

Python for Librarians - Week 2 Workalong

Answers for Week 2 workalong just in case you need them.

Question 1

Describe what the `int()` function does in the following markdown cell

The function `int` will...

turn the value you pass to it into an integer. Useful when changing strings into integers for when we compare them to other integers.

Question 2

Describe what the `input()` function does in the following markdown cell.

The `input` function will...

Prompt the user to to input a value via the keyboard. This values is kept as a string.

Question 3

Fix the following cell so that it prints *They are the same!*. (HINT: there are at least two ways to do this)

```
In [ ]: number = "4"

if int(number) == 4:
    print("They are the same!")
```

First let's reload our Toronto Weather Data from Week 1 and run the `len` function on it. Python tell us how many entries are in that dictionary. Long story short, Python will do a pretty good job of figuring out what you are trying do when you use a function. In the case of `len` in counts the letters if you give it a string, in the case of a dictionary it gives you the number of entries in the dictionary. Neat.

Question 4

Complete the following function that will find the maximum value in a list of integers. (Hint: you only need to modify lines: 11 & 17)

```
In [ ]: numbers = [2, 5, 2, 9]

def findMax(numbers_to_consider):

    # We will start with automatically setting the highest number
    # to the first value in our list.
    # (or in other words: the number at the 0 index in the list)
    max_number = numbers_to_consider[0]

    for n in numbers_to_consider:
        if n > max_number:
            max_number = n

    return max_number

result = findMax(numbers)
print("Largest number is: ")
print(result)
```

Question 5

Complete the following function that will find the smallest value in a list of integers. (Hint: you only need to modify lines: 11 & 15)

```
In [ ]: numbers = [2, 5, 2, 9]

def findMin(numbers_to_consider):

    # We will start with automatically setting the highest number
    # to the first value in our list.
    # (or in other words: the number at the 0 index in the list)
    min_number = numbers_to_consider[0]

    for n in numbers:
        if n < min_number:
            min_number = n

    return min_number

result = findMin(numbers)
print("Smallest number is: ")
print(result)
```

```
In [6]: import pandas as pd
```

```
#Pandas has a function that will open up a csv and load it into a dataframe
sfl_data = pd.read_csv("https://raw.githubusercontent.com/elibtronic/lja_datasets/master/week_2_workalong_san_francisco.csv")

#will show us the top 10 entries in our dataframe using the head function along with the argument 10
sfl_data.head(10)
```

```
Out[6]:
```

	Patron Type Definition	Total Checkouts	Total Renewals	Home Library Definition	Circulation Active Month	Circulation Active Year	Year Patron Registered
0	YOUNG ADULT	0	0	Marina	None	None	2015
1	ADULT	3	12	Richmond	April	2014	2013
2	JUVENILE	11	0	Bayview/Linda Brooks-Burton	June	2014	2010
3	JUVENILE	171	111	Merced	July	2016	2014
4	YOUNG ADULT	14	1	Richmond	October	2013	2011
5	ADULT	5	0	Main Library	May	2013	2013
6	YOUNG ADULT	194	178	North Beach	July	2016	2004
7	ADULT	26	9	Glen Park	July	2016	2003
8	ADULT	14	0	Ingleside	September	2013	2012
9	ADULT	0	0	Main Library	January	2014	2014

To get a general quantitative overview of our dataframe we apply the **.describe()** function to it. It will only apply those calculations to columns that have numeric data.

Question 6

Complete the following cell to find out the average (or mean) of the **Total Checkouts** column?

```
In [ ]: sfl_data["Total Checkouts"].mean()
```

Question 7

Complete the following cell to find out the highest number of **Total Checkout** for a patron.

```
In [ ]: sfl_data["Total Checkouts"].max()
```

Question 8

In the cell below reflect on the two code cells above about average checkouts between 2015 & 2016. What can you say about them?

The difference between 2015 & 2016 is...

People who created accounts in 2015 checked out and renewed more material on average compared to people who created an account in 2016. What did the Library do different in 2015 I wonder?

Question 9

In the markdown cell below describe your observations about Adult versus Juvenile borrowers for the Main Library.

The difference between these two patron groups is...

Adults

- Average amount of accounts created was in 2010

Juveniles

- Average amount of accounts creatwe was in 2012
- Juvenile's on average check out more material and peform more renewals over adults

Question 10

If we want to see if either of two conditional is true (eg. OR) we combine them with `|` . Complete the following code cell so it counts how many patrons are in the main library or Chinatown

```
In [ ]: sfl_data[(sfl_data["Home Library Definition"] == "Main Library") | \
              (sfl_data["Home Library Definition"] == "Chinatown")].count()
```

Question 11

Can you find what the average number of checkouts for adult borrowers at the Glen Park library?

```
In [ ]: sfl_data[(sfl_data["Home Library Definition"] == "Glen Park") & \
              (sfl_data["Patron Type Definition"] == "ADULT")].mean()
```

Question 12

Can you find the the highest number of checkouts for adult borrowers at the Main Library or Chinatown?

```
In [ ]: sfl_data[((sfl_data["Home Library Definition"] == "Main Library") | \
                  (sfl_data["Home Library Definition"] == "Chinatown")) & \
              (sfl_data["Patron Type Definition"] == "ADULT")]["Total Checkouts"].max()
```

Question 14

After reading up on what a correlation matrix is (with the link above). What can you say about the data in the SF Library user data? Add some thoughts in the markdown cell below. (You'll note that much like as when we use `describe()` calculating the correlation matrix is done automatically on just numeric data in the dataset.

The Correlation Matrix Tells me that....

Many things, but for a start:

- *Total Renewals* is positively correlated with *Total Checkouts*
- *Year Patron Registered* is negatively correlated with *Year Patron Registered*
- *Year Patron Registered* is negatively correlated with *Total Renewals*