ISSU0053 Data Science and Big Data Analytics **Report ID:**

UCL International Summer School for Undergraduates 2019

**Assessment I: Computer Practical Work and Write-up (50%)**

|  |  |  |
| --- | --- | --- |
| **Task** | **Mark** | Comments |
| T1: Explore the dataset | B | - didnt rename columns to make more intuitive  - in graphs x and y axes not labeled |
| T2: Linear **/** logistic regression **/** LDA | A+ | - you’re trying to predict is a score calculated from the predictors (taken the average). So you obviously get high accuracy |
| T3: Decision Tree methods | A |  |
| **Overall task:** | A |  |
| Use of Figures | C |  |
| Introduction, Commentary, Summary | D | - very few comments in the code, no summary |
| Coding Technique | C |  |
| Writing Standard | D |  |
| Structure & Presentation | C |  |
| **Overall:** | C  (55%) |  |

**Strengths:**

- you did some good EDA and commented it well, it was easy to follow what you were trying to do

**Focus on improvement:**

- you clearly didnt’t research the dataset enough. you’re trying to predict is a score calculated from the predictors (taken the average). So you obviously get high accuracy. There is no need for model building because if you look into the cato report, you see that personal and economic freedom are average scores of a number of other criteria.

- Also – why is KNN used for a regression task? KNN is used for classification, not regression. Getting the best K of 1 should have shown you this. You should revise KNN

-There are very few comments in the code after EDA

- there is no summary and no comparison of the different models used