits	SS	Portion
Defect type	17	4546.69243	0.4660
Work Station	15	2794.18279	0.2864
Country	14	1584.96213	0.1625
UD code	8	791.891966	0.0812
fault by (ref)	2	38.4684764	0.0039

RSquare	RASE	N	Number of Splits	AICc
0.525	4.548719	426	56	2634.23

- Another **4%** R-Square improve

- Change X factor: **Containment from UD code**



Column Contributions

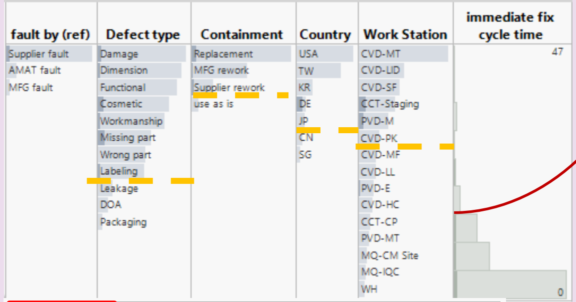
Term	Number of Splits	SS	Portion
Defect type	16	4612.40123	0.4390
Work Station	17	4260.34495	0.4055
Containment	6	894.132282	0.0851
Country	10	603.876596	0.0575
fault by (ref)	3	134.900637	0.0128

RSquare	RASE	N	Number of Splits	AICc
0.566	4.3510406	426	52	2585.74

Top two input (X) factors in rankings: Defect type & Country → Defect type & Workstation

Model **Simplification (R-square improved by 6%)**

- Previous model augment includes all categories & data
  - Plus: considering all scenarios
  - Drawback: too many categories might **dilute prediction power**

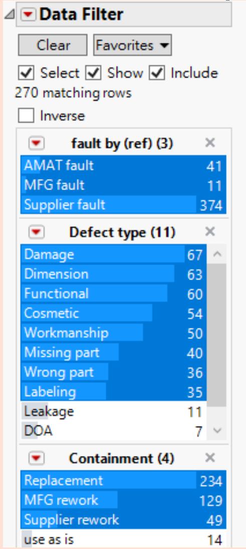


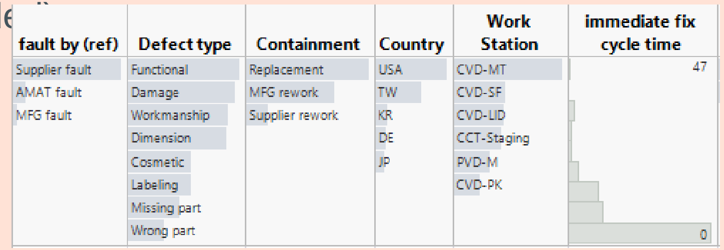
RSquare	RASE	N	Number of Splits	AICc
0.566	4.3510406	426	52	2585.74

Column Contributions

Term	Number of Splits	SS	Portion
Defect type	16	4612.40123	0.4390
Work Station	17	4260.34495	0.4055
Containment	6	894.132282	0.0851
Country	10	603.876596	0.0575
fault by (ref)	3	134.900637	0.0128

- Simplify dataset by filtering out minor categories with fewer counts to improve prediction power
  - **Remove 60% categories of Workstation**
  - Total amount (N) decreases to 270 from 426 (156





RSquare	RASE	N	Number of Splits	AICc
0.623	4.5468444	270	35	1670.14

Column Contributions

Term	Number of Splits	SS	Portion
Defect type	9	4512.26539	0.4900
Work Station	14	2803.04562	0.3044
Containment	3	1150.87421	0.1250
fault by (ref)	3	384.822466	0.0418
Country	6	358.534565	0.0389

Partition Tree Model Optimization – Min & Max QN Cycle Time

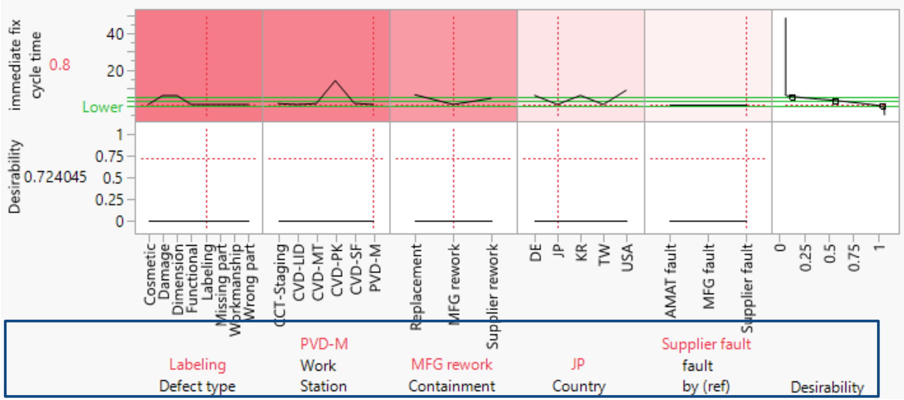
- The major contributor are **Defect type & Workstation ~ 80% (Pareto Concept)**
- According to prediction profiler of the method,
  - The best scenario (min cycle time) :Labeling, PVD-M
  - The worst scenario (max cycle time) :Damage, CVD-MT

Column Contributions

Term	Number of Splits	SS	Portion
Defect type	9	4512.26539	0.4900
Work Station	14	2803.04562	0.3044
Containment	3	1150.87421	0.1250
fault by (ref)	3	384.822466	0.0418
Country	6	358.534565	0.0389

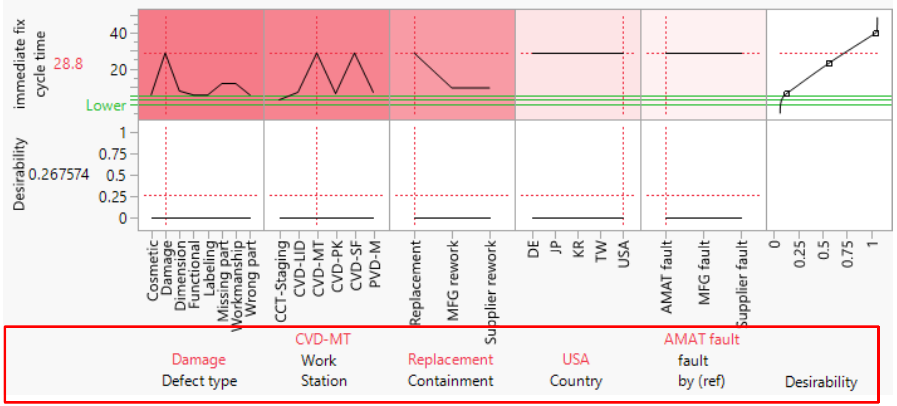
Setting	Defect type	Work Station	Containment	Country	fault by (ref)	immediate fix cycle time	Desirability
Optimal	Labeling	PVD-M	MFG rework	JP	Supplier fault	0.8	0.857144

Prediction Profiler



Setting	Defect type	Work Station	Containment	Country	fault by (ref)	immediate fix cycle time	Desirability
Optimal	Damage	CVD-MT	Replacement	KR	MFG fault	28.8	0.267574

Prediction Profiler



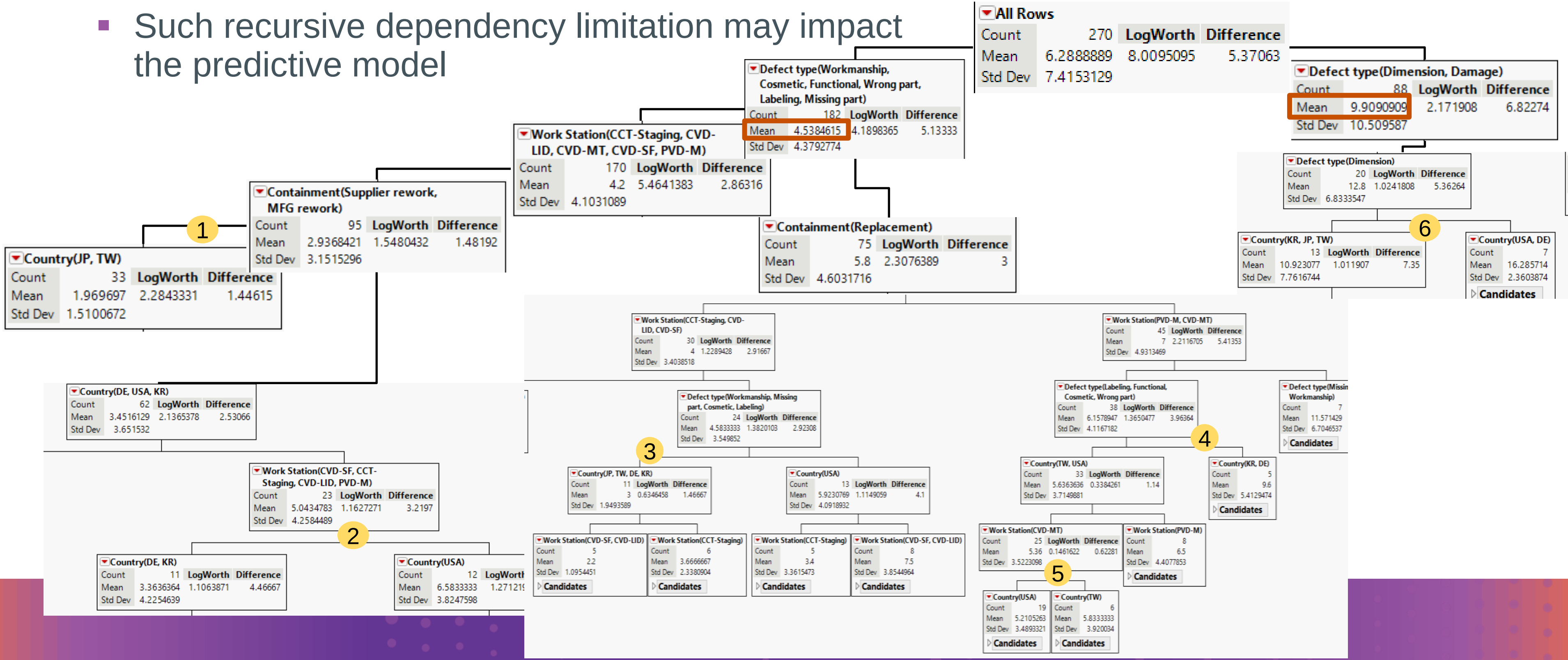
Doesn't country impact QN fix cycle time? Is it right?

Model Limitations: Recursive Partitions

- Recursive partitions (sequential dependency risk)
  - Factor “country” is split 6 times, and only 1 time happened in the higher cycle time cluster.
  - Such recursive dependency limitation may impact the predictive model

Column Contributions

Term	Number of Splits	SS	Portion
Defect type	9	4512.26539	0.4900
Work Station	14	2803.04562	0.3044
Containment	3	1150.87421	0.1250
fault by (ref)	3	384.822466	0.0418
Country	6	358.534565	0.0389



Neural Network (Artificial Intelligence)

- Observe severe **overfit concern** between training and validation R square
  - Overfit: > 20% R-Sq between Training Set (building model), and Validation Set (fitting model)
  - Too aggressive Black-Box transformation to build a model with training set (too good to be true)
- The major contributor : **Workstation**

Training — immediate fix cycle time

Measures	Value
RSquare	0.659755
RASE	4.5389141
Mean Abs Dev	3.0979374
-LogLikelihood	527.69274
SSE	3708.3134
Sum Freq	180

Validation — immediate fix cycle time

Measures	Value
RSquare	0.4449701
RASE	4.7913103
Mean Abs Dev	3.5440675
-LogLikelihood	268.71682
SSE	2066.0989
Sum Freq	90

Variable Importance: Independent Uniform Inputs

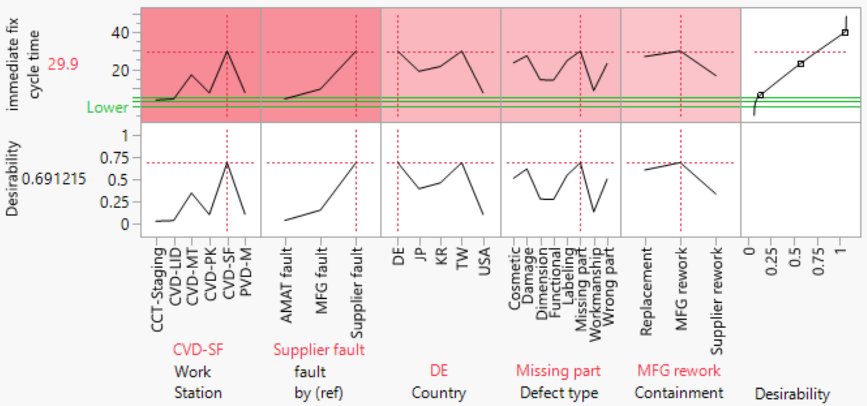
Summary Report

Column	Main Effect	Total Effect
Work Station	0.081	0.566
fault by (ref)	0.05	0.544
Country	0.043	0.352
Defect type	0.108	0.336
Containment	0.104	0.289

Remembered Settings

Setting	Defect type	Containment	Country	Work Station	fault by (ref)	immediate fix cycle time	Desirability
Optimal	Workmanship	Supplier rework	KR	CCT-Staging	Supplier fault	-6.617631	0.999797
Optimal	Damage	Replacement	TW	CVD-SF	Supplier fault	29.019381	0.666909

Prediction Profiler



Model Comparisons and Selection

select model with
lowest prediction error

- Root Cause Analysis: Damage issue (defect type), Replacement (containment), TW (country), CVD-SF (workstation) is the worst scenario with longer QN fix cycle time
  - Neural Model has the identical scenario as the graphical root cause analysis
  - **Only concern on the Overfit risk**
- The 3 models have very close prediction on the worst cycle time within 1.2 Days

Fit Model

Summary of Fit

RSquare	0.303036
RSquare Adj	0.247055
Root Mean Square Error	0.933638
Mean of Response	1.268714
Observations (or Sum Wgts)	270

Remembered Settings

Setting	Defect type	Containment	Country	Work Station	fault by (ref)	immediate fix cycle time 2	immediate fix cycle time 2 Lower CI	immediate fix cycle time 2 Upper CI	Desirability
Optimal	Wrong part	Supplier rework	KR	CCT-Staging	Supplier fault	0.5518602	0.2182356	1.3955084	0.895239
Optimal	Dimension	Replacement	JP	CVD-MT	MFG fault	30.040394	9.7730486	92.338152	0.695992

Partition

RSquare	RASE	N	Number of Splits	AICc
0.623	4.5468444	270	35	1670.14

Setting	Defect type	Work Station	Containment	Country	fault by (ref)	immediate fix cycle time	Desirability
Optimal	Labeling	PVD-M	MFG rework	JP	Supplier fault	0.8	0.857144
Optimal	Damage	CVD-MT	Replacement	KR	MFG fault	28.8	0.267574

Neural

Training

immediate fix cycle time

Measures	Value
RSquare	0.659755
RASE	4.5389141
Mean Abs Dev	3.0979374
-LogLikelihood	527.69274
SSE	3708.3134
Sum Freq	180

Validation

immediate fix cycle time

Measures	Value
RSquare	0.4449701
RASE	4.7913103
Mean Abs Dev	3.5440675
-LogLikelihood	268.71682
SSE	2066.0989
Sum Freq	90

Remembered Settings

Setting	Defect type	Containment	Country	Work Station	fault by (ref)	immediate fix cycle time	Desirability
Optimal	Workmanship	Supplier rework	KR	CCT-Staging	Supplier fault	-6.617631	0.999797
Optimal	Damage	Replacement	TW	CVD-SF	Supplier fault	29.019381	0.666909

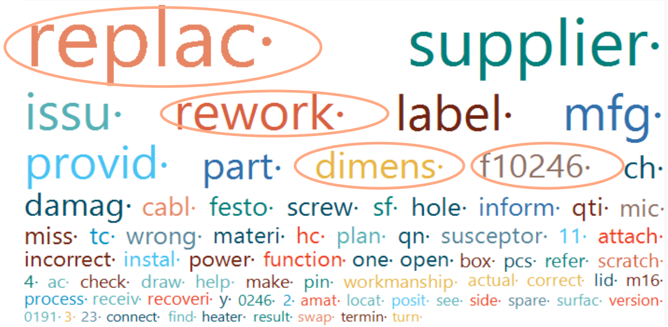
Text Mining and Data Mining Hybrid

- Search keywords from QN Database (Categorical and Text Variables)
- Convert the Keywords information to Binary Indicators
- Conduct the further Data Mining- Root Cause Analysis On F10246 Case

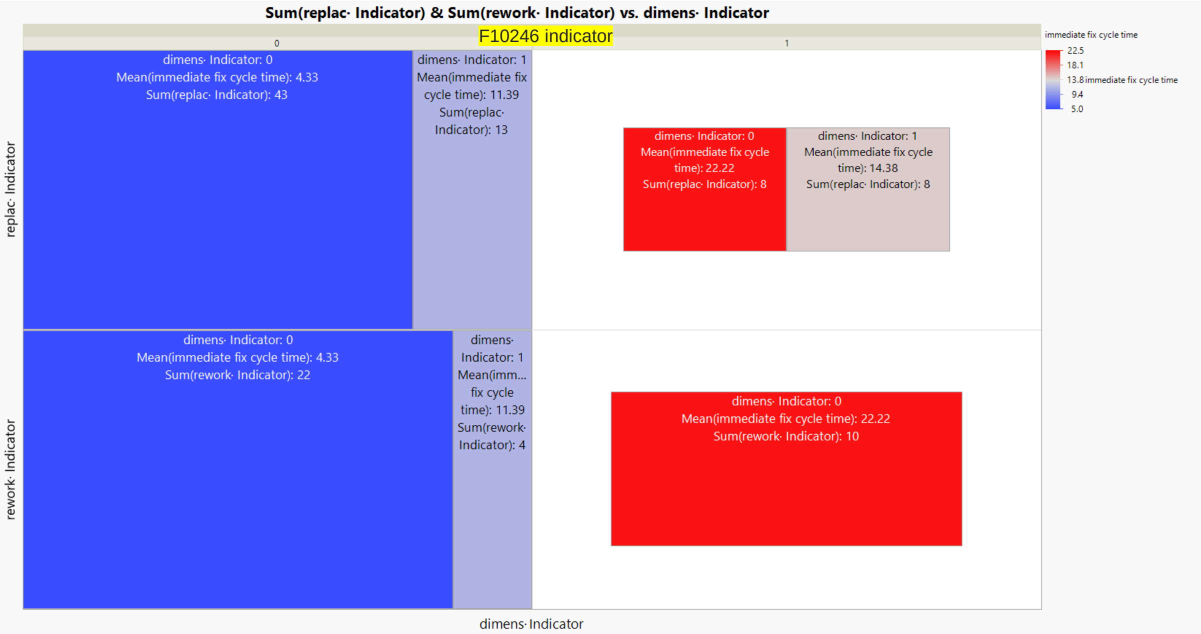
Text Explorer for QN long text

Number of Terms	Number of Cases	Total Tokens	Tokens per Case	Number of Non-Empty Cases	Portion of Non-Empty Cases
737	108	4880	45.1852	108	1.0000

Word Cloud



Over 5 days fix cycle time have strong relationship with "Replace, rework, dimension, F10246"



Take Away Learnings

- JMP Graphical Platforms are powerful to conduct deeper root cause analysis through Engineering, Logical, Data-Driven process
- Compare and Select more appropriate JMP Model from Classical Fit Model to modern Partitions and Neural Network by knowing the model limitations and risks connecting to previous Graphical Root Cause Analysis
- Conduct the Hybrid Text Mining and Data Mining Root Cause Analysis on the Complicated QN Database

- THANK GCI MBB Charles Chen as my Project Mentor

Thank You