# Data Quality Analysis

## Categories

As a data quality assessment system, we will use following categories:

|  |  |
| --- | --- |
| Category | Description |
| Accuracy (0-100%) | The degree to which the data tested is in argument with a standard source, which is assumed to be correct. It is actually a count of number of times an attribute was incorrect. |
| Integrity (0-100%) | The data is not inadvertently deleted, and changes made are applied. You can measure integrity by tracking the number of issues logged or records affected from a call center. |
| Secure (yes/no) | Users and systems that should not have access rights are properly controlled. You may measure this by logging incidents, tracking the number of individuals with write access, or measuring conformance to security standards by role. |
| Completeness (0-100%) | The degree to which all of the required data values appear in the data record or business object. You can measure this by tracking the percentage of records with one or more missing required attributes. |
| Validity (high/low) | The degree to which the tested data conforms to the data validation requirements. This is a count of the number of rules in conformance. |
| Timeliness (0-100%) | The degree to which the data is provided at the time required or specified. You can track this by comparing planned versus actual availability. |
| No redundancy (yes/no) | The degree to which there are multiple databases. To measure this, you can track the number of copies of that record throughout your landscape. |
| No duplicates (yes/no) | The degree to which there are duplicate records for the same item. |
| Relevance (yes/no) | The data should be tied to a business process, metric or objective. |
| Accessibility (high/low) | The ease of use for data consumers to access and consume the data. |
| Availability (high/low) | The degree to which data is available for access to the users. |
| Consistency (high/low) | The ability to use or interpret the data accurately across multiple domains, for example multiple product descriptions for the same product across the application. |

## Requirements

And also, will added the data requirements. To each data is described the format.

|  |  |
| --- | --- |
| **Name** | Requirements |
| Listing ID | Digits without special symbols and letters. Length to 50 symbols. |
| Host ID | Digits and letters without special symbols. Length to 50 symbols. |
| Host total listings count | Digits. Length to 50 |
| Longitude | Digits with comma. Float number. Length to 35 symbols. |
| Accommodates | Digits without special symbols and letters. Length to 50 |
| Bathrooms | Digits without special symbols and letters. Length to 35 |
| Bedrooms | Digits without special symbols and letters. Length to 35 |
| Price | Digits without special symbols and letters. Length to 50 |
| Minimum nights | Digits without special symbols and letters. Length to 2 |
| Maximum nights | Digits without special symbols and letters. Length to 10 |
| Availability 365 | Digits without special symbols and letters. Length to 3 |
| Number of reviews | Digits without special symbols and letters. Length to 3 |
| Review Scores Accuracy | Digits without special symbols and letters. Length to 3 |
| Review Scores Cleanliness | Digits without special symbols and letters. Length to 3 |
| Review Scores Checkin | Digits without special symbols and letters. Length to 3 |
| Review Scores Communication | Digits without special symbols and letters. Length to 3 |
| Review Scores Location | Digits without special symbols and letters. Length to 3 |
| Reviews per month | Digits without special symbols and letters. Length to 3 |

## Assessment data quality

listing\_id int64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | yes |
| Completeness | 100 |
| Validity | high |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | yes |
| Relevance | yes |
| Accessibility | no |
| Availability | no |
| Consistency | yes |

id int64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | yes |
| Completeness | 100 |
| Validity | high |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | yes |
| Relevance | yes |
| Accessibility | no |
| Availability | no |
| Consistency | yes |

date object

|  |  |
| --- | --- |
| Accuracy | 80 |
| Integrity | 100 |
| Secure | no |
| Completeness | 80 |
| Validity | high |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | yes |
| Relevance | yes |
| Accessibility | no |
| Availability | no |
| Consistency | yes |

reviewer\_id int64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | yes |
| Completeness | 100 |
| Validity | high |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | yes |
| Relevance | yes |
| Accessibility | no |
| Availability | no |
| Consistency | yes |

reviewer\_name object

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | yes |
| Completeness | 100 |
| Validity | high |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | no |
| Relevance | yes |
| Accessibility | no |
| Availability | yes |
| Consistency | yes |

comments object

|  |  |
| --- | --- |
| Accuracy | 20 |
| Integrity | 100 |
| Secure | no |
| Completeness | 50 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | No |
| No duplicates | No |
| Relevance | yes |
| Accessibility | Yes |
| Availability | Yes |
| Consistency | yes |

Listing ID int64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | yes |
| Completeness | 100 |
| Validity | high |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | yes |
| Relevance | yes |
| Accessibility | no |
| Availability | no |
| Consistency | yes |

Name object

|  |  |
| --- | --- |
| Accuracy | 80 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | no |
| No duplicates | no |
| Relevance | yes |
| Accessibility | no |
| Availability | yes |
| Consistency | yes |

Host ID int64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | yes |
| Completeness | 100 |
| Validity | high |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | yes |
| Relevance | yes |
| Accessibility | no |
| Availability | no |
| Consistency | yes |

Host Name object

|  |  |
| --- | --- |
| Accuracy | 80 |
| Integrity | 100 |
| Secure | yes |
| Completeness | 80 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | yes |
| Relevance | yes |
| Accessibility | no |
| Availability | no |
| Consistency | yes |

Host Response Rate object

|  |  |
| --- | --- |
| Accuracy | 80 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | yes |
| Relevance | yes |
| Accessibility | no |
| Availability | yes |
| Consistency | yes |

Host Is Superhost bool

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | yes |
| Completeness | 100 |
| Validity | high |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | no |
| Relevance | yes |
| Accessibility | no |
| Availability | yes |
| Consistency | yes |

Host total listings count int64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | yes |
| Completeness | 100 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | no |
| Relevance | yes |
| Accessibility | no |
| Availability | yes |
| Consistency | yes |

Street object

|  |  |
| --- | --- |
| Accuracy | 50 |
| Integrity | 100 |
| Secure | no |
| Completeness | 80 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | no |
| No duplicates | no |
| Relevance | Yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

City object

|  |  |
| --- | --- |
| Accuracy | 50 |
| Integrity | 100 |
| Secure | no |
| Completeness | 80 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | no |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

Neighbourhood cleansed object

|  |  |
| --- | --- |
| Accuracy | 50 |
| Integrity | 100 |
| Secure | no |
| Completeness | 80 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | no |
| Relevance | yes |
| Accessibility | no |
| Availability | yes |
| Consistency | yes |

State object

|  |  |
| --- | --- |
| Accuracy | 50 |
| Integrity | 100 |
| Secure | no |
| Completeness | 80 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | no |
| Relevance | yes |
| Accessibility | no |
| Availability | yes |
| Consistency | yes |

Country object

|  |  |
| --- | --- |
| Accuracy | 50 |
| Integrity | 100 |
| Secure | no |
| Completeness | 80 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | no |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

latitude float64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 80 |
| Validity | high |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | yes |
| Relevance | yes |
| Accessibility | no |
| Availability | no |
| Consistency | yes |

longitude float64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 80 |
| Validity | high |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | yes |
| Relevance | yes |
| Accessibility | no |
| Availability | no |
| Consistency | yes |

Property type object

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | high |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | no |
| Relevance | yes |
| Accessibility | no |
| Availability | yes |
| Consistency | yes |

Room type object

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | high |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | no |
| Relevance | yes |
| Accessibility | no |
| Availability | yes |
| Consistency | yes |

Accommodates int64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | high |
| Timeliness | 100 |
| No redundancy | yes |
| No duplicates | yes |
| Relevance | yes |
| Accessibility | no |
| Availability | no |
| Consistency | yes |

Bathrooms float64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | No |
| No duplicates | No |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

Bedrooms int64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | No |
| No duplicates | No |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

Amenities object

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | No |
| No duplicates | No |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

Price float64

|  |  |
| --- | --- |
| Accuracy | 80 |
| Integrity | 100 |
| Secure | no |
| Completeness | 80 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | No |
| No duplicates | No |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

Minimum nights int64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | No |
| No duplicates | No |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

Maximum nights int64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | No |
| No duplicates | No |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

Availability 365 int64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | high |
| Timeliness | 100 |
| No redundancy | No |
| No duplicates | No |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

Calendar last scraped object

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | No |
| No duplicates | No |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

Number of reviews int64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | No |
| No duplicates | No |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

Last Review Date object

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | No |
| No duplicates | No |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

Review Scores Rating float64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | No |
| No duplicates | No |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

Review Scores Accuracy float64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | No |
| No duplicates | No |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

Review Scores Cleanliness float64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | No |
| No duplicates | No |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

Review Scores Checkin float64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | No |
| No duplicates | No |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

Review Scores Communication float64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | No |
| No duplicates | No |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

Review Scores Location float64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | No |
| No duplicates | No |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

Review Scores Value float64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | No |
| No duplicates | No |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

Reviews per month float64

|  |  |
| --- | --- |
| Accuracy | 100 |
| Integrity | 100 |
| Secure | no |
| Completeness | 100 |
| Validity | low |
| Timeliness | 100 |
| No redundancy | No |
| No duplicates | No |
| Relevance | yes |
| Accessibility | No |
| Availability | yes |
| Consistency | yes |

## Data Quality Analysis in Power BI

