Paper 3877-2019

Predictive Warranty Analytics using SAS® Field Quality Analytics to generate Proactive Insights

Prateek Kumar Singh, Mahindra & Mahindra (Bristlecone); Saket Shubham, Mahindra & Mahindra (Bristlecone); Shiv Kumar, Mahindra & Mahindra; Paresh Peter Rodrigues, Mahindra & Mahindra (Bristlecone)

ABSTRACT

In today's data driven world, unprecedented information related to **product reliability** and **field quality** can be furnished through warranty claims data. The tools and techniques used in Warranty analytics assist the quality management teams in decision making and extraction of meaningful information by analyzing structured as well as non-structured warranty data with statistical techniques.

Predictive warranty analytics helps in **unearthing potential problems early on** and **identifying emerging issues** before they become huge, costly problems, and enables to initiate the problem-solving process months in advance.

SAS® Field Quality Analytics (FQA), a suite for quality analytics, assesses data from warranty, customer service, product and other relevant sources to detect emerging issues sooner and provide actionable insights. This paper intends to showcase how FQA can be used to discover hidden relationships, patterns, and future trends of product performance, claim rates, customer views etc. Data from various sources is manipulated to bring in the desired format using Base and Advanced SAS coding. A combination of techniques like Enterprise Analytic **Early Warning** (for emerging issues identification), **Reliability Analysis** (for claim cost and frequency forecasting) using Weibull distribution, **Text Mining** on comments, inferential analysis etc. have been used. OEMs can benefit from these analyses to support strategic decisions like vehicle recall and maintain customer loyalty.

INTRODUCTION

Manufacturers incur high warranty cost due to in-service product failures, as a consequence of upsurge in failure rates on some of the parts for different models. There is a need for reliable claim forecasts and efficient analysis for strategic decision making and improving product quality. The incumbent warranty analytics processes involve analysis of historical data and generation of reports and dashboards which aids efficient decision making. In the modern era of innovation and technology, predictive and Prescriptive Analytics can take this one step further to generate actionable insights and minimizing the risk involved at various stages.

To enable accurate claim forecasting and root cause analysis on failure data, SAS offers Field Quality Analytics (FQA) which analyzes the event of a premature failure of an item or the inability of the item to perform its intended function within and outside warranty.

SAS FQA can mine, alter, manage and retrieve data from a variety of sources and perform statistical analysis on these sources more reliably than other traditional methods, providing a platform for examining failures and taking impactful decisions.

Early Warning Analysis, one of the key features of FQA, creates alerts for emerging issues which further reduces the cost of poor quality. It includes risk analysis, monitoring & warning and dissemination & communication.

The following workspaces in SAS FQA 6.1 are designed to enrich the experience of monitoring and analysing warranty data:

- **Early Warning workspace**: Early Warning is a system which generates signals based on historical data. One can create enterprise analytic, ad-hoc analytic and ad-hoc threshold early warnings along with deploying the options of assigning, subscribing and managing alerts.

- **Data Selection workspace**: This is used for creation and managing data selections and data selection templates which combines Product, Event and service Data for further Analysis.
- **Analysis workspace**: This workspace enables 14 different types of Analyses. It interacts with analysis results and has the option to export the output in user accessible format.
- **Administration workspace**: To manage administrative tasks like user access, analyses settings etc. in SAS Field Quality Analytics suite.

This paper illustrates the functionalities and usage of SAS FQA 6.1 in automobile industry with relevant examples.

2. BACKGROUND TO DATA USED

2.1 DATA INTRODUCTION

The input data required for FQA includes manufacturing, sales, service and warranty claims information. Several tables from the transaction systems are extracted, processed and combined. Broadly, the data can be categorized as:

2.1.1 Events Data

Events data consists of:

1. A set of transaction tables containing warranty events and basic information about them like the complaint group (failure modes) responsible for an Event, replaced Part-codes, labor-codes, date of failure etc.
2. Master tables containing descriptions of complaint group-codes, part-codes etc.

Following table represents the events data without code descriptions:

VIN Number	Claim Id	Group Code	Comp. Code	Part	Event	Cost
ABC	123	G1	C1	P1	1	100
ABC	123	G1	C2	P4	1	20

Table 1. Claims data sans code descriptions

2.1.2 Manufactured and Sold Data (MnS)

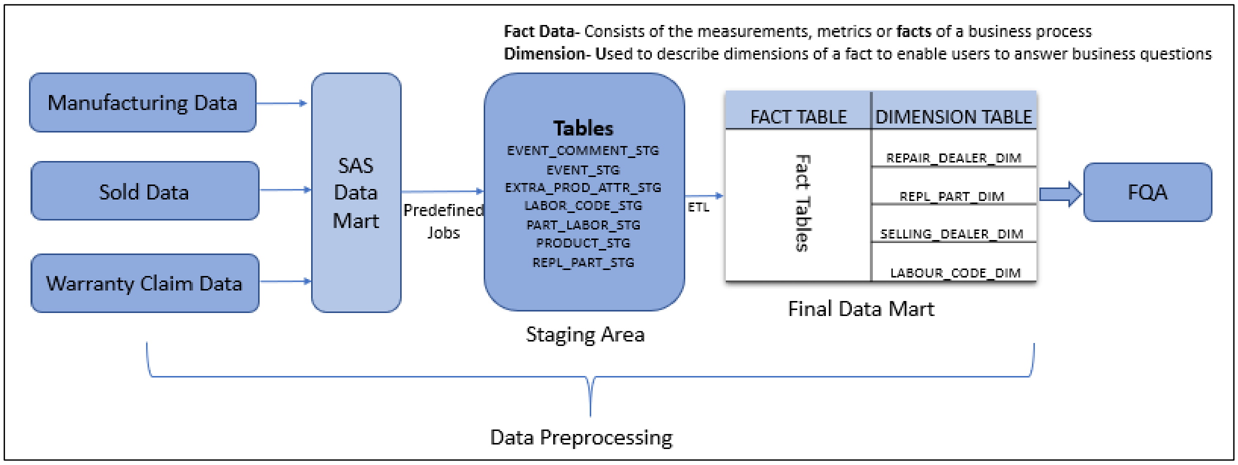
This is another set of transaction and master tables containing the manufactured and sold details of each vehicle that is produced by the organization and various attributes of the vehicles like production dates, plant in which the vehicle was produced, Vehicle Type, Model name, Variant Name, Dealer to which it was billed etc. The total no. of VIN numbers in the data represents the number of vehicles sold.

Following table represents the MnS Data:

VIN Number	Prod. Month	Prod. Year	Model	Dealer	Zone
ABC	Apr	2016	YYY	X123	Zone1
PQR	May	2016	QQQ	X212	Zone2

Table 2. MnS Data

2.2 DATA PREPARATION FOR FQA – PROCESS FLOW DIAGRAM



Display 1: Data Flow Chart

2.3 IMPORTANT KPIS

S. No.	Terms Used	Meaning	Formulas/Notation
1	Production Month	Month in which vehicles are produced	Denoted as "Jan18" for all vehicles produced in the month of Jan
2	Event	Combination of Claim Id, Group Code and Comp-Code is one Event	No. of Events occurred for a Vehicles
3	Manufactured and Sold No. (MnS)	Number of vehicles produced, that have been sold	
4	Warranty Cost	Cost incurred toward the event	
5	Claim Rate (CR)	Estimated no. of events per hundred vehicles sold from that month's produce.	CR= (Event/MnS)*100
6	Warranty Cost Per Vehicle (WCPV)	Average warranty cost incurred per vehicle	WCPV= (Warranty Cost/MnS)

Table 3: KPIs

2.4 TIME IN SERVICE (TIS)

- "Time in Service" is the no. of months a vehicle has completed in service before a failure occurs
- Each event may fall in separate TIS Bucket depending on the no. of days between sale date and failure date
- Some of the TIS Buckets are:
  - *0 TIS* - Contains all events that occur before sale (Pre-Delivery Inspections or PDI)
  - *3 TIS* - Contains all the events that occur within 90 days from sale
  - *12 TIS* - Contains all the events that occur within 360 days from sale

Sample Representation													
TIS	**Months**												
0 TIS	0												
1 TIS	0	1											
3 TIS	0	1	2	3									
6 TIS	0	1	2	3	4	5	6						
12 TIS	0	1	2	3	4	5	6	7	8	9	10	11	12

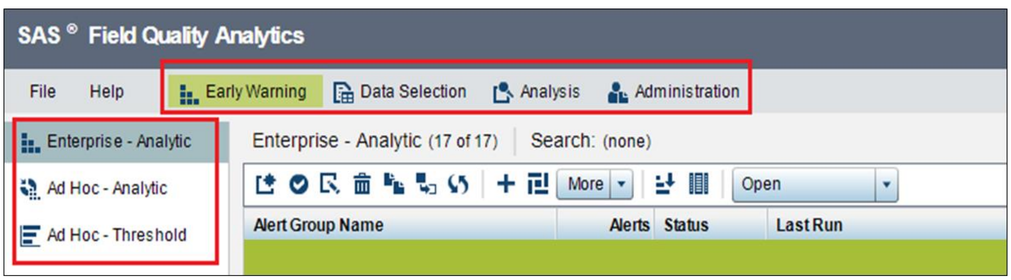
Table 4: Sample Representation of TIS

3. SAS FIELD QUALITY ANALYTICS

SAS FQA is part of the SAS Quality Analytic Suite, which provides an enterprise view of quality performance to help manage the cost of quality, achieve quality excellence and increase customer satisfaction. Display 2 shows the visual representation of SAS FQA Components.

Following are the three main components in SAS FQA:

- **Early Warning Analysis**
- **Data Selection**
- **Analysis**

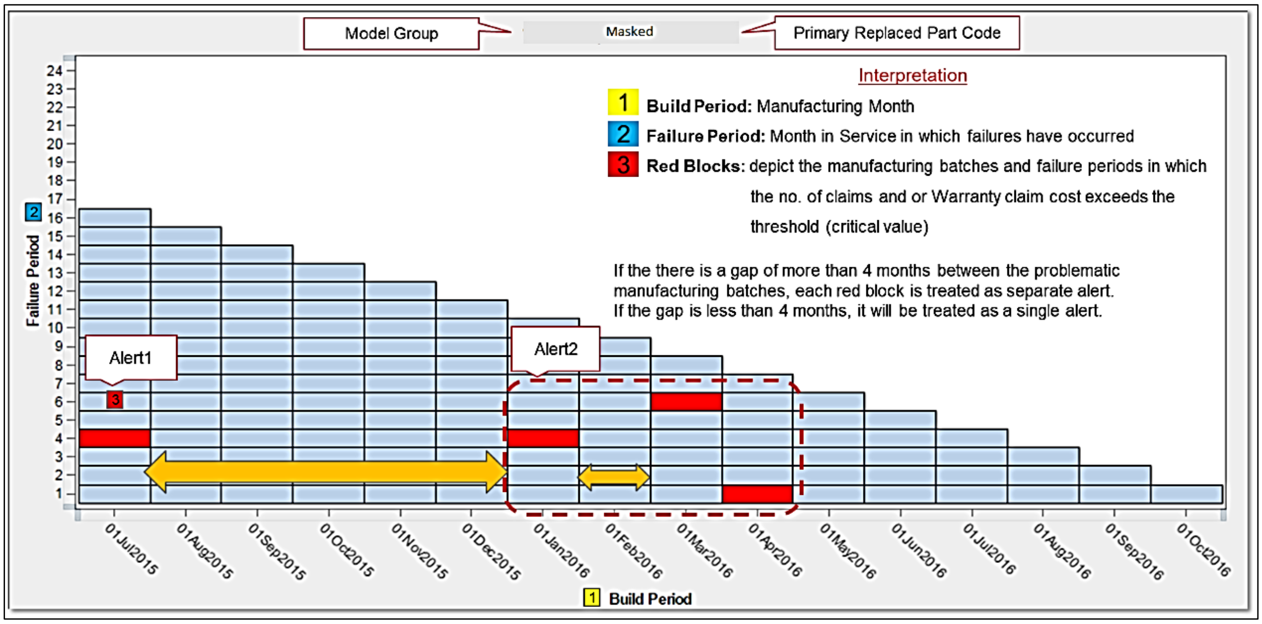


Display 2: FQA Components

3.1 EARLY WARNING ANALYSIS – A BRIEF

Early Warning analysis is a methodology designed to identify and mitigate the risk raised due to manufacturing defects as well as product wear & tear and helps to reduce either warranty cost or recall cost. Any early warning (EW) system works on the principle that “discontinuities do not emerge without warning". These warning signs are represented as "weak signals”, an idea aimed toward detection of these signals that may lead to events that have the potential to jeopardize an organization’s strategy.

SAS FQA Early Warning Analysis deals with dynamic critical values, creating thresholds for each manufacturing part and for every combination of build period[1] and failure period[2], for generating alerts pro-actively which other descriptive analyses (e.g. Pareto Analysis, Trend Charts etc.) are not able to create. FQA can automatically execute early warning analyses and identify emerging issues as soon as new data points are captured and refreshed in the mart.



Display 3: Output of Early Warning Analysis

Display 3 shows the output of Early Warning Analysis (Build Period vs Failure Period). Red cells[3] in the matrix indicate that failures are significantly higher than expected for units in the given build and failure period.

Early Warning Analysis can be performed post selecting the appropriate data in data selection workspace.

3.1.1 Features of Early Warning Analysis

- **Integrated warranty business rules-**
  Early Warning Analysis addresses the complexity of warranty data by applying business rules and algorithms such as sales lag profiles, usage distributions, maturity calculations and seasonality adjustments.
- **Emerging issues**-
  Early Warning Analysis automatically detects anomalies based on violations of analytically driven critical values or manually input thresholds. Automatically determines analytically driven critical values.
- **Alerts notifications, subscription and individual alerts analyses -** Early-warning alerts are prioritized and assigned to appropriate parties for investigation. It has the feature of posting comments with analysis as well as sub setting alert for analysing in analysis workspace.

3.1.2 Statistical Methods for early warning analysis

Two statistical analysis methods are applied in early warning analyses to identify upward shifts in claims activities:

1. **Production period analysis**: The analysis monitors claims activity for the sample size at risk relative to production periods (build period) for different time-in-service periods. The alerts are generated as soon as the actual failure rate crosses the critical value/threshold. The output includes a matrix chart (display-4) that identifies the flagged periods in red.

   This methodology was proposed by Wu and Meeker (2002). They describe the use of an algorithm to compute critical values for the warranty report monitoring procedure. A statistical test is performed for each calendar month to identify when the actual claim count is greater than the critical value.

The Hypothesis for the statistical test is stated as below-
H_0 : Actual Claim rate is less than or equal to Expected Claim Rate ($\lambda_1 <= \lambda'_1$)
H_1 : Actual Claim rate is greater than Expected Claim Rate. ($\lambda_k > \lambda'_k$)
λ_k - Actual Claim Rate during time-in-service period K
λ'_k - Expected Claim Rate for the same period
Here the actual claim rate is greater than expected claim rate when the number of claims generated in that time-in-service period on those units is greater than some critical value.
So, $\mathbf{S_{ijk} > C_{ijk}}$

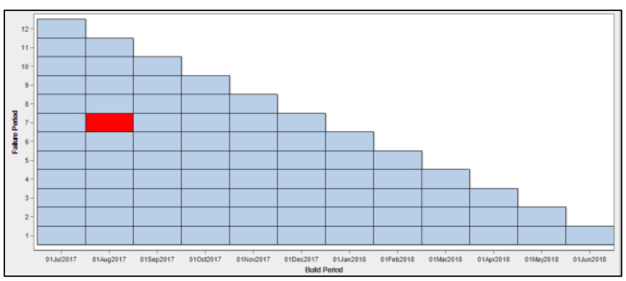
$$S_{ijk} = \sum_{j=1}^{p} (R_{ipk})$$

where R_{ipk} is the number of claims that occurred in the k^{th} claim time-in-service period for units produced in the i^{th} production period and sold in the p^{th} sales period after production.

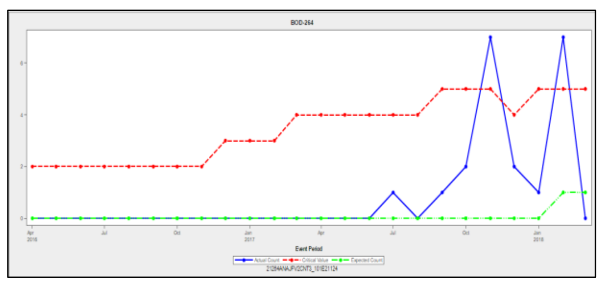
There are four main factors which are considered when **critical value** is to be calculated. The first one being historical claim rate for different TIS, second is the number of units produced in different build period, next is the proportion of units that would go in service for every production period and last is the false alarm probability (alpha).

2. **Event period analysis** monitors claim counts across failure periods and flags an issue when the actual claim count that occurs in a calendar period significantly exceeds what is expected to occur for that calendar period. The output from the claim period analysis (Display 5) is a plot of claim count (y-axis) versus calendar periods (x-axis).

   The Event Period Analysis shows the accumulated claims 'failure month-wise', whereas the production period analysis shows the output based on 'build period'.



Display 4: Production Period Chart



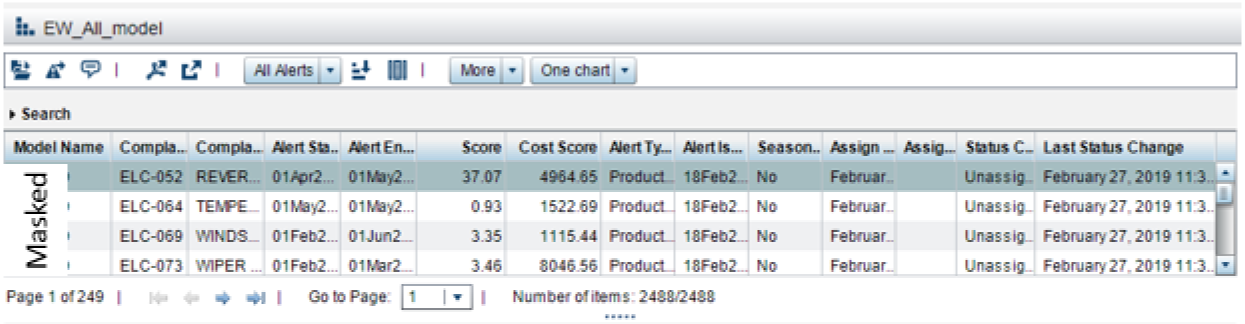
Display 5: Event Period Chart

3.2 THREE TYPES OF EARLY WARNING ANALYSIS

- **Enterprise Analytic**
- **Ad-hoc Analytic**
- **Ad-hoc Threshold**

3.2.1 Enterprise – Analytic Early Warning

It analyses the performance of the historical data and determines whether warranty issues are currently developing. It enables to automate the analysis for all the product models across the enterprise, refreshing the alerts frequently whenever any incremental data is available.

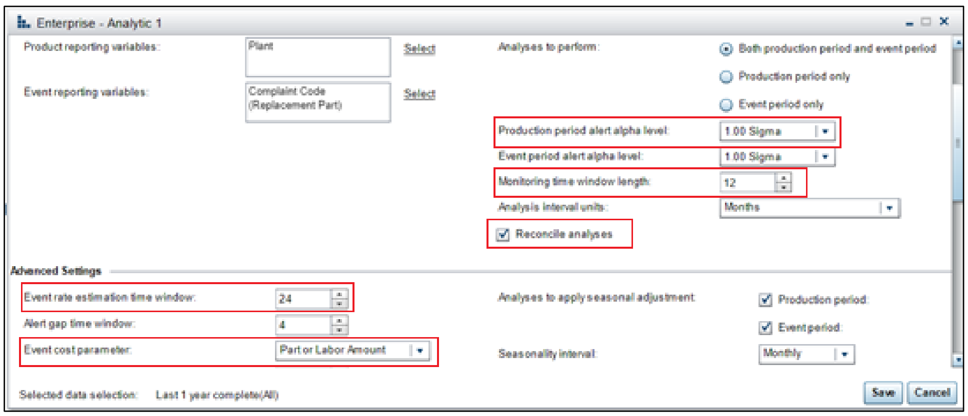


Model Name	Compla...	Compla...	Alert Sta...	Alert En...	Score	Cost Score	Alert Ty...	Alert Is...	Season...	Assign ...	Assig...	Status C...	Last Status Change
Masked	ELC-052	REVER...	01Apr2...	01May2...	37.07	4964.65	Product...	18Feb2...	No	Februar...		Unassig...	February 27, 2019 11:3...
	ELC-064	TEMPE...	01May2...	01May2...	0.93	1522.69	Product...	18Feb2...	No	Februar...		Unassig...	February 27, 2019 11:3...
	ELC-069	WINDS...	01Feb2...	01Jun2...	3.35	1115.44	Product...	18Feb2...	No	Februar...		Unassig...	February 27, 2019 11:3...
	ELC-073	WIPER ...	01Feb2...	01Mar2...	3.46	8046.56	Product...	18Feb2...	No	Februar...		Unassig...	February 27, 2019 11:3...

Display 6: Output table of Enterprise – Analytic EW

Description of columns displayed in Enterprise – Analytic EW output (Display 6)-

- *Model Name*- Product Reporting Variable
- *Complaint code* (*Replacement Part*)- Event reporting variable
- *Alert start Date*- As given while creating Enterprise – Analytic EW
- *Alert End Date*- As given while creating Enterprise – Analytic EW
- *Score* – It is a representation of number of claims crossing the critical value, higher the score higher the claim rate is.
- *Cost Score*- It is the average cost of the event type multiplied by the number of events that occurred during the alert period beyond what was expected.



Display 7: Basic and Advance Settings of Enterprise - Analytic

Some important field settings for Enterprise - Analytic EW under Basic and Advance Settings-

- **Reporting Variable:** Variables whose values are analyzed or included in the analysis output. E.g. Plant and model.
- **Production period alert alpha level-** The Alpha level is used as false alarm level when an increase in event activity is significant. The options available are in terms of a sigma level.
- **Monitoring Time Window Length –** It has the value less than or equal to warranty time length which controls:
    - No. of build period & failure period to be monitored by Production Period Model
    - No. of Calendar period to be monitored by Event Period Model
- **Event Cost Parameter-** Used to calculate the Cost Score that appears in results table.
- **Event Rate Estimation Time Window-** Maximum number of production period to be used for estimating the historical measure of event activity for comparing current level

Golden Tips:

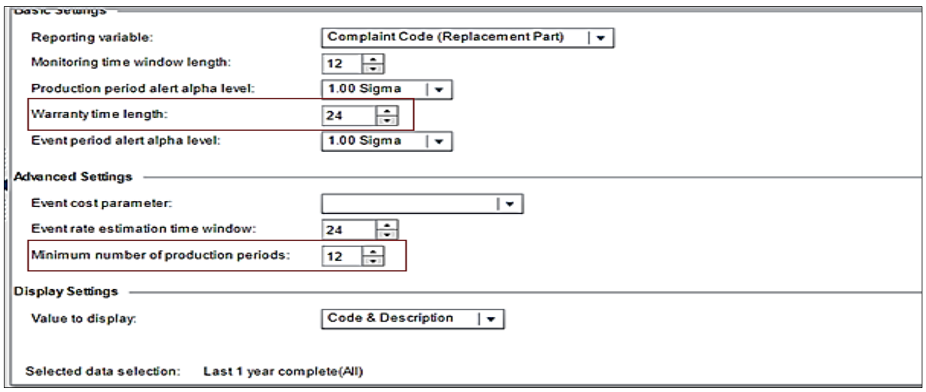
⊠ **Reconcile Analysis -** Enterprise Early Warning has the option to reconcile alerts where it can track existing alerts in Early warning, after every data refresh. Selecting the reconcile option provides previous alerts' issue dates to be the same while updating the refreshed ones with the current date. Thus, it is easy to track the new alerts against old.

3.2.2 Ad-hoc Analytic Early Warning

The ad-hoc analytic process is used to monitor claims activity on a defined subset of data to analyze specific model/build period/parts etc. The only difference with Enterprise analytic is that it can be executed by any user as per their need.

Production Period and Event Period Methodology can be used in Enterprise Analytic as well as Ad-hoc Analytics.

Some important field settings for Ad-hoc analytic EW under Basic and Advance Settings-

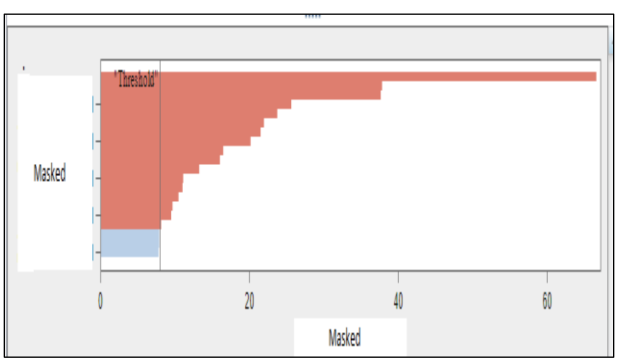


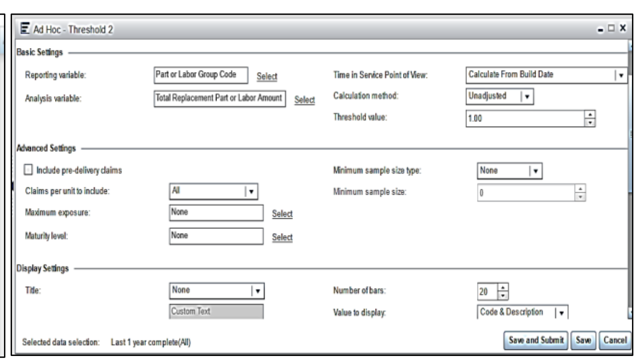
Display 8: Basic and Advance Settings of Ad Hoc – Analytic

- **Warranty Time Length-**Time Duration (in Months) for the units being analysed in the Warranty Program.
- **Minimum number of production Periods-** Minimum number of groups (corresponding to the reporting variable) which can be included in the analysis.

3.2.3 Ad hoc - Threshold

The Ad hoc threshold EW monitors values of a variable (called the *reporting variable*) on a defined subset of data (specified by a data selection) to determine whether values of a calculated quantitative value (called the *analysis variable*) surpass a specified threshold value.





Display 9: Ad Hoc Threshold Output

Display 10: Basic and Advance Settings of Ad Hoc – Threshold EW*

*Detailed description of settings in Ad Hoc - Threshold will be found in table 5 below

Refer below Table for difference of similar settings among above three EW analysis

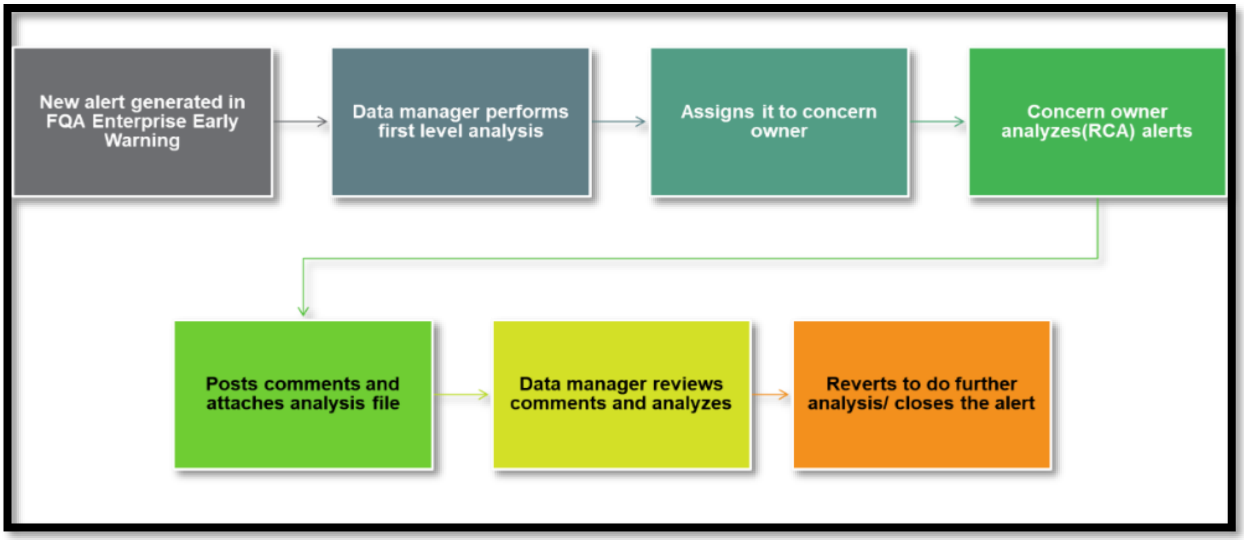
	Enterprise Analysis	Ad hoc Analysis	Ad hoc Threshold
Targeted Users	It is made for interest across enterprise	Configured to be of interest to a specific user or group	
Reporting Variable	Two fields for Reporting Variable- *Product Reporting Variable* (for product attributes) and *Event Reporting Variable* (for event attributes)	Only one field for Reporting variable in which either product or event option can be chosen	Single reporting field like Ad hoc analysis
Data Grouping for Analysis	On the basis of- One or more product reporting variables and One or more event reporting variables	On the basis of- Only one Reporting variable is chosen (either product or event)	
Execution	It refreshes automatically as per pre-defined schedule, if needed it can be force executed by administrator.	Ad Hoc – Analytic analysis is run once when submitted manually, and not updated.	Analytic process is typically made up of multiple runs of each model
Seasonal Factor	Enterprise analysis, the output table includes a column with the value Yes or No indicating whether seasonal factors were applied.	Seasonal factors cannot be applied in an Ad hoc.	Threshold analysis is recommended for safety and regulatory issues, where an appropriate threshold is known
Set up	Must be set up by an administrator	Can be set up by individual users	

Table 5: Options available for different categories of Early Warning

3.3 ADDING/MANAGING SUBSCRIPTION

Early Warning has the option to Add and Manage the subscription of the users who can access and perform the analysis. Adding a subscription consist of two steps where initially a user is selected & second step is to allocate it to the Alert Group. The subscription to user can be deleted/updated using Manage Subscription option. Enabling e-mail alert option in Modify Subscription will send message to the user as soon as the alerts are refreshed. The option for Adding and Managing Subscription is available only for Admin.

The process from New alert generation to closing the alert (including the analysis performed while solving the issue) is shown in the flow chart below.



Display 11: Early Warning Alert Workflow - Assigning Alerts

3.4 DATA SELECTION WORKSPACE

3.4.1 Data Selection- A data selection is a saved description that specifies which data is to be analysed in analysis/early warning workspace. It is not the physical subset of data. When data selection is defined, it creates a view of the data from the FQA data mart. We cannot directly import excel, CSV file or any other data source file in FQA for analysis.

Data selection has the following characteristics:

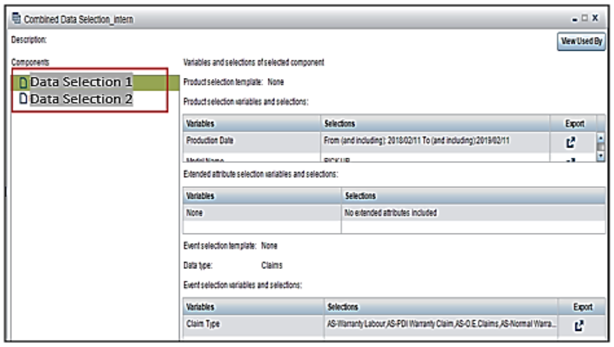
- It can be used with any analysis, including Early Warning analyses.
- It can be shared with other users.
- It can be stored to keep a subset of the mart persisted for later use.

Data Selection can be created of following two types:
3.4.1.1 Simple Data Selection**-** It Combines a product selection and an event selection data with only 'AND clause' between attributes.

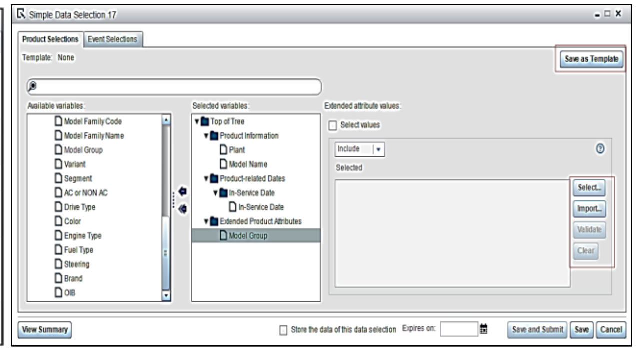
3.4.1.2 Combined Data Selection**-** This combines two or many simple selection, other combined simple selection or both for further analysis (Display 12). Combining existing data selections is more convenient than to create a new one.

Example- Model ABC was introduced in 2018, so there is little data for this model. Model XYZ was in production for several years and is still produced today. Some parts (001-005) of XYZ are common with ABC. We want to analyze the performance of those specific parts (001-005) for ABC only. In this scenario, Combined Data Selection can be used for analyses as:
(2014 – 2017 and Model XYZ and common parts 001-005)
OR (>=2018 and Model ABC and common parts 001-005)

3.4.2 Data Selection Template- It is a saved set of instructions that are used to create a data selection. It includes criteria for either products or events. They are created during data selection process and are optional. With templates, there is no need to store the underlying filtered data for every data selection.



Display 12: Combined Selection



Display 13: Option for creating a new template under new data selection

Data Selection in FQA has the option to import files for filtering values in bulk, which can be validated with system variable values post importing. To enter the new values, tabs or new lines can be used (commas are not allowed). Validate option is shown in Display 13.

3.5 ANALYSIS WORKSPACE

In the Analysis workspace, 14 different types of analyses can be created, submitted and interacted with analysis outputs. The output can also be exported for external usage.

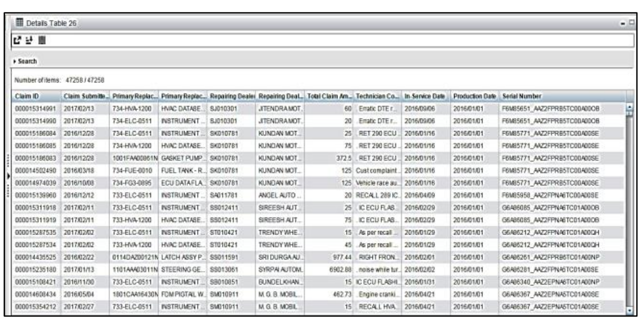
The following fourteen base analyses are available in SAS Field Quality Analytics.

Foundational (Beginner)	Intermediate (Practitioner)	Advance (Expert)
Details Table	Exposure Analysis	Decision Tree
Pareto Analysis	Failure Relationship Analysis	Event Forecasting Analysis
Summary Tables Analysis	Time of Event Analysis	Reliability Analysis
Text Mining Analysis	Trend and Control Analysis	Statistical Drivers Analysis
Geographic Analysis	Trend by Exposure analysis	

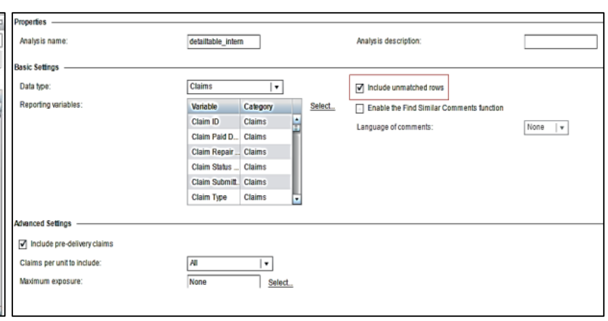
Table 6: Classification of Analysis with respect to Complexity

3.4.1 DETAILS TABLE ANALYSIS: -

Details Table Analysis gives a record by record view for better understating of the data. Detailed product or event (claims) records can be downloaded and viewed using this analysis.



Display 14: Detail Table Analysis Output



Display 15: Details Table Settings

Golden Tips:

⊠ **Include unmatched Row -** This includes data rows from both MnS and event data even in the case of not matching.

⊠ **Enable the Find Similar Comments function -** If one or more comment fields are included in the table, then similar comment can be found out w.r.t selected comments.

3.4.2 RELIABILITY ANALYSIS

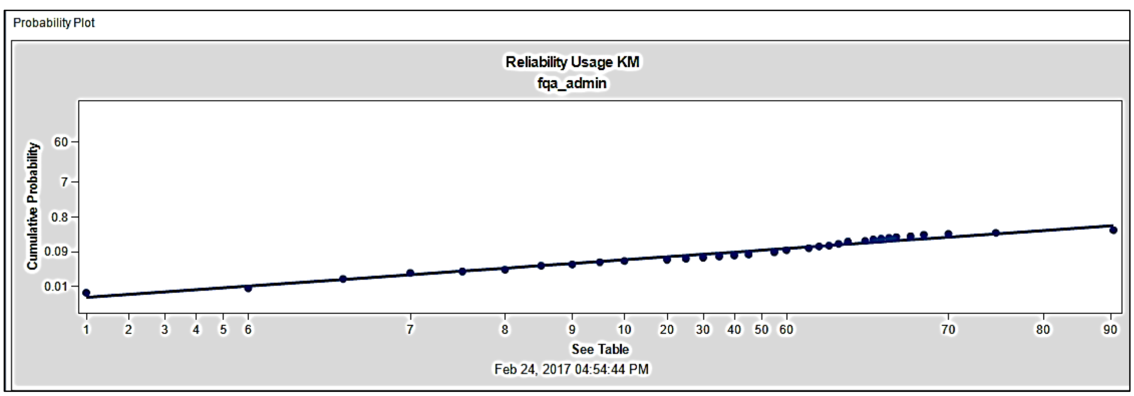
Reliability analysis is used to build parametric statistical distributions to predict event rates over *time-in-service* periods or *usage* intervals. It characterizes how a product functions in the field and durability of a product.

Reliability analysis is a class of statistical methods for which the outcome of interest is time until an event occurs. The process uses information about the units that failed as well as those which did not, to perform a parametric reliability analysis. Analysis models time to the first failure (occurrence of the first event on a unit).

FQA Reliability analysis results include a **distribution table (Display 16),** which provides details about the parametric model used to fit the data-

- *Failures*: Number of units with at least one event within the event data being analyzed
- *Suspensions*: Number of units that did not fail
- *Estimates* and *confidence* limits for the parameters of the parametric distribution
- The failures plus suspensions is equal to the number of units in the *population at risk*

Reliability output always includes a probability plot which displays the cumulative percent failed on the vertical axis and time-in-service or usage values on the horizontal axis. The axes are scaled based on the underlying theoretical distribution being used.



Distribution

Failures	Suspensions	Scale	Lower Probability Limit..	Upper Probability Limit..	Shape	Lower2 (95%)	Upper2 (95%)
1448	128137	4913407.00	3904129.00	6183599.00	0.94	0.89	0.98

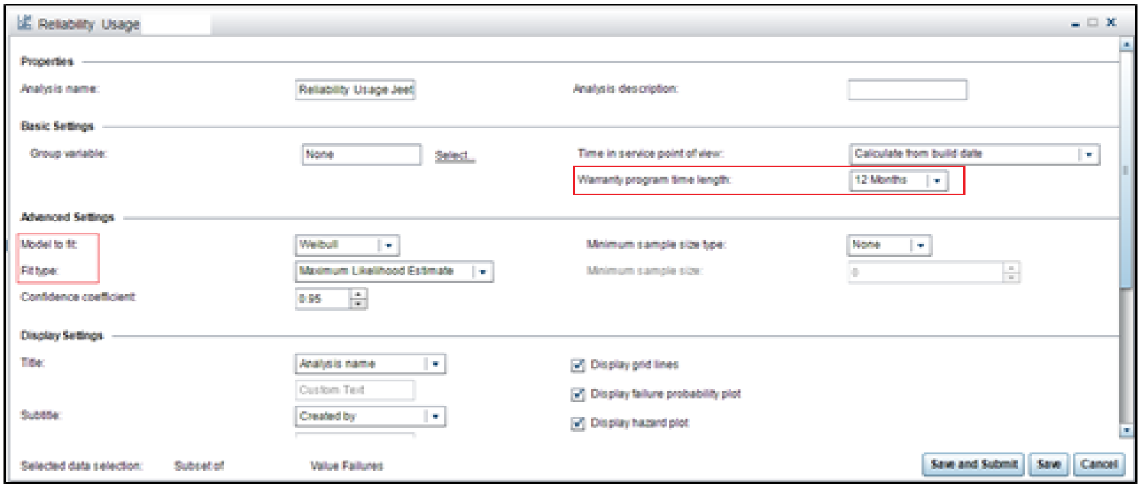
Display 16: Probability Plot and Distribution

The above display shows the output of reliability analysis using Weibull distribution which is one of the most frequently used life-data distributions due to its ability to model different failure modes of a product at different stages of its life.

Golden Tips:

☒ **Censoring –** Reliability Analysis uses the concept of censoring for the in-service products which have never failed. FQA automatically censors the data for suspended units.

Setting Options-



Display 17: Reliability Analysis Options (Model to fit and Fit Type)

Model to Fit: We must choose among four parametric lifetime distributions- Exponential, Lognormal**,** Weibull and Bi-Weibull to fit to the data and choice for model to fit. The Model fit determines the parameters that appear in the distribution table.

The Weibull distributions has a **scale and a shape** parameter while others have only scale parameter. The shape parameter helps define the shape of a distribution. The scale parameter defines where the bulk of the distribution lies, or how stretched out the distribution is.

Display 16 (Distribution) shows the failures are following an early-failure trend (shape <1), implying there is **decreasing failure rate intensity with time/coverage**.

Reliability analysis has following additional options other than Weibull distributions-

- **Exponential** – This distribution is mainly used where behaviour of units has a constant failure rate. Mathematically, it is a simple distribution which gets inappropriately used in some situation. Having $\beta=1$ as shape parameter is considered as special case of Weibull distribution.
- **Lognormal -** The lognormal distribution is commonly used to model the lives of units whose failure modes are of a fatigue-stress nature. As may be surmised by the name, the lognormal distribution has certain similarities to the normal distribution.
- **Bi-Weibull -** The Bi-Weibull distribution is a mixture of two Weibull distributions. It has a scale and shape parameter for each of the two underlying Weibull distributions

Golden Tips:

☒ **Bi-Weibull Distribution:** The Bi-Weibull distribution is beneficial when failures occur due to two modes such as burn-in and wear-out.

☒ **Warranty Program Time Length -** Any events occurring after this time will be excluded when fitting the reliability model as of the current date.

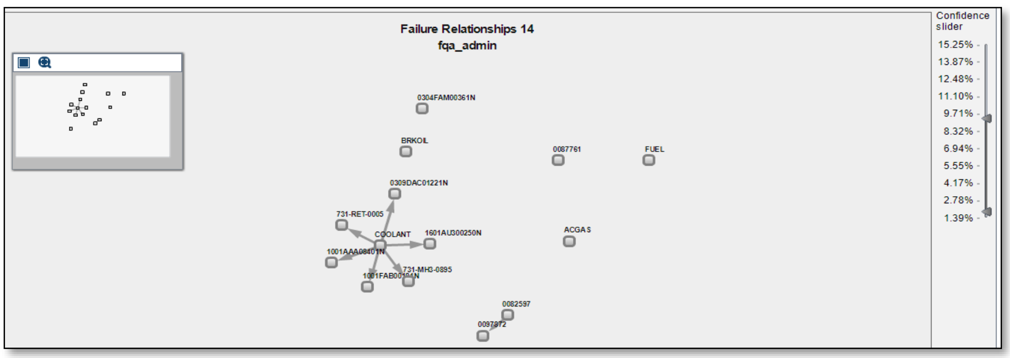
Fit Type: The Fit Type identifies the method that should be used when fitting the lifetime distribution to the data:

- **Maximum Likelihood Estimate** is recommended because it is influenced less by outliers and uses more information about the age of units in the population at risk
- **Least Squares method**: Sometimes the Maximum Likelihood Estimate method does not yield a solution. If this method is chosen, and a solution cannot be reached, SAS FQA will automatically use the **Least Squares method**

3.4.3 FAILURE RELATIONSHIP ANALYSIS-

The failure relationship analysis uses association and sequence analysis to identify relationships between different failure modes. Association analysis identifies relationships that occur at the same time on the same unit while sequence analysis identifies relationships that occur across different points in time for the same unit.

Failure relationship analysis produces a set of rules, each having a left and a right side. For each rule the analysis generates a set of metrices that quantifies the strength of the relationship between the two sides.



Display 18- Failure Relationship Result

Display 18 shows the relationship between Coolant and other parts in which coolant is the cause for failure of other parts.

<u>Key Terminologies</u>

Support: Support is calculated as the percentage of transactions that contain all items in an itemset. Rules with a high support are preferred since they are likely to be applicable to many future transactions.

Consider the relationship D-001 ➡ R-007 in a data selection that has 967 units with claims.

	R-007 Absent	R-007 Present	Total
D-001 Absent	560	154	714
D-001 Present	230	**23**	253
Total	790	177	**967**

Table 7: Illustration for calculation of support

Support = (Number of units with rule / Number of units)
Support = (23 / 967) = 0.024 or 2.4%

Confidence- The probability that a transaction contains the items on both side of the rule i.e., Left- and Right-Hand Side. The higher the confidence, the greater, the return rate you can expect for a given rule.

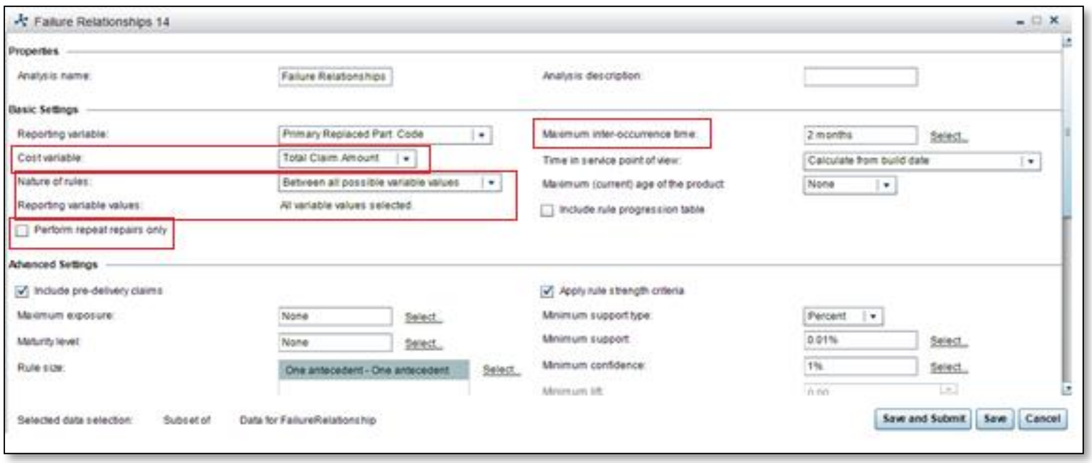
Example- **Rule-** D-001➡ R-007

	R-007 Absent	R-007 Present	Total
D-001 Absent	560	154	714
D-001 Present	230	**23**	**253**
Total	790	177	967

Table 8: Illustration for calculation of Confidence

Confidence = (Number of units with rule / Number of units with antecedent)
Confidence = (23 / 253) = 0.0909 or 9.09%

Setting Options-



Display 19: Failure Relationship Options

> **<u>Golden Tips</u>:**
>
> ⌧ **Cost Variable –** The cost variable is used to calculate a total cost associated with each rule. Also used with the Minimum Rule Cost setting to limit the number of rules displayed in the results.

- **Nature of Rules and Reporting Variable Values**

The following four Nature of Rules setting determines which relationships to include:

Between all possible variable values - Includes rules between all reporting variable values in the data selection
Between Reporting Variable Values - Includes only rules between specified reporting values
Ending with Reporting Variable Values
Starting with Reporting Variable Values

Table 9: Options under Nature of Rules

If user chooses any setting except for the first, the reporting variable values of interest must be specified

- **Perform Repeat Repairs** only setting determines whether all repairs are included or only repeat repairs are included.

- **Maximum inter-occurrence time:** The maximum time required by one replaced part code to follow up with another replaced part code is known as Maximum inter-occurrence time.

For example, consider a set of vehicles where 20 claims have a A-002 replaced part code and a B-006 part code is replaced after some time on those vehicles (an association rule could be A-002 ==> B-006). If the A-002 replaced part code is followed within TWO months by a B-006 replaced part code, then the inter-occurrence time would be 2 months.

3.4.4 TEXT MINING

Text Mining is analytical model to recognize patterns in text. It enables users to **segment observations into clusters based on a text or comment field.** It is useful for reducing the manual effort of reading through numerous comments and for increasing the reliability of finding patterns within comments. It also enables users to identify groups of events that contain similar comments within the data selection and determine how product and event attribute vary across these groups.

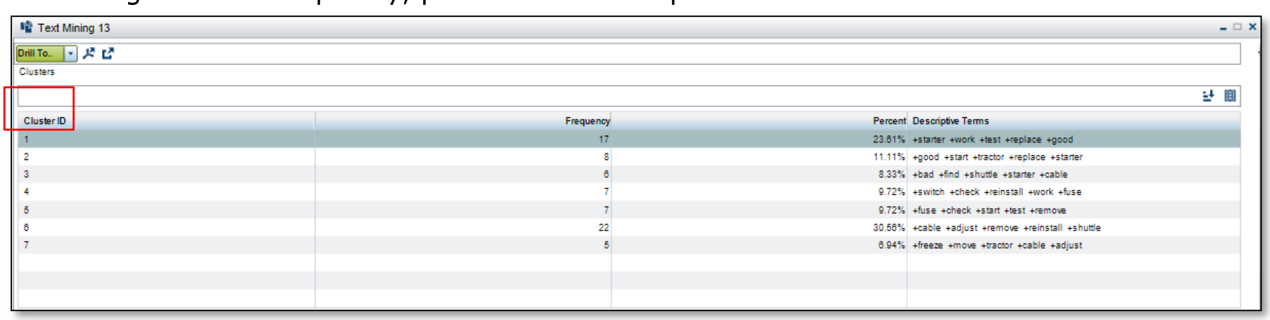
When the comment fields are parsed and are transformed into a numerical representation, clustering is performed. Clustering divides comments into mutually exclusive groups so that the observations for each group are as close as possible to one another and are as far as possible to other groups.

Text Mining results are displayed in three sections:

- Clusters
- Cluster Profile
- Cluster Details

Cluster

The Clusters section identifies the groups of similar comments that were discovered in the data along with the frequency, percent and descriptive terms.



Cluster ID	Frequency	Percent	Descriptive Terms
1	17	23.61%	+starter +work +test +replace +good
2	8	11.11%	+good +start +tractor +replace +starter
3	6	8.33%	+bad +find +shuttle +starter +cable
4	7	9.72%	+switch +check +reinstall +work +fuse
5	7	9.72%	+fuse +check +start +test +remove
6	22	30.56%	+cable +adjust +remove +reinstall +shuttle
7	5	6.94%	+freeze +move +tractor +cable +adjust

Display 20: Cluster Summary Table

Seven clusters are created, and the content of each cluster is characterized with five descriptive terms. 17 events representing about 23% of the events are contained in Cluster 1. The terms starter, work, test, replace, good are most descriptive of this group of comments. The + symbol next to a term indicates that the term has synonyms.

Cluster Profile

The Cluster Profile shows how the reporting variable values are distributed in a given cluster versus the overall population of events when there is a statistically significant difference.

The distribution of Model Codes across the events within Cluster 1 is slightly different from that for all events within the data selection being analyzed.

Display 21: Cluster Profile

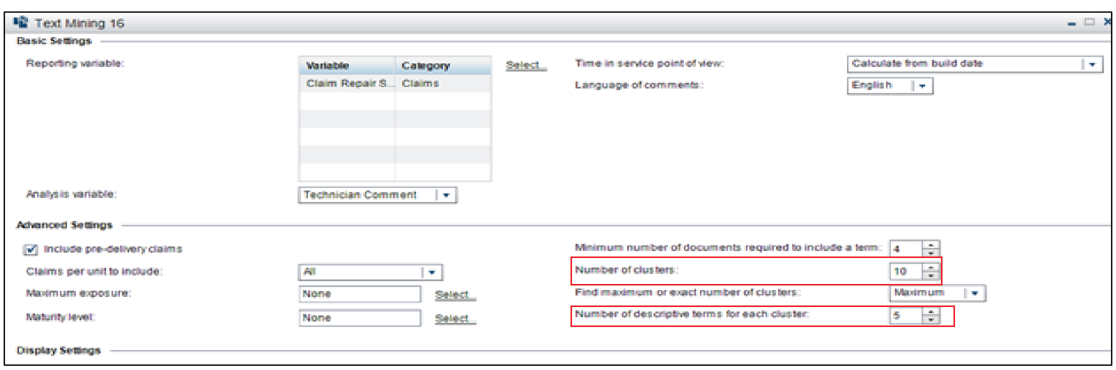
Cluster Details

The Claims Details section is populated when a row is selected in the Clusters section with information for the corresponding cluster. This is showing the reporting and analysis variable (Comment) for the events in the selected cluster.

Cluster ID	Technician Comment	Claim ID	Production Month
1	VEHICLE LOW PICK UP/FOUND INTERNAL DEFECT IN VACCUM MODULATOR ASSY/REPLACED.	000015452298	04/2016
1	. EMS DATASET FLASHING 731	000015179458	03/2016
1	ENGINE NOISE : Found that vaccum modulator noise so replaced both.	000015309401	04/2016
1	. To avoid falling down of the existing balancing weight from radiator fan. We replaced the radiator fan ...	000015046634	03/2016
1	. RADIATOR FAN WEIGHT REPLACEMENT 731/732	000015049809	04/2016
1	. EMS ECU, SOFWATER UPDATE TO FIX ABNORMAL ENGINE NOISE FLASH EMS ECU WITH LATEST DA...	000014903470	03/2016
1	RPM DROP__FOUND VACUUM MODULATOR EGR NOT WORKING PROPERLY	000015211631	03/2016
1	. RADIATOR FAN WEIGHT REPLACEMENT 731/732 RECALL FITMENT.	000015006091	03/2016
1	. ENGINE ROUGH NOISE IS MORE SOMETIMES/FOUND NOISE COMING FROM VACCUM MODULATOR A...	000015203282	03/2016

Display 22: Cluster Details

Setting Options-



Display 23: Settings Available in Text Mining

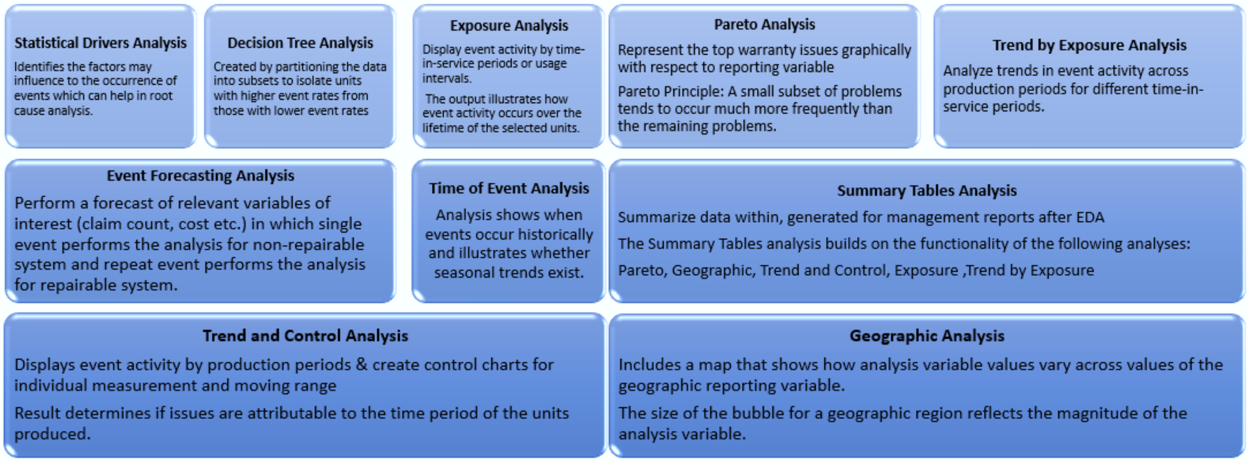
Number of Cluster: This can be given as input based on the requirement which is:

- Maximum number of Cluster **OR**

  Exact number of Cluster

Number of Descriptive Terms for each Cluster: An integer value N is specified for the number of terms used to describe a cluster. The top N terms determined to be most descriptive of each cluster appear in all the Output of Clusters.

3.4.5 OTHER ANALYSES IN FQA ANALYSIS WORKSPACE



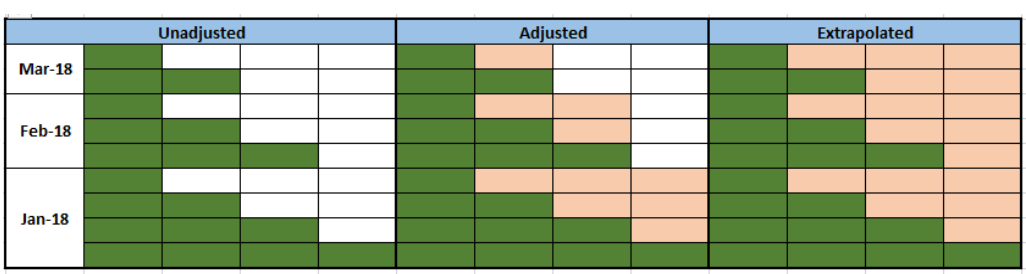
Display 24: Various other Analysis with description

4. IMPORTANT FIELDS IN BASIC AND ADVANCED ANALYSES SETTINGS

Important settings of different type of Analysis is explained. Different Analysis have unique parameters considered on which the output is dependent.

Calculation Methods (refer to **Display 25**, a pictorial representation of all methods)	**Unadjusted**	This option is chosen when we want to know what occurred up to the current data refresh date. For example, how many claims happened so far?
	Adjusted	It does not attempt to estimate what happens for a production period beyond the age of the oldest unit for that period.
	Extrapolated	This Method uses information about older production periods to estimate what happens for newer production periods as they continue to age.
Maturity Level		Maturity Level identifies the minimum time that a product must be in service so that the unit and any event activity on that unit is counted in the analysis.
Maximum Exposure		Maximum Exposure identifies the largest event time-in-service interval to include in the analysis.
Analysis Variable		It is quantitative measure of event activity that is calculated in the analysis.

Table 10: Different Parameters considered in Settings



Display 25: Types of Calculation Methods

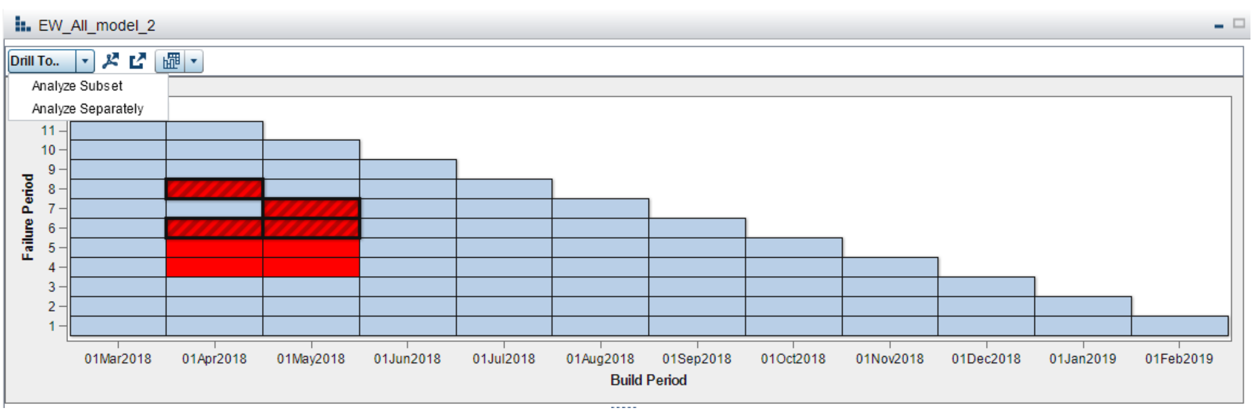
Analysis Settings	Decision Tree Analysis	Detail Table Analysis	Event Forecasting Analysis	Exposure Analysis	Failure Relationship Analysis	Geographic Analysis	Pareto Analysis	Reliability Analysis	Statistical Driver Analysis	Summary Table Analysis	Text Mining	Time of Event Analysis	Trend and Control Analysis	Trend by Exposure Analysis
Basic Setting Options														
Calculation Method	✓			✓		✓	✓		✓	✓			✓	✓
Reporting Variable	✓	✓			✓	✓	✓		✓	✓	✓	✓	✓	✓
Group Variable				✓			✓	✓				✓	✓	
Analysis Variable			✓	✓		✓	✓			✓	✓	✓	✓	✓
Advanced Setting Option														
Maximum Exposure	✓	✓		✓	✓	✓	✓		✓	✓	✓		✓	
Maturity Level	✓	✓		✓	✓	✓	✓		✓		✓		✓	✓

Table 26: Settings available for each Analysis

5. SUB SETTING EARLY WARNING FOR ANALYSIS

The Build Period for which alerts are generated in the Early Warning section can be further taken to perform various detailed analysis. Alerts generated in the process are chosen based on Reporting Variable. A New Data Selection gets created for the selected alert (choose any alert in the available graph and click Analyze in project option). Thus, a Simple data Selection with EI Enterprise type gets created in analysis workspace.

Selecting the targeted area (Red Region) in the newly created EI Enterprise analysis (in Analysis Workspace) and then selecting the Analyze subset in the "Drill to" option available as shown in the Display 27. This enables to perform one or more of the 14 analyses on this subset. The insights from the analyses aid the quality teams to pin-point the exact issues in the specified build periods.



Display 27: Sub setting Early Warning for Analysis

CONCLUSION

Manufacturing industry often struggles to identify and apply the right insights in spite of wealth of data captured continuously. SAS Field Quality Analytics is a tool which uses a mix of predictive and descriptive analytics solutions to identify the failures on the field in advance and enables the quality teams to take correct and informed decisions.

Despite the widely prevalent descriptive dashboards for warranty analytics, there is need for predictive and prescriptive analytical solutions to support strategic decision making.

SAS FQA Early Warning analysis addresses top concerns on priority, before they become big problems for the customers or manufacturers. Deploying alerts from Enterprise Analytic Early Warning can be a game changer in proactively identifying emerging issues and reducing the time to act on emerging issues. Quality teams can prioritize the efforts on warranty issues based on severity and scores generated in early warnings. Claims prediction and reliability analysis can support high impact decisions like vehicle recall maintaining brand reputation and customer loyalty. Several other analyses like Text mining, Pareto, Failure Relationship analysis, Trends analysis etc. enable the quality team to get the complete picture of failure movements, easily identify the root cause and take prioritized actions.

REFERENCES

[1] Huaiqing Wu, William Q. Meeker (23rd Mar 2001) Early Detection of Reliability Problems Using Information from Warranty Databases in Iowa State University

[2] White Paper from Capgemini "Connected Insights through In-Field Quality Analytics" (2015)

[3] SAS FQA User Guide

[4] SAS Paper 2905-2018, "Unleashing the true Power of SAS® For Warranty Analytics for Automobile Industry"

[5] Shaomin WU, Warranty Data Analysis: A Review

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Prateek Kumar Singh
Bristlecone India Ltd.
2nd Floor, Techniplex 1, Techniplex Complex, Opp. V.S Flyover, Goregaon (W),
Mumbai 62, Maharashtra, India
Phone: +91-9987990512
Email: psingh0930@gmail.com

Saket Shubham
Bristlecone India Ltd.
2nd Floor, Techniplex 1, Techniplex Complex, Opp. V.S Flyover, Goregaon (W),
Mumbai 62, Maharashtra, India
Phone: +91-8348522381
Email: saket.shubham@gmail.com

Shiv Kumar
Mahindra & Mahindra, Information Insight Center (IIC).
4th Floor, Mahindra Towers, Kandivali East, Mumbai,
Mumbai 400101, Maharashtra, India
Phone: +91-8451999852
Email: kumar.shiv7@mahindra.com

Paresh Peter Rodrigues
Bristlecone India Ltd.
2nd Floor, Techniplex 1, Techniplex Complex, Opp. V.S Flyover, Goregaon (W),
Mumbai 62, Maharashtra, India
Work: +91-7219663460
Email: paresh590@gmail.com