**1. How to import pandas and check the version?**

**A screenshot of a computer program

AI-generated content may be incorrect.**

**A screenshot of a computer

AI-generated content may be incorrect.**

**2. How to create a series from a list, numpy array and dict?**

**A screen shot of a computer code

AI-generated content may be incorrect.**

**3. How to convert the index of a series into a column of a dataframe?**

A screen shot of a computer code

AI-generated content may be incorrect.

**4. How to combine many series to form a dataframe?**

**A computer code on a white background

AI-generated content may be incorrect.**

**5. How to assign name to the series’ index?**

A screenshot of a computer

AI-generated content may be incorrect.

**6. How to get the items of series A not present in series B?**

**A white rectangular box with black text

AI-generated content may be incorrect.**

**7. How to get the items not common to both series A and series B?**

**A screenshot of a computer code

AI-generated content may be incorrect.**

**8. How to get the minimum, 25th percentile, median, 75th, and max of a numeric series?**

Difficuty Level: L2

Compute the minimum, 25th percentile, median, 75th, and maximum of ser.

Input

ser = pd.Series(np.random.normal(10, 5, 25))

Show Solution

**A screenshot of a computer code

AI-generated content may be incorrect.**

**9. How to get frequency counts of unique items of a series?**

Difficulty Level: L1

Calculte the frequency counts of each unique value ser.

Input

ser = pd.Series(np.take(list('abcdefgh'), np.random.randint(8, size=30)))

Show Solution

A screenshot of a computer program

AI-generated content may be incorrect.

**10. How to keep only top 2 most frequent values as it is and replace everything else as ‘Other’?**

Difficulty Level: L2

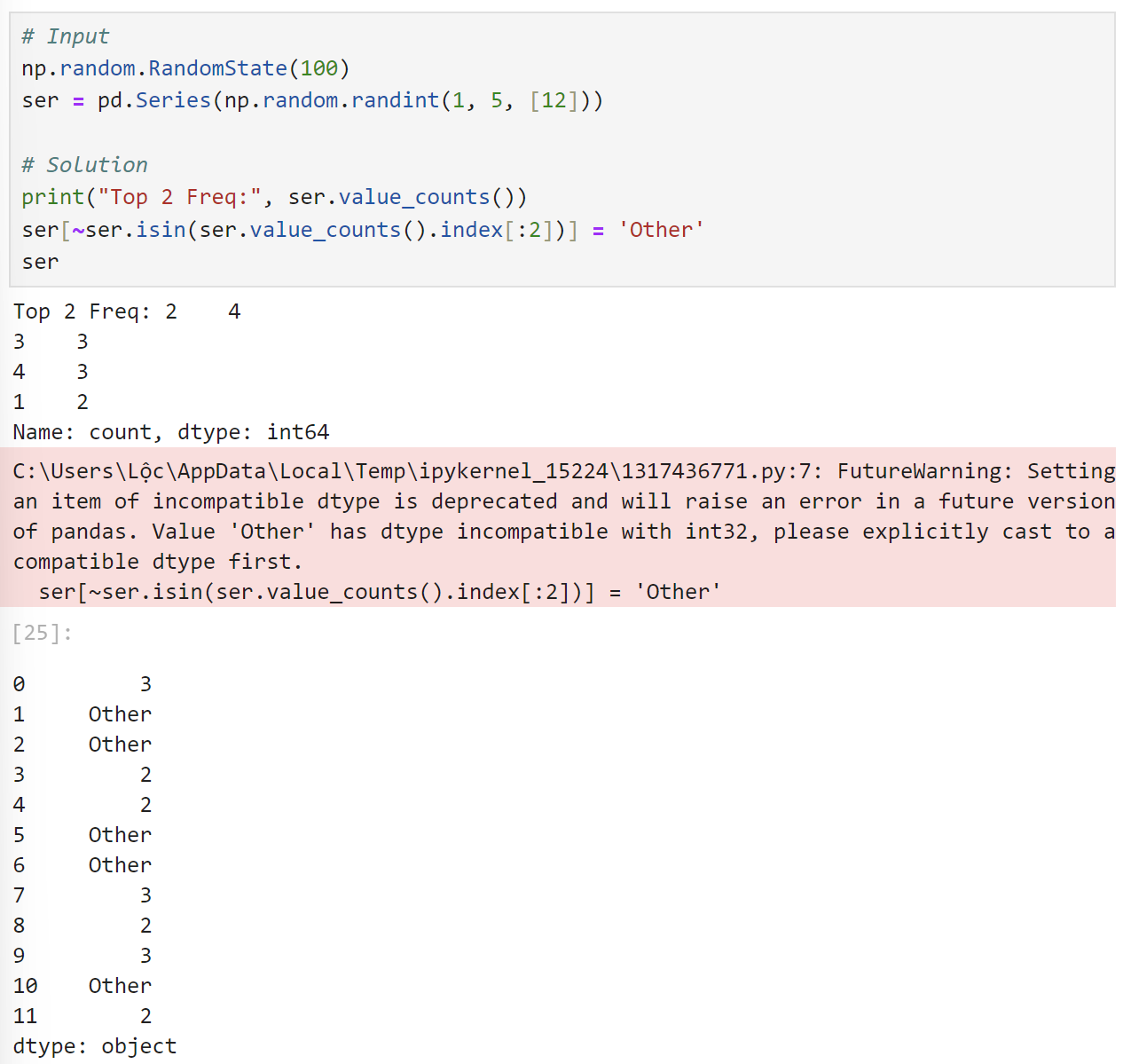
From ser, keep the top 2 most frequent items as it is and replace everything else as ‘Other’.

Input

np.random.RandomState(100)

ser = pd.Series(np.random.randint(1, 5, [12]))

Show Solution



**11. How to bin a numeric series to 10 groups of equal size?**

Difficulty Level: L2

Bin the series ser into 10 equal deciles and replace the values with the bin name.

Input

ser = pd.Series(np.random.random(20))

Desired Output

# First 5 items

0 7th

1 9th

2 7th

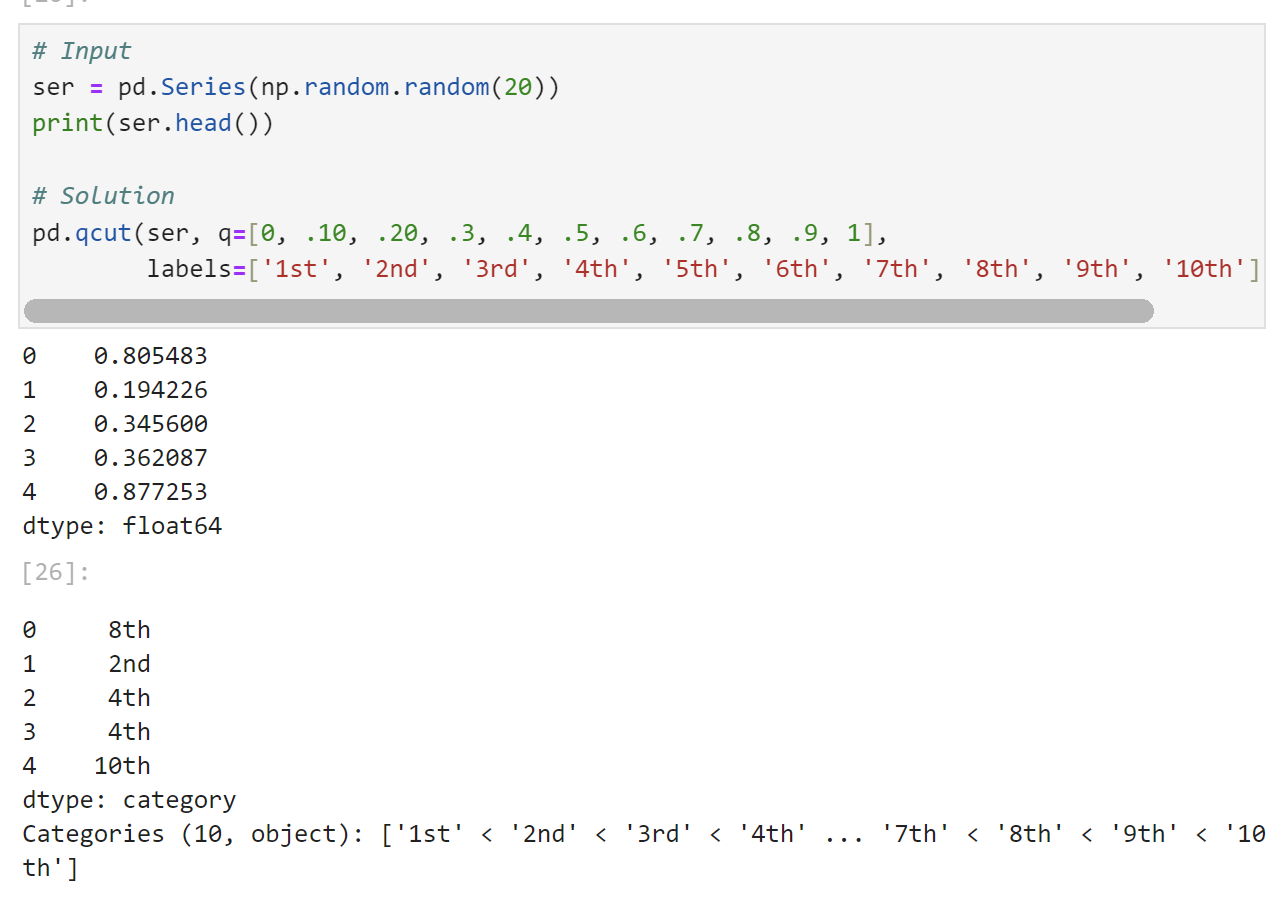
3 3rd

4 8th

dtype: category

Categories (10, object): [1st < 2nd < 3rd < 4th ... 7th < 8th < 9th < 10th]

Show Solution



**12. How to convert a numpy array to a dataframe of given shape? (L1)**

Difficulty Level: L1

Reshape the series ser into a dataframe with 7 rows and 5 columns

Input

ser = pd.Series(np.random.randint(1, 10, 35))

Show Solution

A screenshot of a computer code

AI-generated content may be incorrect.

**13. How to find the positions of numbers that are multiples of 3 from a series?**

Difficulty Level: L2

Find the positions of numbers that are multiples of 3 from ser.

Input

ser = pd.Series(np.random.randint(1, 10, 7))

Show Solution

A screenshot of a computer program

AI-generated content may be incorrect.

**14. How to extract items at given positions from a series**

Difficulty Level: L1

From ser, extract the items at positions in list pos.

Input

ser = pd.Series(list('abcdefghijklmnopqrstuvwxyz'))

pos = [0, 4, 8, 14, 20]

Show Solution

A screen shot of a computer

AI-generated content may be incorrect.

**15. How to stack two series vertically and horizontally ?**

Difficulty Level: L1

Stack ser1 and ser2 vertically and horizontally (to form a dataframe).

Input

ser1 = pd.Series(range(5))

ser2 = pd.Series(list('abcde'))

Show Solution

A screenshot of a computer code

AI-generated content may be incorrect.

**16. How to get the positions of items of series A in another series B?**

Difficulty Level: L2

Get the positions of items of ser2 in ser1 as a list.

Input

ser1 = pd.Series([10, 9, 6, 5, 3, 1, 12, 8, 13])

ser2 = pd.Series([1, 3, 10, 13])

Show Solution

A computer screen shot of a computer code

AI-generated content may be incorrect.

**17. How to compute the mean squared error on a truth and predicted series?**

Difficulty Level: L2

Compute the mean squared error of truth and pred series.

Input

truth = pd.Series(range(10))

pred = pd.Series(range(10)) + np.random.random(10)

Show Solution

A screenshot of a computer program

AI-generated content may be incorrect.

**18. How to convert the first character of each element in a series to uppercase?**

Difficulty Level: L2

Change the first character of each word to upper case in each word of ser.

ser = pd.Series(['how', 'to', 'kick', 'ass?'])

Show Solution

A screenshot of a computer

AI-generated content may be incorrect.

**19. How to calculate the number of characters in each word in a series?**

**A screenshot of a computer

AI-generated content may be incorrect.**

**22. How to get the day of month, week number, day of year and day of week from a series of date strings?**

**A screenshot of a computer program

AI-generated content may be incorrect.**

**23. How to convert year-month string to dates corresponding to the 4th day of the month?**

**A screenshot of a computer

AI-generated content may be incorrect.**

**24. How to filter words that contain atleast 2 vowels from a series?**

**A computer screen shot of a computer program

AI-generated content may be incorrect.**

**25. How to filter valid emails from a series?**

A screenshot of a computer program

AI-generated content may be incorrect.

**26. How to get the mean of a series grouped by another series?**

A screen shot of a computer code

AI-generated content may be incorrect.

**27. How to compute the euclidean distance between two series?**

A white background with numbers and letters

AI-generated content may be incorrect.

**28. How to find all the local maxima (or peaks) in a numeric series?**

Difficiulty Level: L3

Get the positions of peaks (values surrounded by smaller values on both sides) in ser.

Input

ser = pd.Series([2, 10, 3, 4, 9, 10, 2, 7, 3])

Desired output

array([1, 5, 7])

Show Solution

A white rectangular object with text

AI-generated content may be incorrect.

**29. How to replace missing spaces in a string with the least frequent character?**

Replace the spaces in my\_str with the least frequent character.

Difficiulty Level: L2

Input

my\_str = 'dbc deb abed gade'

Desired Output

'dbccdebcabedcgade' # least frequent is 'c'

Show Solution

A screenshot of a computer program

AI-generated content may be incorrect.

**30. How to create a TimeSeries starting ‘2000-01-01’ and 10 weekends (saturdays) after that having random numbers as values?**

**A screen shot of a computer

AI-generated content may be incorrect.**

**31. How to fill an intermittent time series so all missing dates show up with values of previous non-missing date?**

Difficiulty Level: L2

ser has missing dates and values. Make all missing dates appear and fill up with value from previous date.

Show Solution

A screenshot of a computer program

AI-generated content may be incorrect.

**32. How to compute the autocorrelations of a numeric series?**

Difficiulty Level: L3

Compute autocorrelations for the first 10 lags of ser. Find out which lag has the largest correlation.

Input

ser = pd.Series(np.arange(20) + np.random.normal(1, 10, 20))

Desired output

# values will change due to randomness

[0.29999999999999999, -0.11, -0.17000000000000001, 0.46000000000000002, 0.28000000000000003, -0.040000000000000001, -0.37, 0.41999999999999998, 0.47999999999999998, 0.17999999999999999]

Lag having highest correlation: 9

Show Solution

A screenshot of a computer code

AI-generated content may be incorrect.

**33. How to import only every nth row from a csv file to create a dataframe?**

**A screenshot of a computer program

AI-generated content may be incorrect.**

**35. How to create a dataframe with rows as strides from a given series?**

Difficiulty Level: L3

Input

L = pd.Series(range(15))

Desired Output

array([[ 0, 1, 2, 3],

[ 2, 3, 4, 5],

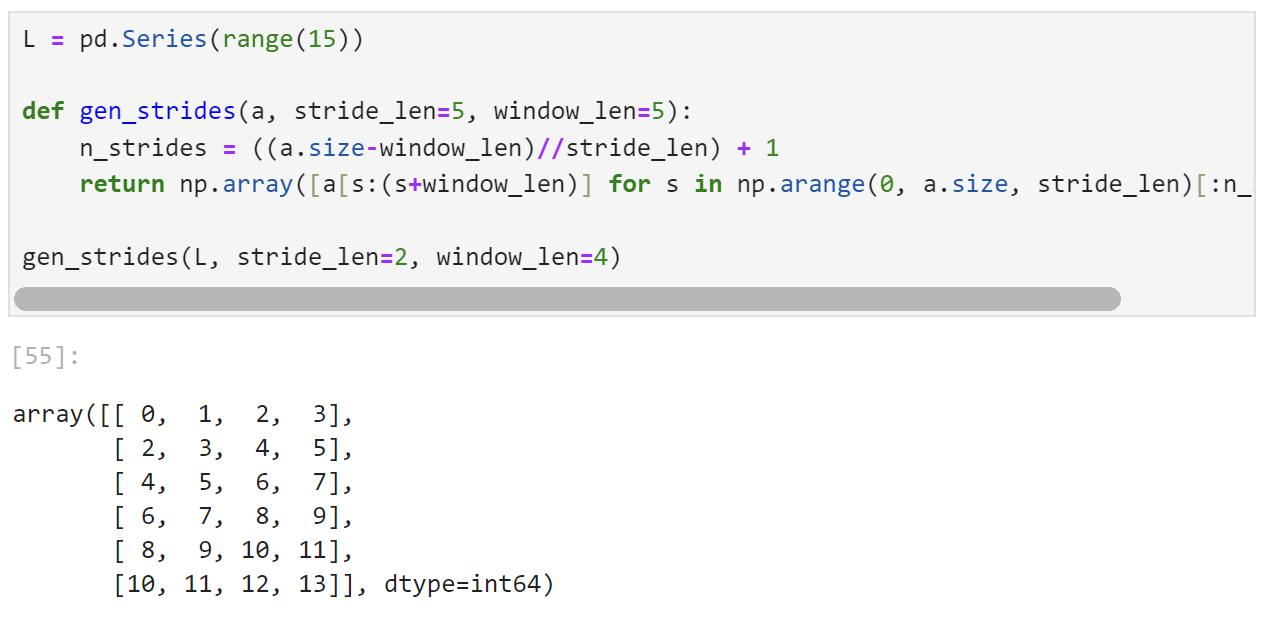
[ 4, 5, 6, 7],

[ 6, 7, 8, 9],

[ 8, 9, 10, 11],

[10, 11, 12, 13]])

Show Solution

**36. How to import only specified columns from a csv file?**

Difficulty Level: L1

Import ‘crim’ and ‘medv’ columns of the [BostonHousing dataset](https://raw.githubusercontent.com/selva86/datasets/master/BostonHousing.csv) as a dataframe.

Show Solution

A close-up of a computer screen

AI-generated content may be incorrect.

**37. How to get the n*rows, n*columns, datatype, summary stats of each column of a dataframe? Also get the array and list equivalent.**

Difficulty Level: L2

Get the number of rows, columns, datatype and summary statistics of each column of the [Cars93](https://raw.githubusercontent.com/selva86/datasets/master/Cars93_miss.csv) dataset. Also get the numpy array and list equivalent of the dataframe.

Show Solution

**39. How to rename a specific columns in a dataframe?**

Difficulty Level: L2

Rename the column Type as CarType in df and replace the ‘.’ in column names with ‘\_’.

Show SolutioA screenshot of a computer program

AI-generated content may be incorrect.

**40. How to check if a dataframe has any missing values?**

Difficulty Level: L1

Check if df has any missing values.

Input

df = pd.read\_csv('https://raw.githubusercontent.com/selva86/datasets/master/Cars93\_miss.csv')

Show Solution

A close-up of a white box

AI-generated content may be incorrect.

**41. How to count the number of missing values in each column?**

Difficulty Level: L2

Count the number of missing values in each column of df. Which column has the maximum number of missing values?

Input

df = pd.read\_csv('https://raw.githubusercontent.com/selva86/datasets/master/Cars93\_miss.csv')

Show Solution

A screen shot of a computer

AI-generated content may be incorrect.

**42. How to replace missing values of multiple numeric columns with the mean?**

Difficulty Level: L2

Replace missing values in Min.Price and Max.Price columns with their respective mean.

Input

df = pd.read\_csv('https://raw.githubusercontent.com/selva86/datasets/master/Cars93\_miss.csv')

Show Solution

A computer screen shot of a computer code

AI-generated content may be incorrect.

**43. How to use apply function on existing columns with global variables as additional arguments?**

**A close-up of a computer screen

AI-generated content may be incorrect.**

**44. How to select a specific column from a dataframe as a dataframe instead of a series?**

Difficulty Level: L2

Get the first column (a) in df as a dataframe (rather than as a Series).

Input

df = pd.DataFrame(np.arange(20).reshape(-1, 5), columns=list('abcde'))

Show Solution

****

**45. How to change the order of columns of a dataframe?**

Difficulty Level: L3

Actually 3 questions.

1. In df, interchange columns 'a' and 'c'.
2. Create a generic function to interchange two columns, without hardcoding column names.
3. Sort the columns in reverse alphabetical order, that is colume 'e' first through column 'a' last.

Show Solution

A computer screen shot of a code

AI-generated content may be incorrect.

**48. How to format all the values in a dataframe as percentages?**

Difficulty Level: L2

Format the values in column 'random' of df as percentages.

Input

df = pd.DataFrame(np.random.random(4), columns=['random'])

df

#> random

#> 0 .689723

#> 1 .957224

#> 2 .159157

#> 3 .21082

Desired Output

#> random

#> 0 68.97%

#> 1 95.72%

#> 2 15.91%

#> 3 2.10%

Show Solution

A screenshot of a computer code

AI-generated content may be incorrect.

**49. How to filter every nth row in a dataframe?**

**A screenshot of a computer program

AI-generated content may be incorrect.**

**50. How to create a primary key index by combining relevant columns?**

Difficulty Level: L2

In df, Replace NaNs with ‘missing’ in columns 'Manufacturer', 'Model' and 'Type' and create a index as a combination of these three columns and check if the index is a primary key.

Input

df = pd.read\_csv('https://raw.githubusercontent.com/selva86/datasets/master/Cars93\_miss.csv', usecols=[0,1,2,3,5])

Desired Output

Manufacturer Model Type Min.Price Max.Price

Acura\_Integra\_Small Acura Integra Small 12.9 18.8

missing\_Legend\_Midsize missing Legend Midsize 29.2 38.7

Audi\_90\_Compact Audi 90 Compact 25.9 32.3

Audi\_100\_Midsize Audi 100 Midsize NaN 44.6

BMW\_535i\_Midsize BMW 535i Midsize NaN NaN

Show Solution

**A screenshot of a computer

AI-generated content may be incorrect.**

**51. How to get the row number of the nth largest value in a column?**

Difficulty Level: L2

Find the row position of the 5th largest value of column 'a' in df.

Input

df = pd.DataFrame(np.random.randint(1, 30, 30).reshape(10,-1), columns=list('abc'))

Show Solution

A computer screen shot of a computer code

AI-generated content may be incorrect.

**52. How to find the position of the nth largest value greater than a given value?**

Difficulty Level: L2

In ser, find the position of the 2nd largest value greater than the mean.

Input

ser = pd.Series(np.random.randint(1, 100, 15))

Show Solution

A computer screen shot of a computer code

AI-generated content may be incorrect.

**53. How to get the last n rows of a dataframe with row sum > 100?**

Difficulty Level: L2

Get the last two rows of df whose row sum is greater than 100.

df = pd.DataFrame(np.random.randint(10, 40, 60).reshape(-1, 4))

Show Solution

A screenshot of a computer code

AI-generated content may be incorrect.

**54. How to find and cap outliers from a series or dataframe column?**

Difficulty Level: L2

Replace all values of ser in the lower 5%ile and greater than 95%ile with respective 5th and 95th %ile value.

Input

ser = pd.Series(np.logspace(-2, 2, 30))

Show Solution

A screenshot of a computer code

AI-generated content may be incorrect.

**55. How to reshape a dataframe to the largest possible square after removing the negative values?**

Difficulty Level: L3

Reshape df to the largest possible square with negative values removed. Drop the smallest values if need be. The order of the positive numbers in the result should remain the same as the original.

Input

df = pd.DataFrame(np.random.randint(-20, 50, 100).reshape(10,-1))

Show Solution

A screenshot of a computer code

AI-generated content may be incorrect.

**56. How to swap two rows of a dataframe?**

Difficulty Level: L2

Swap rows 1 and 2 in df.

Input

df = pd.DataFrame(np.arange(25).reshape(5, -1))

Show Solution

57. How to reverse the rows of a dataframe?

Difficulty Level: L2

Reverse all the rows of dataframe df.

Input

df = pd.DataFrame(np.arange(25).reshape(5, -1))

Show Solution

A screenshot of a computer program

AI-generated content may be incorrect.

**58. How to create one-hot encodings of a categorical variable (dummy variables)?**

Difficulty Level: L2

Get one-hot encodings for column 'a' in the dataframe df and append it as columns.

Input

df = pd.DataFrame(np.arange(25).reshape(5,-1), columns=list('abcde'))

a b c d e

0 0 1 2 3 4

1 5 6 7 8 9

2 10 11 12 13 14

3 15 16 17 18 19

4 20 21 22 23 24

Output

0 5 10 15 20 b c d e

0 1 0 0 0 0 1 2 3 4

1 0 1 0 0 0 6 7 8 9

2 0 0 1 0 0 11 12 13 14

3 0 0 0 1 0 16 17 18 19

4 0 0 0 0 1 21 22 23 24

Show Solution

A screenshot of a computer code

AI-generated content may be incorrect.

**59. Which column contains the highest number of row-wise maximum values?**

**A screenshot of a computer code

AI-generated content may be incorrect.**

**60. How to create a new column that contains the row number of nearest column by euclidean distance?**

Create a new column such that, each row contains the row number of nearest row-record by euclidean distance.

Difficulty Level: L3

Input

df = pd.DataFrame(np.random.randint(1,100, 40).reshape(10, -1), columns=list('pqrs'), index=list('abcdefghij'))

df

# p q r s

# a 57 77 13 62

# b 68 5 92 24

# c 74 40 18 37

# d 80 17 39 60

# e 93 48 85 33

# f 69 55 8 11

# g 39 23 88 53

# h 63 28 25 61

# i 18 4 73 7

# j 79 12 45 34

Desired Output

df

# p q r s nearest\_row dist

# a 57 77 13 62 i 116.0

# b 68 5 92 24 a 114.0

# c 74 40 18 37 i 91.0

# d 80 17 39 60 i 89.0

# e 93 48 85 33 i 92.0

# f 69 55 8 11 g 100.0

# g 39 23 88 53 f 100.0

# h 63 28 25 61 i 88.0

# i 18 4 73 7 a 116.0

# j 79 12 45 34 a 81.0

Show Solution

A screenshot of a computer program

AI-generated content may be incorrect.

**61. How to know the maximum possible correlation value of each column against other columns?**

Difficulty Level: L2

Compute maximum possible absolute correlation value of each column against other columns in df.

Input

df = pd.DataFrame(np.random.randint(1,100, 80).reshape(8, -1), columns=list('pqrstuvwxy'), index=list('abcdefgh'))

Show Solution

A screenshot of a computer program

AI-generated content may be incorrect.

**62. How to create a column containing the minimum by maximum of each row?**

Difficulty Level: L2

Compute the minimum-by-maximum for every row of df.

df = pd.DataFrame(np.random.randint(1,100, 80).reshape(8, -1))

Show Solution

A screenshot of a computer code

AI-generated content may be incorrect.

**63. How to create a column that contains the penultimate value in each row?**

Difficulty Level: L2

Create a new column 'penultimate' which has the second largest value of each row of df.

Input

df = pd.DataFrame(np.random.randint(1,100, 80).reshape(8, -1))

Show Solution

A screenshot of a computer code

AI-generated content may be incorrect.

**64. How to normalize all columns in a dataframe?**

**A screenshot of a computer code

AI-generated content may be incorrect.**

**65. How to compute the correlation of each row with the suceeding row?**

Difficulty Level: L2

Compute the correlation of each row of df with its succeeding row.

Input

df = pd.DataFrame(np.random.randint(1,100, 80).reshape(8, -1))

Show Solution

66. How to replace both the diagonals of dataframe with 0?

Difficulty Level: L2

Replace both values in both diagonals of df with 0.

Input

df = pd.DataFrame(np.random.randint(1,100, 100).reshape(10, -1))

df

# 0 1 2 3 4 5 6 7 8 9

# 0 11 46 26 44 11 62 18 70 68 26

# 1 87 71 52 50 81 43 83 39 3 59

# 2 47 76 93 77 73 2 2 16 14 26

# 3 64 18 74 22 16 37 60 8 66 39

# 4 10 18 39 98 25 8 32 6 3 29

# 5 29 91 27 86 23 84 28 31 97 10

# 6 37 71 70 65 4 72 82 89 12 97

# 7 65 22 97 75 17 10 43 78 12 77

# 8 47 57 96 55 17 83 61 85 26 86

# 9 76 80 28 45 77 12 67 80 7 63

Desired output

# 0 1 2 3 4 5 6 7 8 9

# 0 0 46 26 44 11 62 18 70 68 0

# 1 87 0 52 50 81 43 83 39 0 59

# 2 47 76 0 77 73 2 2 0 14 26

# 3 64 18 74 0 16 37 0 8 66 39

# 4 10 18 39 98 0 0 32 6 3 29

# 5 29 91 27 86 0 0 28 31 97 10

# 6 37 71 70 0 4 72 0 89 12 97

# 7 65 22 0 75 17 10 43 0 12 77

# 8 47 0 96 55 17 83 61 85 0 86

# 9 0 80 28 45 77 12 67 80 7 0

Show Solution

A screenshot of a computer code

AI-generated content may be incorrect.

**68. How to get the n’th largest value of a column when grouped by another column?**

Difficulty Level: L2

In df, find the second largest value of 'taste' for 'banana'

Input

df = pd.DataFrame({'fruit': ['apple', 'banana', 'orange'] \* 3,

'rating': np.random.rand(9),

'price': np.random.randint(0, 15, 9)})

Show SolutionA screen shot of a computer code

AI-generated content may be incorrect.

**69. How to compute grouped mean on pandas dataframe and keep the grouped column as another column (not index)?**

Difficulty Level: L1

In df, Compute the mean price of every fruit, while keeping the fruit as another column instead of an index.

Input

df = pd.DataFrame({'fruit': ['apple', 'banana', 'orange'] \* 3,

'rating': np.random.rand(9),

'price': np.random.randint(0, 15, 9)})

Show Solution

A screenshot of a computer code

AI-generated content may be incorrect.

**70. How to join two dataframes by 2 columns so they have only the common rows?**

**A computer screen shot of a computer code

AI-generated content may be incorrect.**