Notes:

* Dateset 0 – has both key1 and key2. It has some values and response as well
  + Response – 1 or 0
  + Keys and values - integer
* Dataset 1 – has only key2 and a lot of values for it
  + Key2 – integer
  + Values – float or null (nan / na). Majority are NaN
* Dataset 2 – hase only key1 and a lot of values for it
  + Key1 – integer
  + Values – mainly text(string). Some floats and integer. Some columns are (mainly) NaN
* drop\_NA\_only\_columns\_and\_rows – assumes no duplicates in column names

Questions:

1. Should we do inner join of task.app.data with task.cb.1 and task.cb.1? Or outer join?
2. When joining should we use key1+key2 as a unique key? Or we just separately add all values for key1 and then for key2?
   1. We should add separately all values for key1 and key2 because dataset 1 has only key2 and dataset 2 has only key1
3. Response variable is only in dataset0. So, for factor importance analysis and prediction we should use only inner join?