Data Analytics Project

Group of max 3 students, individual mark allocation, value 25%

Analysis of a Data Set

- Deliverables
- Individual Report (.pdf) of your Analysis Findings (Done using Python)
- Python script containing your code.
- Due date: 2 December. Upload on Moodle
- Compulsory Presentation (Individual Marks, 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 - group.
  - What are the questions you want to answer? - individual
- 2. Extraction and Preparation
  - Load the data set into Python - group
  - Clean and filter the data - group
  - What types of data do you have - individual
  - Load the data into suitable Python types – vectors, data frames etc. - individual

What you are expected to do.

- 3. Summarisation and exploratory analysis
  - Using your statistically knowledge decide the statistical test 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 - individual
  - Document why you chose these tests - individual
- 4. Visualisation
  - Select appropriate plots that explore the data and implement them in Python - Group
  - Document each plot and link it to your statistics in part 3 - Individual
- 5. Knowledge extraction
  - Write up your insight into the data set and application area (link to Questions in part 1) - individual

How to undertake the project

- Get in a group
- Review 2nd 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 tutorials
- Read the Python links on Moodle
- Each week complete project tasks listed on Moodle