1. **What is the problem?**

* Current model not accurate enough
* Small number of customers receive recommendations, while a large number of customers receive little to no recommendations at all.

1. **Current state**

* Available is a large unclean training data set, comprising of 1.5 years’ worth of customer data, and a smaller 1-month data set for testing purposes, as well as a sample submission, all are in Spanish.
* Limited working and general knowledge of machine learning theory and algorithms.
* Some of the team go away on holiday on week 2 of the project, limiting resources.

1. **End goal**

* Building a recommendation engine
* Accuracy of recommendation engine tested using MAP (Mean Average Precision), with a target score of at least 0.01
* Recommendation engine will suggest up to 7 products for each customer
* Testing data will be the last 2 months of the training data (testing data provided does not have the products customers bought, so we can use the answers from that.)
* Recommendations will be given in a simple CSV format, like the Kaggle format (any extra time will be used to improve the engine rather than the customer facing aspect of this.)
* The Engine will be highly tuned and specific to this dataset, and not a generic engine as this will prove to be too difficult.

1. **Milestones**

* Learn about Decision Trees and random forest machine learning methods and decide what recommendation approach to use (to be done by Tuesday)
* Clean data to start working with
* 1st recommendation engine
* 2nd improved recommendation engine
* **Submission**