## Data Clean

* Missing data check✓
* Data type check
* Information check
* Skim used ✓
* Explanations and summary given
* Data dictionary
* Plot the three key variables and summarise findings
* Data tables for Publisher and Platform, captioned and explained
* Explanation of why variables will be left out
* Apply changes and skim

## EDA

* Explained the purpose of EDA
* Natural groupings/ parallel coordinate plot
* Principle component analysis of key variables
* Categorical relationships
* Numeric relationships

## Data pre-processing

* Reasonable pre-processing step selection
* Seed set
* Remove rows that lead to over fitting
* Create training set and test set
* Create model specs

Year – Might factor in early adopters to new tech and prevent overfitting

## Model fitting

* Set seed
* Outline the process
* Tune model
* Cross validation prediction, accuracy and roc\_auc

## Model evaluation

* Outline process
* Vip importance plot
* Fit true values and graph against model
* RSME and R^2 of model
* Summarise important variables
* Comment on model adequacy
* Make prediction