**Data Exploration of Learning Analytics**

**Designing the Data Pipeline**

This data was analysed within the context of helping a company to apply learning analytics to improve their educational provision. The chosen routes of enquiry were to consider the decline in participants to interpret a drop-out rate, the influence of the introduction of videos, and the variation in attainment across the different runs of the course.

This was done by preparing the data into data frames at each stage to select and sometimes merge specific group data. This started with analysis of the step activity, to consider the number of people recorded as starting and finishing at each step. This aspect included a mistaken assumption: that non finishing a step meant that an individual had dropped out of the course. This assumption would be challenged further within a future section.

Due to the doubts surrounding the counting of ‘finishers’ as an estimation of course success, an alternative route was chosen; attainment. This was considered through the number of correct questions within a cohort, as a measure of the success of the teaching in the only deconstructed form of assessment. The impact of videos was also considered alongside this dataset, to consider if they are a popular method to invest in, and if their introduction has influenced the success of the cohorts.

Finally, the flaws in the step activity dataset were investigated to consider what was meant by an unfinished step, particularly in comparison to the leavers survey introduced in the fourth run of the course. This was designed to question the validity of the data.

The entirety of this analysis was informed by the CRISP – DM Framework.

**Intricacies of CRISP- DM**

All routes of analysis were influenced by Business Objectives. All data files were examined to provide accurate responses, whilst ensuring that the focus was on improving learning analytics to improve the course for future years.

The process of data preparation ensured that the data was cleaned and easy to use within analysis.

The most beneficial aspect of the CRISP-DM cycle was the exploration.

Throughout the report as a whole, a key strength of CRISP – DM was its fluidity. This process was supported by the constant switching between ‘data preparation’ and ‘data exploration’, to constantly be refining the findings that were produced.

However, a limitation of this fluidity is that, when in use, the categories have the potential to become constraining, in the assumption that the analysis must contain all stages. This particular exploratory process lacked modelling, because it did not apply to this dataset within the goals that had been set for its analysis.

**Best Practice**

R Markdown and Project Template – clearly see each stage of the exploration, can reproduce all code

Github – clear process of building the code.