

Re: Data analysis – Report on Data Quality Assessment.

Dear Manager,

I have evaluated the datasets based on the dimensions you have mentioned in the email.

- Accuracy – correct values
- Completeness – data fields with values
- Consistency – values free from contradiction
- Timeliness – values up to date
- Relevancy – data item with value meta-data
- Uniqueness – records that are not duplicated
- Validity – data containing allowable values

During the process, I have found out a lot of inconsistencies in the data. The summary table below highlights key data quality issues we have discovered in the data cleaning process.

Dataset	Accuracy	Completeness	Consistency	Timeliness	Relevancy	Uniqueness	Validity
(Dataset 1) Customer Demographics	Erroneous: DOB	Missing: last name, job title, job industry, tenure	Format: gender	Out of date: deceased customer	Corrupted: default column deleted		Unreliable: age
(Dataset 2) Customer Address			Format: state				
(Dataset 3) Transactions	Missing: profits (cost - price)	Missing: customer ID (3501-4999), online order, brand, product line/class/size, standard cost, first sold date	Format: list price, standard cost		Filtered: cancelled orders		Format: product sold date
(Dataset 4) New Customer List		Missing: last name, DOB, job title, job industry			Irrelevant: unnamed columns	Duplications: customers	

Please let us know if you have comments or queries about the same as I would be happy to discuss to ensure that all assumptions applied align with Sprocket Central Ltd.'s understanding.

Kind regards,

Gifty Cheruvallimalayil