

## Cs2065 Spring 2022

### Discussion: Groups, Joins, Pivots (Project 1 Lab)

Welcome to Project 1 Lab! This week we will be discussing groups, joins, and pivots. The group function allows us to aggregate the unique entries in one or more columns, while the join function allows us to merge data from two tables into a single table. Alternatively, pivot allows us to aggregate over the unique values of two columns. All of these methods are vital to creating simple yet powerful tables that assist in analyzing data.

**Question 1.** Ian has opened up a chocolate store where he sells small boxes of chocolates in groups of different sizes and colors. His table `chocolates` is as follows:

Color	Shape	Amount	Price (\$)
Red	Round	4	1.30
Green	Rectangular	6	1.20
Blue	Rectangular	12	2.00
Red	Round	7	1.75
Green	Rectangular	9	1.40
Green	Round	2	1.00

Notice that the table contains multiple rows containing information about chocolates of the same color. We would like to figure out how many chocolates of each color he has for sale in total, and what the cost would be to purchase all chocolates of each unique color.

- Write a line of code that will return a new table which displays the total number of boxes for each color.
- Write a line of code which will return a new table with the total number of chocolates and the total cost for each unique color. For example, the row for “Red” should have a total of  $4+7=11$  chocolates, and a total cost of  $\$1.30 + \$1.75 = 3.05$ .

**Question 2.** The table below, called `weights`, contains information about the weights of the chocolates that are sold. The weights of the chocolates differ depending on the shape, and round chocolates have two different sizes.

Shape	Weight(g)
Round	3.1
Round	4.25
Rectangular	3.6
Triangular	2.9

The following line of code has been executed in a blank cell. Take a moment to discuss with your neighbors what the resulting table will look like. Then, write the number of columns and rows in the resulting table, and describe the information in the table in 1-2 sentences.

Hint: It may help to draw a sketch of the resulting table!

```
chocolates.join('Shape', weights)
```

**Question 3.** We will continue with the same table as before, copied below for your convenience.

Color	Shape	Amount	Price (\$)
Red	Round	4	1.30
Green	Rectangular	6	1.20
Blue	Rectangular	12	2.00
Red	Round	7	1.75
Green	Rectangular	9	1.40
Green	Round	2	1.00

Write code to create a pivot table on the colors and shapes of chocolates, finding the average price for each color-shape combination. Then, fill in the blank table in the image of the resulting table.

*Hint:* You can use the `np.average` function to find the average of an array of inputs. The average of no values is marked as zero.

	Rectangular	Round
Blue		
Green		
Red		