Table 1 Evaluation results. To evaluate the DQI procedure suggested by the system, we measure DQ dimension, ML model performance, and K-S tests. The table at the top displays the assessment results for DQ issue ratios, while the table at the bottom shows the results based on the types of DQ issues. Note that the DQ issues(s) and DQI method(s) columns are labeled with the icon texts in the DQI view.

Dataset	Rate	DQ issues(s)	DQI method(s)	Completeness		Outlier		Homogeneity		Duplicate		Correlation		Relevance		RMSE		Model	K-S test
				Before	After	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After		
Bike sharing	10	NONE		98		99		100		100		71		71		653.8587		RF	1
	25	OUT-MIS-HIG-INC	IQR-REM-PEA-REM	95	100	99	100	100	100	100	100	57	100	79	100	332.4573	748.3867	RF	≥ 0.05
	30	OUT-INC-MIS	IQR-REM-MEN	86	100	99	100	100	100	100	100	57	57.14	71	78.57	455.3126	72.7218	MLP	≥ 0.05
Forest fires	10	MIS-OUT-INC	MED-IQR-REM	92	100	94	100	100	100	100	100	100	84.62	100	100	57.2133	4.0292	MLP	≥ 0.05
	25	MIS-INC-OUT	MED-REM-IQR	100	100	95	100	100	100	98	100	85	100	100	100	38.2917	3.8258	LR	≥ 0.05
	30	OUT-LOW	IQR-KEN	88	100	91	100	100	100	100	100	100	84.62	100	100	80.3145	4.9414	RF	≥ 0.05
Diabetes	10	NONE		98		98		100		100		100		100		0.3925		RF	1
	25	OUT	IQR	96	100	97	100	100	100	100	100	100	100	100	100	0.4711	0.4591	LR	≥ 0.05
	30	MIS-OUT-INC	REM-IQR-REM	88	100	97	100	100	100	100	100	100	100	100	100	0.5052	0.4487	LR	≥ 0.05

Dataset	Type	DQ issues(s)	DQI method(s)	Completeness		Outlier		Homogeneity		Duplicate		Correlation		Relevance		RMSE		Model	K-S test
				Before	After	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After		
Bike sharing	OUT	NONE		100		99		100		100		57		79		24.7773		RF	1
	MIS	LOW-MIS	KEN-EM	100	100	99	99	100	100	100	100	57	82	79	100	24.7773	587.9532	RF	≥ 0.05
	MIS, OUT	MIS	MIN	100	100	99	99	100	100	100	100	57	57	79	79	24.7773	24.7773	MLP	≥ 0.05
Forest fires	INC	OUT-HIG	IQR-KEN	100	100	95	99	100	100	98	99	85	100	100	100	38.2917	3.8258	LR	≥ 0.05
	MIS	MIS-OUT-HIG	MED-IQR-PEA	100	100	95	99	100	100	98	99	85	100	100	100	38.2917	3.8258	LR	≥ 0.05
	MIS, OUT	MIS-OUT	MAX-IQR	100	100	95	99	100	100	98	99	85	100	100	100	38.2917	3.8258	LR	≥ 0.05
Diabetes	MIS	MIS	MEN	100	100	98	98	100	100	100	100	100	100	100	100	0.3907	0.3907	RF	≥ 0.05
	OUT, INC	NONE		100		98		100		100		100		100		0.3907		RF	1
	MIS, INC	OUT-MIS	Z-MIN	100	100	98	99	100	100	100	100	100	100	100	100	0.3907	0.3931	LR	≥ 0.05

APPENDIX: STATISTICAL ANALYSIS

Table 1 shows the DQI procedures proposed by the system for the poisoning data and the evaluation results. Most data have higher DQ dimension scores due to DQI. In the table above, for the bike sharing dataset 30%, data completeness, outlier detection, correlation, and data relevance change from 86, 99, 57, and 71 to 100, 100, 57.14, and 78.57, respectively. In the table below, for the bike sharing dataset with missing values, data correlation and relevance detection change from 57 and 79 to 82 and 100, respectively. Additionally, most of the data enhances the ML model's performance due to DQI. In addition, since all data have a p-value of over than 0.05 in the K-S test, the system only proposes the appropriate DQI procedures. Note that since the data without the DQI procedure are the same before and after DQI, the p-value of the K-S test shows *1*.

Some data also have lower DQ dimension scores. In the table above, for the forest fires dataset 10% and 30%, data correlation changes from 100 to 84.62. As mentioned in Section VI, this problem occurs because the system evaluates the new DQ based on the quality-improved data. Therefore, the score is not lowered. Also, depending on the data, the degree to which the ML model performance is enhanced due to DQI is different. For example, the forest fires dataset shows a significant decrease in RSME with DQIs, whereas the diabetes dataset shows little difference. Therefore, the user should analyze the effect of the ML model performance enhancement while improving the DQ.

The table placed at the top shows the results of the poisoned data to have different DQ issue ratios. The DQI procedure differs in the degree of improving the ML model performance according to the ratio of DQ issues. For example, the forest fires dataset 10% has a 53.1841 decrease in RSME, while 30% has a 75.3731 decrease. Also, for the bike sharing dataset and diabetes dataset 10%, none of the DQI procedures enhance the ML model performance compared to the original data. The reason is that since the lower the ratio of DQ issues, the less impact they have on the ML model performance, they are not addressed in the DQI procedure. Therefore, it is important to show high ML model performance even if DQ issues exist in the data. Notably, for the bike sharing dataset 25%, RMSE increases 332.4573 to 748.3867 after applying DQI. This issue is attributed to the batch application of DQI procedures for usability. Such cases highlight the necessity of customizing DQI pipelines based on context, as discussed in Section V.G.

The lower table summarizes the results of poisoned data, which exhibit various types of DQ issues. The system proposes different DQI procedures depending on the type of issue. For example, in the bike sharing dataset with missing values, the DQ dimension scores for correlation and relevance improved after DQI. However, the corresponding procedure involved removing several features, which reduced the volume of data available for model training. This led to an increase in RMSE, indicating a potential trade-off between improving data quality and maintaining model performance. These findings underscore the need for customizing DQI procedures based on the specific task context, as discussed in Section V.G.

We evaluate whether the DQ dimension, K-S test, DQ issue ratio, and DQ issue types have statistically significant effects on the ML model performance from the ANOVA test. We randomly generate 100 DQI procedures per data for the ANOVA test and compute the DQ dimension, K-S test, DQ issue ratio, DQ issue type, and ML model performance. Afterward, we perform the ANOVA test by calculating the difference between the data before and after DQI for each independent variable. From the ANOVA test, all independent variables do not affect the ML model performance. As mentioned in Section II, this result is shown because the DQ issue that affects performance is different depending on the data and ML model. Therefore, the user needs to figure out the DQI methods and configure the order to perform them through the visual analytics system.