Edit with the Docs app

Make tweaks, leave comments, and share with others to edit at the same time.

NO THANKSUSE THE APP

Data Science Africa Challenge

Fully-Funded Trip to Attend the Data Science Africa Conference in Abuja (12th - 16th November, 2018): Competition Guidelines

Requirements:

1. A statement of interest (not more than 200 words)

2. Solution to the problem set below

3. Valid passport (at least six months before expiry)

Send your statement of interest and solution to the problem as an email to code@developersinvogue.org by 11:59pm GMT on Sunday 14th October, 2018.

Only competition finalist will be contacted.

This competition is brought to you by Inclusive and Developers in Vogue.

Problem Set:

DataFrames Basics [ Remember to import numpy using: import numpy as np ]

Consider the following Python dictionary data and Python list labels:

data = {'animal': ['cat', 'cat', 'snake', 'dog', 'dog', 'cat', 'snake', 'cat', 'dog', 'dog'],

'age': [2.5, 3, 0.5, np.nan, 5, 2, 4.5, np.nan, 7, 3],

'visits': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],

'priority': ['yes', 'yes', 'no', 'yes', 'no', 'no', 'no', 'yes', 'no', 'no']}

labels = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j']

1. Create a DataFrame df from this dictionary data which has the index labels.

2. Display a summary of the basic information about this DataFrame and its data.

3. Return the first 3 rows of the DataFrame df.

4. Select just the 'animal' and 'age' columns from the DataFrame df.

5. Select the data in rows [3, 4, 8] and in columns ['animal', 'age'].

6. Select only the rows where the number of visits is greater than 3.

7. Select the rows where the age is missing, i.e. is NaN

8. Select the rows where the animal is a cat and the age is less than 3.

9. Select the rows the age is between 2 and 4 (inclusive).

10. Change the age in row 'f' to 1.5.

11. Calculate the sum of all visits (the total number of visits).

12. Calculate the mean age for each different animal in df.

13. Append a new row 'k' to df with your choice of values for each column. Then delete that row to return the original DataFrame.

14. Count the number of each type of animal in df.

15. Sort df first by the values in the 'age' in descending order, then by the value in the 'visit' column in ascending order.

16. The 'priority' column contains the values 'yes' and 'no'. Replace this column with a column of boolean values: 'yes' should be True and 'no' should be False.

17. In the 'animal' column, change the 'snake' entries to 'python'.

18. For each animal type and each number of visits, find the mean age. In other words, each row is an animal, each column is a number of visits and the values are the mean ages (hint: use a pivot table).