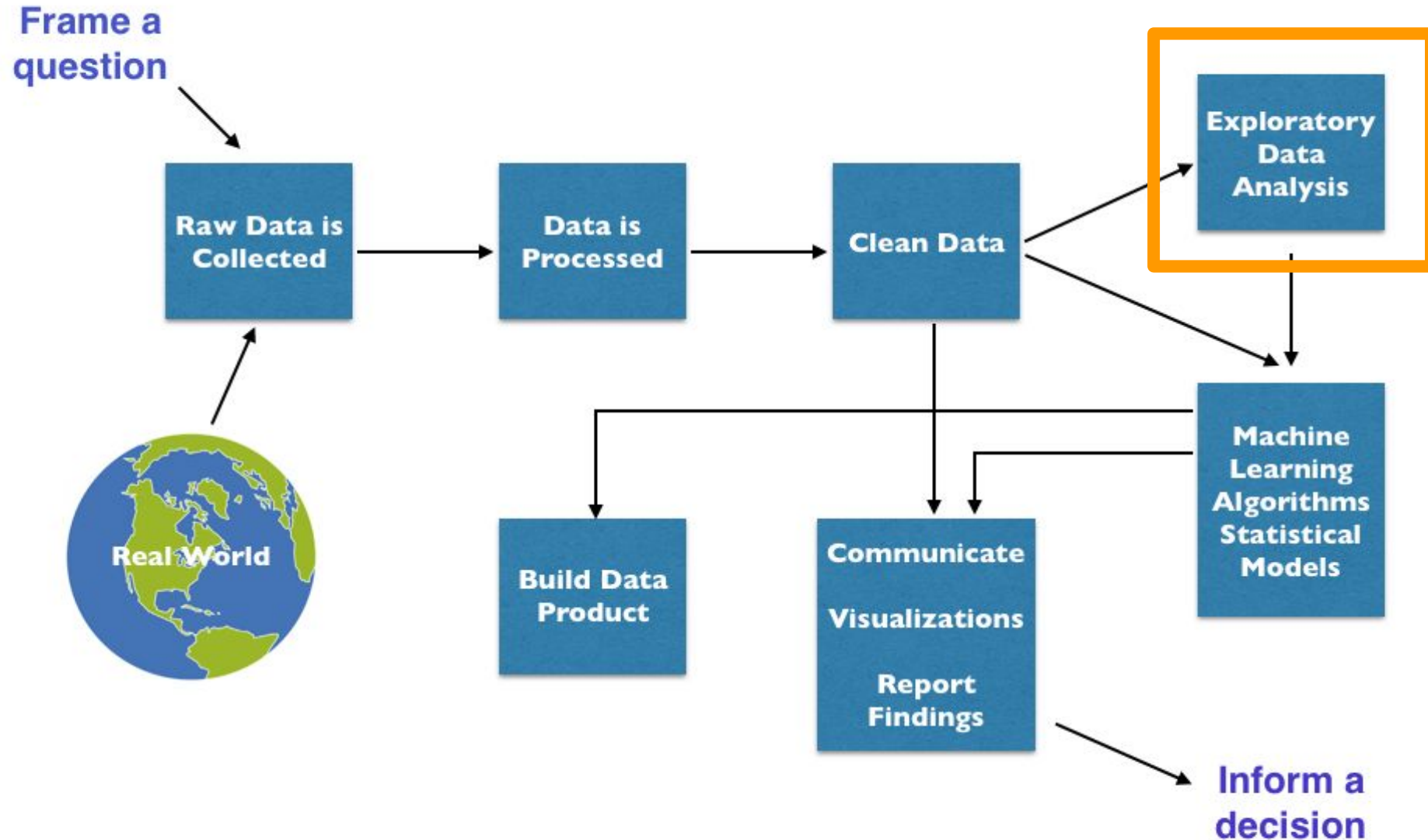


Exploratory Data Analysis

Data Science Process



Exploratory Data Analysis (EDA)

- Systematic Exploration of a Dataset
- Somewhat of an Art

An iterative process:

1. Generate questions about your data.
2. Search for answers by visualising, transforming, and modelling your data.
3. Use what you learn to refine your questions and/or generate new questions.

Exploratory Data Analysis

Exploratory analysis, by its very nature will result in following a lot a dead ends. Just continue to explore.

EDA is necessary for any project involving data.

At a minimum, you can determine if the data meets your expectations (data types, missing values, etc.)

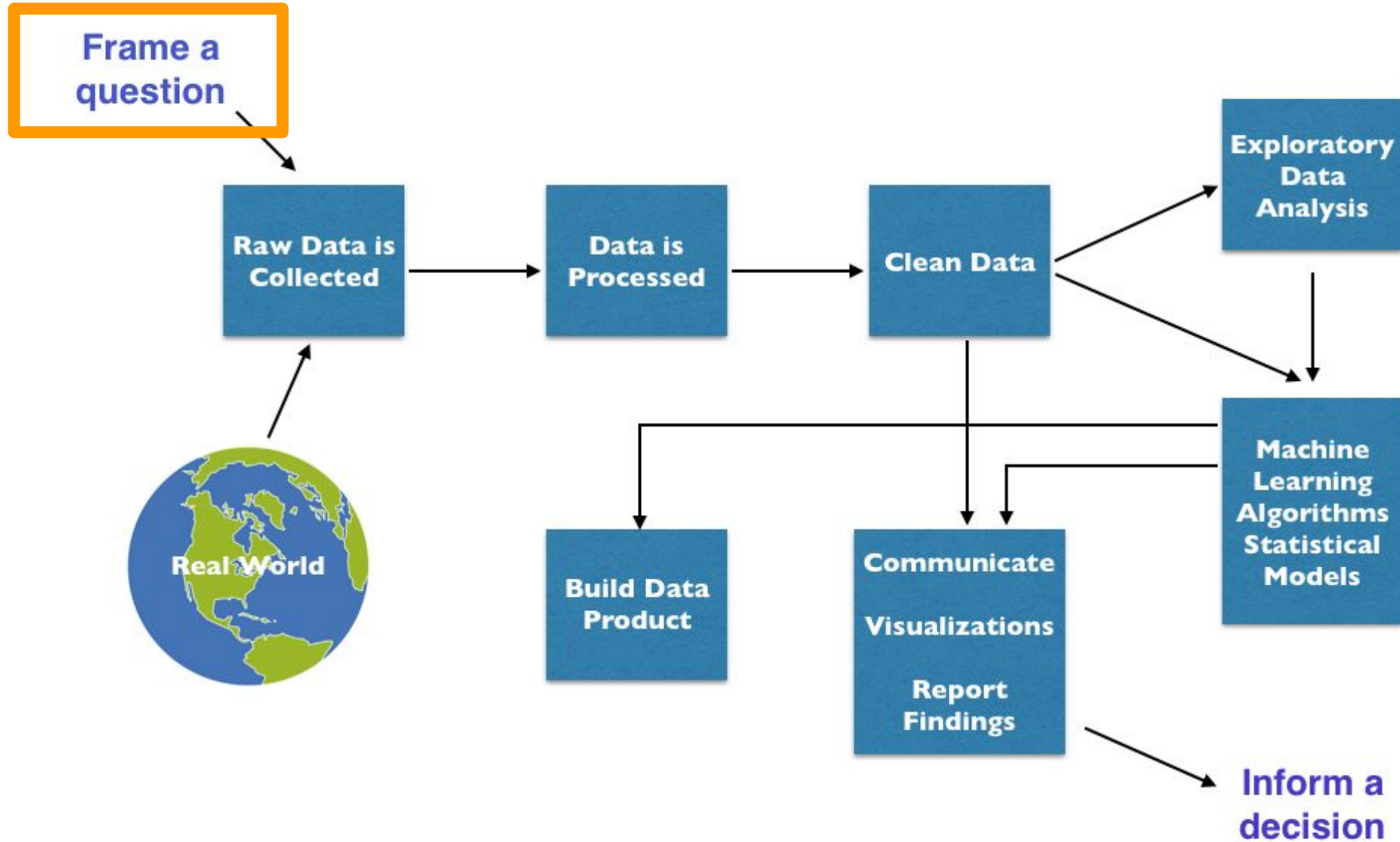
Exploratory Data Analysis

Ask a lot of questions. As you answer those questions, new questions or ideas will develop.

Starter questions:

- What variables and data types do I have?
- How are those variables distributed?
- What relationships are there between variables?
- Do I have missing values? If so, can I explain why they are missing?

Data Science Process



Exploratory Data Analysis

When working with data, you will almost always have a goal/objective in mind. This can guide the beginning of your exploration.

Starter questions:

- What variable am I most interested in?
- How are other variables related to the variable(s) that I care about?

Exploratory Data Analysis Example



Now, we'll do an EDA walkthrough using the [NASA Exoplanet Archive](#).

Our main objective is to understand, through the lens of the available data, the various methods of planet discovery.