

CS2065 Spring 2022

Discussion: Data Types and Table Manipulation (Lab03)

In lecture, you have been introduced to various *data types* in Python such as integers, strings, and arrays. These data types are particularly important for manipulating and extracting useful information out of data, an important skill for data science. In this section, we'll be analyzing some of the behavior that Python displays when dealing with particular data types.

Discuss each of the following questions with the people around you.

1. What Would Python Do?

For each of the following examples, presume that the code was run in an empty cell. Write down what Python would output. If the code results in an Error, explain why an error would occur.

a. `"I love " + "Data 8"`

b. `"I love Data " + 8`

c. `np.arange(1, 4) + np.arange(2, 7, 2)`

d. `make_array(3, 4, 5) + np.arange(2, 7, 1)`

2. Fun with Arrays

Suppose we have executed the following lines of code. Answer each question with the appropriate output associated with each line of code.

```
odd_array = make_array(1, 3, 5, 7)
even_array = np.arange(2, 10, 2)
```

a. `odd_array + even_array`

b. `odd_array.item(1)`

c. `even_array.item(3) * odd_array.item(1)`

d. `odd_array*3`

In this section, we will practice working with tables. In particular, we will take a look at how to use Python and functions to manipulate tables for them to show what you are interested in.

3. eBay Auctions

Your friend Al is curious to see whether or not it's cheaper to buy his favorite items on eBay rather than through some other platform! Al stumbles upon some auction data from eBay, and decides to use his newly developed Table skills to do some data-crunching. However, Al is making a few mistakes and needs your help. For the following questions, identify why the code won't work as is.

The table below is called `ebay` and contains more than just the 3 rows displayed.

Auction_ID	Item	Opening_Bid	Closing_Price
1	Jacket	50	75
2	Smartwatch	100	150
3	Tablet	350	600

- a) # Use comments to describe the code you write
`ebay.where('Opening_Bid', are.above(60)) / ebay.num_rows`
- b) `ebay.column('Closing_Price') - ebay.select('Opening_Bid')`
- c) Al really wants a new jacket, but his budget is only \$150. To see whether eBay has had good deals on jackets historically, Al tries to filter the auction data such that it only contains jackets that were sold for a closing price less than \$150. He writes the following code to do so, but hasn't realized the mistake he's making. Help Al fix it so that your friend can hopefully get the jacket he deserves!

```
only_jackets = ebay.where('Item', 'Jacket')
jackets_under_price = ebay.where('Closing_Price', are.below(150))
```

4. Violence in California

You are working on a project related to arrest-related violence in California. After a bit of searching, you finally find some relevant data and import it into your Jupyter Notebook. The table `arrests` is shown below. (Dataset source: <https://www.kaggle.com/sohier/arrest-related-violence-in-california>)

	Incident_Date_Str	City	Zip_Code	Num_Involved_Civilians	Num_Involved_Officers
0	7/3/2016	Hayward	94544	1	2
1	11/20/2016	San Leandro	94578	1	1
2	5/30/2016	Dublin	94568	1	2
3	6/30/2016	Dublin	94568	1	2
4	5/25/2016	Hayward	94545	1	1

- a) Before you start your exploration, you want to understand your data better. For each of the columns **City**, **Num_Involved_Civilians** and **Zip_Code**, identify if the data contained in that column is numerical or categorical.
- b) Suppose you were only interested in arrests that involved two, three or four civilians. Assign `range_civilians` to a table that only contains rows from the `arrests` table that correspond to arrests involving two, three or four civilians.

```
range_civilians = _____
```

- c) You're curious about finding out when the arrest involving the most officers occurred in Hayward. Assign `arrest_date` to a string that represents the date of the arrest that occurred in Hayward that involved the most officers. For this question, assume that the maximum number of officers involved in an arrest in Hayward is unique. It is okay for your code to wrap along the next line for `hay_arrests_sorted`.

```
hay_arrests_sorted = _____
```

```
arrest_date = _____
```