

```
1 # === binning.ipynb ===
2 # just a convenience way to categorize your data w/o using if then else stmt
3 # a good example is converting numerical grades to letter grades
4
5 #%%
6 # Import Dependencies
7 import pandas as pd
8
9 #%%
10 raw_data = {
11     'Class': ['Oct', 'Oct', 'Jan', 'Jan', 'Oct', 'Jan', 'Nov'],
12     'Name': ["Cyndy", "Logan", "Laci", "Elmer", "Crystle", "Emmie", "test"],
13     'Test Score': [90, 56, 72, 88, 98, 67, 59]}
14 df = pd.DataFrame(raw_data)
15 df
16
17 #%%
18 # Create the bins in which Data will be held
19 # Bins are 0, 60, 70, 80, 90, 100
20 bins = [0, 60, 70, 80, 90, 100]
21
22 # Create the names for the four bins
23 group_names = ["F", "D", "C", "B", "A"]
24
25 #%%
26 # -----
27 # When using the `pd.cut()` method, three parameters must be passed in:
28 #     1. the Series that is going to be cut
29 #     2. a list of the bins that the Series will be sliced into
30 #     3. a list of the names/values that will be given to the bins.
31 #
32 # It is important to note that, when creating the list for bins, Pandas will
33 # automatically determine the range between values. This means that, when
34 # given the list [0, 60, 70, 80, 90, 100], Pandas will create bins with
35 # ranges between those values in the list.
36 #
37 # The labels in the pd.cut() method must be one fewer element than those in an
38 # the bin (as the lowest possible value, otherwise an error will be returned:
39 #     `Bin labels must be one fewer than the number of bin edges`
40 # -----
41
42 df["Test Score Summary"] = pd.cut(df["Test Score"], bins, labels=group_names)
43 df
44
45
46 #%%
47 # What makes binning so powerful is that, after creating and applying these
48 # bins, the DataFrame can be grouped according to those values and thus a
49 # higher-level analysis can be conducted.
50
51 # Creating a group based off of the bins
52 df = df.groupby("Test Score Summary")
53 df.max()
54
55 #%%
```