Students should be familiar with the TUD General Assessment Regulations before starting work on the assignment.

Assignment Specification

Module: Data 9910 – Working with Data

Title:	Lecturer:	Class Group:
35% CA	David Leonard	TU59 / TU60

Name:	Worth:	Due Date/Time:
Data Wrangling Assignment	35% 100 marks	Friday 20th November 2020/ 23:55

Submission Mechanism:	Late Submission Penalty:
Upload to link available in Brightspace	10 marks deducted per day late

Description of Tasks:

This assignment gives you the opportunity to demonstrate what you have learned about how to import, clean, explore and manipulate data. The tasks are as follows:

Find data in publicly accessible online sources e.g. UCI, data.world, Kaggle, webpages, APIs etc..

Part 1

Import the data and perform any cleaning and merging to produce a final dataframe. (33 **marks**)

Part 2

Carry out exploration of this dataframe to develop an overall understanding of the data. (33 **marks**)

Part 3

Focus on a particular subset of the dataframe and drill down into it extracting details to answer a series of questions that are of interest to you as an analyst - Ideally the motivation for such questions would be framed within the context of a hypothetical use case scenario. (33 **marks**)

Submission Requirement:

Zip File containing:

Report: 1 word processed document e.g. Microsoft Word or OpenOffice (pdfs will not be accepted, as they can be generated if required). 5 pages would be an indicator of the expected length with equal amount of text written on each part. The examiner will stop reading reports after 6 pages. Plots and tables should be put in the appendix and referenced in the main text. The total number of pages in the report should not exceed 8 pages.

Students are encouraged to present their reports in a professional manner, however it is best to focus on the structure of the paragraphs i.e. points made (what was done), flow of reasoning (motivation for steps), depth of analysis and conclusions reached.

Code: 1 script file (written in R or Python). Please ensure that variable names are informative and appropriate comments are placed before different sections (line-by-line commenting obscures the structure of the code and is not necessary). Code sourced from elsewhere must be **clearly cited** (bear in mind your programming skills are being examined). The code should work without issue when run on the examiner's machine.

N.B. data should **not** be uploaded.

Demo Requirement:

None, though please bear in mind that the examiner reserves the right to conduct video interviews with students about their work to ensure that learning outcomes have been met.

Rubric:

The grading rubric for the assignment may be found in the accompanying file Rubric.pdf