

Taimoor Ali (2k20/ITE/136)
Fayaz Ali (2k20/ITE/39)
Mubeen Gul (2k20/ITE/152)
Rahul Dev (2k20/ITE/102)
Ali Kolachi (2k20/ITE/17)

Assignment : Data Science

BSIT-PART 3RD SEMISTER-6TH

To improve data quality, start by assessing your data and defining what is acceptable quality. Organizations also need to eliminate data silos, correct errors upfront, and promote a data-driven culture. The best way to improve data quality is to use an automated data quality management solution, such as DataBuck.

How can we assess our data quality :-

Data quality meets six dimensions: accuracy, completeness, consistency, timeliness, validity, and uniqueness.

How to Improve Data Quality in our Organization

1-Assess Your Data.

2-Define Acceptable Data Quality.

3-Correct Data Errors Up Front.

4-Eliminate Data Silos.

5-Make Data Accessible to All Users.

6-Use the Correct Data.

7-Impose a Defined Set of Values for Common Data.

8-Secure Your Data.

How can we improve the quality of data in an experiment?

You can increase the validity of an experiment by controlling more variables, improving measurement technique, increasing randomization to reduce sample bias, blinding the experiment, and adding control or placebo groups.

What are the 4 categories of data quality?

There are five traits that you'll find within data quality: accuracy, completeness, reliability, relevance, and timeliness.

Github link (Code)

```
Import pandas as pd  
carbs = ['pizza', 'hamburger', 'rice']  
pd.Series(carbs)
```

```
0 'pizza'
```

```
1 'hamburger'
```

```
2 'rice'
```

```
dType: object
```

➤ We can also create a series with label attributes

```
Foods = pd.Series(['pizza', 'hamburger', 'rice'], index=['Italy', 'USA', 'Japan'])
```

```
Italy 'pizza'
```

```
USA 'hamburger'
```

```
Japan 'rice'
```

```
dType: Object
```

➤ Query a Series

```
foods.iloc[2]
```

```
foods.loc['USA']
```

```
'hamburger'
```