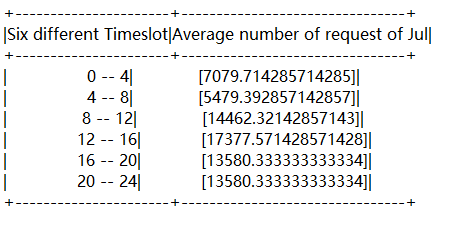
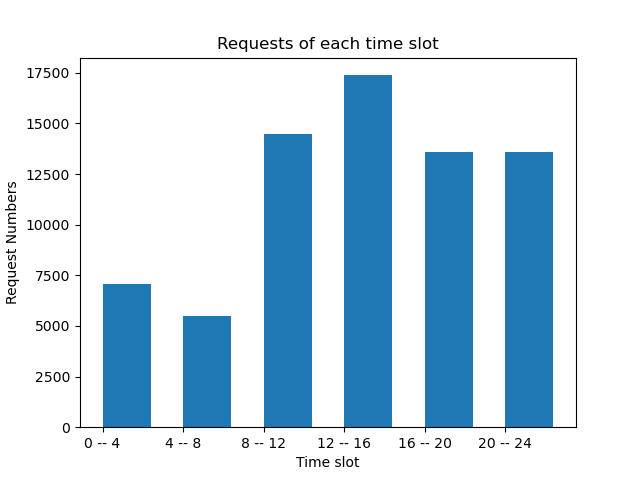
Question 1. Log Mining :

**Q1A:**

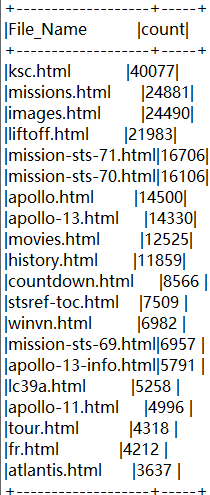
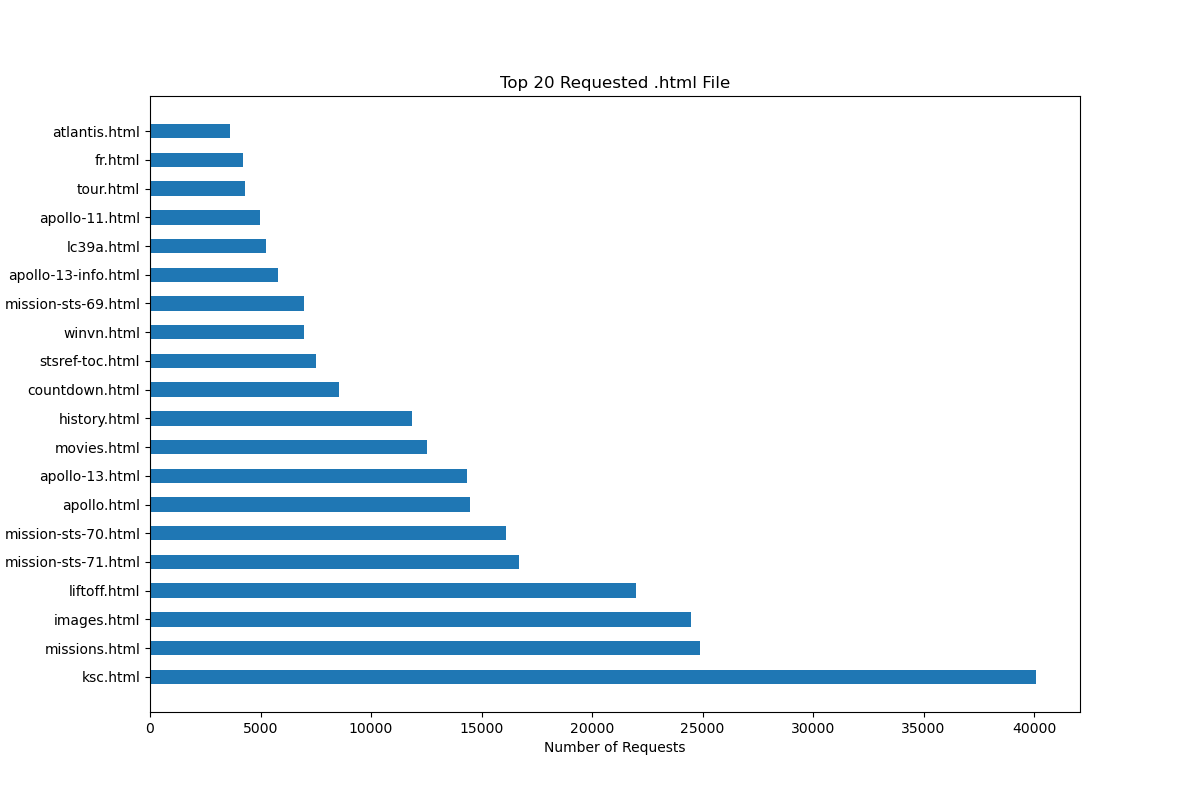
**The average number of requests on each four hours of a day of July 1995 by the local time:**





As we can see in the figure, most of requests happen in the from 8 o'clock to 24 o'clock. From 12 to 16 time slot, it is the time slot that has the largest amount of requests. Between 0 to 8 o'clock, the request are significantly less.

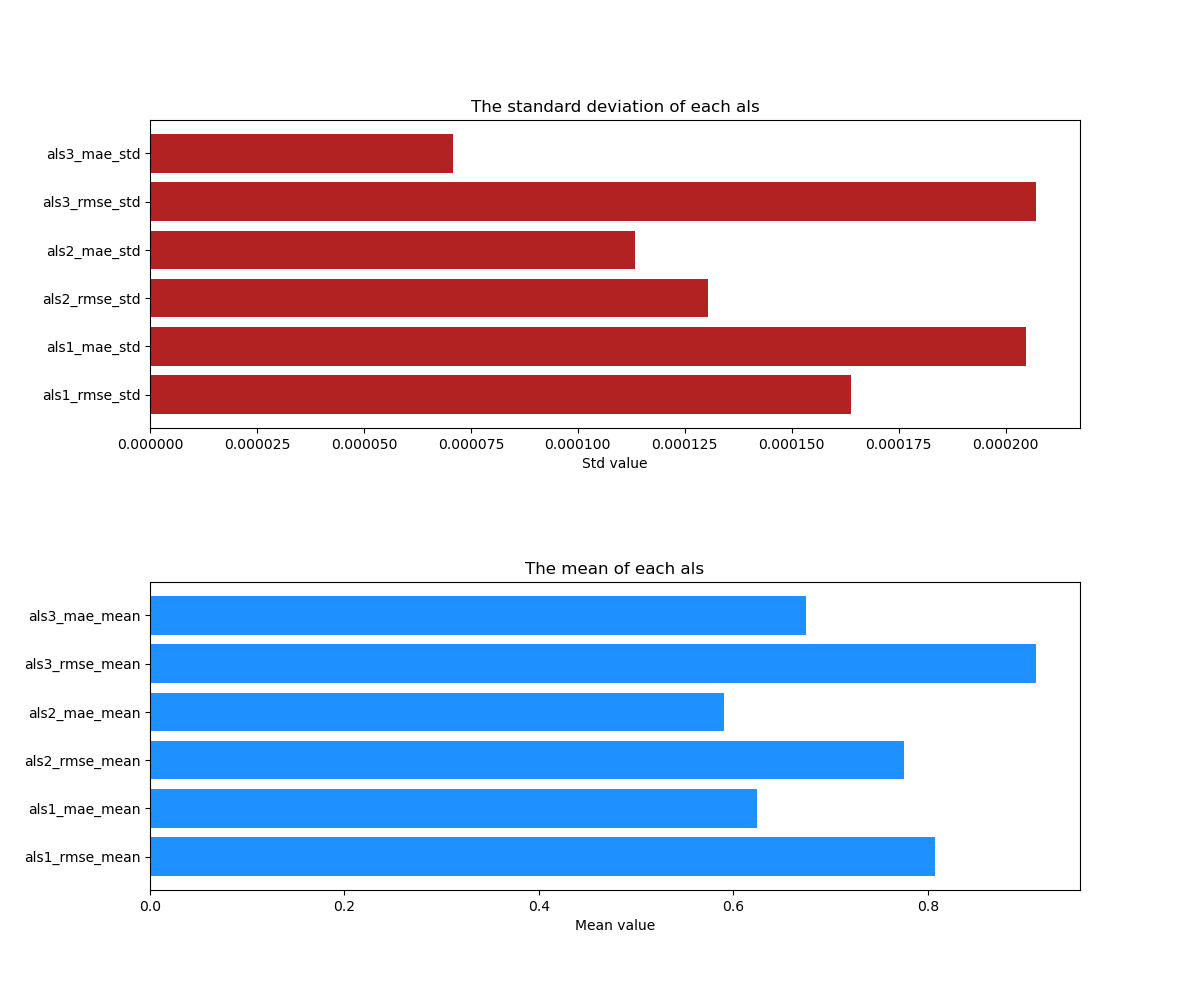
**Q1C:**

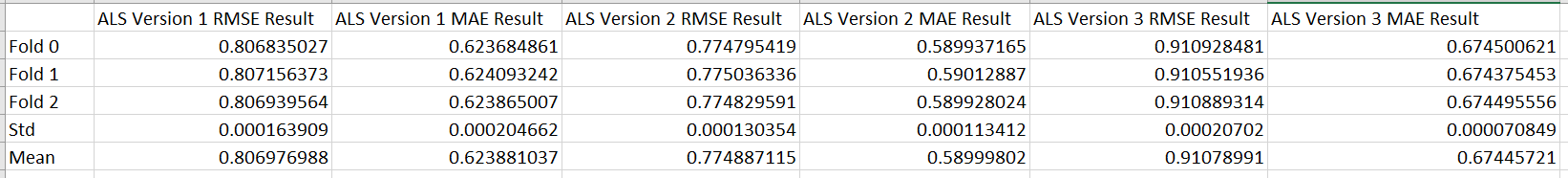
**The top 20 most** **requested.html files :**

The .html file with largest request is ksc.html. Many .html requested files are related to Apollo-13 and various mission-sts. People are also interested in images.html and missions.html

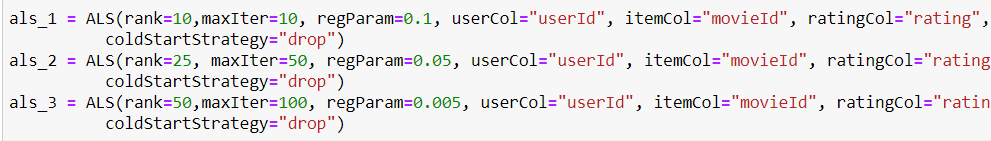
**Question 2. Movie Recommendation**

**Q2A：**



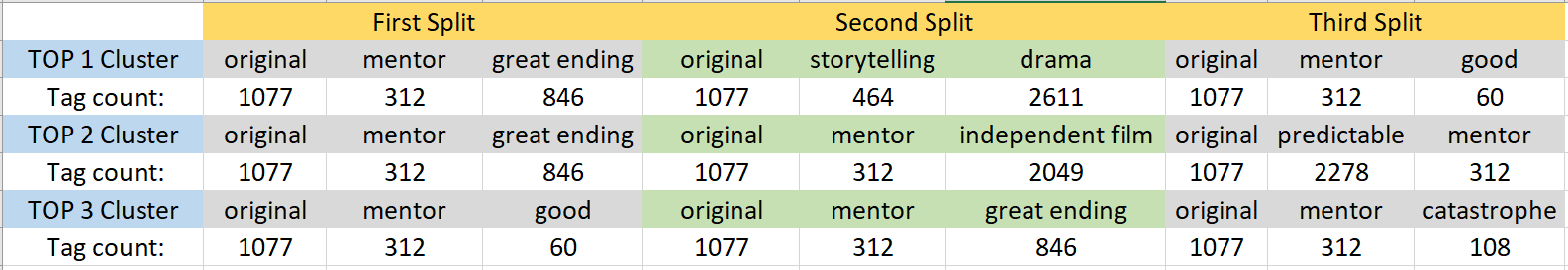
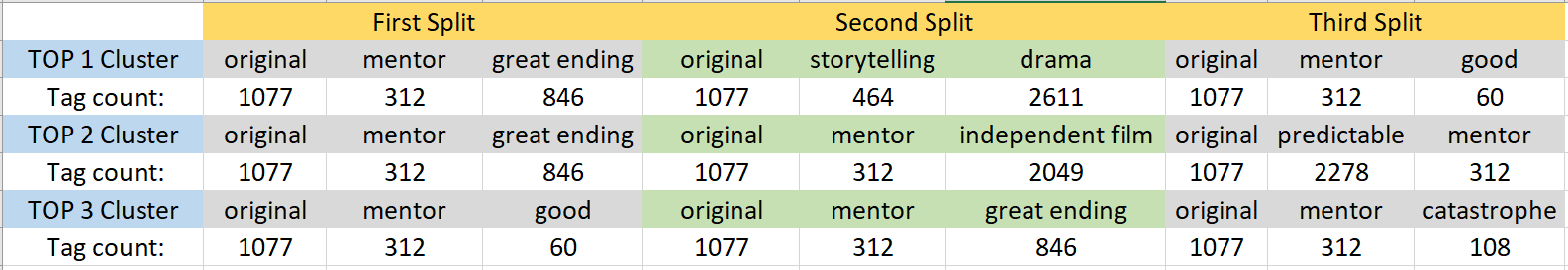


**The numbers and bar charts are generated by using 3-fold cross validation based on 3 different ALS settings shown below:**



Compared to the result from first version of ALS setting and second version of ALS setting, the rank and maximum iterations are increased and regParam(regularization) is reduced and the rmse and mae value are getting smaller which mean the second version is a better fit with higher rank and more iterations. However, in the third version of ALS setting, the model seems start to become over fit. Even though it has more rank and iterations but it gives higher RMSE and MAE value which indicates it start to overfit.

**Q2C：**



**This table is generated from performing 3-fold** **cross validation with the first version of ALS setting in Lab 3.**

The tags are the top 3 tags from each cluster in each split of cross validation. As we can see, the most popular tag in the movie tags dataset is 'original'. In all top 3 clusters in 3 splits, the word 'original' is always the top tag. Also, tags like 'great ending' and 'mentor' are also very popular among the top tags.