University of Macau

Faculty of Science & Technology

Department of Computer & Information Science

**CISC7201 INTRODUCTION TO DATA SCIENCE PROGRAMMING**

**Project for 1st semester 2018/2019**

**[Objective]**

In this project, we need you to design some data analytical process on a dataset (>100 MB or above).

**[Basic requirements]**

1. Write your codes to collect the data (e.g., from web) and / or
2. Write your codes to clean the data (e.g., removing noisy records) and / or
3. Write your codes to process the data (e.g., transform data in a easy-to-access form) and / or
4. Write your codes to analyze / recommend / forecast / visualize your data

We need you provide at least **TRHEE “contributions”** in this project. For instance, if your dataset is fully cleaned, then you may focus on requirement (3) (e.g., reduce the dimensionality) and (4) (e.g., visualization + forecasting).

Every team is required to try at least one new Python library (not the basic libraries being introduced in the course) that is helpful in your project. You need to present the reason why you choose to use these libraries.

**[Marking]**

1. If your project can meet the basic requirements, your project grade is at least B ~ B+.
2. You are always welcome to tell us your new findings (e.g., observation, methodologies, skills) from the data. More you proposed, more marks you will get (but in a sigmoid like scale).
3. Certain marks are given based on
   1. Understanding your data
   2. Reasonable objective of your analytical process
   3. Novelty
   4. Data difficulty (volume, type, velocity, etc.)

**[References]**

<https://www.edureka.co/blog/data-science-projects/>

**[Submissios]**

4 page report + presentation slides (5 minutes)