Data Analytics Project

Individual project, value 25%

Analysis of a Data Set

- Deliverables

- Individual Report (.pdf) of your Analysis Findings (approx. 20/25 Pages & Tests Done using Python)

- Python script containing your code.

- Due date: 3,10,17,24 November. Uploads via Moodle

- Compulsory Presentation (Evidence of Critical Thinking)

What you are expected to do.

- Select a data set from the Moodle links and follow the Data Analytics cycle
- 1. Business Case Evaluation
  - Research the data set area.
  - What are the questions you want to answer?
  - Define a research question and a hypothesis to test.
- 2. Extraction and Preparation
  - Load the data set into Python
  - Clean and filter the data
  - What types of data do you have
  - Load the data into suitable Python types – vectors, data frames etc.

What you are expected to do.

- 3. Summarisation and exploratory analysis
  - Using your statistically knowledge decide the statistical tests you want to carry out (location aka measures of central tendency and and variability aka measures of spread, distribution, hypothesis testing etc.)
  - Implement your statistics in Python
  - Document why you chose these tests
- 4. Visualisation
  - Select appropriate plots that explore the data and implement them in Python
  - Document each plot and link it to your statistics in part 3
- 5. Knowledge extraction
  - Write up your insight into the data set and application area (link to Questions in part 1) – individual.
- Reference all sources

How to undertake the project

- Do the Data Camp Labs

- Review 1st year statistics

- Read the essential reading on Moodle

- Read in detail the example Basic Statistics and Plots of the Titanic Data given on Moodle

- You will need to teach yourself basic Python

- Do the Python tutorial https://docs.python.org/3/tutorial/

- Read the Python links on Moodle

- Each week complete a project task.