OSTBAYERISCHE TECHNISCHE HOCHSCHULE REGENSBURG

# ASSIGNMENT ONE

## Semester 1, 2020

## Data Science

| | |
|---|---|
| **Marks:** | 35 |
| **Length:** | |
| **Due Date:** | 19th November 2020 |
| **Time:** | **8.15am** by Turnitin |

### INSTRUCTIONS:

Students are to attempt all tasks. All work submitted must be original and entirely your own work, except where you use ideas, quotations, tables, diagrams, code, or any other material from other writers. In such cases you must acknowledge the source using the APA referencing style. A maximum of 30% of your entire assignment may be cited/quoted material.

Unless you have prior approval from your lecturer, no part of the work submitted may be used as part of any assessed work for any other academic course. All work submitted must be word-processed and the presentation of a professional standard.

Assignments need to be uploaded on this paper's Moodle website using the appropriate upload link.

This assignment comprises one component, which needs to be thoroughly analysed. You need to find articles, magazines, books, and even newspapers that have also been looking at this same (or similar) topic and then discuss what their opinions are on the topics.

You may not engage in a trivial discourse on the subject, but your discussion of other people's opinions needs to be your own interpretation. Your report should not exceed six A4 pages of text. All diagrams, pictures, and tables may not exceed 30%. A template has been provided for this assessment.

## TASK ONE – Write a function that take a filename as input

1. Go to moodle and download file **bill_of_rights.txt** under the **Activity files for Lecture** folder Write functions that takes filename as input and find total number of lines, words and most frequent word in the file and write these statistics in a file named '**stats.txt**'. (10pts) (20min)

## TASK TWO – PARSE a file to produce a list

2. Go to moodle and download file **gps.csv** under the **Activity files for Lecture** folder. Parse the file and produce a list of coordinates (**lat** and **lon**) along with the **unix timestamp** of their visit. (10pts) (15mins)

## TASK Three– Convert DataFrame

1. Download the test_assignment_data.csv file from Moodle under the **Activity files for Lecture** folder.
2. Lets say we are only interested in population (POP) and total GDP (tcgdp), convert the data frame with "country" as index and POP, tcgdp are columns.
3. Rename POP -> Population and tcgdp-> Total GDP (1pt)
4. Add a new column 'GDP Percap' to the existing data frame. (4pt)

| | country | POP | tcgdp |
|---|---|---|---|
| 0 | Argentina | 37335.653 | 295072.218690 |
| 1 | Australia | 19053.186 | 541804.652100 |

HINT: GDP per capita can be calculated as ratio of Total GDP and Population.

## TASK Four – Find Data

Go to Moodle and download forest fire data "forestfires.csv" and "ReadMe_forestfire.txt" under the **Activity files for Lecture** folder.

1. Find the month and day where "rain" measurement is NOT Zero. (2pts)
2. Find the descriptive statistics of "temp" and "wind" (2pts)
3. Find mean temperature for each month and plot it (2pts)
4. Find the month and day with the highest and lowest temperature. (2pts)
5. Find the mean temperature at different unique locations (x,y) (2pts)

Data Source: http://archive.ics.uci.edu/ml/datasets/Forest+Fires