Advanced Business Decision Support Systems Professor Deepu Philip

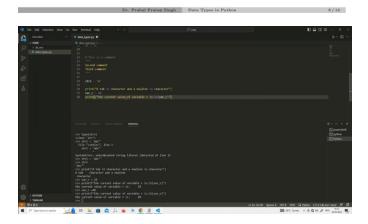
Department of Industrial Engineering and Management Engineering

Indian Institute of Technology, Kanpur Professor Amandeep Singh Imagineering Laboratory Dr. Prabal Pratap Singh Indian Institute of Technology, Kanpur Lecture 29

Data Types (Part 2 of 2)

Hello everyone, I welcome you all to the lecture series on Advanced Business Decision Support Systems and I am Prabal Pratap Singh from IIT Kanpur. We are discussing about the implementation aspects of decision support systems using Python. So, as you all know, till now we have covered the Elements and Classifications of programming languages and discussed what is Python programming language. Next, we also created our development environment using the Python executable file downloaded from its official site and we created our development environment on windows. So, today we will be discussing the data types in python. So, we already discussed few things about data types.





So, moving to our next data type that is Strings, which is the most common type of data type. The very first special thing about strings is that, strings are a kind of sequence data types. So, what are the Sequence data types? So, we until now, we have seen that, if we are defining a variable let us say, a is equal to 40. So, the memory storing one particular value in a particular location, but what happens, if we need to store multiple values in a sequential order like 40, then 20, then 30.

So, let us say that how many kind of fruits are available in a particular warehouse. So, let us say, in the first warehouse, there are 40 fruits, in the second, there are only 20 kinds, in the third warehouse, there is 30. So, this can be stored in a particular memory location and this will become a Sequential data type. So, strings also are classified as sequential data type instead of using numbers, they are using characters to store in sequences. So, these are nothing but collection of single or multiple characters.

So, here we can see on the right, that a single character string can look like this. Here, we are only using a particular character, this can be a, b, c or anything and we are storing it in a variable string 1 (str1). So, if again, I am repeating that if we will write here it as 1 str, then this will not be a valid variable, but str1 is a valid variable name. The next thing is Multi character string which can be written as here. So, this complete str2 is a single variable that is holding abcd space efgh right and to define a particular character as a string in python, we will use single quotes or double quotes or triple quotes.

So, single quote and double quotes are interchangeable, if you are using single quotes to start a string, then you need to end the single quote to end the string. If you mix and match these two quotes like if you start with double quote and end with single quote, then this will be a syntax error. So, let me write this as well, Python identifies text inside a quote, quote can be single, double, triple as a string. We can use triple quotes to create Multi line strings. So, we can see, this in our environment like, to create a multi line string.

So, we have seen that, a single line comment is like this is a comment. Now, to create multiple line comments, we can use second comment or you can write third comment. Otherwise, what you can do is, you start with a triple quote and remove these. So, this is a valid multi—line string and the interpreter will not throw an error here. Again, let us define a simple string str1 = 'a'.

So, if you run this, it stores the interpreter, now stores str1 as a string and you can check the type of this variable using the type function str1 and it shows that it is using a class string. Now, if you write string str1 = 'abc''. So, this is a syntax error here unterminated string literal. So, the interpreter is saying that, you started it with a single quote, but you have not yet terminated it. So, if you just change the other thing with a single quote.

Now, this is a valid string. So, you can see, the value of str1, what holds is abc with a double quote. So, this whole abc and double quote is the value of this variable str1. Now, I have already shown you that using type function, we can see that the variables are that string variables, are objects of class string str, which confirms that everything is an object in python. The last thing about strings is that, we can use special characters like /n or /t, these are the escape sequences to insert new lines or tag characters that is, white spaces in a string.

So, here we can see, this string 4, it is using this tab character /t and /n. So, what happens is, /t will be a tab character. So, after this tab, there will be a blank space and then a and d will start and here after line space, then a new line will start. So, we can check this, how the interpreter will print the statement. See, when I ran the above statement using shift plus enter, the interpreter says that this /t is missing from this output.

Why? Because /t was interpreted as a tab character here. So, there is a large gap between tab and character. Similarly, new line and character should be on the same line, but after new line, we have inputted a /n character. So, that is why, the interpreter moved to the next line. So, this is how you can use, you can customize your output using print statements.

Also, I have shown you in the presentation that, we can use f strings, these are f strings to use a dynamic variable into a string. So, that we can dynamically provide an output to the screen when the programmer wants. So, let us say, the current value of variable r is and I will use tab character /t and open and close curly braces and then write the variable name here where _ r. So, to use where _ r, we first need to define it. So, let us say, var _ r has a value of 24.

So, we first run var _ r is 24. Now, we write this. So, what happens in this print statement is after variable r is a colon /t will provide a tab character and then we should get the value of var_r, the value stored in var r as an output to the string. So, let us print this. So, here it is showing that the current value of variable r is 24.

Now, if I change the value of var r in my interpreter as 48 and then again run this command. So, now it will print current value of variable r is 48. So, where this thing will be useful is when, if you want to check the value of a particular variable throughout its computation during the progression of your code and you want to check, whether the value of variable is changing as per your computations or not, you can create these kind of print statements. Although, you should use logging statements for these, but that will be an advanced stage. So, before that, we can use these kind of print statements to check whether your computation is going well or not.

So, let us move further. So, until now, we have defined strings and saw how to create different kinds of print statements using strings, but since, string is an object and a class, as we will see in the advanced lectures of this python programming that class has data and methods inside it. So, every object has data and associated methods with it. So, since, string is an object, so string will also have this kind of behavior. So, strings will have predefined methods here.



So, some of the most used methods in strings are, we can use find to search inside a string. So, let us see, this as an example here. We are defining a new string which is month is August and year is 2023. Now, we can either find a single character s or we can find a complete word like year. So, the output of this first statement will be the location of the first occurrence of s in this string and similarly, for the complete word, if the word exists in this, then the interpreter will give you the output of the first occurrence of the location of the word year.

So, we can see this year, we will create another section by writing a new comment string methods and save the file and now create year, 1 is August and year is 2023. Now, we will find the occurrence of s character. So, it is showing 11. So, what this 11 hold is? As I have told you in earlier lecture that, the indexing starts with 0 in python and now, we know that string is a sequence data type. So, the indexing will start from 0.

So, the t in the memory location, it is not stored in the first location, it is stored in the 0th location. So, counting from here, the t is at the 0th location, h is at the first location. So, 0, 1, 2 the space is also holding a position. So, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11. So, s is at the 11th position in the memory.

So, this is what this find method is showing that the first occurrence of s is at 11th. Now, similarly we can find the location of year, if it is available in the string. So, it is showing 24. So, until s, it was 11 then, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24. So, wherever the word year starts, it will give you the location of the first character of that word.

So, it is giving the output as 24. So, let us save the script the next useful method is count. We can use count the occurrence of a particular string. So, here we can see that, in our string, there are two is. So, we can use a count method to find the number of occurrences of the word is and it should give the output as 2 for year, it should give the output as 1.

So, let us see what is. So, it is showing 2 and similarly, for the word year, str is 41. count year, it is only 1. So, you can use any word, any character using this count method. Next, useful method is Capitalize, you can use capitalize to make the text in capital case. So, these kind of methods are useful when we are trying to change the cases of our long text. So, this looks trivial for short strings, but let us say, you are changing the whole text of a particular e-book.

So, you can use this to change a huge amount of text in that complete e-book and see, whether if you need to capitalize all the sentences, then just use this single function and it will do the job. Similarly, let us say, if you have a complete paragraph and you want to change it into upper case that is, in all caps. So, you can use this method upper. So, what it will do is, it will change the complete text into upper case, next thing is, you can use Title method to create the text in title case. So, title case is let us say, we have a string world of wonder.

So, this is not in title case right now. So, title case will be, the first character of every word should be in capital letters. So, this function will do it. The next useful thing is, we can also check, whether the string contains alphabets, alphanumeric text or only digits using functions like isalpha(), isalnum ()or isdigit(). Now, this is very crucial when we are taking the input from users.

So, this we can see with the example. I have also written this example here. So, you can see that we can use this input function that is predefined in python to get some input from the user and user can do anything while providing the input. It can provide you just the number, just the text or the mixture of both and it does not matter what you want from the user, it can provide anything. So, you want to first check, whether user has

given the correct input or not. So, to do that, we can use these kinds of methods in strings.

So, let us see, how this works. So, we can create a new variable user _ input and we will use the predefined input method and this input method, we can use just like that without putting any string inside it or we can create a string. So, that user will see the string while giving the input. So, let us say, that we want a digit input from the user. So, we can ask for enter the age in numbers /t. Now, the user should give an input in numbers.

But what if the user is not giving it in numbers? So, we can first run this statement. So, as soon as we run this statement, the interpreter has seen that there is an string provided inside the input with a /t character. So, it has created a tab character here and is asking for a input after the tab character. So, I can write third t, which is not a correct input, but user can do that. So, we can see, what user input variable holds user _ input.

So, it holds a string, but we need a number. So, when you are creating a long form of code. So, you have not the capability to check all these things by going to the interpreter and writing all this stuff. So, what happens is, you can use the prebuilt functions like, user_input. Now, these are all the methods that are available.

So, we can use is digit and run this. So, it is showing false here. So, Python is saying that, this input does not contain numbers, it contains something else. So, let us run this line again and give it a correct input as 30. Sorry, here 30 and now run this check method again and now it is showing true.

So, this is how we can use these things. So, let us say, if the user has given something very different: 30, thirty both in numbers and in words. So, this is a alphanumeric text and we can also check this whether the string contains an alphanumeric text or not. So, we used here the method is lnum is alphanumeric and it is showing true. So, this is how these string methods are useful. So, we can also use replace to find and replace the provided substring.

So, we all have done it at least once in our life that, we use excel or in Microsoft word that, we need to find something and then replace it in the whole document. So, Python can also do it with this replace method. So, here I have shown you an example like in this string, there is this end and I want to replace it with this character ampersand. So, I will just write str1 . replace the thing that I need to find and the final thing that I want to replace it with. So, I will first provide the end and then the ampersand operator and it will do the job, we can see this quickly.

So, let us first check whether our string contains this statement or not yes. Now, we can write str1 . replace a and d and replace it with ampersand characters. So, you can see

here that, our string has been changed and end is replaced by the ampersand character, but keep in mind that, when we are using these methods. So, these are not overwriting the string. So, str1 is still the same in the memory, but the output of this operation is here, it has not yet overwritten the original string.

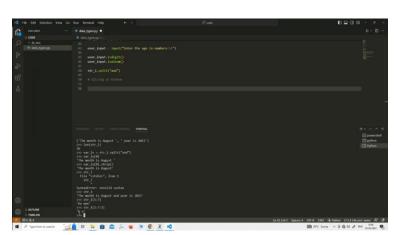
So, some methods, in some kind, some different data types can overwrite things, but this replace method is not that kind. So, we will also look those methods that can replace things in the original string or list or other kind of data types. The other method is, we can use split to divide a string based on a particular character or string in the available text. So, here I have shown you the example that let us say, this complete sentence contains two different sub sentences like month is August and year is 2023. So, what if we want to divide this complete string into two on the basis of this end.

So, we can use the split function and provide this end as a string, then it will create a new data type. We are using strings here, but it will create a new data type, which is list and it will store the first sub sentence and the second sub sentence in the list. So, we can see here and it will also be an introduction to a list str1 . split on the basis of string a and d. So, you can see here that, our complete string is has been split into the month is August and the other one is space year is 2023. So, why these two spaces? Because when we wrote our string so, the python used this end and it gave as a output, this complete string including this space character.

Similarly, this space character is included in the second sub string and this is the notation of a list with a square bracket and a comma d limits different quantities or different elements of a list. So, this list has two elements. Similarly, we have also this one useful function to check the length of string. So, what if we need to find the length of a particular paragraph or text or complete book. So, we can use this simple function len str _ 1 and it is showing that, it contains 36 characters.

So, this is also useful function which we will use in our further coding. One more useful function is strip, use strip to remove the white spaces from the string. So, what are the white spaces? So, let me show you here that when we are running this split command, it





is giving these two sub strings and this has a white space. So, we can remove it using the strip command. So, let me store this into a variable and I can get the first element of this list using the square notation and at the zeroth index.

So, it is showing me that at the zeroth index of the list it contains this string. Now, since, this is a string. So, I can use the method strip and it has removed the white space from it. So, this is how we can use these things. So, there are many more string methods, but these are some of those that are very useful and we will use and the programmers also use this very frequently.

So, moving on to our next concept that is, slicing in python. So, this is useful. Since, we are talking about sequential data types. So, sequence data types act like containers that can store multiple values in a deterministic order. So, let us understand this statement. So, what happens is, I have already shown you that, if we are using a single variable, then it is storing, then it gets stored in a single location in memory.

Let us say, this is a container and it stores 40. Now, sequential data types has this kind of array like behavior, which can store multiple quantities sequentially. So, the first location stores 40, second may store 50, third store 7, 4, 3 like this. So, storage is ok, but how to retrieve or modify these things? So, what python developers did, they configured it in the memory using this deterministic order. So, what they say is that, any sequential data type will start its location using the zeroth index. So, the first location is 0, second is 1, third is 2, similarly, 3 and 4.

So, this sequential data type which contains only integers has the address of the first value at 0, then 1, 2, 3 and 4. So, these are like containers that can store multiple values in a deterministic order. So, whenever you want to retrieve the seventh value, you will say that the name of this container let us say, bd and you will ask the interpreter to go to this variable bd and search at the second location, what is the value. This second location is actually 3, but as per the python interpreter, it is 2 and then, retrieve that value.

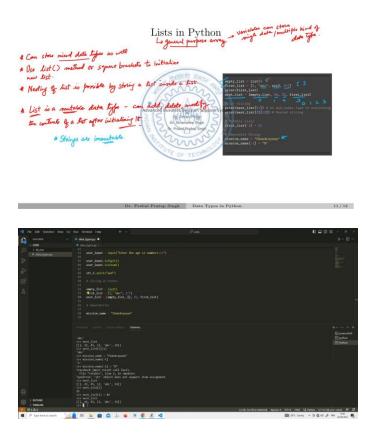
So, interpreter will give you the output as 7 right. So, string is one out of many sequence types available in python and programmers often need to retrieve or modify particular parts of sequences which comes under this slicing feature. So, either a programmer wants a single value or it can also ask for a series of values from 1 and going until the third location. So, this is getting a slice from this sequential data type and the interpreter will provide these three values 57 and 4. So, we will learn here how to extract or retrieve this kind of series of values from a sequential data type, what is the correct syntax for that. So, let me again mention this that, a sequence container assigns integer index to each unit in the sequence starting from 0.

Now, let us see the syntax of the slicing. So, let us say that, we have a variable with the name var _ name then, we will use the square brackets, we will open the square bracket then, we will write the start index, which is nothing, but this integer value starting from 0 until the length of the sequence, you can use any one of it then, you will use a colon operator. After that, you will write the end index. After that, one more optional parameter is to provide the step. So, what happens is, this step can skip some values while counting in the sequential data type.

So, let us say, we provide a step of 2 starting from 0. So, what happens is that, interpreter will give first value is 40, then it will skip two values and give the second value is 4 and since, there are no more values, so it will stop here. So, this kind of slicing with step of 2 will only provide two values if we start from 0. So, we can see this in our Visual Studio code by creating a new section. So, we have our string available str _ 1 and we can use the slicing here as well. So, let us say, str _ 1, open the bracket and write starting from first, we need to check until the seventh value as the end index.

Since, the step is a optional parameter, we can skip it also. So, what it will do is, it will go to the first index and until the last seventh index, it will print all the values and here let us see, zeroth is t. So, it skipped that after that, he 1, 2, 3, 4, 5, 6, 7. So, the last value that is the end index, is not included in the slicing operator in Python. So, that is why, the t, which is at the seventh location, will not get printed.

Now, you can also use this with a step command of 2. So, now, what we are saying is, skip two values by default, it was skipping nothing. So, here it will say, the first value is h then, it will skip e and space character and then it will give m and similarly, it will move forward. So, let us see the output h and o after that, the string completed at the seventh location. So, there are only two strings sorry, it created an output with three characters. The first is h, the other is the space character, after skipping the two values, the second value and then the o.



So, now, let us move to our next data type list, which is a general purpose array. There are no arrays in Python, but we call them list here they provide a similar functionality. So, these variables can store single data or multiple data, multiple kind of data types. So, it can store can store mixed data types as well. So, the first sequence data type, we saw was string, but string only includes characters here, we can include characters numbers, even list as well inside list, that we call them as nested list. So, we can use mixed data

types in strings which makes it general purpose and we can create an empty list using the list method or we can initialize a new list using square brackets.

So, this we can see here that, we are generating a new list by using just this list method and after that, we can also initialize a different list by using these square brackets and this string contains a number, a string, a variable itself which we have defined earlier and it can also do inline computations like this. So, 2 into 7 will be stored as 14 when the interpreter will run this code. So, nesting of list is also possible by storing a list inside a list. So, the example for this is here, the nest list variable will contain an empty list, which we defined here then, the first list which we defined here and these two numbers 45 and 22.

So, this list contains number data type, an empty list and a list as well. This list contains strings. So, this is called nesting of list and this variable holds different kinds of data types in itself and the numbering starts here as 0, 1, 2, 3. Now, since this is again a list. So, the next numbering starts from at the third, we start again with a square bracket and the numbering starts from 0, 1 for this abc, then 2 for the string, 2 variable and 3 for 2 into 7, 14.

So, this is how it works. Now, we can see the example of this quickly. So, let us first define empty list equals list. So, we can run this one by one. So, now, our interpreter has ran all the three commands and we have these three kind of list in our memory. So, we can print this like, first list and we can assess the different kinds of elements using the square notation. As we have discussed earlier, like first since, there are three numbers, but the indexing starts from 0.

So, the first value is available at the 0th location. So, 0, 1, 2. So, to get the string value from the list, we will write 1 here. I have not used the correct name of the list variable first list. So, it is giving the output as abc. Now, let us see what next list contains.

So, this contains an empty list, then two number data types and the other list. So, to assess abc again from this nested list, we will write next list square bracket and we will first give the locations available in this next list. So, that is 0th location for this empty list 1, 2 and at the third location, the other list is available.

So, we will go to the third location first, then we will close the square brackets and then we will again start a new square bracket to assess the location of abc in the second list and we have already seen that, this is available at 0 and 1 first location. So, this also gives abc. So, to assess nested list, we use these kind of multiple square notations to assess a particular element from the next list.

So, next important thing is list is a new table data type that is, it can add delete, modify the contents of a list after initializing it. So, this list sequence data type is a new table

data type, it can modify anything in the list, but the string data type which we have seen earlier strings are immutable. So, we can look at a example of this here. So, let us say that, the mission name is Chandrayan and we want to change the last n to the capital N.

So, the first thing which we should know is that, we can also in the last slide, I have shown you that we can initialize Sequence data type and then, start the indexing from 0, but what if a programmer wants to assess the last location of the data contained in a sequence data type?

So, to do that, Python programmers define that, we can also assess the last location by using the negative sequencing. So, negative sequencing does not start with 0 from the right, it starts from minus 1, minus 2, minus 3, minus 4, minus 5. So, we can see here that, to assess this last n, we can use mission _ name and square notation with the minus n and to overwrite that value, we can use this capital N as a string. So, let us see this and we will provide a new section here with immutable t and let us initialize our variable mission name of course, Chandrayan, since it is a sequence data type. So, we can assess the last location by using minus indexing and it will provide you the last character as n.

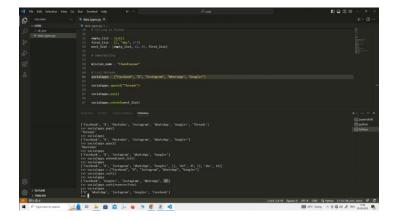
Now, to overwrite this character in string, you can say that, assign this n character as a capital N, but the interpreter should throw an error here like this, type error string object does not support item assignment.

Why? Because strings are immutable, but let us say, we want to change this value in nest list 22 to 44. So, we can do this. Let us first check what nest list contains.

So, it contains the same things. Since, we have not modified anything. So, nest list and assess this 22 location as 0, 1. So, the value 22 is available at the first location at the index 1. So, to overwrite this, we just need to assign it as 44.

Now, there are no errors. So, let us again check nest list. Now, it has been overwritten with 44. So, this way, you can mutate a list, but cannot mutate a string. You can also mutate this same location with a string and it will work as well. So, now, the same thing is towards the string. So, this is how a string can handle different kinds of data types and that is different kinds of data types and that is why, we usually call it a general purpose data type.





Now, similar to the other sequence data types like, string, list also contains its own methods. So, we can use append method to add an element at the end of the list. We can also use insert method to add an element at a specified location. So, you may ask both of these methods are adding something to a list that a user has defined. So, what is the difference? Now, the difference here is at a specified location or at the end of the list. So, you can see here that, we have defined a new list named as social apps and it contains these different kinds of social applications available.

So, if we use the append method, then this newly available threads, social application will be appended to the end of the list. However, if we insert at a particular location that is here, the programmer is providing at the index location 2, we should insert this string mastodon. So, the list will contain this at the location 2.

So, we can quickly see. We will first initialize our variable by providing the values. So, now we can check the difference between append and insert method. So, social apps append, we can append threads and it should go to the end of the list. So, let us see. Now, you can see that, threads is available at the end and as I have already told you that, in strings, there are some functions, there are some methods that overwrite the contents or there are some methods that do not overwrite the contents.

Similarly, in list also, this method will overwrite the content. This will not just provide you a new output. It has overwritten our original parent list. Now, use the insert method and see the Facebook is available at the zeroth location. The next is 1, 2. So, we need to provide this new string at the second location. So, we will first give the location as 2 and then, we will provide our string which is master and this will insert the this string before Instagram.

So, let us see again what happened. So, you can see that, at the second location 0, 1, 2, the mastodon string is available before Instagram. So, next important method available in list is, we can use pop method to remove the element from the end of the list. So, by default, it will remove from the end, but if you want to remove a particular element, then we can also specify the location. Also, we can use extend to add any kind of iterable.

Iterable is nothing but sequence data types like list or we will see dictionaries as well in our further lectures.

We can also use sort; to sort the list in a in a particular order and this sort method has two arguments with it like reverse and key. key. So, what reverse do is, it is a bool means, it is a binary argument it will be either true or false. So, if true, it will sort the complete list in a reverse order. Similarly, if we are using a key, then we can provide a function which we will discuss in our further lectures, how to write a function and how to return values from it and if it returns some values based on those outputs, the sort method can sort the complete list on the returned output from that function.

So, we can also see this on our code. So, we can use the pop method to first see, what it performs by default, social x . pop. If there is a method, then it should use a opening and closing parenthesis. Let us run. So, it printed the last value from the list and should have removed from the container, that is list.

So, let us see whether it has removed or not that is, whether it has overwritten the values or not. So, now you can see that, the interpreter does not contain threads in its original list, but what if we want to remove a particular value at a particular location using the pop method? So, we can do that as well, social apps . pop and provide the particular location like, two where we inserted mastodon earlier and every time we use pop, it will always provide the return output that has been removed from the list.

So, we can see here that, earlier it showed threads, now it is showing mastodon and we can check, whether it has removed or not. So, now you can see that there is no mastodon in the list. Similarly, we can extend this list by using the extend function, using the extend method social apps .extend and earlier we have a nested list.

So, we can use that list here, let us run this and see what happened social apps. So, now our original social apps list has been extended with a new list, nested list which we used earlier and you can see after this, google plus string, we have the empty list which was available in nested list. Then, def, then 45 and other another list which was contained in the nested list. So, this way, we can extend a list with another list. The last thing which we discussed is to sort.

So, if we try to run this again social apps original social apps and then try to sort this with default input. So, now these are following a particular alphabetical sorting method in ascending order like, f (facebook) is the first, google plus is the next, f, g, i, w, x, but originally it was something different f, then x, then instagram, then whatsapp, then google plus. So, this way, what if we want to provide this sorting order in reverse. So, we can use . sort the argument reverse equals true and then, it should sort this in reverse order.

So, now you can see the last element earlier x has become first, then WhatsApp, then Instagram, google plus and Facebook. So, these are some of the major data types that python contains and we will discuss further data types like dictionaries, sets, tuples in the upcoming lectures, but before that to understand few very beautiful characteristics of the list and other data types, we need to first learn what are the control flow statements? what are the looping statements and other things in python.

So, that we can use features like, list, comprehensions and other things. So, in the next lecture, we will discuss these control statements first and then we will complete our data types like, dictionaries and other available data types in python. So, thank you very much. I hope you are practicing all these things in your development environment as well. If you find any difficulties, please reach out us to in forums and we will help you out. Thank you.