Building Science Institute, Ltd. Co. Procedure G-2023 Data Visualization & Analysis

Data Visualization

Process

1. Value Map
    a. Value System: aggregation of value chains that deliver value to customer & enable competitive advantage
    b. Value Chain: sequence of activities that provide specific service/product delivery to customer
2. SIPOC
    a. Suppliers
        i. Who supplies the inputs?
        ii. Written down AFTER "Inputs" are written down
    b. Inputs
        i. Information
        ii. Equipment
        iii. Materials
        iv. Supplies
        v. What inputs are needed to successfully complete the "Process step"?
        vi. Written down AFTER "Process steps" are defined
    c. Process steps
        i. Process is broken down into 5-7 high level activities (steps) in sequence
        ii. SIPOC analysis starts with "Process steps"
    d. Outputs
        i. What has been transformed in this "Process step"?
        ii. Written down AFTER "Process steps", "Inputs", and "Suppliers" are defined
    e. Customers
        i. Who gets the output from this "Process step"?
        ii. Who needs to get them?
        iii. Written down AFTER "Outputs" have been described
3. Process Map
    a. Activities
        i. Process steps from SIPOC
    b. Roles
        i. Suppliers
        ii. Customers
        iii. Stakeholders
        iv. Practitioner
    c. Information
        i. Process inputs
        ii. Process outputs
    d. Technology
        i. All tech needed to complete process step

Graphical

Attributes

1. Check sheets
    a. Measure counts (defects)
    b. Identify defect types
    c. Requires understanding of Process Map to know what to measure & types of defects
2. Pareto charts
    a. Measure factors, defects, counts
    b. Summarizes defect data from Check Sheet
    c. Identifies critical few vs trivial many defect types

Variables

1. Histogram
    a. Understand distribution
        i. Central tendency

ii. Variation
iii. Outliers

b. 2 or 3 humps means additional processes mixed in
c. Leverages knowledge of Process Map variables & metrics to collect data on

Data Analysis

Root Cause Analysis

1. Cause and Effect diagram
    a. Understand root cause & contributing factors
    b. Requires knowledge from Process Map, Check Sheet, Pareto Chart, Histogram to help identify root causes
    c. "Bones" for Cause & Effect Diagram
        i. Generic
            1. Measurement
            2. Environment
            3. Materials
            4. Methods
            5. Machines
            6. People (always last)
        ii. Manufacturing
            1. Machine
            2. Method
            3. Material
            4. Measurement
            5. Mission
            6. Management
            7. Maintenance
            8. Manpower (always last)
        iii. Marketing
            1. Product
            2. Price
            3. Place
            4. Promotion
            5. Process
            6. Proof
            7. Performance
            8. People (always last)
        iv. Service
            1. Surroundings
            2. Suppliers
            3. Systems
            4. Safety
            5. Skill (always last)
        v. Data Quality
            1. Organizational changes
            2. Manual data entry
            3. Data integration
            4. Loss of expertise
            5. Data corruption by hackers
            6. Aging of data / data decay
            7. Inadequate validation in the data capture process
            8. Inefficient business process management and design
            9. Lack of common data standards, data dictionary, and metadata
            10. Data purging
            11. Data migration
            12. System upgrades
            13. Data cleansing programs
            14. Multiple uses of data and lack of shared understanding of data
            15. Business data ownership and governance issues

Improve & Control

1. Run charts
    a. See trends
    b. Identify improvements
    c. Investigate from Measurements & Analysis
        i. Factors
        ii. Defects
        iii. Root causes
2. Control charts
    a. See patterns
    b. Identify out of control conditions (assignable causes)
    c. Ensure stable processes
    d. Identify when process improves
    e. Leverages process knowledge
        i. Factors/variables
        ii. Metrics
        iii. Defects
        iv. Root causes

Statistical

1. Simple Ratio
    a. 1 - (number of undesirable outcomes / total outcomes)
        i. Measures ratio of positive outcomes to total outcomes
        ii. Useful to show continuous improvements over time
    b. 1 - (number of desirable outcomes / total outcomes)
        i. Measures ratio of negative outcomes to total outcomes
        ii. Useful to show reduction in negative outcomes over time
    c. Quality dimensions measured with simple ratio
        i. Free of Error
            1. 1 - (Number of units without error / total units) OR
            2. 1 - (Number of units with errors / total units)
        ii. Completeness
            1. Schema completeness: degree to which entities & attributes are not missing from schema
            2. Column completeness: function of missing values in a column
            3. Population completeness: degree to which expected population is not missing from column
            4. 1 - (number of incomplete items / total items)
        iii. Consistency
            1. Consistency of redundant values across tables
            2. 1 - (violations of specific consistency type / total number of consistency checks)
        iv. Concise Representation
        v. Relevancy
        vi. Ease of Manipulation
2. Min or Max Operation
    a. Handles quality dimensions that require aggregation of multiple data quality variables
    b. Compute the minimum or maximum value from the normalized values of individual data quality variables
    c. Min Operator
        i. Conservative
        ii. Assigns to the dimension an aggregate value no higher than the value of its weakest data quality variable which has been evaluated & normalized to between 0 and 1
        iii. Quality dimensions measured with Min Operator:
            1. Believability
                a. Reflects assessment of
                    i. credibility of data source
                    ii. comparison to common standard
                    iii. previous experience
                b. Each variable rated on scale of 0 to 1
                c. Overall Believability assigned minimum value of the 3 OR
                    i. Overall Believability computed as weighted average of individual variables

2. Appropriate Amount of Data
    a. Minimum of 2 Simple Ratios
        i. number of data units provided / number of data units needed
        ii. number of data units needed / number of data units provided

d. Max Operator
    i. Liberal
    ii. Individual variables may be measured by Simple Ratio
    iii. Quality dimensions measured with Max Operator:
        1. Timeliness
            a. Reflects how up-to-date the data is with respect to the task it's used for
            b. Maximum value of 2 terms:
                i. 0 or
                ii. 1 - (currency / volatility)
                    1. currency = age + delivery time - input time
                        a. age = age of the data when first received by system
                        b. delivery time = time when data is delivered to user
                        c. input time = time when data received by system
                    2. volatility = length of time data remains valid
        2. Accessibility
            a. Reflects ease of data attainability
            b. Maximum value of 2 terms:
                i. 0 or
                ii. 1 - (time interval from request by user to delivery to user / time interval from request by user to point at which data is no longer useful)

3. Weighted Average
    a. Appropriate if organization has good understanding of importance of each variable to overall evaluation of the quality dimension.
    b. Each weighting factor must be between 0 and 1
    c. Sum of weighting factors must equal 1

Approved by the Building Science Institute, Ltd. Co. Quality Council on May 22, 2023.
Approve: Kevin Burk, Erik Straite, Brian Christensen, Amber Wood
Reject: None
Not Voting: Wes Davis, Brett Dillon (Chair)