

How to run the code?

The code should be executed from the command line in the following manner (Assuming the PATH is pointing to spark and Java):

```
spark-submit --class <className> <JarFileName.jar> <input_file> <output_file>
```

So to get the result for task one, the following command must be executed:

```
spark-submit --class Task1 Bashar_Alhafni.jar /Users/alhafni/Desktop/stack-overflow-2018-developer-survey/survey_results_public.csv  
/Users/alhafni/Desktop/inf553_data_mining/Bashar_Alhafni/OutputFiles/Bashar_Alhafni_task1.csv
```

for task two:

```
spark-submit --class Task2 Bashar_Alhafni.jar /Users/alhafni/Desktop/stack-overflow-2018-developer-survey/survey_results_public.csv  
/Users/alhafni/Desktop/inf553_data_mining/Bashar_Alhafni/OutputFiles/Bashar_Alhafni_task2.csv
```

for task three:

```
spark-submit --class Task3 Bashar_Alhafni.jar /Users/alhafni/Desktop/stack-overflow-2018-developer-survey/survey_results_public.csv  
/Users/alhafni/Desktop/inf553_data_mining/Bashar_Alhafni/OutputFiles/Bashar_Alhafni_task3.csv
```

Note:

If one of the **<input_file>** or **<output_file>** directories is missing, the program won't execute, and the following message will appear:

the input and/or output argument(s) are missing ... aborting