**George Brown College**

**Centre for Arts, Design and Information Technology**

**Machine Learning Process – Working with pandas (Assignment 04 – Lab Exercise)**

Your Name:

Your Student ID:

The goal of this lab-exercise is to practice with pandas functions.

Please follow the following steps using the Jupyter Notebook or any other environment to implement steps. At the end you need to present the codes to the instructor and upload your ipynb file to designated dropbox.

Please note that this is an individual lab-exercise, and everyone needs to finish the lab-exercise and submit their tasks individually.

The individual submission is mandatory, while everyone will open and run their assignments in the class to demonstrate what they have learned from it.

**Important:** Please add #question number for every question, e.g. #01 for your answer to question 1.

**Let’s start the lab-exercise:**

1. Download the csv file from the following URL and store it in a DataFrame named, myDF.

<http://tiny.cc/lab-exercise-02>

1. Show the **first three rows** of the DF.
2. Delete the ‘Model’ column from the DF temporarily.
3. Delete the ‘Model’ column permanently, using the assignment (Not inplace parameter). Is there any output when you run the permanent deletion? So, how can you make sure it is deleted?
4. Now, let’s delete rows 1, 3, and 5 permanently, using inplace. Then show the first 4 Rows of myDF.
5. Calculate the mean of ‘Range’ column.
6. How many ‘F’ do we have in the ‘Type’ column? Replace them permanently with ‘AAA’ and show the first 5 rows.
7. Get rid of the [, $ and ] from the ‘Cost’ column permanently and display the first 5 rows. What is the type of ‘Cost’ column after this?
8. Change the ‘Cost’ column data type to integer and display the result.
9. Calculate the average of ‘Cost’ column. Then, calculate the average of ‘Cost’ for each ‘Continent’.
10. Calculate the min, max and standard deviation of the ‘Cost’ column, group by each ‘Continent’, all in one command.
11. Calculate the min, max and standard deviation of all numerical columns, group by each ‘Continent’, all in one command.
12. Illustrate the mean of all numerical columns, group by the ‘Continent’, using a ‘bar’ chart.
13. Count the number of each continent in the ‘Continent’ column.
14. Show the percentage of each continent in the ‘Continent’ column.
15. Show the unique values that we have in the ‘Continent’ column. Then show the number of unique values that we have in the ‘Continent’ column.
16. Use the value\_count method for the ‘Cost’ column and describe the output as a comment in the next cell, using the # for the comment.
17. Plot a histogram for the ‘Cost’ column.
18. Bonus question: Try to change the number of bins to 50 to show the result more accurately.
19. How many missing values do we have in the ‘Zip Code’ column? How many missing values in the ‘Type’ column? Provide a brief summary of null values for the entire Data Frame.
20. If we drop all the rows that have ‘any’ missing value, how many rows will remain in myDF?
21. Show the statistics for the ‘Type’ column using the value\_count method, including the NaN values in the list. How many missing values do we have?
22. Replace the ‘Type’ column missing values with ‘XXX’ and show the result again.
23. Drop the rows that have any missing value permanently and show the number of rows after running the drop command.
24. Set the ‘ID’ column as the new index and delete the index name. Then show the first 5 rows of the DF. Show the value of the cell that its ID is 70 and its column header is ‘Value’.
25. Show the index, columns and the shape of the current DF. Bring back the ID column to the DF and reset the index. Again, show the index, columns and the shape of the current DF.
26. Use the describe method to see the statistics of myDF and fetch the value of the first row and the third column of the result.
27. Sort the value\_counts() of the ‘Continent’ column of the DF. Then, sort the value\_counts() of the ‘Continent’ column, based on the alphabetic order of the continents.
28. Show all the columns of the following rows in one command. [1,3,5,7] and show the first 7 rows of the following columns: ‘Type’ and ‘Continent’
29. Show the ‘Value’ of the ‘Type’ that are equal to ’XXX’.

Thank you and good luck,

Reza

What you seek is seeking you. ~Rumi