17.6.2

## **Encode Labels With Scikit-learn**

**Pandas**, as you have seen, offers tools to encode your data. Scikit-learn offers another way to encode your labels.

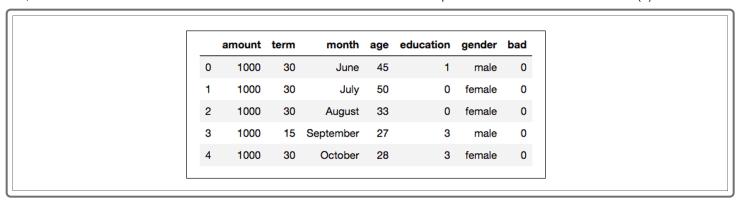
Scikit-learn's <u>LabelEncoder</u> module can also transform text into numerical data. Let's look at an example. Continue down the notebook from the preceding section:

```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
df2 = loans_df.copy()
df2['education'] = le.fit_transform(df2['education'])
```

The code includes the following elements:

- After importing the module, an instance of the label encoder object is created and assigned the variable le
- A copy of the original <u>loans\_df</u> is created for this example, but this step is not necessary for using label encoder.
- The label encoder's <u>fit\_transform()</u> method is used to first train the label encoder, then convert the text data into numerical data.

The result is a numerical encoding of the education column. In contrast to pd.get\_dummies(), the label encoder assigns a number between 0 and 3 for each of the education categories. The applicant in the first row, for example, has the value 1, which represents high school or below:





© 2020 - 2022 Trilogy Education Services, a 2U, Inc. brand. All Rights Reserved.