

# More on Missing Data

[\\_ \(https://github.com/learn-co-curriculum/dsc-more-on-missing-data\)\\_](https://github.com/learn-co-curriculum/dsc-more-on-missing-data)[\\_ \(https://github.com/learn-co-curriculum/dsc-more-on-missing-data/issues/new/choose\)](https://github.com/learn-co-curriculum/dsc-more-on-missing-data/issues/new/choose)

## Introduction

Now that you've seen various methods of how to deal with missing data, its time to further discuss how to choose an appropriate methodology given a particular scenario. Commonly, many people will immediately turn to imputing the mean or median of a feature with missing values. This can be a valid and effective methodology, hence why it is standard, but does have caveats. For example, doing so will reduce the overall variance of your dataset which should be taken into account when performing subsequent analyses or training a machine learning algorithm on the dataset.

## Objectives

You will be able to:

- Evaluate and execute the best strategy for dealing with missing, duplicate, and erroneous values for a given dataset
- Determine how the distribution of data is affected by imputing values
- Inspect data for duplicates or extraneous values and remove them

This tool needs to be loaded in a new browser window

Load More on Missing Data in a new window



Help

How do you feel about this lesson?



Have specific feedback?

[Tell us here!](https://github.com/learn-co-curriculum/dsc-more-on-missing-data/issues/new/choose) ➞ [\(https://github.com/learn-co-curriculum/dsc-more-on-missing-data/issues/new/choose\)](https://github.com/learn-co-curriculum/dsc-more-on-missing-data/issues/new/choose)