* **What is the scientific goal?**

The goal is to predict the income of the individuals and create a classification and correspondence in the people.

* **What would you do if you had all the data?**

Datasets are a collection of instances that all share a common attribute. Machine learning models will generally contain a few different datasets, each used to fulfill various roles in the system.

For machine learning models to understand how to perform various actions, training datasets must first be fed into the machine learning algorithm, followed by validation datasets (or testing datasets) to ensure that the model is interpreting this data accurately.

Once you feed these training and validation sets into the system, subsequent datasets can then be used to sculpt your machine learning model going forward. The more data you provide to the ML system, the faster that model can learn and improve.

Therefore if I have all the data I can prepare a more efficient and accurate model.

* **What are the Features or Attributes of data set**

Attribute is an important type of semantic properties shared among different objects or activities. It is a representation in a higher level than the raw feature representation directly extracted.

These are generally the rows/ columns contained in the dataset.

* **What do you want to predict or estimate?**

The project will make use of machine learning to estimate the average salary of an individual using the given features and dataset.