**Introduction**

**an introduction/overview/executive summary section that describes the dataset and summarizes the goal of the project and key steps that were performed**

key steps

**Method**

**a methods/analysis section that explains the process and techniques used, including data cleaning, data exploration and visualization, insights gained, and your modeling approach**

**1. Data cleaning**

edx data set contains 6 columns (userId, movieId, rating, timestamp, title and genres).

[head(edx)]

In order to facilitate movie rating prediction modelling, pre-process data cleaning is applied to edx data set prior to data partition creation. Title column is split to title and year. Timestamp which is number of second after 1-Jan-1970 00:00:00 is converted to date. Using createDataPatition function of caret package to create train set and test set with percentage of 80% and 20% correspondingly. Semi-join by movieId and userId is applied to test set to avoid #NA situation when joining is applied to test set in validation stage.

[head(train\_set)]

**2. Data exploration**

Talk about the 4 effects in train set.

1. Movie also with chart
2. User also with chart
3. Genre also with chart
4. Rate per Year also with chart

Visualization (included in data exploration)

Insights gained

Modeling approach

**Result**

**a results section that presents the modeling results and discusses the model performance**

RMSE result

**Conclusion**

**a conclusion section that gives a brief summary of the report, its limitations and future work**

further discussion on