**Workflow**

1. ~~Data cleaning for dataset~~
   1. ~~Train data - EDA & Cleaning – 01a~~
   2. ~~Spray data - EDA & Cleaning – 01b~~
   3. ~~Weather data - EDA & Cleaning – 01c~~
2. Merge data & EDA – 02
   1. More EDA
   2. Explain feature engineering  
      DARION
3. Modelling – 03
   1. Baseline prediction / score
   2. Collate all the models (LogReg, RF, GaussianNB, AdaBoost, GradientBoosting, SVC)
   3. TimeSeries predictions JOSEPH
   4. Interpretations of model
   5. Select best model for predictions, based on which score?
   6. Export model to .pkl, REBECCA
4. Predictions, Conclusions – 04
   1. Kaggle Test data transformation
   2. Kaggle predictions
   3. Cost-benefit Analysis - This should include annual cost projections for various levels of pesticide coverage (cost) and the effect of these various levels of pesticide coverage.

* Presentation Slides
* GitHub submission

**Extras**

1. Notebooks structure in each notebook. (Contents)
2. Problem Statement (becca did, to check)
3. Executive Summary