COMP1007 Assignment 4

Submission Deadline: 23:59 PM, 2 April 2019

This assignment helps you review the fundamental operations of Pandas DataFrame. Please download the data file **automobile_data.csv** from Moodle, and develop a Python program that performs the following tasks:

1. Create a Pandas DataFrame by loading the file **automobile_data.csv**.

2. Show the shape of the DataFrame.

3. Show the first 8 rows and the last 8 rows of your DataFrame.

4. Show the data types of every column of your DataFrame.

5. There are some rows with missing data. Please drop these rows from the DataFrame.

6. Show the shape of the DataFrame again.

7. Create a new DataFrame that consists of all the rows of Toyota cars. You can first use **pandas.groupby()** function to generate a GroupBy object, and then use **GroupBy.get_group()** function to generate the DataFrame object. You can learn from the **Grouping** example at https://pandas.pydata.org/pandas-docs/stable/user_guide/cookbook.html ([114], [115], [117], [118]).

8. Find the average mileage of each car company. You can make use the GroupBy object created in Step 7.

9. Sort all cars by the **price** column in descending order, and save the sorted table in a new CSV file automobile_data_new.csv. You can use the **pandas.to_csv()** function.

10. Show the statistics information (such as mean, std, min, max, and 25/50/75% percentiles) of the final DataFrame.