DESCRIPTION OF COURSEWORK

Course Code	SWE404
Course Name	Big Data Analytics
Lecturer	Lu Yang
Academic Session	2020/04
Assessment Title	Assignment 5

A. Introduction/ Situation/ Background Information

This assignment evaluates the students' understanding of Lecture 13, which are mainly about data visualization.

B. Course Learning Outcomes (CLO) covered

At the end of this assessment, students are able to:

CLO 1 Describe the concept and understand broad knowledge of Big Data.

CLO 2 Demonstrate appropriate knowledge of Hadoop, Spark, Storm and Map Reduce framework and apply them to build a VM-based environment.

CLO 3 Integrate knowledge and understanding of the basic principles, techniques and methodologies of organizing and searching Big Data, and apply them to create value with business insight.

CLO 4 Identify the awareness of the wide applicability of Big Data for real-world practical purposes.

CLO 5 Demonstrate the need to continually follow Big Data development trends.

C. University Policy on Academic Misconduct

1. Academic misconduct is a serious offense in Xiamen University Malaysia. It can be defined as any of the following:
   i. **Plagiarism** is submitting or presenting someone else's work, words, ideas, data or information as your own intentionally or unintentionally. This includes incorporating

published and unpublished material, whether in manuscript, printed or electronic form into your work without acknowledging the source (the person and the work).

ii. **Collusion** is two or more people collaborating on a piece of work (in part or whole) which is intended to be wholly individual and passed it off as own individual work.

iii. **Cheating** is an act of dishonesty or fraud in order to gain an unfair advantage in an assessment. This includes using or attempting to use, or assisting another to use materials that are prohibited or inappropriate, commissioning work from a third party, falsifying data, or breaching any examination rules.

2. All the assessment submitted must be the outcome of the student. Any form of academic misconduct is a serious offense which will be penalised by being given a zero mark for the entire assessment in question or part of the assessment in question. If there is more than one guilty party as in the case of collusion, both you and your collusion partner(s) will be subjected to the same penalty.

D. Instruction to Students

This assignment is an **individual** assignment. Each student should submit Jupyter Notebook file on Moodle, named as "Assignment5_YourStudentID.ipynb".

The deadline is **18:00, 13th July**. Overdue penalty will be given to the assignment that is submitted after the deadline.

* Your codes will be sent to a **Plagiarism** detection system for duplication checking. Please write your codes independently. (Modify your code if you copy some fragment from the Internet because your classmates may copy the same fragment.)

E. Evaluation Breakdown

No.	Component Title	Percentage (%)
1.	Assignment 5	100
	TOTAL	**100**

F. Task(s)

The attached csv file shows the percentage of persons between 25 and 29 years old with selected levels of educational attainment, by race/ethnicity and sex. You are required to do the following data visualization tasks with **pandas** and **seaborn**:

1. Plot the total percentages of all people of at least bachelor's degree with a line chart over years. Use different colors to show the attribute “min degree”. (the degree level of associate’s is lower than bachelor’s) (30 marks)
2. Plot the total percentages of women, men, and total people with a minimum education of high school degrees with a bar chart in the year 2009. Do you think the bar chart is an effective data visualization for this question? Explain in 1-2 sentences why or why not. (30 marks)
3. Plot the results of how the percent of Hispanic individuals with degrees has changed between 1990 and 2010 (inclusive) for high school and bachelor's degrees with a chart of your choice. Explain in a few sentences why you chose this type of plot over others. (40 marks)