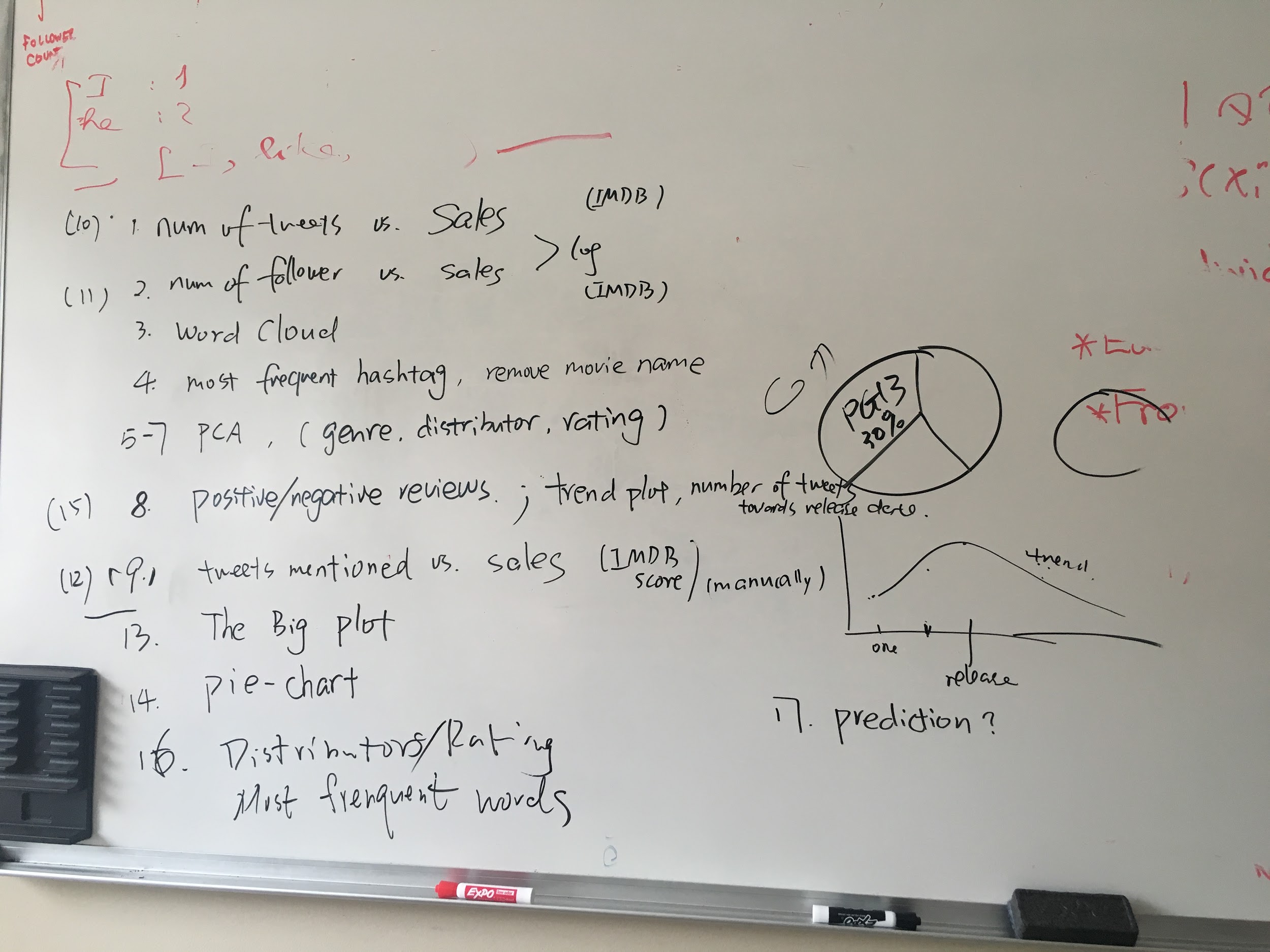
2/3 Meeting Summary

Here’s the plots that we can think of :



1. we all need to work on the rough plots and start putting plots into the slides; done by Thursday Night.
2. slides: make a clear flow (Sici/Lu)
3. improve the plots after Thursday
4. Assignments:
   1. Lynn: 1,2,9
   2. Sici: 3,4
   3. Hoang: 5-7,8
   4. Lu: retrieve data for 8;
   5. Cenzhuo: retrieve data for 8, 13-16

5. Probably meet on Friday afternoon to improve the plots and go over the flow (slides) together?

6. Probably another meeting on Sunday for rehearsal of the presentation

1/31

Things we already done:

1. retrieved the data (timeline and screenname) of 89 movies that are usable, see info.py

the resulting files are: Movie\_timeline.json; movie-list.py; Movie\_names.json

1. transfer the Movie.csv to dataframe in python, see dataframe.py or dataframe.ipynb
2. Pca analysis and term-doc Matrix, see ProcessData.py under Hoang folder

Assignments/ things need to do

1. retrieve data with time period of 2 month around the release data (Cenzhuo)
2. remove the three outlier movie (which have spam messages) (Cenzhuo)
3. plot the movies with different color depends on Genre, Studio, Rating (Lu)
4. K-means clustering, fix pca (Hoang)
5. add common stop word to the stoplist (Sici)
6. data cleaning (Sici)
7. Plot the number of tweet vs sales (Luyun)
8. plot the number of followers vs sales (Luyun)
9. prepare for the presentation slides and think about the entire flow/logic of our project that what questions we answered based on what we did and think more subquestions that we can explore. (Thibault)
10. Word Cloud (Sici)