TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

VICTORIA
UNIVERSITY OF WELLINGTON

EXAMINATIONS – 2019

TRIMESTER 2

COMP 132

PROGRAMMING FOR THE NATURAL AND SOCIAL SCIENCES

Time Allowed: TWO HOURS

OPEN BOOK

Permitted materials: Notes, calculators and dictionaries are permitted.
You are not allowed to use phone, online chat, email or any other mechanism to communicate with anyone.
You are not allowed to post any questions or any answers.

Instructions: **Before the examination starts:** You may login to a lab computer, create a file called `exam.ipynb`, and write your name and ID on top of the file as comments. You can download the data files from the "Exam" page. You can access the "Exam" page by going to the course home page, clicking on "Exam" from the green menu on the left hand side. Do not open this examination paper yet.

During the examination: Save your file frequently to avoid data loss. Do not spend too much time on a question if you get stuck.

At the end of the examination: Save your file and do not modify your file after the cut-off time. Submit your file immediately after the examination using our online submission system. You can access the submission system by going to the "Exam" page and then clicking on one of the "submit" links.

The examination will be marked out of 120.

Questions

1.	Programming Basics	[15]
2.	Numpy	[45]
3.	Matplotlib	[8]
4.	Pandas	[52]

Question 1. Programming Basics [15 marks]

(a) **[7 marks]** Write a Python program that first asks a user for the number of days for renting a car, and then calculates and prints the car rental fee according to the following rules:

- The car rental costs $100 per day.
- A user gets $80 discount off the total if renting for more than one week.
- A user gets $30 discount off the total if renting for more than 3 days.
- Please note that a user only gets one discount, not both.

(b) **[8 marks]** Write a Python program that first asks a user for a string, then calculates and prints the number of upper case letters and the number of lower case letters in the string.

Question 2. Numpy [45 marks]

Download the data file `rental.txt`. This file contains the typical rental price that households pay for house rentals in Canterbury collected by Stats NZ for each month from 2007 to 2018. The first number in the file is the rental price in January 2007, the second number is that of February 2007, etc. up to December 2018.

(a) **[2 marks]** Read `rental.txt` into a NumPy array, and name it `rental`.

Accessing the elements:

(b) **[3 marks]** Print the 20th element of the array.

(c) **[2 marks]** Print the last five elements of the array.

(d) **[7 marks]** Define a function `get_index` with two parameters: `year` (as an integer) and `month` (as an integer 1-12), and make it return the index of the rental array corresponding to the year and month.

For example:
`get_index(2007,1)` should return 0
`get_index(2007,2)` should return 1
`get_index(2018,12)` should return 143

(e) **[3 marks]** Use your function to print the rental price of March 2010.

Calculating the average:

(f) **[3 marks]** Find the average rental price across the whole period.

(g) **[5 marks]** Find all the elements that are larger than the average price. Print these elements and their indexes.

(h) **[10 marks]** Find and display the average price for each year. Our data array contains rental price for 12 years from 2007 to 2018.

Doing more analysis:

(i) **[10 marks]** Find the highest rental price in September in the 12 years.

Question 3. Matplotlib **[8 marks]**

Use the same data in the previous question to answer this question. Please note that this question is independent of the previous question.

(a) **[2 marks]** Create two numpy arrays to store the rental price data of the year 2007 and 2008 respectively. You may use the `get_index(year, month)` function from the previous question if you want to, but it is not necessary.

(b) **[6 marks]** Plot the two arrays generated in (a) in the same figure with two different lines. Use different colours and line styles to show the two arrays. Add a legend to label the two line plots.

Question 4. Pandas **[52 marks]**

Download the data file `automobile.csv`. The file contains a list of automobiles giving various characteristics and their assigned insurance risk ratings.

(a) **[5 marks]** Load the data to a dataframe and print the first 5 rows.

(b) **[5 marks]** Print the number of rows, the number of columns of the dataframe and the names of all columns.

(c) **[5 marks]** Extract the "Brand" and "BodyStyle" information of rows indexed 50-55 (inclusive) into a variable and print it.

(d) **[5 marks]** The "Losses" column contains some NaN values which represent missing values. Calculate the number of missing values in the "Losses" column.

(e) **[5 marks]** Calculate the average "Risk" rating for each different brand of automobile. Print the brand name and the average rating for each brand.

(f) **[5 marks]** Make a scatter plot using "Risk" and "Price" as the axes.

(g) **[6 marks]** Find and print the information about the cheapest automobile for which the "Brand" is "toyota" and the "BodyStyle" is "sedan".

(h) **[6 marks]** Sort the automobiles from the highest price to the lowest price. Out of the top 50 automobiles in the sorted list, print all the ones for which the insurance "Risk" ratings are negative.

(i) **[10 marks]** Add a new column "RegFee" to the dataframe to show the registration fee for each automobile. Calculate the registration fee using the following rule:

If the "EngineSize" is smaller than 150 or the "Cylinders" is "four", then the registration fee is 140. Otherwise, the registration fee is 240.

* * * * * * * * * * * * * * *