DISCUSSION 20th Jun after Gerard

We have decided to run each data set separately

And we found out from Gerard he is just bothered about how many ticket registrations happen, not how many people turn up.

We want to be able to have a prediction model which says how many people are going to turn up at the end at given points in time.

Gerard wants to click a button and predict end registrations

* Train the model and then input just 25% of one at the end to see if it accurately predicts what the rest of the data set/ registrations are likely to be AT THE END

Data pre-processing

* Take out attended column
* Take out cancelled

Can we find

* Where the different data sets had advertising?
* Is there a difference between each target audience

TRAINING THE MODEL

* Train the model on the whole data set of a few of our data sets then leave a few which we test at different stages of 25%, 50%, 75%

Things to do for this week

1. Pre-process all the data consistently – Niv
2. Find a trend of advertising? Does this existing – Lucy
3. Is there a difference between each target audience - Awais
4. Get to know some of the libraries which are compatible with our project (sales, timeseries) and list them out so we can assign members to work on them - Kelvin
5. Write report plan – Julie
6. Initial Evaluation of data Umer