**Things we need to data with the coursework:**

1. Data
   1. Describe the dataset
   2. Show that there is more data than 13 thousand something that he pointed out in the document
   3. Describe how to split the data, how to validate the data
2. Formulate machine learning problem
   1. Loss function
   2. How we will measure the accuracy
   3. The error function
   4. The techniques
   5. Basically, Need to define the input data, the ML problem, the ML objective, loss function, algorithm to use, performance metric. In mathematical formulas wherever possible.
3. The methods
   1. Baseline (KNN)
   2. Possible K-means?
   3. Investigate KNN
   4. Optimisation
4. Rank error understanding
5. https://docs.google.com/document/d/1PTPuIAenCW95\_5EA4PNDtiawR0YdEdEmstwpcJiTgdQ/edit