Assignment 8

**Problem Statement 1:**

1) How-to-count-distance-to-the-previous-zero

For each value, count the difference of the distance from the previous zero (or the start

of the Series, whichever is closer) and if there are no previous zeros, print the position

Consider a DataFrame df where there is an integer column {'X':[7, 2, 0, 3, 4, 2, 5, 0, 3, 4]}

The values should therefore be [1, 2, 0, 1, 2, 3, 4, 0, 1, 2]. Make this a new column 'Y'.

import pandas as pd

df = pd.DataFrame({'X': [7, 2, 0, 3, 4, 2, 5, 0, 3, 4]})

**Code:**

df = pd.DataFrame({'X': [7, 2, 0, 3, 4, 2, 5, 0, 3, 4]})

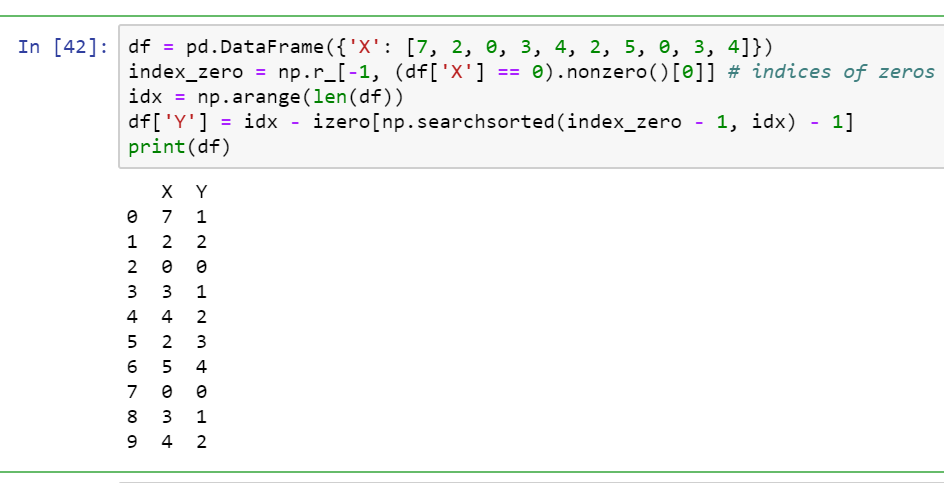
index\_zero = np.r\_[-1, (df['X'] == 0).nonzero()[0]] # indices of zeros

idx = np.arange(len(df))

df['Y'] = idx - izero[np.searchsorted(index\_zero - 1, idx) - 1]

print(df)

**Screenshot**



2) Create a DatetimeIndex that contains each business day of 2015 and use it to index a

Series of random numbers.

**Code:**

import pandas as pd

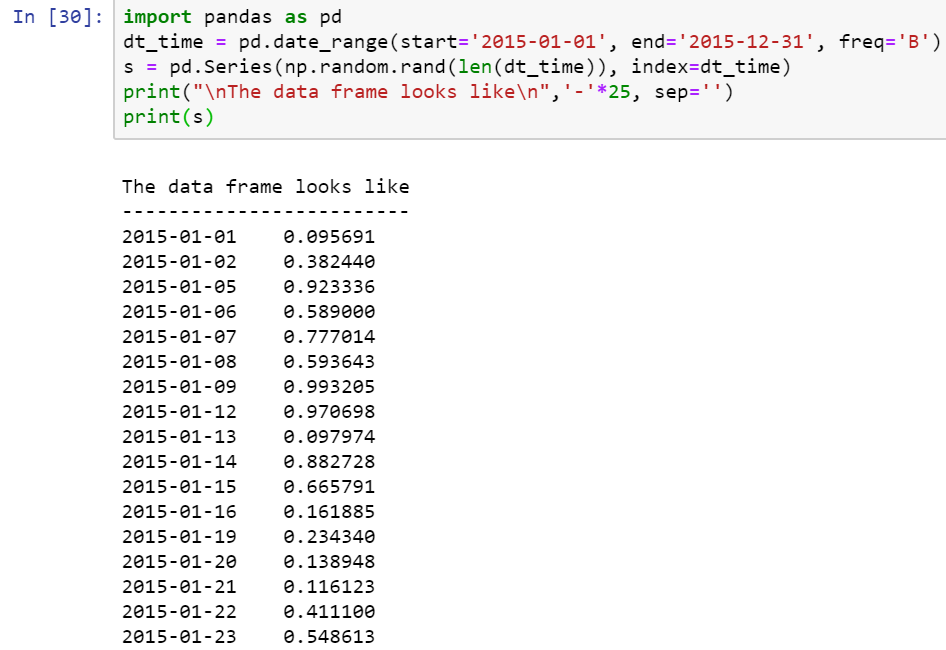
dt\_time = pd.date\_range(start='2015-01-01', end='2015-12-31', freq='B')

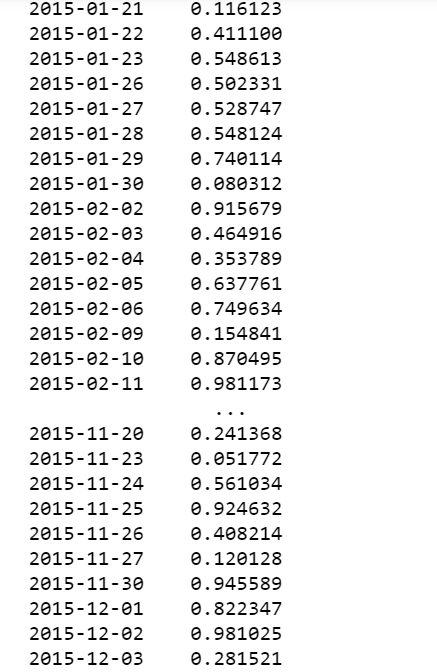
s = pd.Series(np.random.rand(len(dt\_time)), index=dt\_time)

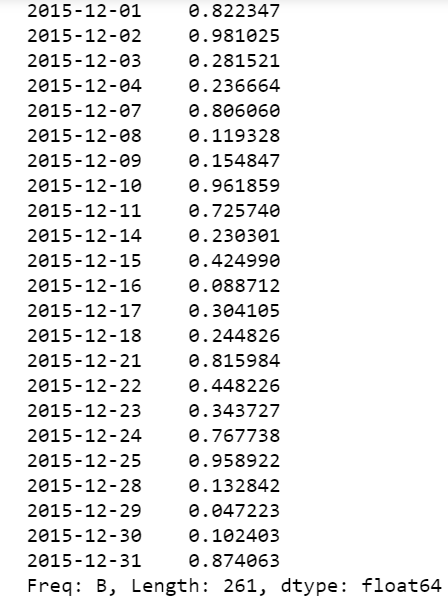
print("\nThe data frame looks like\n",'-'\*25, sep='')

print(s)

**Screenshot**







3) Find the sum of the values in s for every Wednesday

**Code:**

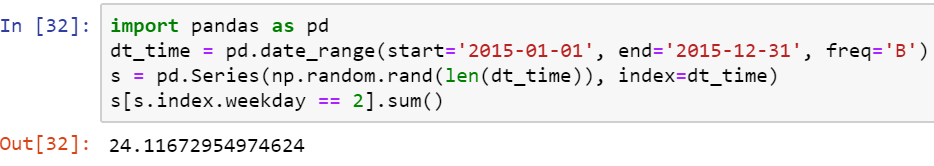
import pandas as pd

dt\_time = pd.date\_range(start='2015-01-01', end='2015-12-31', freq='B')

s = pd.Series(np.random.rand(len(dt\_time)), index=dt\_time)

s[s.index.weekday == 2].sum()

**Screenshot**



4) Average For each calendar month

**Code:**

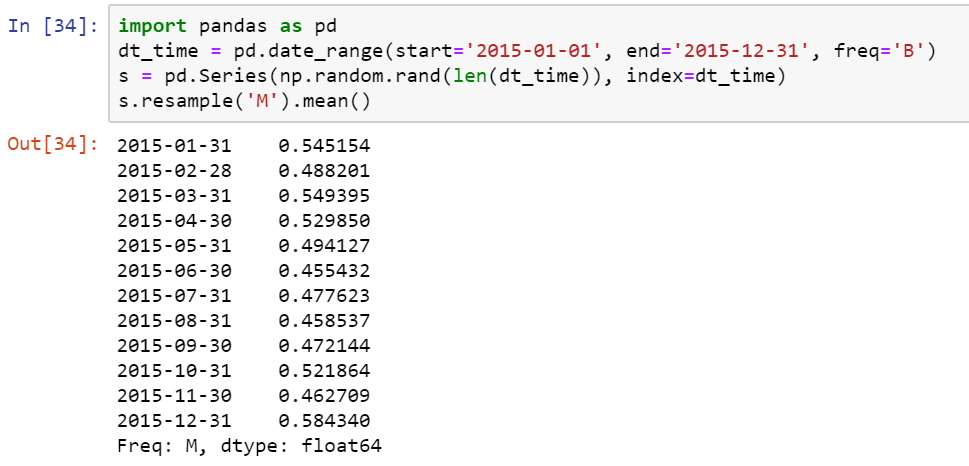
import pandas as pd

dt\_time = pd.date\_range(start='2015-01-01', end='2015-12-31', freq='B')

s = pd.Series(np.random.rand(len(dt\_time)), index=dt\_time)

s.resample('M').mean()

**Screenshot**



5) For each group of four consecutive calendar months in s, find the date on which the

highest value occurred.

**Code:**

import pandas as pd

dt\_time = pd.date\_range(start='2015-01-01', end='2015-12-31', freq='B')

s = pd.Series(np.random.rand(len(dt\_time)), index=dt\_time)

s.groupby(pd.TimeGrouper('4M')).idxmax()

**Screenshot**

