**Instructions & Brief**

The aim of this assignment is to investigate and visualise data using various data science tools.

This is an individual assignment and worth 15% of your total mark for this unit.

**Task 1: Unemployment Rate and House Prices**

In the task, you are required to visualise the relationship between the unemployment rate, the house price index, and the population of different Australian states, and gain insights from how these relations and trends change over time using motion chart in python.

**The data files**

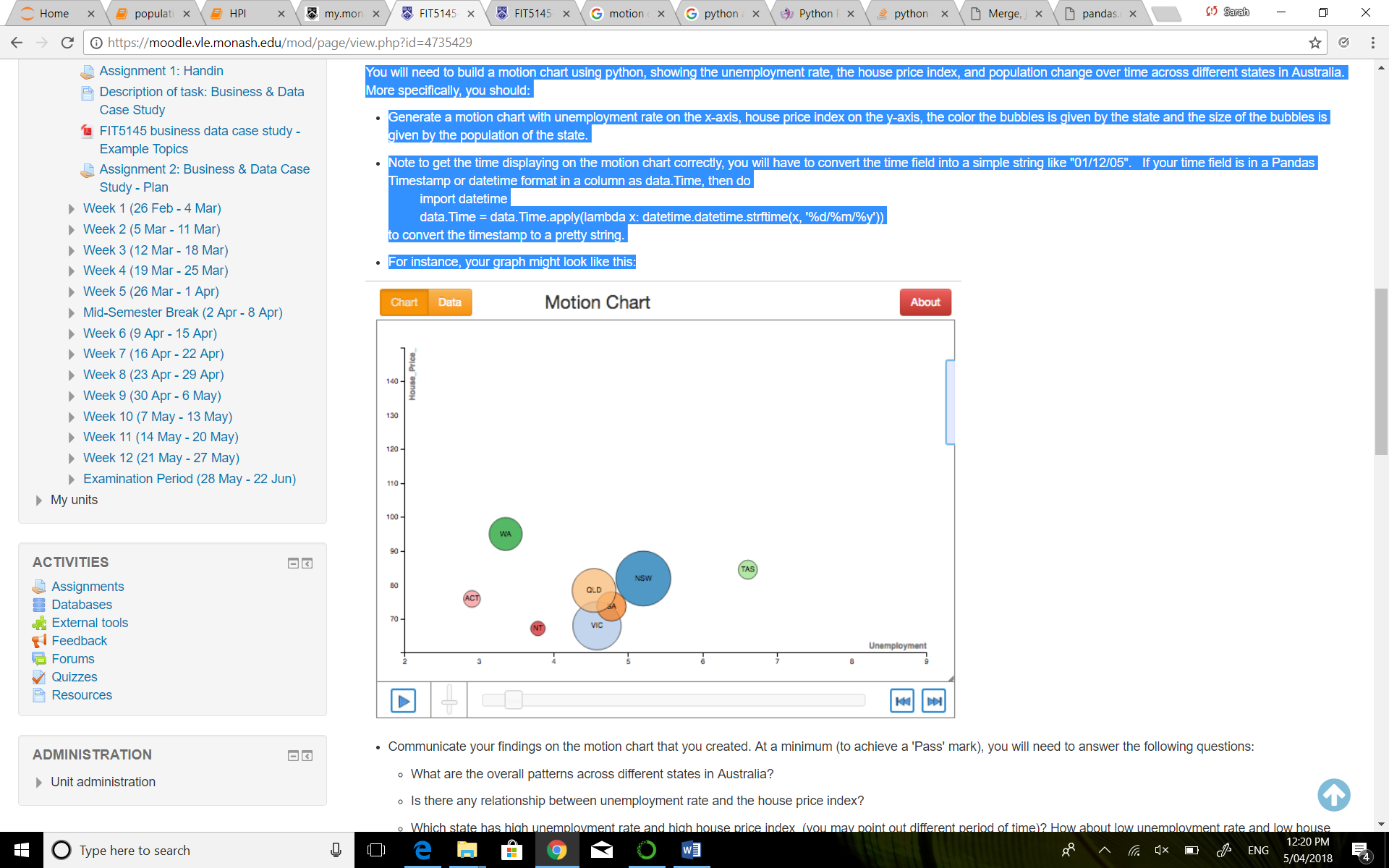
The data files we used in this part are originally downloaded from [the Department of Employment](https://employment.gov.au/) and [Australian Bureau of Statistics](http://www.abs.gov.au/). We have extracted the data from the original downloaded files and put into a simpler format. You can download the data from the Assessments Moodle page.

Alternatively, you can search for and download the original or related data from the above mentioned agencies, however, be warned that you will then have to do more preparation of the data for your motion chart.

* **ERP\_by state and gender.csv**: This data file contains quarterly data about the estimated resident population, grouping by state and gender, between 1/12/2005 and 1/6/2015.
* **SA4 Time Series - October 2016.xls**: This data file contains monthly data of employment and unemployment rate across different Australian states, for the period between 01/02/1978 and 01/10/2016.
* **House Price Index:** This data contains quarterly data of the Residential Property Price Index of the capital cities in different states, between 1/12/2005 and 1/6/2015.

You will need to build a motion chart using python, showing the unemployment rate, the house price index, and population change over time across different states in Australia. More specifically, you should:

* Generate a motion chart with unemployment rate on the x-axis, house price index on the y-axis, the color the bubbles is given by the state and the size of the bubbles is given by the population of the state.
* Note to get the time displaying on the motion chart correctly, you will have to convert the time field into a simple string like "01/12/05".   If your time field is in a Pandas Timestamp or datetime format in a column as data.Time, then do  
           import datetime  
           data.Time = data.Time.apply(lambda x: datetime.datetime.strftime(x, '%d/%m/%y'))  
  to convert the timestamp to a pretty string.
* For instance, your graph might look like this:



* Communicate your findings on the motion chart that you created. At a minimum (to achieve a 'Pass' mark), you will need to answer the following questions:
  + What are the overall patterns across different states in Australia?
  + Is there any relationship between unemployment rate and the house price index?
  + Which state has high unemployment rate and high house price index  (you may point out different period of time)? How about low unemployment rate and low house price index?
  + What is the performance of the most populated states like NSW and VIC?
* Notice that:
  + The data files are not yet in the right format to visualise using motion chart. You’ll need to aggregate the data together in different data files so you can put them in a single dataframe for visualisation. You will need to submit your python notebook which details the process of data preparation and visualisation. Python code to help you conduct data preparation is all available in "basic python support.zip" file on the Assignment Moodle page and [motion chart activity](https://www.alexandriarepository.org/syllabus/fit5145-introduction-to-data-science-online/63843/) is available in Alexandria.
  + Once the data is prepared in the right format, you will also need to output the data as a CSV file and include that in your submission.

**Task 2: Repeat the Analysis on Another Dataset/Topic**

There is a huge amount of public data available online. For example, the Australian, US, UK, Singapore and Indian governments all provide websites with links to datasets:

-<https://www.data.gov.au/>

-<https://www.data.gov/>

-<https://data.gov.uk/>

-<https://data.gov.sg/>

-<https://data.gov.in/>

There are also many other websites provide data set for the public, e.g., Kaggle, Google Public Data Sets, UCI Machine Learning Repository, etc.

Your task is to find some interesting data and do an analysis similar to Task 1 in this Part.  However, to make it easier on yourself, it is better if its a single file with a time column.  It is not necessary to do all the data merges as is the case above.  So find some data suitable for display with a motion chart.

Look for some data with a temporal component and:

1. Prepare the data as required for visualising it using motion chart in Python
2. Generate Motion Chart.
3. Communicate your findings.