ISSU0053 Data Science and Big Data Analytics **Report ID:** 1043356

UCL International Summer School for Undergraduates 2019

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

|  |  |  |
| --- | --- | --- |
| **Task** | **Mark** | Comments |
| T1: Explore the dataset | A++ |  |
| T2: Linear **/** logistic regression **/** LDA | A++ |  |
| T3: Decision Tree methods | A++ |  |
| **Overall task:** | A++ |  |
| Use of Figures | A |  |
| Introduction, Commentary, Summary | A++ |  |
| Coding Technique | A++ |  |
| Writing Standard | A++ |  |
| Structure & Presentation | A++ |  |
| **Overall:** | A++  (100%) |  |

**Strengths:**

- excellent report

- each step commented with appropriate introduction and proper conclusion which sums everything up and compares the different models

- dealt with ACTUAL big data

- feature engineering/EDA done in a separate file, which makes it easy to follow and follows good coding practices

**Focus on improvement:**

- the choice of problem is not really interesting since the fare amount OBVIOUSLY depends on the distance (thats how taxometers work)

- more research about taxis in NY would have also revealed that the cabs can hold more than 4 people (like the black ones in London) so it’s not surprising that the fare amount doesnt different between “small” and “big” groups of people