## **Abstract of Apply Data Science**

This Abstract will introduce the work that we have down about the Smart Home project and what we will plan to do for next step.

Firstly, we dedicated time to understand how the data was collected and how the used collecting devices (camera, accelerometer and PIR) work. From this task we got parameters and information with regard to operation ranges, sample frequency and location of the devices. Secondly, based on the information gathered, we worked on getting rid of noise an erroneous data within the dataset, discarding data that was outside operation ranges or semantically inconsistent. After this data cleaning we focus on how we could visualise and interpret each one of the datafiles generated by different devices. As an example, we describe with more detail the process followed for video datafiles.

**Video**

We firstly computed some features for the 2D bounding box and 3D bounding box, and then deleting some useless data based on the meaning value of these features. These feature included width, length, height and area. After that, resampling the video’s data can assist us view this dataset clearly. We aggregated the each second’s dataset that contains more than one row into another one row only. Finally, the three kinds of datasets (e.g. original data, cleaning data and resampling data) will be visualized in order to compare and analyse the effect of these operations.

**Further work**

After visualising and interpreting each datafile separately, we are now in a good position to integrate all datafiles to infer new information from the entire dataset.

As part of the cleaning process for the dataset, a next step would be to create validation rules to help us to make the dataset consistent with really, this is, making sure for example the PIR dataset do not indicate that the person is in two places at the same time.