



Practical Data Science with Python
COSC 2670/2738
Assignment 1 (Part 2)

	Assessment Type	Individual
	Due Date	23:59, the 3rd of May, 2020
	Marks	20

Please read all the following information before attempting your assignment. This is an *individual* assignment. You may not collude with any other people, or plagiarise their work. Each student is expected to present the results of his/her own thinking and writing. Never copy other student's work (even if they "explain it to you first") and never give your written work to others. Keep any conversation high-level and never show your solution to others. Never copy from the Web or any other resource. Remember you are meant to generate the solution to the questions by yourself. Suspected collusion or plagiarism will be dealt with according to RMIT policy.

In the submission (your PDF file) you will be required to certify that the submitted solution *represents your own work only* by agreeing to the following statement:

I certify that this is all my own original work. If I took any parts from elsewhere, then they were non-essential parts of the assignment, and they are clearly attributed in my submission. I will show we I agree to this honor code by typing "Yes":

A sample format for this requirement is provided, and please find it in **Canvas** – > **Assignments** – > **Assignment1Part2**.

Tasks

This is the part 2 of Assignment 1, and it includes two tasks. This is independent to your assignment 1, so your current assignment 1 will not affect this part 2.

Task 1: An oral presentation of the work in Assignment 1 (10%)

The presentation should briefly describe

- How to prepare the data?
- How to explore the data?
- What are the results from your analysis?

The presentation should be a maximum of 10 minutes. Your presentation slides should be:

- Microsoft PowerPoint slides (with audio inserted for each slide by using: **Insert** – **> Audio** – **> Record Audio**).
- or you can create your own presentation slides (e.g. PDF version) and please submit your own record of your presentation as well.

Task 2: Short answer question (10%)

The questions in the survey can be divided into two parts:

- one is about people's attitude or opinion about Start War movies, including:
 - Have you seen any of the 6 films in the Star Wars franchise?
 - Do you consider yourself to be a fan of the Star Wars film franchise?
 - Which of the following Star Wars films have you seen? Please select all that apply. (Star Wars: Episode I The Phantom Menace; Star Wars: Episode II Attack of the Clones; Star Wars: Episode III Revenge of the Sith; Star Wars: Episode IV A New Hope; Star Wars: Episode V The Empire Strikes Back; Star Wars: Episode VI Return of the Jedi)
 - Please rank the Star Wars films in order of preference with 1 being your favorite film in the franchise and 6 being your least favorite film. (Star Wars: Episode I The Phantom Menace; Star Wars: Episode II Attack of the Clones; Star Wars: Episode III Revenge of the Sith; Star Wars: Episode IV A New Hope; Star Wars: Episode V The Empire Strikes Back; Star Wars: Episode VI Return of the Jedi)
 - Please state whether you view the following characters favorably, unfavorably, or are unfamiliar with him/her. (Han Solo, Luke Skywalker, Princess Leia Organa, Anakin Skywalker, Obi Wan Kenobi, Emperor Palpatine, Darth Vader, Lando Calrissian, Boba Fett, C-3P0, R2-D2, Jar Jar Binks, Padme Amidala, Yoda)
 - Which character shot first?
 - Are you familiar with the Expanded Universe?
 - Do you consider yourself to be a fan of the Expanded Universe?
 - Do you consider yourself to be a fan of the Star Trek franchise?
- the other is about people's demographics, including
 - Gender
 - Age
 - Household Income
 - Education
 - Location (Census Region)

We would like to build a classifier (or some classifiers, for example one classifier per demographic feature), which can classify people's demographics (gender, age, household income, education, location (census region)) based on their attitude or opinion about

Start War movies. Please describe how to build this classifier (or these classifiers) by using the data collected in the survey (the data provided in Assignment 1).

Please note that this is a short-answer question, and no coding work is required. Your submission must be in PDF document, and must be **at most 6 (in single column format) pages (including figures and references) with a font size between 10 and 12 points**. Penalties will apply if the report does not satisfy the requirement.

What to Submit, When, and How

The assignment is due at

23:59, the 3rd of May, 2020 .

Assignments submitted after this time will be subject to standard late submission penalties.

You need to submit the following files:

- Your presentation slides and the oral audio presentation as required in Task 1.
- Your `Assignment1_Part2.pdf` file includes your answers to Task 2.

They must be submitted as ONE single zip file, named as your student number (for example, 1234567.zip if your student ID is s1234567). The zip file must be submitted in Canvas:

Assignments/Assignment 1 (Part 2).

Please do NOT submit other unnecessary files.