**Derived and Base Measure for Veracity**

**M(Ver) : Weighted sum of accuracy, completeness, currentness and availability**

| **Derived measure or indicator: M(ver)** | | | |  | | |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| # 1 | Derived measure or indicator | Formula  M(ver) = Accuracy(MDS)\*Wacc + Completeness(MDS)\*Wcomp + Currentness(MDS)\*Wcurr + Availability(MDS)\*Wavail  Each weight is set to ¼ by default. | | | | | | | |
| M(Ver) : Weighted sum of accuracy, completeness, currentness and availability |
| Link with the measurement goal (which goal)  Veracity | | | Responsible (who analyzes)  Developer  Data Analyst  Data Engineer  Data Scientist | | | Stakeholder (who uses)  Project Manager  Data Scientist  Senior Management | | Frequency (when)  Veracity of dataset can be calculated on monthly, quarterly or yearly basis. | |
| Data source (where the measurement data will be extracted from)  Credit Card classification - https://www.kaggle.com/datasets/samuelcortinhas/credit-card-classification-clean-data | | | Storage of the result (where data will be stored after the extraction)  The data will be stored in excel file or database.  In our case we will be storing the result in jupyter notebook for reporting purpose. | | | Data interpretation rules  Veracity can range between 0 and 1, higher veracity means better accuracy and trustworthiness of data whereas low veracity means unreliable data.  Veracity >= 0.8 means that the data quality is good and it can be used for machine learning model.  Veracity < 0.8 means that the data quality is good and it can be used for machine learning model.  The weights for each of the sub measures defines the importance of each sub measure in the calculation of veracity. | | | |
|  | | |  | | |  | | | |
| Analysis procedure   1. Retrieve recent accuracy value available at that timestamp from database. 2. Retrieve recent completeness value available at that timestamp from database. 3. Retrieve recent currentness value available at that timestamp from database. 4. Retrieve recent availability value available at that timestamp from database. 5. Use the weighted sum formula to calculate veracity 6. Analyze and interpret the results and make decisions | | | | | Presentation of the results (sketch illustrating what it looks like): | | | | |
|  | | | | |
| Potential decision making depending on the results  Veracity of data can be crucial for decision making process, for machine learning model to perform well veracity of data should be good enough, poor data from unreliable source can lead to failure of machine learning model. In business big data is used to study customer behaviour and if the veracity of data is low then this could lead to wrong decision making, however good quality data can be beneficial for the growth of the company. | | | | |

**Accuracy:**

| **Derived measure or indicator: Accuracy** | | | |  | | |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| #1 | Derived measure or indicator | Formula  Availability(MDS) = | | | | | | | |
| Degree to which data attributes represent the true value in a specified context of use. |
| Link with the measurement goal (which goal)  Veracity | | | Responsible (who analyzes)  Developer  Data Analyst  Data Engineer  Data Scientist | | | Stakeholder (who uses)  Senior management  Project manager  Data scientist  Data analyst | | Frequency (when)  The accuracy of data set can be measured on monthly, quarterly or yearly basis. | |
| Data source (where the measurement data will be extracted from)  Credit Card classification - https://www.kaggle.com/datasets/samuelcortinhas/credit-card-classification-clean-data | | | Storage of the result (where data will be stored after the extraction)  The data will be stored in excel file or database.  In our case we will be storing the result in jupyter notebook for reporting purpose. | | | Data interpretation rules  Successful request is categorized as a request which returns the correct result.  Every query to a database is considered as a request.  Accuracy = 1 - means that the data attributes always represent truth value. This is a desired value for implementation of a successful machine learning model.  Accuracy = 0 means that data attributes do not hold true value .  Accuracy >= 0.90 means that 90% of the data attribute holds true value which can be useful to train our machine learning algorithm.  Accuracy could increase or decrease depending upon the dataset size increasing or decreasing. | | | |
|  | | |  | | |  | | | |
| Analysis procedure   1. Dataset is loaded using the analyses tool, excel file or jupyter notebook. 2. Lbd is counted using COUNT function to get number of records 3. Pj is calculated to get the total number of duplicate items and their specific count in each dataset using the function like COUNT() 4. Hacc and Hmax are calculated using the formula. 5. Accuracy of the dataset will be calculated using the formula. 6. The value will be interpreted according to the decision making rules and appropriate decision will be taken. | | | | | Presentation of the results (sketch illustrating what it looks like):  Accuracy of the data will be presented as a single numerical value. | | | | |
|  | | | | |
| Potential decision making depending on the results  Accuracy of the data attributes can give the overview about truthfulness of the data. This is an important measure in order to get the Machine Learning model trained with the correct data. If the accuracy value is more it will give the confidence to stakeholders in order to trust the results produced by the machine learning algorithms. | | | | |

| **Base Measure: pj** | | |  | | |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| #2 | Measure (what: entity, attribute)  Measures the total number of duplicate items and their specific count in each dataset  Entity: Dataset  Attribute: Total number of duplicate items | | | | Scale type  Absolute | Applicability  Helps us to understand how many duplicate records are there in the dataset and also their count in each dataset. Better understanding of data quality. | | |
|  | | | |  |
| Who measures?  Developer  Data Analyst  Data Engineer  Data Scientist | | Source of measurement  Credit Card classification - https://www.kaggle.com/datasets/samuelcortinhas/credit-card-classification-clean-data | | Where to store the result  CSV File  Database | | Tool  Excel  Jupyter Notebook  Python libraries for data analysis like pandas , numpy etc. | Time (when to measure)  This metric could be measured on a monthly, quarterly or yearly basis to calculate the accuracy trend of the database. | |
|  | |  | |  | |  |  | |
| Collection procedure (how to collect the data)  This can be calculated by using the excel built in function or python various data processing libraries. | | | | Notes or comments: | | | | |
|  | | | |

| **Base measure: Lbd(MDS)** | | |  | | |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| #3 | Measure (what: entity, attribute) | | | | Scale type | Applicability  Total number of records in data sets acts as a fundamental unit of measurement which can be used to calculate other derived measures. It also gives the idea about the sizeof the dataset. | | |
| Measures the total number of records in multiple datasets.  Entity: Dataset  Attribute: Number of records | | | | Absolute |
| Who measures?  Developer  Data Analyst  Data Engineer  Data Scientist | | Source of measurement  Credit Card classification - https://www.kaggle.com/datasets/samuelcortinhas/credit-card-classification-clean-data | | Where to store the result  CSV File  Database | | Tool  Excel  Jupyter Notebook  Python libraries for data analysis like pandas , numpy etc. |  | |
|  | |  | |  | |  |  | |
| Collection procedure (how to collect the data) | | | | Notes or comments:  The number of records will be counted for each dataset for all time periods. I.e Length of records will be counted for a dataset as a whole and for each time period separately  E.g if we have dataset D1,D2 for time T1,T2 then number of records will be lbd(D1) = lbd(D1T1) + lbd(D2T2) | | | | |
| The data is loaded into excel sheet or database and the total number of records can be retrieved from query the database or using inbuilt functions of excel. | | | |

| **Derived measure or indicator: Hacc** | | | |  | | |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| #1 | Derived measure or indicator | Formula | | | | | | | |
| Entropy of the given dataset |
| Link with the measurement goal (which goal)  Veracity | | | Responsible (who analyzes)  Developer  Data Analyst  Data Engineer  Data Scientist | | | Stakeholder (who uses)  Developer | | Frequency (when)  The Hacc of data set can be measured on monthly, quarterly or yearly basis. | |
| Data source (where the measurement data will be extracted from)  Credit Card classification - https://www.kaggle.com/datasets/samuelcortinhas/credit-card-classification-clean-data | | | Storage of the result (where data will be stored after the extraction)  The data will be stored in excel file or database.  In our case we will be storing the result in jupyter notebook for reporting purpose. | | | Data interpretation rules  This measure will be used to calculate the accuracy measure. If the value of Hacc is greater than Hmax, accuracy value will be more and hence the data attribute has truth value. | | | |
|  | | |  | | |  | | | |
| Analysis procedure   1. Dataset is loaded using the analyses tool, excel file or jupyter notebook. 2. Lbd is counted using COUNT function to get number of records 3. Pj is calculated to get the total number of duplicate items and their specific count in each dataset using the function like COUNT() 4. Hacc is calculated using the formula mentioned above. | | | | | Presentation of the results (sketch illustrating what it looks like):  Hacc of the data will be presented as a single numerical value. | | | | |
|  | | | | |
| Potential decision making depending on the results  This value is being used to calculate accuracy. | | | | |

| **Derived measure or indicator: Hmax** | | | |  | | |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| #3 | Derived measure or indicator | Formula | | | | | | | |
| Maximum Entropy of the given dataset |
| Link with the measurement goal (which goal)  Veracity | | | Responsible (who analyzes)  Developer  Data Analyst  Data Engineer  Data Scientist | | | Stakeholder (who uses)  Developer | | Frequency (when)  The Hmax of data set can be measured on monthly, quarterly or yearly basis. | |
| Data source (where the measurement data will be extracted from)  Credit Card classification - https://www.kaggle.com/datasets/samuelcortinhas/credit-card-classification-clean-data | | | Storage of the result (where data will be stored after the extraction)  The data will be stored in excel file or database.  In our case we will be storing the result in jupyter notebook for reporting purpose. | | | Data interpretation rules  This measure will be used to calculate the accuracy measure. If the value of Hmax is less than Haccw, accuracy value will be more and hence the data attribute has truth value. | | | |
|  | | |  | | |  | | | |
| Analysis procedure   1. Dataset is loaded using the analyses tool, excel file or jupyter notebook. 2. Lbd is counted using COUNT function to get number of records 3. Hmax is calculated using the formula mentioned above. | | | | | Presentation of the results (sketch illustrating what it looks like):  Hmax of the data will be presented as a single numerical value. | | | | |
|  | | | | |
| Potential decision making depending on the results  This value is being used to calculate accuracy. | | | | |

**Completeness:**

| **Derived measure or indicator: Completeness** | | | |  | | |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| # 2 | Derived measure or indicator | Formula | | | | | | | |
| Completeness : Degree to which subject data associated with an entity has values for all expected attributes and related entity instances in a specific context of use. |
| Link with the measurement goal (which goal)  Veracity | | | Responsible (who analyzes)  Developer  Data Analyst  Data Engineer  Data Scientist | | | Stakeholder (who uses)    Project Manager  Data Scientist | | Frequency (when)  Completeness of the data can be calculated at the start of the project, periodically at certain time intervals or it could be calculated eack time a new data is loaded into the system. | |
| Data source (where the measurement data will be extracted from)  Credit Card classification - https://www.kaggle.com/datasets/samuelcortinhas/credit-card-classification-clean-data | | | Storage of the result (where data will be stored after the extraction)  The data will be stored in excel file or database.  In our case we will be storing the result in jupyter notebook for reporting purpose. | | | Data interpretation rules  Successful request is categorized as a request which returns the correct result.  Every query to a database is considered as a request.  Completeness = 1 - means that the subject data associated with an entity has values for all expected attributes and related entity instances. This is a desired value for implementation of a successful machine learning model.  Completeness = 0 means that data attributes hold null value .  Completeness >= 0.90 means that 90% of the data attribute holds non null value which can be useful to train our machine learning algorithm.  Completeness could increase or decrease depending upon the dataset size increasing or decreasing. | | | |
|  | | |  | | |  | | | |
| Analysis procedure | | | | | Presentation of the results (sketch illustrating what it looks like):  Completeness of the data will be presented as a single numerical value. | | | | |
| 1. Dataset is loaded using the analyses tool, excel file or jupyter notebook. 2. Lbd is counted using COUNT function to get number of records 3. Rec\_no\_null is calculated to get the total number of no null items and their specific count in each dataset using the function like COUNT() 4. Completeness of the dataset will be calculated using the formula. 5. The value will be interpreted according to the decision making rules and appropriate decision will be taken. | | | | |
| Potential decision making depending on the results  Completeness of the data attributes can give the overview about absoluteness of the data. This is an important measure in order to get the Machine Learning model trained with the correct data. If the completeness value is more it will give the confidence to stakeholders in order to trust the results produced by the machine learning algorithms. | | | | |

| **Base measure: Rec\_no\_null (MDS)** | | |  | | |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| #1 | Measure (what: entity, attribute) | | | | Scale type | Applicability  Total number of no null records in data sets acts as a fundamental unit of measurement which can be used to calculate other derived measures. | | |
| Measures the total number of records with no null values.  Entity: Dataset  Attribute: Number of records | | | | Absolute |
| Who measures?  Developer  Data Analyst  Data Engineer  Data Scientist | | Source of measurement  Credit Card classification - https://www.kaggle.com/datasets/samuelcortinhas/credit-card-classification-clean-data | | Where to store the result  CSV File  Database | | Tool  Excel  Jupyter Notebook  Python libraries for data analysis like pandas , numpy etc. | Time (when to measure)  Length of the data set can be measured each time new data is loaded into the database. | |
|  | |  | |  | |  |  | |
| Collection procedure (how to collect the data) | | | | Notes or comments:  None | | | | |
| The data is loaded into excel sheet or database and the total number of no null records can be retrieved from query the database or using inbuilt functions of excel. | | | |

| **Base measure: Lbd(MDS)** | | |  | | |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| #2 | Measure (what: entity, attribute) | | | | Scale type | Applicability  Total number of records in data sets acts as a fundamental unit of measurement which can be used to calculate other derived measures. It also gives the idea about the sizeof the dataset. | | |
| Measures the total number of records in multiple datasets.  Entity: Dataset  Attribute: Number of records | | | | Absolute |
| Who measures?  Developer  Data Analyst  Data Engineer  Data Scientist | | Source of measurement  Credit Card classification - https://www.kaggle.com/datasets/samuelcortinhas/credit-card-classification-clean-data | | Where to store the result  CSV File  Database | | Tool  Excel  Jupyter Notebook  Python libraries for data analysis like pandas , numpy etc. | Time (when to measure)  Length of the data set can be measured each time new data is loaded into the database. | |
|  | |  | |  | |  |  | |
| Collection procedure (how to collect the data) | | | | Notes or comments:  The number of records will be counted for each dataset for all time periods. I.e Length of records will be counted for a dataset as a whole and for each time period separately  E.g if we have dataset D1,D2 for time T1,T2 then number of records will be lbd(D1) = lbd(D1T1) + lbd(D2T2) | | | | |
| The data is loaded into excel sheet or database and the total number of records can be retrieved from query the database or using inbuilt functions of excel. | | | |

**Currentness:**

| **Derived measure or indicator: Currentness** | | | |  | | |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| # 4 | Derived measure or indicator | Formula  Currentness(MDS) = | | | | | | | |
| Currentness : Degree to which data has attributes that are of the right age in a specific context of use. |
| Link with the measurement goal (which goal)  Veracity | | | Responsible (who analyzes)  Developer  Data Analyst  Data Engineer  Data Scientist | | | Stakeholder (who uses)    Project Manager  Data Scientist | | Frequency (when)  Currentness of the data can be calculated at the start of the project, periodically at certain time intervals or it could be calculated eack time a new data is loaded into the system. | |
| Data source (where the measurement data will be extracted from)  Credit Card classification - https://www.kaggle.com/datasets/samuelcortinhas/credit-card-classification-clean-data | | | Storage of the result (where data will be stored after the extraction)  The data will be stored in excel file or database.  In our case we will be storing the result in jupyter notebook for reporting purpose. | | | Data interpretation rules  For counting total number of records, every record should be considered for counting without any filters on data.  Data older than 10 years will be considered old data.  Currentness of data will be measured based on threshold value meaning that the currentness of the dataset should be above a certain value.  Currentness(Dataset) >= 0.7 - relevant for use in machine learning model  Currentness(Dataset) between 0.5 and 0.7 - relevant for use in machine learning model with some caution.  Currentness(Dataset) <= 0.5 - can be used for training machine learning model after checking the relevancy of data. | | | |
|  | | |  | | |  | | | |
| Analysis procedure | | | | | Presentation of the results (sketch illustrating what it looks like):  Currentness of the data will be presented as a single numerical value. | | | | |
| 1. Dataset is loaded using the analyses tool, excel file or jupyter notebook. 2. Total number of records are calculated using inbuilt COUNT() function or its equivalent 3. Number of records within acceptable range will be calculated by applying filter over timestamp of data record. 4. Currentness of the dataset will be calculated using the formula. 5. Currentness of each dataset will be added to get the total currentness of MDS at various stages of data processing. 6. The value will be interpreted according to the decision making rules and appropriate decisions will be taken. | | | | |
| Potential decision making depending on the results  If the currentness of the data is within the acceptable range then the data can be used to train machine learning models to identify recent trends in data. If the data is too old then the decision derived from the data would not be relevant to the current scenario. Data needs to be updated if the dataset currentness value is too low. | | | | |

| **Base measure: Lbd(MDS)** | | |  | | |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| #1 | Measure (what: entity, attribute) | | | | Scale type | Applicability  Total number of records in data sets acts as a fundamental unit of measurement which can be used to calculate other derived measures. It also gives the idea about the sizeof the dataset. | | |
| Measures the total number of records in multiple datasets.  Entity: Dataset  Attribute: Number of records | | | | Absolute |
| Who measures?  Developer  Data Analyst  Data Engineer  Data Scientist | | Source of measurement  Credit Card classification - https://www.kaggle.com/datasets/samuelcortinhas/credit-card-classification-clean-data | | Where to store the result  CSV File  Database | | Tool  Excel  Jupyter Notebook  Python libraries for data analysis like pandas , numpy etc. | Time (when to measure)  Length of the data set can be measured each time new data is loaded into the database. | |
|  | |  | |  | |  |  | |
| Collection procedure (how to collect the data) | | | | Notes or comments:  The number of records will be counted for each dataset for all time periods. I.e Length of records will be counted for a dataset as a whole and for each time period separately  E.g if we have dataset D1,D2 for time T1,T2 then number of records will be lbd(D1) = lbd(D1T1) + lbd(D2T2) | | | | |
| The data is loaded into excel sheet or database and the total number of records can be retrieved from query the database or using inbuilt functions of excel. | | | |

| **Base measure: Rec\_acc\_age (MDS)** | | |  | | |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| #2 | Measure (what: entity, attribute)  Provides the total number of records with ages that fall within the acceptable range based on the upper and lower quartiles of the Box and Whisker.  Entity : Dataset  Attribute: number of records within acceptable age range | | | | Scale type  Absolute | Applicability  The number of records in acceptable range brackets is a relevant metric for calculating currentness of data and it also tells helps us analyze the age of dataset and its relevancy of future use. | | |
|  | | | |  |
| Who measures?  Developer  Data Analyst  Data Engineer  Data Scientist | | Source of measurement  Credit Card classification - https://www.kaggle.com/datasets/samuelcortinhas/credit-card-classification-clean-data | | Where to store the result  CSV File  Database | | Tool  Excel  Jupyter Notebook  Python libraries for data analysis like pandas , numpy etc. | Time (when to measure)  Total number of records with acceptable age range should be measured before calculating currentness of data. This could happen before the start of the project, periodically at certain time intervals to keep the track of currentness of data or whenever a new data is loaded into system. | |
|  | |  | |  | |  |  | |
| Collection procedure (how to collect the data)  The data can be collected by applying control limits for the acceptable ranges over time attribute and filtering the number of records within acceptable ranges. | | | | Notes or comments:  The number of records within acceptable ranges will be calculated for each dataset D(i) and not for each time stamp. | | | | |
|  | | | |

**Availability:**

| **Derived measure or indicator: Availability** | | | |  | | |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| #5 | Derived measure or indicator | Formula  Availability(MDS) = | | | | | | | |
| Currentness of the data will be presented as a single numerical value. |
| Link with the measurement goal (which goal)  Veracity | | | Responsible (who analyzes)  Developer  Data Analyst  Data Engineer  Data Scientist | | | Stakeholder (who uses)  Senior management  Project manager  Data scientist  Data analyst | | Frequency (when)  The availability of data set can be measured on monthly, quarterly or yearly basis. | |
| Data source (where the measurement data will be extracted from)  Credit Card classification - https://www.kaggle.com/datasets/samuelcortinhas/credit-card-classification-clean-data | | | Storage of the result (where data will be stored after the extraction)  The data will be stored in excel file or database.  In our case we will be storing the result in jupyter notebook for reporting purpose. | | | Data interpretation rules  Successful request is categorised as request which returns correct result.  Every query to a database is considered as request.  Availability = 1 - means that the database is available at all times, for every request a successful result has been returned. This is a desired value for implementation of successful machine learning model.  Availability = 0 means that database does not return result for any request.  Availability >= 0.99 means that 99% of the request were successful. This is a acceptable value for training a machine learning and the model will train in time.  Availability >=90 means that 90% of the request were successful and this could increase the training time of machine learning model significantly as data might not be available for 10% of the cases. More number of requests are required to fetch data.  Availability could increase or decrease depending upon the number of successful requests to database. If number of successful requests fall then the availability is expected to go down. | | | |
|  | | |  | | |  | | | |
| Analysis procedure   1. Dataset is loaded using the analyses tool, excel file or jupyter notebook. 2. Total number of requests and successful requests are retrieved from the query log or issue log. 3. Availability of the dataset will be calculated usng the formula. 4. The value will be interpreted according to the decision making rules and appropriate decision will be taken. | | | | | Presentation of the results (sketch illustrating what it looks like):  Availability of the data will be presented as a single numerical value. | | | | |
|  | | | | |
| Potential decision making depending on the results  Availability of the dataset can give an overview of the resiliency of the system infrastructure, low availability could lead to decrease in confidence of stakeholders in the system leading to abandoning of the system whereas high availability could increase the confidence of stakeholders which is preferred by stakeholders for training machine learning model. | | | | |

| **Base measure: N\_succ\_req (MDS)** | | |  | | |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| #1 | Measure (what: entity, attribute)  Measures the number of successful request from an API server, database etc.  Entity: Dataset  Attribute: number of successful requests | | | | Scale type  Absolute | Applicability  The number of successful request gives us the metric to calculate availability and gives us the intuition about the likelihood of success of a request to an API or database. | | |
|  | | | |  |
| Who measures?  Developer  Data Analyst  Data Engineer  Data Scientist | | Source of measurement  Credit Card classification - https://www.kaggle.com/datasets/samuelcortinhas/credit-card-classification-clean-data | | Where to store the result  CSV File  Database | | Tool  Excel  Jupyter Notebook  Python libraries for data analysis like pandas , numpy etc. | Time (when to measure)  This metric could be measured on a monthly, quarterly or yearly basis to calculate the availability trend of the database. | |
|  | |  | |  | |  |  | |
| Collection procedure (how to collect the data)  Generally API request or queries are logged for future/audit references therefore count the number of requests for which the correct responses have been returned from the database or dataset. | | | | Notes or comments:  In case of static dataset count the number of successful queries on the database. | | | | |
|  | | | |

| **Base measure: N\_req (MDS)** | | |  | | |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| #2 | Measure (what: entity, attribute)  Measures the total number of requests to a database within a given timeframe.  Entity: Dataset  Attribute: Total number of requests to dataset. | | | | Scale type  Absolute | Applicability  The number of request to a database could be considered as the fundamental unit of database which gives us the idea about the frequency of usage an importance of the dataset. More number of requests means that the database is usage is high and it is important. | | |
|  | | | |  |
| Who measures?  Developer  Data Analyst  Data Engineer  Data Scientist | | Source of measurement  Credit Card classification - https://www.kaggle.com/datasets/samuelcortinhas/credit-card-classification-clean-data | | Where to store the result  CSV File  Database | | Tool  Excel  Jupyter Notebook  Python libraries for data analysis like pandas , numpy etc. | Time (when to measure)  This metric could be measured on a monthly, quarterly or yearly basis to calculate the availability trend of the database. | |
|  | |  | |  | |  |  | |
| Collection procedure (how to collect the data)  Generally API request or queries are logged for future/audit references therefore count the number of requests/queries performed on database or dataset. | | | | Notes or comments:  In case of static dataset count the number of queries on the database. | | | | |
|  | | | |