# Results

## Data sourcing

\*Tom’s section\*

## Visualisation of structure within data

* Why visualise structure of data
* What data was used, bedrock, superficial, both, artefacts etc?

### Principal Component Analysis

* Introduction to what PCA is used for
* Do groups resolve well in both techniques? Which groups resolve best/worst? Dies this link back to domain knowledge?

### t-distributed stochastic neighbour embedding

* Introduction to what t-SNE is used for
* Do groups resolve well in both techniques? Which groups resolve best/worst? Dies this link back to domain knowledge?

## Feature selection

* Why do feature selection and what is feature selection?
* Explain method used
* Which features used and why based on graph

## Model comparisons

## Novelty detection

* Outline problem aka only sampled from number of sites within UK, what if artefacts came from sites not sampled?
* With any of these models all artefacts would be classified into the sites sampled from so need method of identifying which artefacts cannot be classified into these sites.
* Explain what local outlier factor model is
* How many artefacts could be classified and what proportion is this of total artefacts
* The non-classifiable artefacts were classified as ‘other’

## Final model evaluation

* Which model used and why
* How was it evaluated
* Overall F1-score
* Class-specific F1 scores
* Feature importances, link in with domain knowledge?

## Classification of artefact sourcing

* Final model built and artefacts classified
* \*Tom how deal with probabilities at end\*

## Implications for archeology

\*Tom’s section\*