Assignment 1

You are given a dataset about individuals with multiple features. Source of data:

[<https://archive.ics.uci.edu/ml/datasets/Census+Income>]

Your target is to predict if the income of a given individual is <=50K or >50K (USD).

Recommendations:

1. Since this is a classification problem, you could use the Random Forest model.
2. You need to remove the remove the “target” feature from the training set.
3. Draw graphs where relevant to examine the data.
4. Split the data and target into training and test datasets using “train\_test\_split”.
5. Use “OrdinalEncoder” to convert non-numerical features such as “workclass”, “occupation” etc. into numerical features.
6. Use “RandomizedSearchCV”