Data retrieval and preprocessing of Python quiz

TOTAL POINTS 8

1. The read_csv () function in the pandas module is commonly used. In addition to reading a csv file and converting the result into a DataFrame, it can also read other formatted text files. Assume that each line of a text file contains the same number of values, and the data is separated by a #, like this: 1 point

```
12#34#5.67#1234
12#346#5.67#77
...
12#3.4#67#67.89
```

In the read_csv () function, which one of the following options needs to be added?

- ```
  names='#'
  ```

- ```
  sep='#'
  ```

- ```
  index_col='#'
  ```
  ```

```
engine='#'
```

2. Complete the following program, fill in the missing code in the program, and use a # to connect the two answers. The program functions are: read the score data in the file score.csv, calculate the average score and count the scores of each course for the students whose Chinese score is 80 or more and English score is 85 or more (sort from big to small), Output the result to the file result.csv and plot a bar chart of the average results of the students who meet the conditions as shown in the figure. 1 point

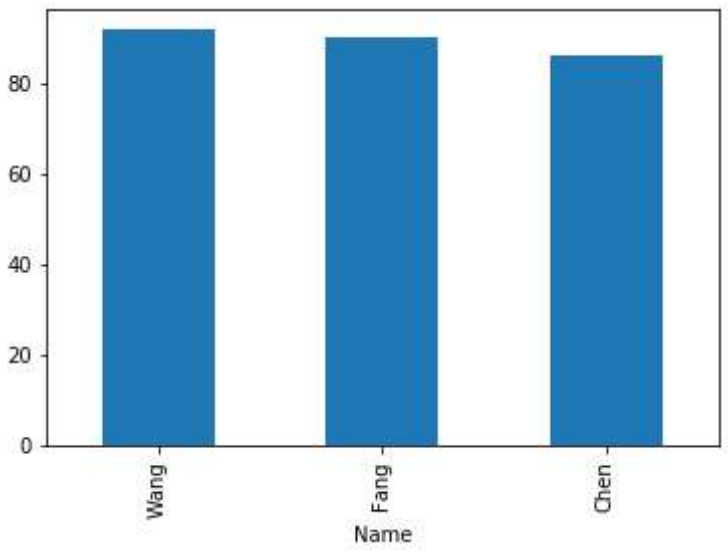
[Test data and results]

Content of score.csv:

```
Name,Chinese,Maths,English
Chen,88,87,85
Fang,93,88,90
Wang,82,99,96
Peng,77,94,84
Ding,80,94,76
```

Content of result.csv and the bar chart:

```
        Chinese  Maths  English  Avg
Name
Wang         82     99       96   92
Fang         93     88       90   90
Chen         88     87       85   86
```



[Code to be added]

```
import pandas as pd

df = pd.read_csv('score.csv', index_col = 'Name')
df['Avg'] = (df.Maths+df.Chinese+df.English) // 3
# & represents condition and, that is, both conditions must be met
data = df[(df.Chinese >= 80) & ([1])]
data = data.sort_values(by = 'Avg', ascending = 0)
data.to_csv('result.csv')
data_parts = data.iloc[:, -1]
data_parts.[2](kind = 'bar')
```

Enter answer here

3. For a Series or DataFrame object, which of the following options is incorrect? 1 point

- The isnull() method can be used to determine missing values.
- The drop() method can be used directly to delete missing rows.
- The fillna() method can be used to fill missing rows.
- The fillna() method can be specified by the parameter of method to fill in missing values with the first non-missing value above or below.

4. If you want to quickly observe the brief statistics of a DataFrame object in the form shown below, please write the corresponding method name on the horizontal line. 1 point

```
>>> iris_df_length._____()
       sepal length (cm)  petal length (cm)
count         150.000000        150.000000
mean            5.843333          3.758000
std             0.828066          1.765298
min             4.300000          1.000000
25%             5.100000          1.600000
50%             5.800000          4.350000
75%             6.400000          5.100000
max             7.900000          6.900000
```

Enter answer here

5. Normalization is an important method in data transformation. Please select the following options which are commonly used as the normalization method. 1 point

- [x] Min-Max
- [ ] Binning
- [x] Z-Score
- [x] Decimal Scaling

6. Data preprocessing is an important stage of data analysis and mining. sklearn is a well-known machine learning package. Please write the name of the module commonly used in sklearn for data preprocessing. 1 point

preprocessing

7. The following normalization results may belong to the classic Min-Max normalization. Is this statement correct? 1 point

```
array([[-0.14421743,  0.41367189, -0.12001342],
       [-0.74026221,  0.19427445,  0.36716642],
       [-0.74026221,  1.28271368, -0.26581176],
       ...,
       [ 0.15812412,  0.98496002,  0.79744934],
       [ 0.15812412,  0.72567214,  0.73699637],
       [ 0.15812412, -0.36276709,  0.43473151]])
```

- ( ) True
- (•) False

8. Please select the wrong option in the following two forms of the data reduction. 1 point

- ○ Feature reduction is a reduction of the attributes of a dataset, the purpose of which is to obtain a reduction representation of a representative number of data columns.
- ○ PCA is an important feature reduction method.
- ◉ Boxplots are often used to implement value reductions.
- ○ Sampling is a common method of value reduction. Common sampling methods include random sampling, cluster sampling and stratified sampling.

☑ I, **Soujatya Bhattacharya**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

Learn more about Coursera's Honor Code