Pandas Assignments (1ed) ___Hongyu

July 20, 2019

Imagine that you are working in a market department for a grocery company. You are required to do some analyses on the custom data that you have (customer_data.csv).

0.1 *1. Simple import and saving*

0.1.1 First, you need to import data from the csv file into the data frame in Pandans and print out its first 3 records in the last 10 records for testing. Then, save this data frame into a excel file without index.

0.2 *2. Import non-standard CSV*

0.2.1 However, sometimes the data we have may not always standard CSV files that data are separated by commas. For example, the file "customer_data_tap" has exactly the same data as the CSV file but its data are separated by taps. Please import this data and do the same operation as the first question (you should get the same result as the first question).

0.3 *3. Sorting Data*

0.3.1 The data we now have is sorted by the CustomerID. However, this may not eligible for analysis.

0.3.2 Please print out all the attributes as a list in the data and sort the data by CustomerAge as the highest priority in ascending order and SpendingScore as the second priority in descending order.

0.4 *4. Changing Data*

0.4.1 Please add a new column, TotalScore, which represents the total score of the CreditScore and the Spending score for the first 10 records. Print out the first 20 records.

0.5 *5. Clean Data*

0.5.1 Now you may notice that there are many NaN value in the TotalScore column which are not clear. Clean the data by replacing the NaN with "Ignore".

0.6 *6. Filtering Data && Conditional Changes*

0.6.1 In order to do some evaluation on the Spending score, add a column called SpendingEvaluation in which the value is "High" if the corresponding SpendingScore is greater and equal to 60, and if not, the value is "Low". Then, change the CustomerID to " * " if its SpendingEvaluation is "High".

0.7 *7. Aggregate Statistics*

0.7.1 Group all data by the CustomerGender as the highest priority and the CustomerCity as the second priority. Then, print out the name and the corresponding number of customers for each group.

0.8 *8. Plotting data*

0.8.1 Using matplotlib to plot two figures described below and save the first one as a png file and save the second one as a pdf file.

0.8.2 (1) Make a stacked bar chart for the data frame that you created in question 7;

0.8.3 (2) Make a line chart for two columns: CreditScore, SpendingScore (y-axis) against the CustomerID (x-axis).