Friday, 11 September 2020

1 Histogram of Sum of Student Score (10 mins)

1.1 Objectives

1. Can use pandas to read csv file
2. Can modify the column name of a dataframe
3. Can add a new column to a dataframe
4. Can plot a simple histogram using pandas

1.2 Description

Same as in the lecture, this time you need to read the `score.csv` file and calculate all related data. Finally, please show the histogram of total score of the students. The format of the file is same as other experiments as shown below.

```
5995778521,4,0,4,6,7
5974831121,6,5,0,6,4
5992211621,9,2,0,0,4
5987519321,3,5,1,4,3
5979431521,5,7,1,8,2
```

1.3 Procedure

1. Read the `score.csv` file and store it into a dataframe
2. Change the name of columns
3. Add score from all quizzzes into a new column, `total`.
4. Plot the histogram of the total score

1.4 Sample Output

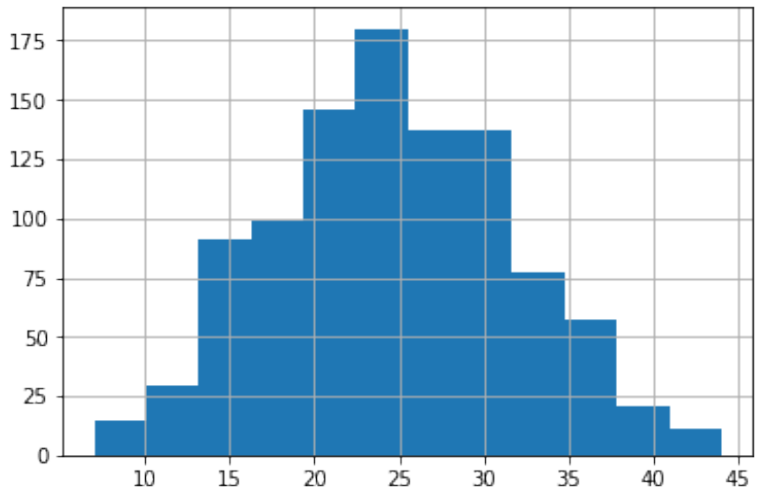

Figure 1: *Output histogram*

2 Titanic(20 mins)

2.1 Objectives

1. Can use pandas to read csv file
2. Can select the wanted row according to some conditions
3. Can plot the bar chart using pandas

2.2 Description

Plot the number of survived and not survived passengers separated by the gender of passengers. The desired output is shown below.

2.3 Procedure

1. Read the `train.csv` file and store it into a dataframe
2. Use the pivot_table that align the value of `Survived` to the x axis (use it as index) and use the value of `Sex` as the

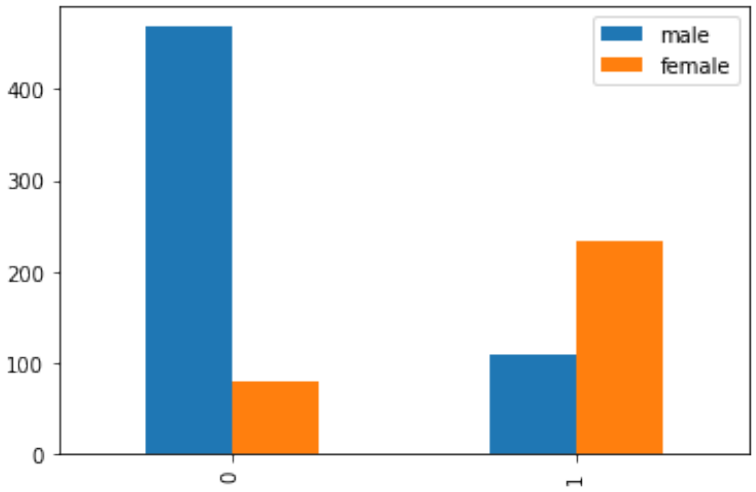

Figure 2: *Output chart*

categories (use it as the column). Please do not forget to use 'count' as the `aggfunc`.

3. Plot the dataframe

3 Other Idea (30 mins)

3.1 Description

This experiment is your turn. You have to create an interesting question and try to answer that question using visualization tools from pandas. You can choose the `pm25.csv`, `titanic`, or any other data from online resources.