Proceedings of the 6th European Conference of the Prognostics and Health Management Society 2021 - ISBN – 978-1-936263-34-9

The Impact of Data Quality on Maintenance Work Order Analysis: A Case Study in HVAC Work Durations

Anna Conte[1], Coline Bolland[2], Lynn Phan[3], Michael Brundage[4], and Thurston Sexton[5]

[1,2,3,4,5] *National Institute of Standards and Technology, Gaithersburg, MD, 20899, United States*
michael.brundage@nist.gov
nestor@nist.gov

Abstract

Historical data from maintenance work orders (MWOs) is a powerful source of information to improve maintenance decisions and procedures. However, data quality often impacts an analyst's ability to calculate important Key Performance Indicators (KPIs) and analyze trends within a facility. Data quality itself can be impacted by missing data, inaccurate data, or lack of appropriate data. We recommend several strategies to investigate data quality issues. First, the end goal of an analysis (e.g., calculated KPIs) should dictate data quality requirements, and therefore, quality reduction investigation. Second, basic Exploratory Data Analysis (EDA) techniques can be powerful tools to discover signs of quality reduction. We contextualize these techniques to the maintenance management domain with examples. Since analysts rarely have access to "baseline" high-quality data (i.e., to compare against), we first develop a technique based on Survival Analysis that corrects for censored *MWO Closing date* entries. We then use this synthesized baseline to investigate the impacts of missing data on KPI calculations. Further investigation into the use of Technical Language Processing to find common human errors is needed, along with more community-driven techniques to perform quality corrections when discovered.

1. Introduction

A maintenance work order (MWO) documents maintenance activity in a facility. Refer to Table 1 for example MWO data. MWO data in combination with Technical Language Processing (TLP) techniques are increasingly used to calculate key performance indicators (KPIs) to inform decision making in an organization. However, MWO datasets commonly have poor data quality. During data entry, fields may be left empty on certain MWOs, rendering a MWO ineffective for analyses, or in extreme cases, meaningless. Other data quality issues include missing features (not having needed data fields at entry-time) or improperly or inaccurately recorded data in existing fields.

These issues are rooted in the fact that many fields in the MWOs are manually entered by maintenance personnel. These personnel are often under significant time constraints, or asked to adopt data-entry systems that mesh poorly with their needs. This environment leads to an increased chance for MWOs to be entered into the system incompletely or not-at-all. As a result, KPIs calculated from the MWOs can be inaccurate or misleading.

Several types of missing data can occur, including missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) (Kang, Hyun, 2013). Missing data can be classified as MCAR when data is missing in a way that is unrelated to the overall dataset features (e.g. what kind of data is being recorded, how often, etc) *and* unrelated to the values being recorded. MCAR is generally considered an unrealistic assumption for missing data except in some cases where data is missing by design or accident. MAR, meanwhile, implies that data goes missing dependently on dataset features, but is unrelated to specific missing values that were supposed to be present. Missing data rates may depend on the set of records that was collected, but not the content of those records (e.g. daily recording may lead to fewer missing record entries than weekly, etc.) If neither condition is met, the missing data is MNAR. In this case, data goes missing with probability that depends on the dataset features, *and* the intended contents of the record. MNAR data are especially concerning for analysts due to assessment and mitigation difficulty. Kang, Hyun explains that the only way one can attain an unbiased estimate of model parameters using MNAR data is to first model the missing data itself (e.g. the mechanism for going missing), before incorporating that model downstream.

Quantifying the effect of missing data can be difficult, since its very nature stems from what goes *unobserved*. Analysts do not generally have access to "clean" versions of this data, preventing them from measuring deficiencies in their KPI calculations. In this paper, we discuss several basic techniques to start inves-

Table 1. Example Maintenance Work Orders. Further elements beyond the columns illustrated here would typically be included, such as location, cost, labor-hours, etc. A version of this table first appeared in Navinchandran, Madhusudanan and Sharp, Michael E and Brundage, Michael P and Sexton, Thurston B (2021)

Asset ID	Problem	Open	Closed	Remarks
162545	HP and LP pumps INOP	2/09/07 07:57	2/13/07 06:23	Checked / No Problem Found
150428	Broken door clamp -hook bolt	2/09/07 08:34	2/11/07 13:19	camera ordered. Delivery 7/14
156997	St#5 motor inop/humming	-/--/-- --:--	-/--/-- 10:22	camera ordered. Delivery 7/14
150428	Saw blade spun on hub	2/12/07 06:12	2/11/07 13:52	
150428	Speed limit @ Spindle A exceeded	2/12/07 08:27	2/12/07 --:--	Complete
164243	Broken chain on loader	2/12/07 09:49	-/--/-- --:--	
156551	Encoder coupling broken	2/12/07 --:--	2/12/07 13:35	Remove Vacuum Plug
150428	Emergency retract solonoid failure	2/12/07 13:45	2/24/07 13:45	Replaced Spray Nozzles

tigating MWO quality issues. Then, we use these techniques to construct an artificial baseline, letting us examine how KPI calculations are impacted by these data inaccuracies. Our findings indicate that human-caused data quality errors are likely to be non-random. Instead, they may be inadvertently bound-up with the resultant KPIs in unexpected ways. Our case study serves as an example of how data quality issues can impact an analyst's ability to calculate reliable and relevant KPIs, which in turn affect decision making.

While data quality issues seem like an inevitable predicament, methods to combat these issues and minimize the effects of human error exist. One approach to decreasing errors in data input is to have technicians perform data entry using a graphical user-interface (GUI) that has predetermined functional categories. Many Computer Maintenance Management Systems (CMMS) use this technique to enforce strict categories, but datasets still frequently have errors. In fact, restricted-entry systems can increase error probability under some circumstances (Sexton, Hodkiewicz, & Brundage, 2019). Possible mitigations include designating a time for data-entry throughout the day, establishing a "buddy system" for new hires and assets, encouraging technicians to explain more details in MWOs when they are unsure, or using an interface for data entry that reduces learning curve steepness. Additionally, other studies illustrated the importance of ensuring data quality through steps that first ensure accurate analyses, especially concerning KPI calculations (Lukens, Sarah and Naik, Manjish and Saetia, Kittipong and Hu, Xiaohui, 2019). Using the framework described by Lukens, Sarah and Naik, Manjish and Saetia, Kittipong and Hu, Xiaohui while collecting data can minimize data quality issues, allowing for more reliable and accurate analyses. While these methods improve data quality, issues with MWO data still inevitably exist, and measurements for data quality's impact on KPI calculations are needed.

Maydanchik, Arkady (2007) describes data profiling as one approach to investigating data quality. Before the quality of a dataset is actually assessed, analysts profile the data. This can involve finding basic statistical information of the dataset, using distribution charts or value frequencies. By doing this, analysts can obtain a better perspective of the dataset's contents, which can inform data quality assessment decisions to be more optimal, compared to performing a data quality assessment with no prior information. Data profiling methods can be used to observe frequently occurring values and the distributions of the values, which is referred to as "attribute profiling". We will apply this concept in Section 2 with the use of Exploratory Data Analysis (EDA) techniques. This will demonstrate ways that MWO data quality might be diagnosed and provide a stepping-off point to determine KPI prioritization.

This paper provides an overview of strategies to detect issues, a novel technique to rectify issues using a representative dataset, and an initial foray into measuring the impact of missing data on KPI calculations. The paper structure is as follows. Section 2 provides an overview on investigating the quality of a dataset. Section 3 illustrates this methodology with a case study using Heating Ventilation and Air Conditioning (HVAC) maintenance data. Section 4 discusses how to measure the impact of these data quality problems and how to rectify these issues. Discussions are provided in Section 5 and conclusions and future work are described in Section 6.

2. Investigating Dataset Quality

There is no established assessment to tell whether a certain dataset is of high quality. A perfect dataset only exists in a utopian world; the "ideal" dataset for an analyst really depends on the specific needs of that analyst. In other words: data quality, as we use here, is only well defined with respect to an end goal. As the data is intended for use by an analyst, even perfectly recorded data that fails to meet the analyst's needs might be considered "of low quality" for that use-case. Therefore, understanding data quality requires a context-oriented approach.

Multiple users of a particular dataset may use it for different purposes, and thus data that is deemed "sufficient" quality for one user may not be appropriate for another. Even within the same company, a technician, for instance, may not need to use the same information as a business analyst. The quality of data can be impacted at different times in a workflow. Lukens, Sarah and Naik, Manjish and Saetia, Kittipong and Hu, Xiaohui (2019) specify the qualities that make a generally useful MWO dataset.

The quantity of data is also important: when statistical or Machine Learning (ML) techniques are used, the amount of "good" data often affects the quality of the overall estimation. Too little sample data can lead to generalizing assumptions that are biased, and misinterpretations may occur. Density is also another aspect to consider; MWOs are easier to categorize when similar maintenance issues recur. Despite it being virtually impossible to create a dataset that is perfect for every context, these characteristics set guidelines that data collectors can follow to make their dataset closer to "ideal."

2.1. Prioritize KPIs

As foreshadowed by the above mention of statistics, ML, and dataset size, it is clear that a dataset's quality is only well-defined with respect to what you want to *do* with it. The first step in investigating the dataset is determining the goal: what analyses are intended, and how they will improve maintenance operations (Helu, Moneer and Libes, Don and Lubell, Joshua and Lyons, Kevin and Morris, Katherine C, 2016). KPIs are calculations used to guide maintenance decisions using different metrics, like cost, time spent on a work order, or frequency of occurrences of a MWO type. Keep in mind that different stakeholders within an organization often have different needs. ASTM International published E3012-20 Standard Guide for Characterizing Environmental Aspects of Manufacturing Processes to help manufacturers identify procedures to find the appropriate KPIs for a facility by taking into account the different stakeholder perspectives (Astm E3012-20, 2020; Kibira, Deogratias and Brundage, Michael P and Feng, Shaw and Morris, KC, 2018).

In maintenance management, several KPIs derived from data elements in MWOs are described in Brundage, Michael P and Morris, KC and Sexton, Thurston and Moccozet, Sascha and Hoffman, Michael (2018). Each KPI is calculated from data "elements," such as date, time, or raw text elements (Iso, 2014). An element can be defined as the "relevant measurements for use in the formula of a key performance indicator" (Iso, 2014). For example, a time element could be a *Machine Down time-stamp*, while a raw text element could be a free text description of the MWO. Most KPI calculations require input from multiple elements. Understanding important connections between data elements and KPIs can help prioritize KPIs for specific analysis. Since a specific data element may feed into multiple KPIs, it is advantageous to find the most "important" data elements.

Even when data elements are not directly associated in a desired KPI formula, they may have indirect relationships with each other, revealing important patterns in the maintenance workflow (Brundage, Michael P and Bernstein, William Z and Morris, Katherine C and Horst, John A, 2017). When entire data elements are missing, the possibility of uncovering these relationships is reduced. For example, the particular technician performing a work order will have his/her own set of skills and experience which impact how efficiently their work is performed. In turn, this could affect time elements and time-related KPIs. However, since the technician's skill and experience are not generally recorded in the data, the relationship between this and the aforementioned KPI elements may be difficult to quantify and analyze (Navinchandran, Madhusudanan and Sharp, Michael E and Brundage, Michael P and Sexton, Thurston B, 2021). Investigating these key limitations and communicating analysis assumptions with stakeholders can aid in discovering underlying data quality issues.

Ultimately, there will be occasions where a previously unknown or irrelevant KPI becomes relevant in the course of operation. Understanding this possibility and planning for dataset adaptation going forward can also be considered a part of investigating dataset quality.

2.2. Exploratory Investigation Techniques

Within the MWOs, common patterns may reveal problematic data entry practices. For example, the systematic omission of a specific MWO element may indicate that technicians are purposely ignoring an element, rather than it actually being "missing". Another problematic data entry practice could involve the entry of a certain data element into the incorrect location, resulting in misrepresented data. Yet another example could be when certain numerical data points are missing, such as cost or asset IDs. Without understanding the impact of missing data elements, any related decisions made with these KPIs is uncertain.

A common technique for quickly investigating a dataset is Exploratory Data Analysis (EDA) (Tukey, John W and others, 1977). EDA is a broadly-defined set of methods to visualize and manipulate data in a way that can highlight underlying, latent patterns between elements. It includes numerical techniques like simple regression and dimensionality reduction, but even simple graphical tools like box-and-whisker plots, pairwise scatterplots, and histograms can be powerful aids in assessing data quality.

While many of these EDA methods are basic exploratory data analysis principles and strategies proposed by Ishikawa, Kaoru (1985), elaborating on their application to maintenance management will serve both as a guide for the reader to extend into other EDA techniques, and as a discussion aid for our subsequent case study. Below are some examples of EDA strategies that support MWO quality investigation:

Day-of-Week Histograms Bin time-stamps into day-of-week (or similar) histograms. This may reveal unexpected patterns indicating the data entry practices. For instance, if most *WO Completion time* timestamps occur on specific days, these times likely reflect some aspect of a management schedule and not the actual amount of time spent physically working on the maintenance task.

Category Value Counts Generate frequency tables or bar plots for categorical data columns. Doing so may reveal patterns that indicate how MWOs are entered. For example, if nearly all MWOs are being assigned the same priority level, this may just be a default data entry that is not being entered by the maintenance personnel and will not provide meaningful insight for analysis.

Visualize Numerical Data Generate box-and-whisker plots (or histograms, etc.) for numerical data columns. This is a simple way to visualize data distributions, and could reveal a pattern of how the data is entered. For instance, if a significant portion of a data element is "zero" or "null", the analyst may want to further investigate if these reflect a real value or are simply the default value from the CMMS. Similarly, if multiple columns have identical distributions, such as a *Planned WO cost* and an *Actual WO cost*, the underlying data entry technique should be investigated.

Duration and Difference Frequencies Calculate the value- or time-differences between data elements, then examine patterns. Box-and-whisker plots could again be useful here, or pairwise scatter-plots. For instance, if the time difference between the start and end of a work order is consistently zero or even negative (implying the work ended before it started), then one or both of these timestamps is inaccurate.

There are certainly many further techniques that can be borrowed from EDA; extrapolating such patterns of analysis can aid in investigating data quality.

2.3. Rectifying Data Quality Issues

Now that several methods of investigating data quality have been identified, it is important to discuss the next steps: what should an analyst do after evaluating the quality of a dataset? How can an analyst rectify some of these data quality issues?

Since missing data is a frequent problem that arises for MWO data quality, analysts could address this issue by modeling the missing data (Kang, Hyun, 2013). The assumptions used to create the model could come from interviews with maintenance management or personnel, personal expertise within the organization, or the results of EDA. These models enable an analyst to point out KPI calculations that are skewed by missing data. In some circumstances, these data quality issues may be mitigated or corrected through predictive or inferential models, such as imputation (Kang, Hyun, 2013). This is especially applicable if the KPI of interest is aggregated over a large set of observations (e.g. a long period of time), which provides a more robust estimate of uncertainty.

Another source of insight into the data are written comments or descriptions from the MWO. These natural-language (text-based) fields are often underutilized because they are prone to many quality issues. However, they can be a rich source of insight when treated with care (Hodkiewicz, Melinda and Ho, Mark Tien-Wei, 2016; Michael P. Brundage and Thurston Sexton and Melinda Hodkiewicz and Alden Dima and Sarah Lukens, 2021; T. Sexton and M. P. Brundage and M. Hoffman and K. C. Morris, 2017). To better understand a potential pattern in the missing data, an analyst can look for correlations between MWOs that have elements missing and the associated content of text-based columns. Annotation is one technique that helps structure free-form MWO text, which helps in the calculation and identification of these correlations. One tool that is specifically designed to overcome some of the quality issues that may arise in natural text is Nestor[1], which is a TLP toolkit that allows a user to extract structured data from raw maintenance text (Sexton, Thurston B and Brundage, Michael P, 2019). The output is a set of tags: each tag is a unique token that represents a specific concept in the maintenance data, and each MWO can be assigned multiple tags. By grouping different spellings, formats, or abbreviations of maintenance terms into consistent terms, or "aliases", users can organize ideas that are the same, even when they appear differently, or as typos, in unstructured text data (Brundage, Michael P and Weiss, Brian A and Pellegrino, Joan, 2020).

Semi-structuring MWO text data through annotation and aliasing allows for improved coordination between the needs of maintainers (who are the source of MWO datasets) and the needs of analysts (who perform data analyses and interpret their results). For example, if date information is commonly missing with a particular solution tag, it may be worth investigating why this is the case. Correlations between occurrences of missing data, data quality issues, and the type of work being performed could potentially reveal underlying problems with maintenance management and data collection workflow. These findings could then be used to improve workflows to collect higher quality data.

The next section illustrates the application of these techniques for investigating data quality issues in MWOs using real-world MWO data.

3. Case Study

The case study makes use of historical HVAC MWO data of mixed use laboratory and office space, spanning ten years (2009-2019). The dataset contains a combination of scheduled maintenance, ad hoc maintenance, and occupant-submitted requests, such as temperature complaints. These work orders

[1]`https://www.nist.gov/services-resources/software/nestor`

primarily document maintenance of air handling units and exhaust fans. The MWO dataset had multiple free-text columns with written descriptions of the problem and completed work. These fields were tagged using Nestor (Sexton, Thurston B and Brundage, Michael P, 2019) and the output tags were subsequently used for data analysis in lieu of the free-text columns. This was done because, as described in previous sections, the semi-structured aliases are much more conductive to statistical analysis.

3.1. Case Study KPI Prioritization

The goal of the exploratory process demonstrated in this case study is to showcase methods which are applicable to a wide array of users. To that end, we understand that many analysts will not calculate all possible KPIs available to them, but only ones deemed relevant through a process of prioritization. Sometimes, a KPI's priority may be obvious to an analyst through organizational goals. However, one can also prioritize KPIs with data-driven techniques or by determining stakeholder needs through a consensus-based process (Astm E3012-20, 2020; Hester, Patrick and Ezell, Barry and Collins, Andrew and Horst, John and Lawsure, Kaleen, 2017). This type of KPI prioritization was the focus of Brundage, Michael P and Morris, KC and Sexton, Thurston and Moccozet, Sascha and Hoffman, Michael (2018), which presents time-related KPIs.

Another way to conceptualize the priority of a KPI — and therefore its importance — is to consider how calculating it will support the future calculation of other KPIs. This idea of importance-through-relationship, while possibly less directly relevant to short-term decision making, can be used for demonstration when direct stakeholder input is unavailable.

In this fashion, our case-study examines the relationships among KPIs to approximate their importance. We start with the KPIs and data elements defined in Brundage, Michael P and Morris, KC and Sexton, Thurston and Moccozet, Sascha and Hoffman, Michael (2018). The KPI dependencies on particular data elements can be represented as a bipartite network graph, where each "node" is a KPI or data element and an "edge" between nodes indicates the KPI depends on that data element (depicted in Figure 1). This graph, when represented as an adjacency matrix, can be thought of as a Design- or Dependency-Structure Matrix (DSM), which are common techniques for determining system dependency and component importances (Eppinger, Steven D and Browning, Tyson R, 2012). To determine the best KPI to use to demonstrate our procedure, we begin by performing a bipartite projection of our DSM with simple weighting, which means that our new graph of KPI nodes will be connected more strongly when they share more data elements in common (Zhou, Tao and Ren, Jie and Medo, Matúand Zhang, Yi-Cheng, 2007).

In network analysis, calculating a node's "centrality" is how

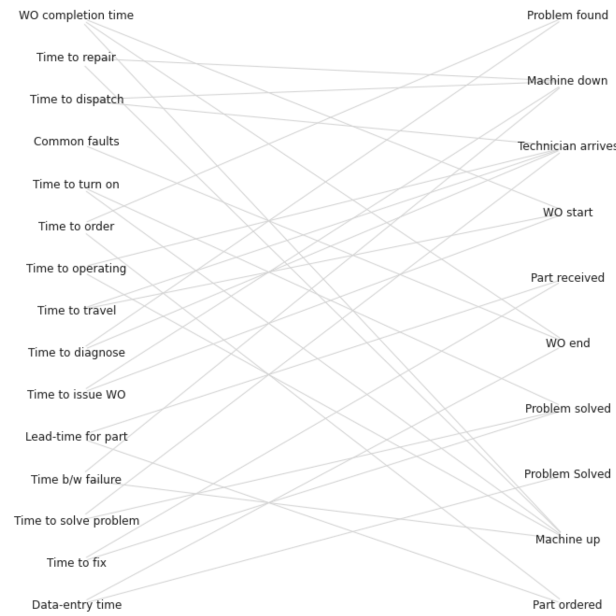

Figure 1. Bipartite Graph of Time-related KPI Elements. Nodes on the right are time-related data elements (timestamps of different stages of a MWO). Nodes on the left are KPIs, connected to their composing elements.

we go about determining its importance relative to other nodes. There are many techniques for performing this task, but for the sake of illustration we select a centrality measure that rewards highly connected nodes (i.e., the KPI deserves more attention if it requires data elements that can be used by many other KPIs) (Katz, Leo, 1953). Table 2 shows the resulting KPI rankings. These results were calculated by creating a bipartite network connecting the KPIs with their elements, projecting this graph to a network of just the elements, and then calculating their centralities.

The first two data elements listed (*Time-to-operating* and *Time-to-dispatch*) are certainly important to maintenance management operations, in general. However, in our representative dataset, the time stamps to calculate them were not available, as may often be the case. *Work Order (WO) completion time* also has a high centrality, so we proceed as discussed previously by investigating the data elements needed to calculate it: *MWO Reporting date* and *MWO Closing date*. *Work Order (WO) completion time* is the difference between these two elements (Brundage, Michael P and Morris, KC and Sexton, Thurston and Moccozet, Sascha and Hoffman, Michael, 2018).

To address the business needs of an organization, an analyst may report a KPI on a regular basis, such as monthly, weekly, or yearly. For demonstration purposes, we chose to examine KPIs on a monthly basis. We first aggregate the KPI by which month it would have been reported in and calculate *Average MWO Duration per Month*. In addition to the time-based KPIs

Table 2. Brundage, Michael P and Morris, KC and Sexton, Thurston and Moccozet, Sascha and Hoffman, Michael (2018) KPIs ranked by importance (Katz centrality on DSM projection).

Data Element	**Importance**
Time to operating	0.39
Time to dispatch	0.34
WO completion time	0.34
Time b/w failure	0.32
Time to repair	0.32
Time to travel	0.30
Time to turn on	0.29
Time to solve problem	0.28
Time to issue WO	0.27
Time to diagnose	0.23
Time to fix	0.12
Data-entry time	0.09
Common faults	0.09
Time to order	0.07
Lead-time for part	0.05

provided in Brundage, Michael P and Morris, KC and Sexton, Thurston and Moccozet, Sascha and Hoffman, Michael (2018), we considered other measures that are commonly found in maintenance datasets. Due to general importance within operations, and the typical inclusion of *cost* in MWOs, we also report the *Average Cost per Month*, which is, many times, of primary concern to business analysts.

3.2. Determining Data Quality

Using *Day-of-Week Frequencies*, we notice MWO reports are initiated on dates distributed approximately evenly throughout the work-week (around 20% for each weekday). However, the MWO closures mostly occur at the end of the week. Indeed, 19,597 MWOs out of 21,107 (around 93%) happened on Fridays, as opposed to the 4,000-5,000 we would expect from the uniformly distributed start-days. (See Table 3).

These counts imply that there is a maintenance management workflow where personnel close-out several MWOs at once. We often see batches of MWOs being marked as “complete” at regular or weekly intervals. Therefore, it is highly unlikely that these dates correspond to the “true” time a work order was completed in the field, potentially leading to incorrect estimates of work order duration. As discussed, a possible way to identify this effect in data is to compare the weekday distributions of the *MWO Reporting date* and *MWO Closing date*. A strong bias of *MWO Closing date* occurring on a single day of the week, especially when *MWO Reporting date* shows no such bias, is an indication that the *MWO Closing date* data entry does not exactly reflect the duration of labor hours associated with an MWO.

4. Measuring Data Quality Impacts

Because we applied the EDA strategies outlined in Section 2, it is possible to model the cause of our data quality reduction,

Table 3. Relative proportions of MWO opening and closing times for HVAC dataset, by weekday.

	MWO-open	**MWO-close**
Monday	0.22	0.01
Tuesday	0.24	0.01
Wednesday	0.21	0.01
Thursday	0.19	0.03
Friday	0.13	**0.93**
Saturday	0.01	0.01
Sunday	0.01	0.00

and for instructive purposes, further investigate what *would* happen if data were missing from our dataset.

Our primary observation is that MWO closing dates are likely inaccurate. Since they are being closed out approximately once every week, it is not reasonable to assume that all work orders happened to be completed on the same day. Instead we can place upper- and lower-bounds on when work on a given MWO was likely completed: namely, sometime between its reported close-out date and the close-out date immediately preceding it. We assume that the MWOs are not closed before the maintenance task is complete, so the close-out date would be the upper bound of when the work was likely completed.

These observations about the possible upper- and lower-bounds on MWO durations will allow us to model the KPI within a formal framework for duration-type data: Survival Analysis (Miller Jr, Rupert G, 2011).

4.1. Survival Analysis

To more accurately model the *Average MWO Duration*, we estimate the MWO “survival” functions. Survival functions represent the probability that, at time t, a random event at time T has not yet occurred:

$$S(t) = 1 - F_T(t) = P(T \geq t), \tag{1}$$

where $F(t)$ is the cumulative distribution function (CDF). This originates in healthcare statistics, where an “event” was an observed death (thus, “survival” function). In our case, the event of interest can be the time at which an MWO ends.

This can be done with both parametric models, like Weibull and Log-Normal, or non-parametric models, of which the Kaplan-Meier estimate is quite popular. Non-parametric models do not make any assumptions about the underlying form of the mapping function. Kaplan-Meier (KM) is a non-parametric estimator which does not use modeling assumptions for the underlying distribution $P(T = t)$, but rather estimates a probability of survival $S(t)$ up to each time, empirically:

$$\widehat{S}(t) = \prod_{i:\ t_i \leq t} \left(1 - \frac{e_i}{s_i}\right), \tag{2}$$

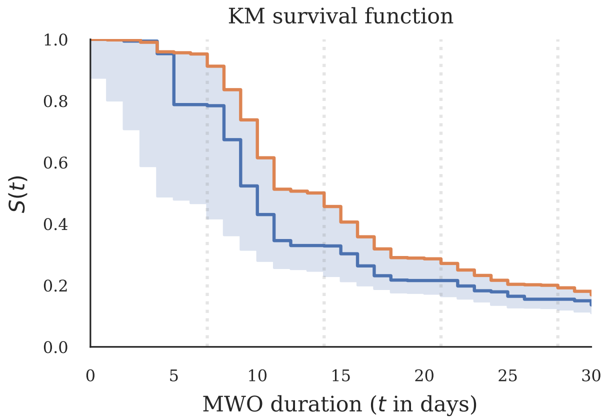

Figure 2. KM survival curves for MWO durations. Interval-censored estimate (blue line) is often significantly shorter-lived than the estimate that would have been produced without modeling data-censoring (orange line). The shaded region is the range of possible durations between the upper and lower bounds.

where e_i are the number of observed events at time t_i, and s_i are the "surviving" individuals remaining at t_i.

Since we do not know the true MWO closing times, but only an upper- and lower-bound for each, our data is *interval censored*. Consequently, we can compare the survival curve estimated by a KM model both with and without the censoring knowledge gained from our prior EDA. As mentioned, we assume that the real date when the MWO ends is sometime between the day reported in the dataset and the previous date when MWOs were reported to have ended (recall there are only 520 such days in ten years of data). The recorded *MWO Closing date* is the latest date the physical work of an MWO could have been completed, resulting in the longest possible MWO duration. The previously recorded end date for any MWO is assumed to be the first possible date the physical work of an MWO could have been completed, resulting in the shortest possible MWO duration.[2] In terms of data manipulation, this is a simple instance of finding unique dates and shifting the timeline to the previous entry date.

We apply the interval-censored KM estimator described by Giolo, Suely Ruiz (2004). All survival models were estimated using the `lifelines` Python package (Davidson-Pilon et al., 2021). When calculating the difference in time, date entries were rounded to the nearest day: if the difference is 0, the MWO duration is assumed to be between 0 and 1 days, etc. Resulting models can be seen in Figure 2.

At the starting point of the survival function, t = 0, no MWO experienced MWO closure. The further we advance in the timeline, the greater the probability that MWO closure will occur. After 20 days, almost 80 percent of the MWOs will have been closed. The difference between the survival function with and without interval censoring shows an overestimation of MWO duration if we had not used the insight from the initial EDA. For example, the non-censored estimate for $S(t)$ implies the median MWO duration should be 14 days (i.e. 50 percent of the MWOs are closed within this time), while the censored estimate (accounting for the our data-entry model) shows that the median duration is only 9-10 days This is a more-than 30% reduction from the non-censored median duration, which adds up to a significant difference in time spent. When direct estimates of labor hours are not available, simply using the MWO close date to calculate MWO duration would overestimate machine downtime or labor hours by a significant margin.

4.2. Rectifying Data Quality

There are many ways to address data quality for missing data, depending on what assumptions can be made about the missing data and what models can be applied (Kang, Hyun, 2013; Pigott, Therese D, 2001). While system level or design-based solutions to data quality are out of scope for this paper, it is sometimes possible for an analyst to use their knowledge of a KPI or the data elements to attempt rectifying data quality numerically.

Given the function $S(t)$, for instance, it is possible to approximate reasonable corrections to the *Average MWO Duration per Month*. For instance, one could use the distribution $P(T=t)$ to sample simulated durations for MWOs, and provide uncertainty bounds. In this example, the KM model that provided useful insights into non-smooth MWO entry behavior (e.g., $S(t)$ flattens close to week-markers), does not have an underlying probability density function for event occurrence. Instead, one could approximate it by applying inverse-transform sampling (Luc Devroye, 2006) to an interpolated version of the CDF, giving an empirical estimate for the "corrected" probability density function for MWO durations.

To rectify individual MWO durations with appropriate estimations, we take a similar approach using linear interpolation on an inverted CDF to approximate the MWO duration *quantile function*:

$$F^{-1}(p) = \min\{t \in \mathbf{R}^+ : F(t) \geq p\}, \quad p \in \{0, 1\} \qquad (3)$$

The quantile function maps a given probability of an event occurring, and returns the time t where the event T has that probability, i.e., $P(T=t)$.[3] The original probability $p_i = F_T(t_i)$ can be estimated with a non-censored KM model, and corrected using a quantile function derived from the interval-

[2]This assumes that any time a technician "sits down" to enter completed work, they enter all completions available during that session.

[3]This assumes $F(t)$ is continuous and monotonically increasing, else it returns the *minimum* time t where this is true. This can be mitigated using linear interpolation of the KM approximation.

censored data.

Using this point estimate as a form of central tendency, along with the original upper- and lower- bounds, we implement this idea in the form of a triangle distribution.

4.3. Baseline Generation and Missing Data's Impact

Together, the correction technique with an estimate of uncertainty about it (using a triangle distribution) can be used to create a baseline dataset. We sample "ideal" datasets repeatedly from what is now a distribution over corrected durations for every MWO.

To examine how missing data impacts data analysis, we use this "ideal" dataset to test how removing data affects the calculation of KPIs from it.

We use the correction and uncertainty technique described above to generate 30 different sets of "corrected" durations. These were used as part of a Monte Carlo (stochastic simulation) estimate of the *Average MWO Duration per Month* and *Average MWO Cost per Month*, with the monthly aggregates occurring on the now-corrected MWO close-dates.

To test how missing data affects KPIs, we first must model the mechanisms by which data *goes missing*. As a benchmark, we say that data is MAR with some probability. As we increase this probability, we expect the variation between the Monte Carlo experiments to increase, since there are fewer samples from which to correctly estimate our KPI mean value. Here we repeat the Monte Carlo experiment by varying the fraction f of "missing data": remove data in 10% increments, resulting in separate experiments for data ranging from 0% missing to 90% missing. For each of these experiments, we calculate the KPI of *Average MWO Duration per Month* and *Average MWO Cost per Month*.

More realistically, data will be MNAR. There are many potential mechanisms for this type of data loss, but to create our model, we assume that data entry is less likely to occur if the maintenance job being performed is relatively rare or unknown. This could correspond to a CMMS having a limited selection of drop-down menu options, for instance. Additionally, rarer maintenance events and work to address emergency situations may involve more follow-up testing or other irregular facets of the workflow. Closing out these types of jobs may be more tedious and therefore more likely to be incomplete or left open longer than necessary. With this scenario in mind, we make use of the Nestor tags described previously: a "missing data" fraction of f corresponds to MWOs containing a tag in the bottom f-percentile frequency, which are considered "rare". These MWOs are dropped. Once again, we calculate both example KPIs for each experiment. The results for both mechanisms of data loss can be found in Figure 3.

Note that when data is randomly missing, the calculated KPIs do not vary significantly; however, the calculated KPIs stray from the actual value when missing data is associated with particular tags. We discuss this result further in the next section.

5. Discussion

The Monte Carlo simulations show how missing data affects the reliability of KPI calculations. When data is randomly lost throughout the dataset, the KPI uncertainty increases when less data is available, as expected. However, the mean remains relatively stable and reliable. On the other hand, when data loss is associated with particular types of work (MNAR), the mean drastically changes with data loss, as seen with KPI calculations from datasets with tag-based data loss. If data is missing as a function of MWO content or of the type of work performed, then even with large sample sizes, a calculated KPI is likely to be inaccurate. Notably, the KPI values do not change consistently with a reduction in data quality; when data is MNAR, the KPI calculations depend on which types of data are missing. To ascertain the reliability of a KPI calculation, an analyst must also understand how the data quality may be related to the type of work being performed within the MWOs. Non-random, human-caused data quality issues can impact KPIs in significant ways. Assuming that a systemic bias against certain issues being recorded does occur, the end result would be that any KPI measured would inherently be inaccurate and possibly misleading. Textual information within the MWOs is integral to uncovering these patterns in a dataset.

As seen in the experiments where missing data is based on tag frequencies, a dataset may have underlying data quality patterns related to the type of work being performed in the field. TLP techniques can be helpful to further understand the nuances of these patterns in a particular dataset. For instance, an analyst can examine if certain tags (or topics) are more likely to be associated with missing or inaccurate time or cost data. In order to analyze the concepts within historical MWOs from textual data, and not simply the numerical data, a strategy to structure the free text data is needed. One such method to process a dataset's text is Nestor, as described previously. By using this toolkit, more data can be effectively mined to inform patterns detected in numerical data, increasing analysts' and stakeholders' understandings of patterns seen in KPI calculation results.

TLP techniques, like Nestor tagging, can also be used to explain why data goes missing and which topics are related to the missing data. Nestor orders tags by TF-IDF score, putting aliases in order from the most common, or highest-scoring tags to the least common, or lowest-scoring tags. Through Nestor tagging, since similar words and letters are grouped together in aliases, analysts can look for associations between rare aliases and missing data. If this association exists, this

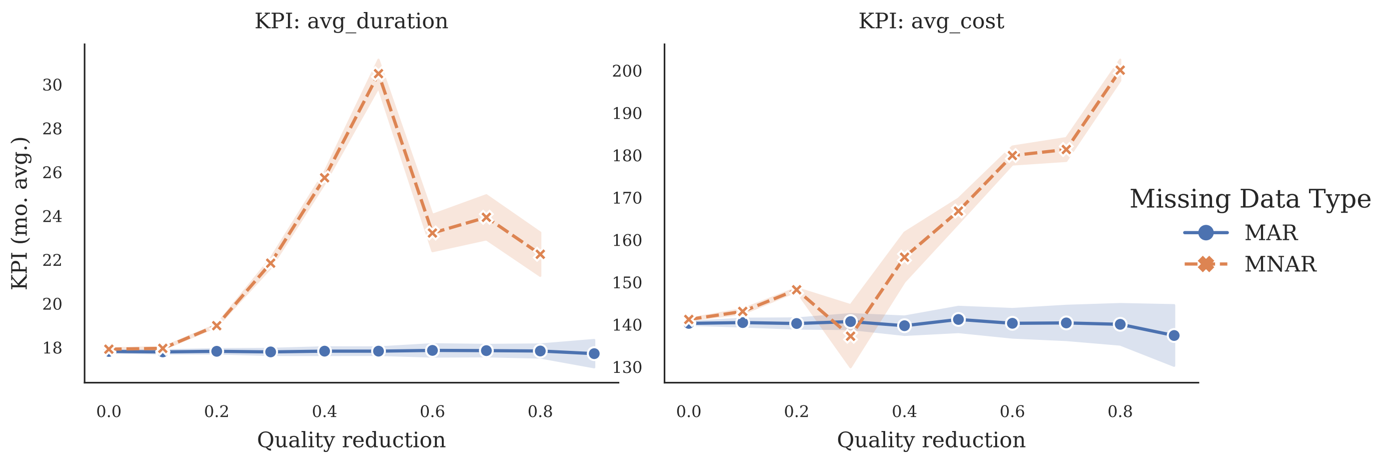

Figure 3. Effect of Missing Data on KPI Calculation. *Average MWO Duration per Month* and *Average MWO Cost per Month* are impacted as a greater fraction of the data is missing (*Quality reduction*). When data is randomly missing (MAR), the calculated KPIs do not vary much from their actual mean. However, when missing data is MNAR, and related to certain MWO content (a tag-based data quality reduction), the KPIs stray from the actual mean, and their accuracies do not devolve consistently with a reduction in data quality.

may indicate that at the point of data entry, technicians may be entering MWOs with missing data because they are rare, and technicians may be unsure how to describe the rare issue. Associations like these can help to identify why data is seemingly missing, but is actually a result of technicians choosing to withhold information due to a lack of understanding of a rare topic.

While the methods presented in this paper are not the only methods that can evaluate the effects of missing data, they provide a replicable strategy that future analysts can use to gauge whether or not they have a major amount of missing data, and if so, the magnitude of the impact of missing data on a MWO dataset. The same logic of the dataset evaluation methods presented in this paper can be applied to other facilities as well. For example, analysts in another facility could evaluate their dataset using one of our proposed methods (such as Day-of-Week histograms) and may notice that their MWOs tend to start mainly on a specific day of the week. This would alert the analysts to be aware of potential inaccuracies in analytic results related to the KPI element of *MWO Reporting date*. As discussed in this paper, this result may simply be a representation of facility MWO scheduling. Analysts can then perform the same actions as proposed in this paper and may even find it helpful to follow the next steps as illustrated in the case study to assess data quality. The methods in this paper present a specific example to show how heavily influencing missing data can be to a KPI, as well as the implications this has for the validity of other KPIs an analyst may be tasked to calculate.

6. Conclusions and Future Work

As the results of this study illustrate, the data quality problem of missing data has a major impact on the knowledge generated from these datasets. KPI calculations determined from MWO datasets that are riddled with missing data may be inaccurate and unrepresentative of actual organizational activities. As a result, decision making using these KPIs will not be as optimally informed. However, missing data is an inevitable occurrence, since it is a result of inevitable human error and the manual nature in which MWOs are recorded. We recommend analysts to evaluate their dataset for missing data before beginning analyses, to ensure awareness of this and other problems that may be hidden in the dataset that are not obvious at first glance. This further demonstrates the need for Technical Language Processing (TLP) tools to discover data quality issues, along with more community-driven techniques to perform quality corrections when discovered.

While we have provided an application in the maintenance domain to quantify data quality aspects and raise awareness of certain impacts it has on KPIs, several areas for exploration still remain from this work:

- Assess data quality over time: the data quality checks in this paper can be used and adapted to a more extended length of time, looking for inflection points where facility procedures may have changed and impacted data quality.
- Cross-tabulate to fill in missing data by other methods not described in this paper, such as identifying other data sources containing the complementary desired information.
- Infer relationships among data entities using graph databases, for instance if two different problem tags con-

cern the same machine, a rule can be written to deduce they are in the same location and building to fill in these values if they are missing.

- Combine information extraction techniques to deal with incomplete data. For instance, a MWO may include the problem tag "fan broken" and the solution tag "room hot". In other MWOs, "room hot" itself may lead to several effects, like the solution "new cooling machines". A system that could deduce that a broken fan should be associated with specific actions would be incredibly useful. It would also be useful to be able to weight and quantify these kinds of connections.

Regardless, the usefulness of going through the process of KPI calculation should not be discounted. Dataset quality is important because it allows us to place more appropriate confidence in the resultant KPIs. Similarly, the process of calculating KPIs can in turn elucidate key data quality issues. Using techniques established by past studies as well as this study focusing on KPI specification and relation, may improve the usefulness of datasets plagued with missing data. An approach to do so is to use the methods outlined in this paper, which analysts can use to assess data quality to mitigate the problem of missing data. When analysts can quantify missing data and understand patterns of missing data, data collection techniques can be improved to establish a mutually-beneficial ecosystem of data exchange and use between technicians and analysts.

NIST Disclaimer

The use of any products described in this paper does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that products are necessarily the best available for the purpose.

References

Astm E3012-20. (2020). *Standard Guide for Evaluation of Environmental Aspects of Sustainability of Manufacturing Processes*. ASTM International.

Brundage, Michael P and Bernstein, William Z and Morris, Katherine C and Horst, John A. (2017). Using graph-based visualizations to explore key performance indicator relationships for manufacturing production systems. *Procedia Cirp*, *61*, 451–456.

Brundage, Michael P and Morris, KC and Sexton, Thurston and Moccozet, Sascha and Hoffman, Michael. (2018). Developing maintenance key performance indicators from maintenance work order data. In *International manufacturing science and engineering conference* (Vol. 51371, p. V003t02a027).

Brundage, Michael P and Weiss, Brian A and Pellegrino, Joan. (2020). Summary report: Standards requirements gathering workshop for natural language analysis.

Davidson-Pilon, C., Kalderstam, J., Jacobson, N., Reed, S., Kuhn, B., Zivich, P., ... Golland, D. (2021, March). *CamDavidsonPilon/lifelines: 0.25.10*. Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.4579431` doi: 10.5281/zenodo.4579431

Eppinger, Steven D and Browning, Tyson R. (2012). *Design structure matrix methods and applications*. MIT press.

Giolo, Suely Ruiz. (2004). Turnbull's nonparametric estimator for interval-censored data. *Department of Statistics, Federal University of Paraná*, 1–10.

Helu, Moneer and Libes, Don and Lubell, Joshua and Lyons, Kevin and Morris, Katherine C. (2016). Enabling smart manufacturing technologies for decision-making support. In *ASME 2016 international design engineering technical conferences and computers and information in engineering conference.*

Hester, Patrick and Ezell, Barry and Collins, Andrew and Horst, John and Lawsure, Kaleen. (2017). A method for key performance indicator assessment in manufacturing organizations. *International Journal of Operations Research*, *14*(4), 157–167.

Hodkiewicz, Melinda and Ho, Mark Tien-Wei. (2016). Cleaning historical maintenance work order data for reliability analysis. *Journal of Quality in Maintenance Engineering*.

Ishikawa, Kaoru. (1985). *What is total quality control? The Japanese way*. Prentice Hall.

Iso. (2014). *ISO 22400 Automation systems and integration — Key performance indicators (KPIs) for manufacturing operations management — Part 1: Overview, concepts and terminology*. Geneva Switzerland.

Kang, Hyun. (2013). The prevention and handling of the missing data. *Korean journal of anesthesiology*, *64*(5), 402.

Katz, Leo. (1953). A new status index derived from sociometric analysis. *Psychometrika*, *18*(1), 39–43.

Kibira, Deogratias and Brundage, Michael P and Feng, Shaw and Morris, KC. (2018). Procedure for selecting key performance indicators for sustainable manufacturing. *Journal of Manufacturing Science and Engineering*, *140*(1).

Luc Devroye. (2006). Chapter 4 Nonuniform Random Variate Generation. In S. G. Henderson & B. L. Nelson (Eds.), *Simulation* (Vol. 13, pp. 83–121). Elsevier. doi: https://doi.org/10.1016/S0927-0507(06)13004-2

Lukens, Sarah and Naik, Manjish and Saetia, Kittipong and Hu, Xiaohui. (2019). Best Practices Framework for Improving Maintenance Data Quality to Enable Asset Performance Analytics. In *Annual Conference of the PHM Society* (Vol. 11).

Maydanchik, Arkady. (2007). *Data quality assessment*. Technics publications.

Michael P. Brundage and Thurston Sexton and Melinda Hodkiewicz and Alden Dima and Sarah Lukens. (2021). Technical language processing: Unlocking maintenance knowledge. *Manufacturing Letters*, *27*, 42–46. Retrieved from `https://www.sciencedirect.com/science/article/pii/S2213846320301668` doi: https://doi.org/10.1016/j.mfglet.2020.11.001

Miller Jr, Rupert G. (2011). *Survival analysis* (Vol. 66). John Wiley & Sons.

Navinchandran, Madhusudanan and Sharp, Michael E and Brundage, Michael P and Sexton, Thurston B. (2021). Discovering critical KPI factors from natural language in maintenance work orders. *Journal of Intelligent Man-*

ufacturing, 1–19.

Pigott, Therese D. (2001). A review of methods for missing data. *Educational research and evaluation*, *7*(4), 353–383.

Sexton, T., Hodkiewicz, M., & Brundage, M. P. (2019). Categorization errors for data entry in maintenance work-orders. In *Proceedings of the annual conference of the phm society* (Vol. 11).

Sexton, Thurston B and Brundage, Michael P. (2019). Nestor: A Tool for Natural Language Annotation of Short Texts. *J. Res. NIST*, *124*.

T. Sexton and M. P. Brundage and M. Hoffman and K. C. Morris. (2017). Hybrid datafication of maintenance logs from AI-assisted human tags. In *2017 ieee international conference on big data (big data)* (pp. 1769–1777). doi: 10.1109/BigData.2017.8258120

Tukey, John W and others. (1977). *Exploratory data analysis* (Vol. 2). Reading, Mass.

Zhou, Tao and Ren, Jie and Medo, Matúand Zhang, Yi-Cheng. (2007). Bipartite network projection and personal recommendation. *Physical review E*, *76*(4), 046115.