


```
import pandas as pd
dataset = "/Student_bucketing.csv"
df = pd.read_csv(dataset, header = 0)
```

Double-click (or enter) to edit

Data discretization is the process of converting continuous data into discrete buckets by grouping it. Discretization is also known for easy maintainability of the data

Once we load the dataframe, display the first five rows of the dataframe. Add the following code to do this:

```
df.head()
```



	Student_id	Age	Grade	Employed	marks
0	1	19	1st Class	yes	29
1	2	20	2nd Class	no	41
2	3	18	1st Class	no	57
3	4	21	2nd Class	no	29
4	5	19	1st Class	no	57

Double-click (or enter) to edit

Perform bucketing using the `pd.cut()` function on the marks column and display the top 10 columns. The `cut()` function takes parameters such as `x`, `bins`, and `labels`. Here, we have used only three parameters. Add the following code to implement this:

```
df['bucket']=pd.cut(df['marks'],5,labels=['Poor','Below_average','Average','Above_Average','Excellent'])
df.head(10)
```

	Student_id	Age	Grade	Employed	marks	bucket
0	1	19	1st Class	yes	29	Poor
1	2	20	2nd Class	no	41	Below_average
2	3	18	1st Class	no	57	Average
3	4	21	2nd Class	no	29	Poor
4	5	19	1st Class	no	57	Average
5	6	20	2nd Class	yes	53	Average
6	7	19	3rd Class	yes	78	Above_Average
7	8	21	3rd Class	yes	70	Above_Average
8	9	22	3rd Class	yes	97	Excellent
9	10	21	1st Class	no	58	Average

Double-click (or enter) to edit

In the preceding code, the first parameter represents an array. Here, we have selected the marks column as an array from the dataframe. 5 represents the number of bins to be used. As we have set bins to 5, the labels need to be populated accordingly with five values: Poor, Below\_average, Average, Above\_average, and Excellent.

[Colab paid products](#) - [Cancel contracts here](#)

