



Pandas

Assignment 1 Question

Pandas: Assignment 1

Table of Contents

1.Introduction

2.Problem Statement

3.Output

1.Introduction

This assignment will help you to consolidate the concepts learnt in the session.

2.Problem Statement

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]})
```

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

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

4) Average For each calendar month

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

Problem Statement 2:

Read the dataset from the below link

https://raw.githubusercontent.com/guipsamora/pandas_exercises/master/06_Stats/US_Baby_Names/US_Baby_Names_right.csv

Questions:

1) Delete unnamed columns

2) Show the distribution of male and female

- 3) Show the top 5 most preferred names
- 4) What is the median name occurrence in the dataset
- 5) Distribution of male and female born count by states

3. Expected Output

This assignment consists of 200 marks and needs to be submitted in the upcoming session for evaluation.