

Data similarity

AI feature description	This feature suggests alternative but similar data to the data being viewed. The feature leverages metadata and profiling statistics of tables and columns to compute a similarity score.
Feature optional?	Yes
Model	Model type: K-nearest neighbors clustering Gen AI?: No Owner: Collibra
Intended use	This feature is intended to provide suggestions as to alternative tables to a table being viewed. User discretion is required as to whether the suggestions are appropriate and should be used instead of the original table.
Model training/retraining data	Open source publicly available relational database data has been used for the first trained iteration of the model. Live retraining is not currently planned.
Model input	An irreversibly hashed form of the metadata data ingested through the Collibra profiling functionality will be sent to the model.
Model output	<ul style="list-style-type: none"> Given a table, tables which are similar based on table name, column name, and statistical properties of the data List of asset IDs with associated similarity scores
Is input data used to train the model for general product improvement purposes?	No
Data storage and retention	All Collibra storage locations are referenced here . Retention is concurrent with customer environment retention.
Human in the loop?	Yes. Usage of suggested data is within the user's discretion.
Is personal data processed?	None. Model input data is irreversibly hashed.
Other considerations	None