Match the Google Apps with Mturk results and compare results

Jenny Zeyang

What was done-1. Preliminary matching

- **Method:**
- Every word to lowercase
- use pandas.str.contains function

- **Result:**
- Preliminary matching: only 48/79 were matching successfully
- After manually matching: 76/79 Mturk apps were found in Google Play store

- **Main reason:**
- 1. Google dataset is not big enough, need to crawl more data. (already find a way, will do this week)
- 2. matching algorithm not good enough, need to revise.

- **Matching tips(not yet realize)**
- 1. delete words after hyphen
- 2. delete 'free'

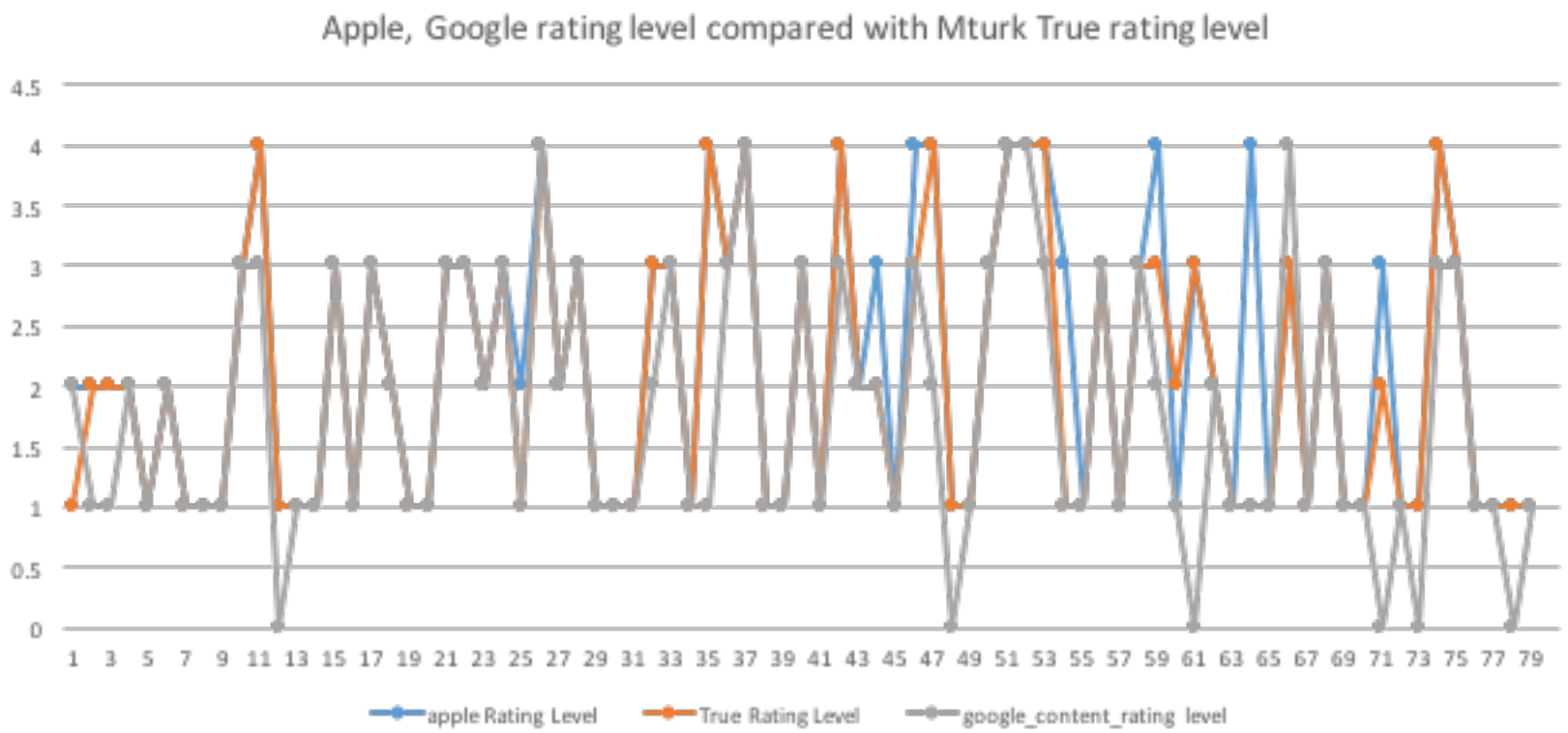
What was done 2. Comparison analysis

- In the following analysis, use levels instead of actual rating names to be consistent

Level	Apple rating names	Google rating names
1	4+	Everyone
2	9+	Everyone 10+
3	12+	Teen
4	17+	Mature 17+

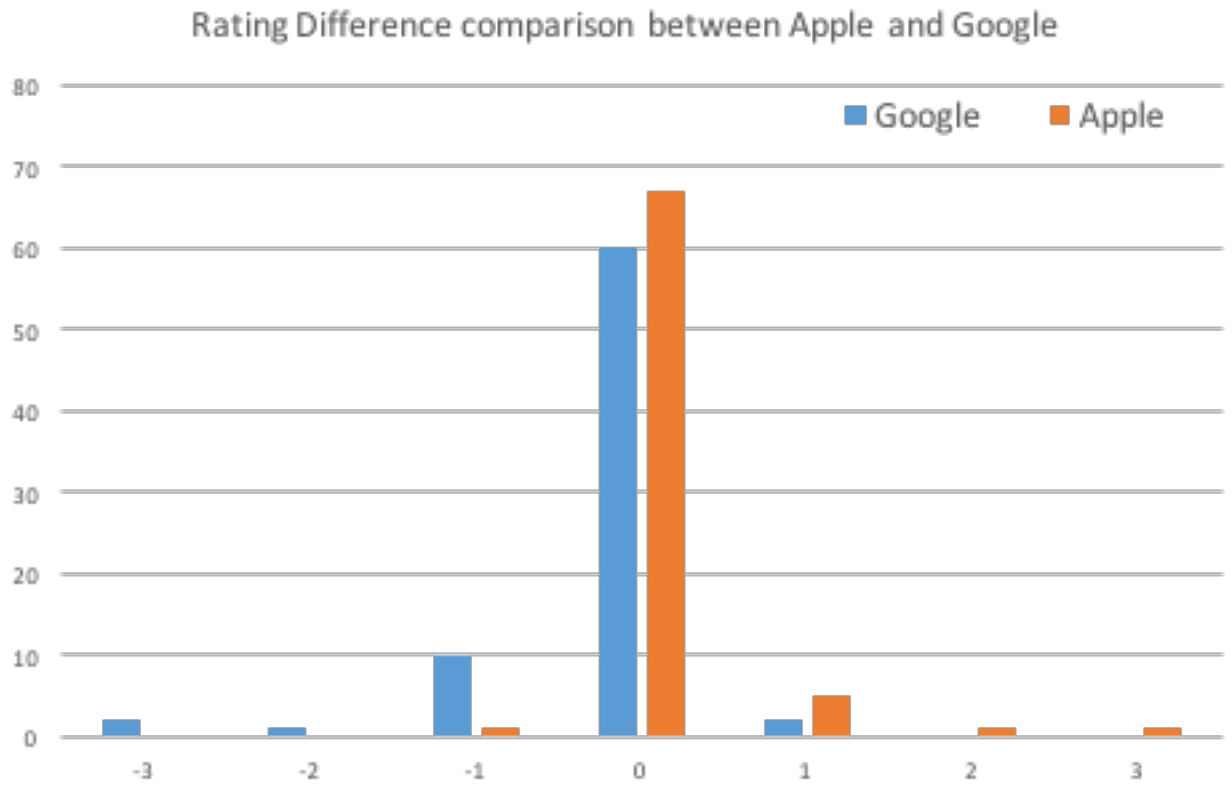
Note: Maybe there are some difference between Apple and Google rating rules, it's not accurate to treat them the same need further research on Apple and Google rating criteria.

Finding 1- Apple tend to overate, while Google tend to underrate



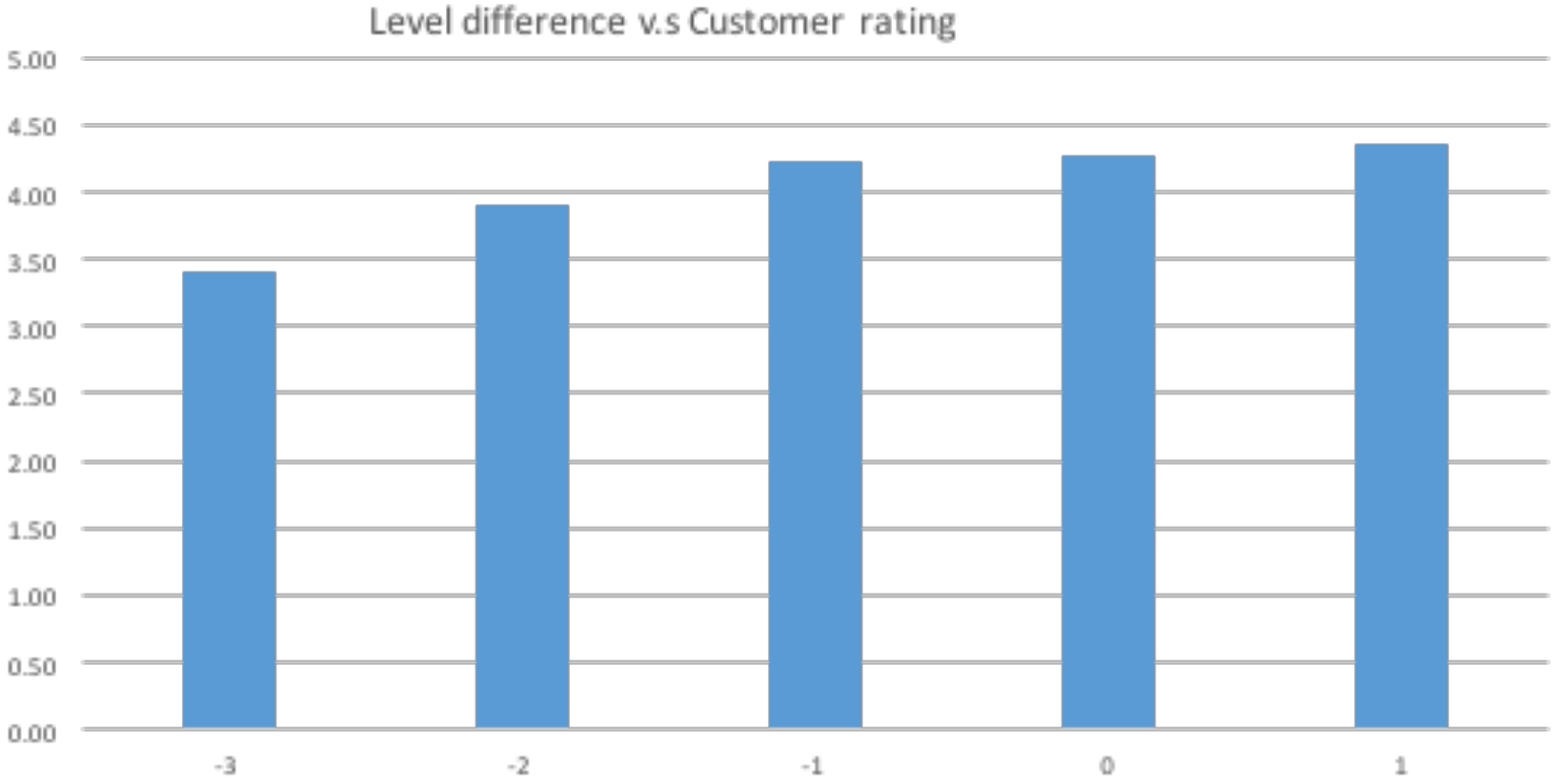
- Apple rating levels tend to be higher than the true rating levels
- Google rating levels tend to be lower than the true rating levels

diff_level	Google	Apple
-3	2	0
-2	1	0
-1	10	1
0	60	67
1	2	5
2	0	1
3	0	1



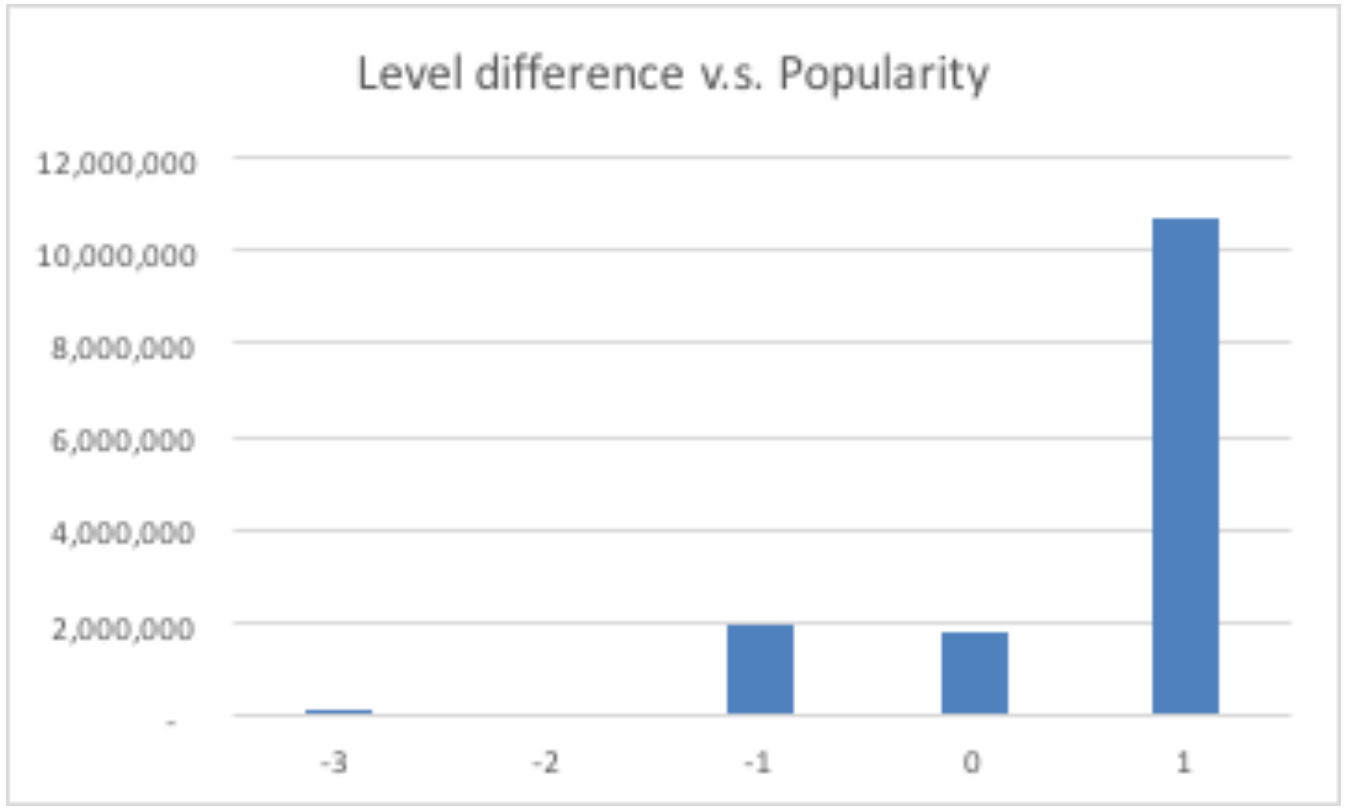
- Proved that Apple tend to overate, while Google tend to underrate

Finding 2- Google tend to underrate low customer rating apps



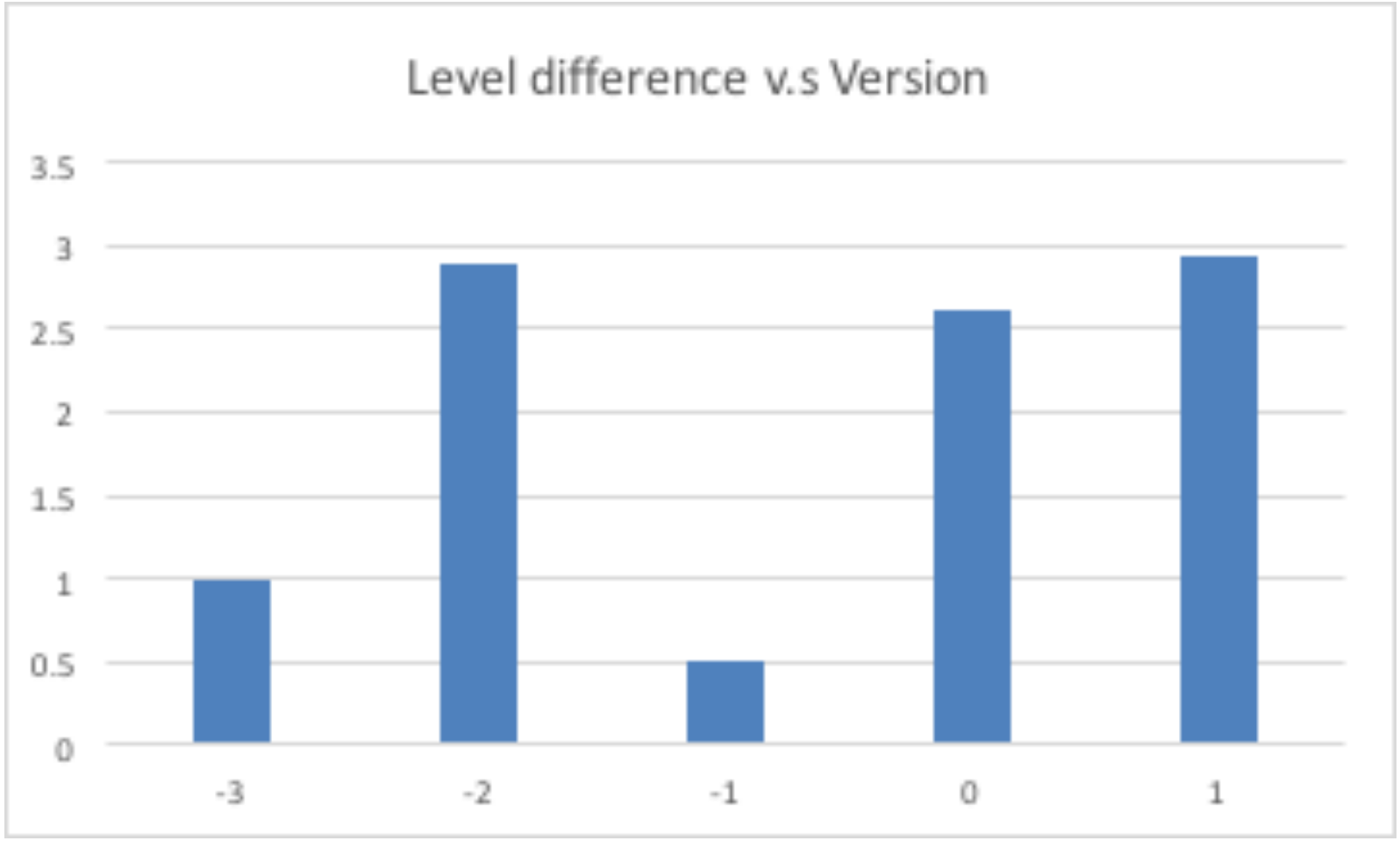
- Avg number

Finding 3- Popular apps get overrating more



- Avg number
- This finding is consist with Mengchuan’s Apple finding

Finding 4- High version apps get overrating more



- Column ‘-2’ is outlier, it only has one data point.
- Major tend is still higher version apps get overrating more, ignore the outlier.

What to do next

1. Crawl deeper for Google play app store

- already find a way: crawl 'similar apps' and 'same developer apps' by several levels to acquire more apps, then delete duplicates

2. improve the matching algorithm

3. text analytics on 'description’

4. research on original rating rules of Apple and Google platform, see if they are similar.

5. get more true ratings on MTurk