Let's now look into the basic steps

to be repeated with new information.

And that's the science in Data Science.

in the Data Science process.

We will keep applying some or all of these steps

in all of the case studies we discuss in this course.

After this video, you will be able to

identify the steps in the data science process

and understand what each step involves.

We have already seen a simple linear form of

data science process including five distinct activities

that depend on each other.

Let's summarize each activity further

before we go into the details of each.

Acquire includes anything that makes us retrieve data

including finding, accessing, acquiring, and moving data.

It includes identification of an authenticated access

to all related data, transportation of the data from

different sources, and ways to subset and match the

data to regions or times of interests.

Sometimes we refer to this as geospatial query.

We divide, prepare data into two steps based on

the nature of the activity.

First step in data preparation involves literally

looking at the data to understand its nature, what it means,

it's quality, and format.

It often takes a preliminary analysis of data

or samples of data to understand this.

This is why this step is called Prepare.

Once we know more about the data through

exploratory analysis, the next step is pre-processing

of data for analysis.

It includes cleaning data, subsetting or filtering data,

and creating data that programs can read and understand

via modeling raw data into a more defined data model

or packaging it using a specific data format.

If there are multiple datasets involved, this step also

includes integration of data from

different data sources or streams.

The prepared data then will be passed onto

the analysis step which involves selection of

analytical techniques to use, building a model of the data,

and analyzing results.

This step can take couple of iterations on its own

or might require a data scientist to go back to steps

one and two to get more data or package data

in a different way.

Step four is for communicating results.

It includes evaluation of analytical results,

presenting them in a visual way,

and creating reports that include an assessment of results

with respect to success criteria.

Activities in each step can often be referred to the terms

like interpret, summarize, visualize,

and post-process.

The last step is to bring us

back to the very first reason we do data science for.

Its purpose.

Reporting insights from analysis and determining

actions from insights based on the purpose

you initial define is why we refer to as the Act Step.

We have now seen all of the steps in a typical

data science process.

Please note that this is an iterative process and findings

from one step may require previous steps