



# SUMMER ANALYTICS 2020



*Welcome to the Data Engineering assignment of Summer Analytics 2020.*

**IMPORTANT:** After completing the assignment , it is very important that the following form is attempted <https://forms.gle/n7AjWfbVhkU5GrkM9> . If not attempted you will not be graded for the assignment. There is no need of submitting the notebook to our team , the grading will be done on the basis of your performance in the submission form.

**REMEMBER :** IT IS ADVISED THAT THE ASSIGNMENT IS ATTEMPTED ONLY AFTER COMPLETION OF THE 5 DAYS OF WEEK 1.

- To begin with, please download the ipynb file to the default directory of your Jupyter Notebook (Generally this will be C:\Users\your\_username)
- Once copied to the default directory , you can browse the file when you first open Jupyter , click on the assignment notebook which will be visible in the file search window. Now you may begin solving the assignment.
- The Assignment requires the knowledge of Python , Pandas , Numpy, Data Analysis Skills and Python Web Scraping( **Some Additional Information is given in the next page**)
- Certain questions will require searching the web for additional Python/ Pandas/Scrapin concepts which may not be part of Summer Analytics 2020. It is very important that you study it on your own as this forms a vital part of the learning process.
- This Assignment is graded and will form a vital part of the final grade.
- Please follow the honor code and violation of the same will lead to harsh actions being taken.

Stay Safe , Stay Healthy and Enjoy the Course

Regards  
Team CnA, IIT Guwahati



# SUMMER ANALYTICS 2020



## Additional Material

- `soup.find_all('tag')` -> This beautiful soup command is used to find all the occurrences of a particular tag which is passed as an argument.
- `get_text()` is a method which may be used to get the text associated with a particular tag. *(Refer to the web for more information)*
- The `strip()` method removes any leading (spaces at the beginning) and trailing (spaces at the end) characters (space is the default leading character to remove) from a string.
- A lambda function is a small anonymous function. A lambda function can take any number of arguments, but can only have one expression.

*A lambda function that adds 10 to the number passed in as an argument, and print the result:*

```
x = lambda a : a + 10
```

```
print(x(5))
```

***Please don't restrict yourself to this document, everything here is explained very briefly but a in-depth knowledge of the above would help you solve the assignment and also benefit you in the future.***