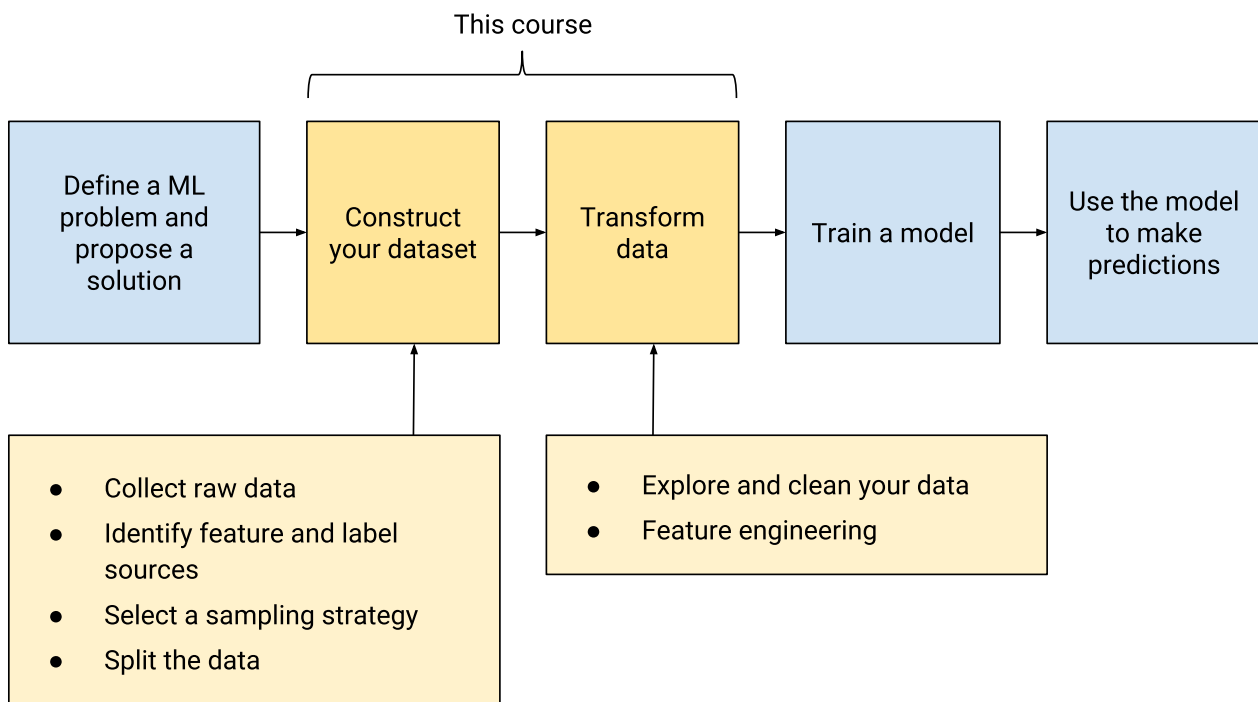


The Process for Data Preparation and Feature Engineering

What's the Process Like?

As mentioned earlier, this course focuses on constructing your data set and transforming your data.



Keep in mind:

- The figure shows a *typical* process, which might not be ideal for every project. This course applies primarily to linear regression and neural nets.
- The process shown is not always sequential. You might, for example, split your data *after* you transform it. You might need to collect more data. You might need to modify the feature set, even after training begins, as you learn empirically what works and what doesn't.

How Much Time Does it Take?

For the following question, click the desired arrow to check your answer:

Take a guess: In your machine learning project, how much time will you typically spend on data preparation and transformation?

Less than half of the project time



More than half of the project time



[Previous](#)

← [Introduction](#) (/machine-learning/data-prep)

[Next](#)

[Introduction to Constructing Your Dataset](#) →
(/machine-learning/data-prep/construct/construct-intro)

Except as otherwise noted, the content of this page is licensed under the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) (<https://creativecommons.org/licenses/by/4.0/>), and code samples are licensed under the [Apache 2.0 License](https://www.apache.org/licenses/LICENSE-2.0) (<https://www.apache.org/licenses/LICENSE-2.0>). For details, see the [Google Developers Site Policies](https://developers.google.com/site-policies) (<https://developers.google.com/site-policies>). Java is a registered trademark of Oracle and/or its affiliates.

Last updated 2021-02-05 UTC.