Data Engineering Assignment Submission

Total points 15/15

Welcome to the submission form for the Data Engineering Assignment. The Form consists of 15 questions from multiple topics , answers to which may be found by working on the assignment. Some questions may require studying certain topics from the web , no support material will be provided for the same, it is upto the participants to study it on their own.

General Guidelines:
1) All Questions are compulsory and thus should be attempted.
2) Each Question has some weightage and will contribute in the final grading of the course
3) Please attempt this if you have completed all the 5 days of Week 1
4) Violation of the honor code will lead to harsh actions being taken.

Cheers!

0 of 0 points

IMPORTANT!!!
Attempt this form only after you receive the User ID Pass and Password.

You'll recieve by 16th April 10 am

If not recieved by then, mail us at caciitg@gmail.com

Honor Code

1) You can submit the form only once , using a single ID Pass. Usage of multiple accounts for submission of quizzes will lead to harsh actions being taken.

2) Your answers to the questions must be your own work.

3) You may not share your solutions with anyone else unless explicitly permitted by the mentor. This includes anything written by you, as well as any official solutions provided by the course .

4) You may not engage in any other activities that will dishonestly improve your results or dishonestly improve or damage the results of others.

You can report Honor Code violations by contacting any of the members of Consulting and Analytics Club, IIT Guwahati.

Please Enter you Name *

_____

Please enter your User ID Pass correctly in the following text field (This will be needed for grading) *

_____

Please enter the password alloted along with the User ID by our Team. *

_____

Please type " I Accept the Honor Code and will not violate it in any possible way" in the following text field. *

_____

SUMMER ANALYTICS 2020

Data Engineering Assignment Submission

15 of 15 points

✓ How many 'p' tags are present in the webpage ? * 1/1

- ○ 67
- ◉ 85 ✓
- ○ 12
- ○ 48

✓ How many 'table' tags are present in the webpage ? * 1/1

- ◉ 27 ✓
- ○ 19
- ○ 15
- ○ 36

✓ What is the text associated with the h1 tag? * 1/1

- (●) 'Harvard University' ✓
- ( ) 'Wikipedia'
- ( ) 'Harvard'
- ( ) 'History'

✓ Of all the h3 tags find the text associated with the 4th h3 tag from the output(Hint: You will get a list of tags as output once you use soup.find_all()) * 1/1

- ( ) 'Cambridge'
- (●) '21st century' ✓
- ( ) '20th century'
- ( ) 'Colonial'

✓ Each table has a class associated with it (class is an attribute of the table tag). Of all the classes from all the tables , find the 6th class from the output which is received on listing out all the classes of all the tables. * 1/1

- ( ) ['toccolours']
- ( ) ['infobox']
- ( ) ['wikitable', 'sortable', 'collapsible', 'collapsed', 'floatright']
- (●) ['wikitable'] ✓

✓ Convert the soup object created in the assignment to a string format and assign it to a new variable named soup_s. Now apply the method .split() on this string(soup_s) and store it in variable ls. What is the value of ls[10]? * 1/1

- (●) 'University' ✓
- ( ) 'Julian–Gregorian'
- ( ) 'short'
- ( ) 'semi-protected'

✓ What is the value of ls[50]? (Refer to the last question for instructions on how to create ls) * 1/1

- ( ) 'the'
- (●) 'against' ✓
- ( ) 'in'
- ( ) '500,000+'

✓ What is the datatype of column 'Undergrad' in df? * 1/1

- (●) int ✓
- ( ) object
- ( ) boolean

✓ Using df.corr() find from the correlation table the value showing the correlation between 'Undergrad' and 'Grad/prof' columns * 1/1

- ○ 1
- ◉ 0.730634 ✓
- ○ 0.848438
- ○ 0.930511

✓ Using df.corr() find from the correlation table the value showing the correlation between 'Undergrad' and 'US census' columns * 1/1

- ○ 1
- ○ 0.730634
- ◉ 0.848438 ✓
- ○ 0.930511

✓ Find the mean for the column 'Undergrad' from the cleaned dataset(df_clean)? * 1/1

- ◉ 16.33 ✓
- ○ 11.5
- ○ 9.5
- ○ 18.75

✓ Find the standard deviation for the column 'Grad/prof' from the cleaned dataset(df_clean). * 1/1

- ○ 18.75
- ○ 15.78
- ◉ 14.96 ✓
- ○ 13.24

✓ Find the range for the column 'Grad/prof' from the cleaned dataset(df_clean). * 1/1

- ◉ 35 ✓
- ○ 38
- ○ 33
- ○ 36

✓ Calculate X = df_clean/(df_clean.sum()). Using X find out the race in the Undergrad column which has the maximum value. * 1/1

- ○ Black
- ◉ White ✓
- ○ Asian
- ○ International

✓ Calculate X = df_clean/(df_clean.sum()). Using X find out the race in the Grad/prof column which has the value = 0.051020 * 1/1

- (●) Black ✓
- ( ) White
- ( ) Asian
- ( ) International

SUMMER ANALYTICS 2020