

Analysis of Apps in Apple and Google Play

Group 18:
Ruochen Zhong
Qing Gao
Shiying Wang





Focus

- 1. Review**
- 2. User Preference**
- 3. Market Comparison**

Overview of the Dataset for Apple Store and Google Play

In [4]: data_apple

	id	Name	Size	Price	rating_count_tot	rating_count_ver	Rating	user_rating_ver	Version	Content_Rating	Category	sup_devices.num	ip
0	281656475	PAC-MAN Premium	96.119141	3.99	21292	26	4.0	4.5	6.3.5	4+	Games	38	
1	281796108	Evernote - stay organized	151.232422	0.00	161065	26	4.0	3.5	8.2.2	4+	Productivity	37	
2	281940292	WeatherBug - Local Weather, Radar, Maps, Alerts	95.867188	0.00	188583	2822	3.5	4.5	5.0.0	4+	Weather	37	
3	282614216	eBay: Best App to Buy, Sell, Save! Online Shop	122.558594	0.00	262241	649	4.0	4.5	5.10.0	12+	Shopping	37	

- Features: Price, Rating, Category, etc..
- Original Datasets from two Kaggle datasets

In [9]: data_android

	Name	Category	Rating	Reviews	Size	Installs	Type	Price	Content_Rating	Genres	Last Updated	Version	Source
0	Photo Editor & Candy Camera & Grid & ScrapBook	ART AND DESIGN	4.1	159.0	19.0	10,000+	Free	0.0	4+	Art & Design	2018-01-07	1.0.0	Android
1	Coloring book moana	ART AND DESIGN	3.9	967.0	14.0	500,000+	Free	0.0	4+	Art & Design;Pretend Play	2018-01-15	2.0.0	Android
2	U Launcher Lite – FREE Live Cool Themes, Hide ...	ART AND DESIGN	4.7	87510.0	8.7	5,000,000+	Free	0.0	4+	Art & Design	2018-08-01	1.2.4	Android
3	Sketch - Draw & Paint	ART AND DESIGN	4.5	215644.0	25.0	50,000,000+	Free	0.0	12+	Art & Design	2018-06-08	Varies with device	Android
4	Pixel Draw - Number Art Coloring Book	ART AND DESIGN	4.3	967.0	2.8	100,000+	Free	0.0	4+	Art & Design;Creativity	2018-06-20	1.1	Android
5	Paper flowers instructions	ART AND DESIGN	4.4	167.0	5.6	50,000+	Free	0.0	4+	Art & Design	2017-03-26	1.0	Android

What people tend to say when they try to rate an App?

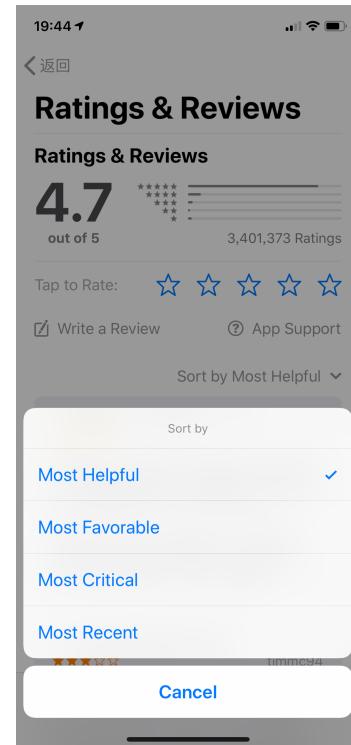
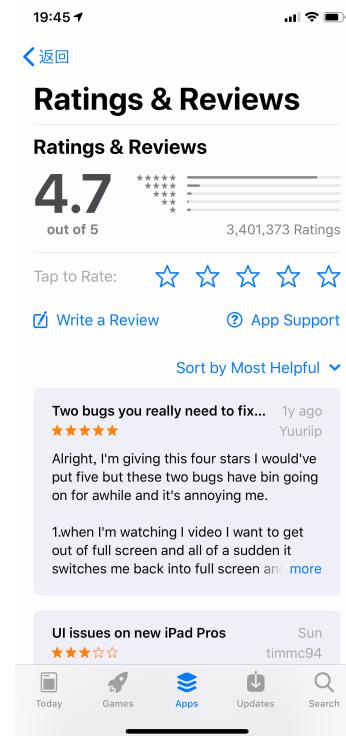
Web Scrapping of reviews:

- For each App: 50 most helpful reviews defined by Apple.
- From iTunes RSS Generator API

```
In [12]: review_apple.head()
```

```
Out[12]:
```

	appid	title	rating	review
0	281796108.0	Come back in - the water is fine	5.0	Having been a long time and premium user of Ev...
1	281796108.0	Customer Service	1.0	If you are looking for a company that has acce...
2	281796108.0	Lost photos I uploaded for recipes	1.0	I used to love Evernote until today. I bragged...
3	281796108.0	No customer service	2.0	So I've been using Evernote for over three yea...
4	281796108.0	It's a great tool	5.0	It's a great tool for various different reason...



Word Cloud for 5 Star Reviews:

Most of them are praises:

- Awesome
 - Amazing
 - Excellent



Word Cloud for 1 Star Reviews:

After we filter words like “Terrible”,
“Horrible”, “Bad”.

We find some major complains:

- not interesting
 - Waste time or money
 - Many ads
 - Bugs, Crashes, Rip off, Scam
 - No sound

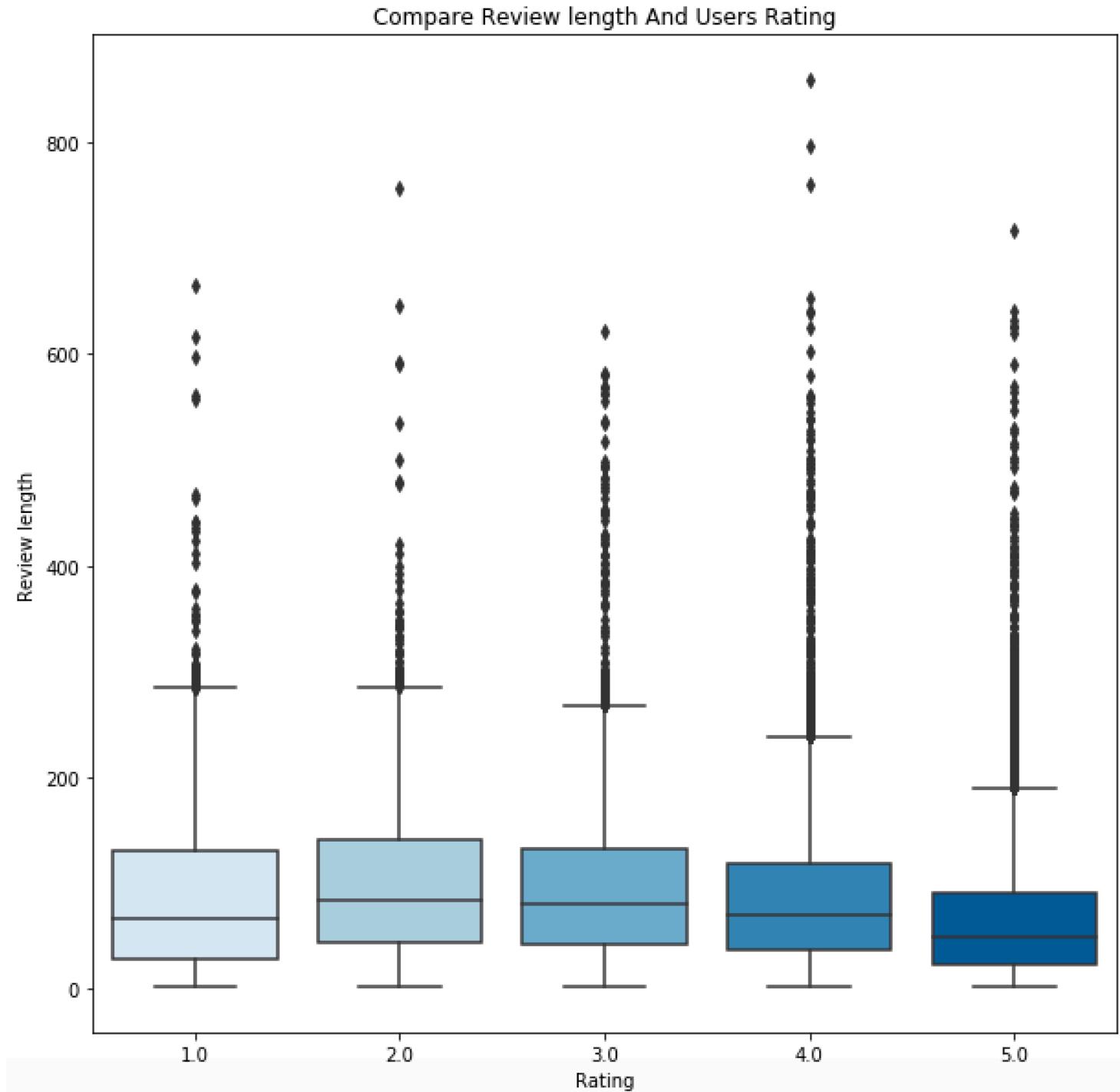
1 star Review's Word Cloud



How many words people prefer to type in each rating group?

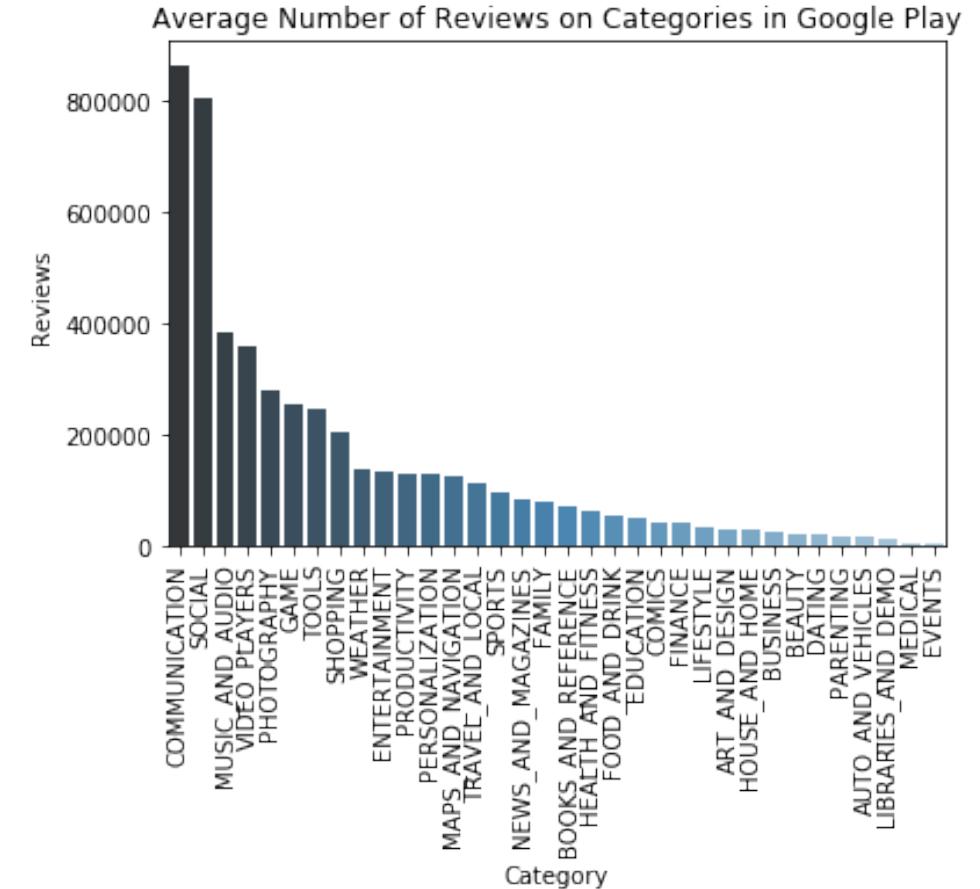
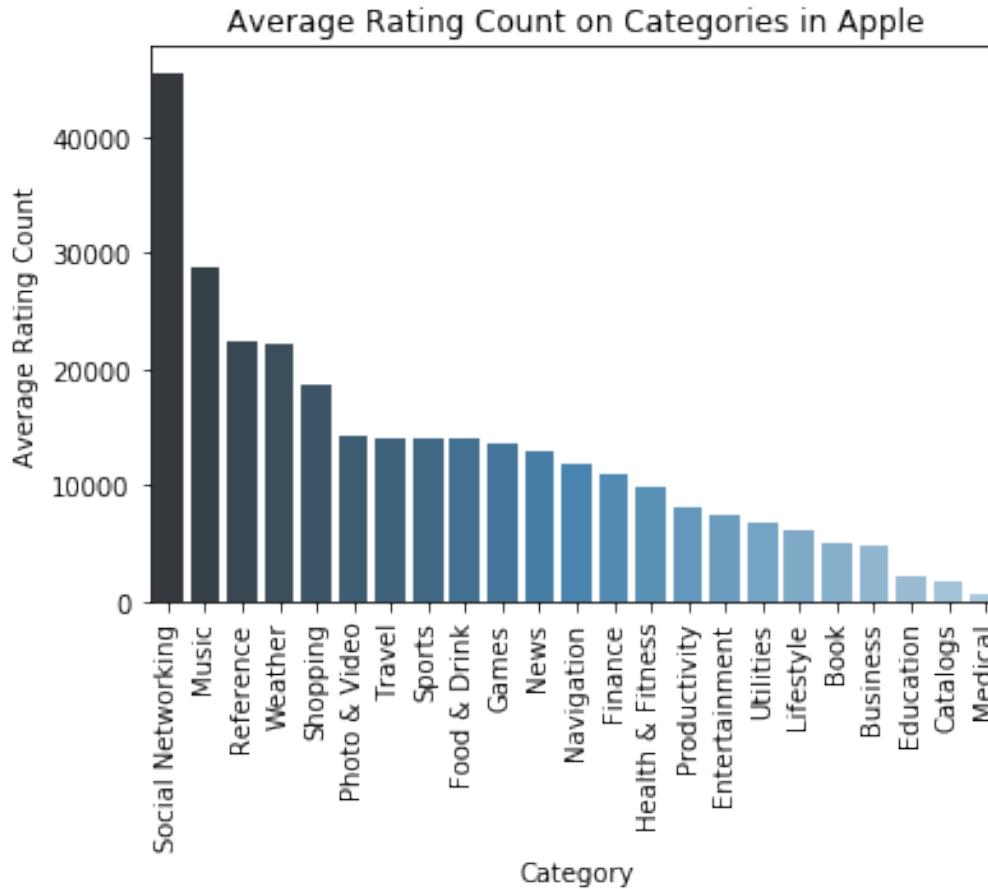
Rating Group	Review length median
1	65
2	83
3	79
4	69
5	48

For those people who give 5 stars, they tend to type less words. We think this is because they are very satisfied and don't have too much suggestions to give. For people who give 1, 2, or 3 stars, their range of upper 50 percent are wider, this means those people tend to give more suggestions in their review.

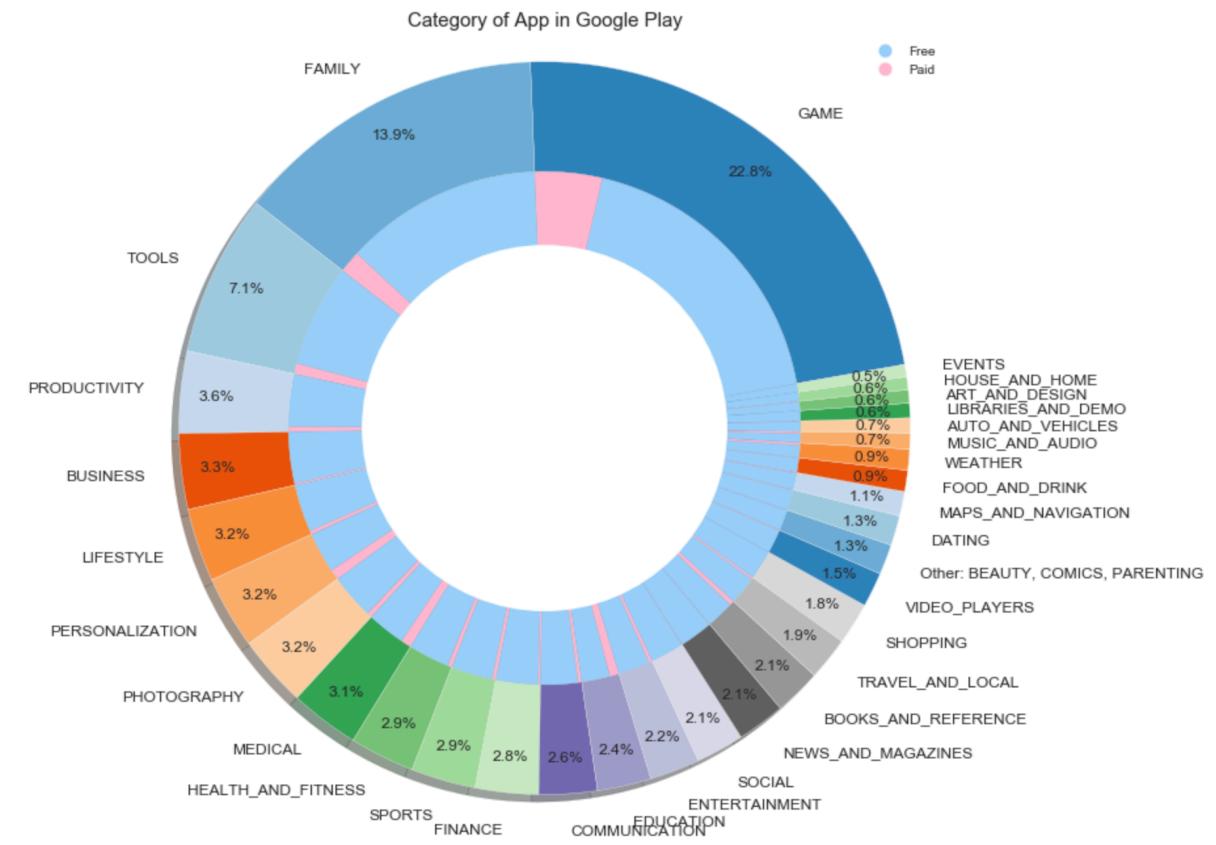
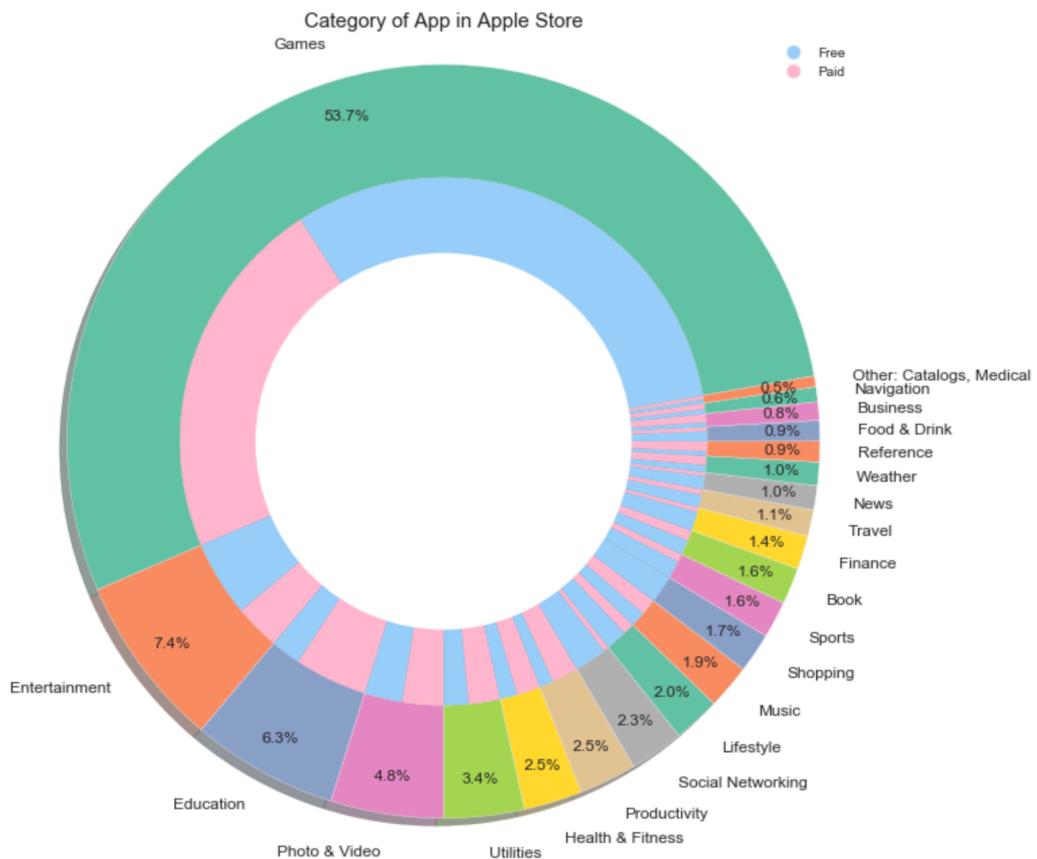


Average Rating/Reviews in both market

- People are more likely to rate/comment on Social and Music Categories for both markets



Category & Price for Apple/Google Play



Price Distribution

- Both stores' interquartile range is 1.99 to 4.99 dollars
- Google play has much higher standard deviation
- Apple has more paid App

Apple Price Distribution

Price	
count	3141.000000
mean	3.955298
std	8.315882
min	0.990000
25%	1.990000
50%	2.990000
75%	4.990000
max	299.990000

Google Play Price Distribution

Price	
count	1637.000000
mean	8.532028
std	40.321031
min	0.990000
25%	1.990000
50%	2.990000
75%	4.990000
max	400.000000



I'm Rich - Trump Edition

Rich Studios XXL Lifestyle

Everyone

⚠ You don't have any devices.

Add to Wishlist

\$400.00 Buy



The most expensive App in this Google Play store dataset

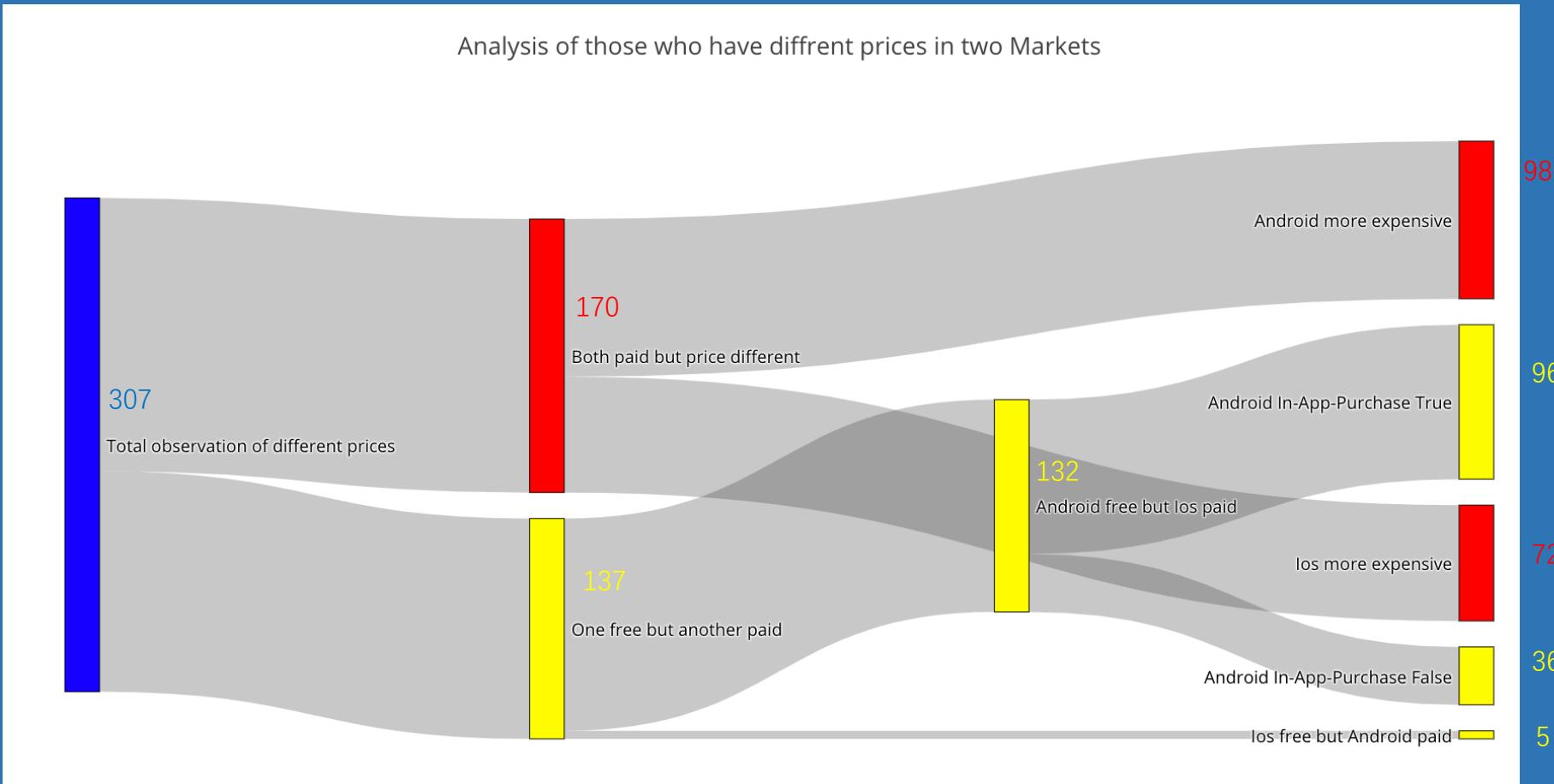
After roughly analyzing Apps in IOS and Android market separately, we are interested in whether the same App performs differently in both market?

Therefore, we scrape more Google play data base on the name in our Apple dataset to find some apps in our apple dataset which also released in Android market, there are 1600 of them. We combine them:

Andro_name	ios_name	Category	Content_Rating	Andro_price	ios_price	Andro_size	ios_size	Andro_version	ios_version	Andro_rating	ios_rating	Andro_type	ios_type	Andro_description	ios_description	Price_Consistency	
0	Ms. PAC-MAN	Ms. PAC-MAN	Games	4+	3.99	3.99	23M	66.779297	2.5.0	4.0.4	4.5	4.0	Paid	Paid	What good is it to be the retro old school kin...	Now with MFi controller support!\n\nMs. PAC-MA...	True
1	Facebook	Facebook	Social Networking	4+	0.00	0.00	Varies with device	371.818359	Varies with device	95.0	4.0	3.5	Free	Free	Keeping up with friends is faster and easier t...	Keeping up with friends is faster than ever. \...	True
2	iQuran	iQuran	Reference	4+	1.99	1.99	29M	67.432323	2.5.4	3.3	5.0	4.5	Paid	Paid	*****\n*****\n*****\n*****\n*****	** iPray for Mac OS X is OUT **\n\n***** ...	True
3	Cleartune - Chromatic Tuner	Cleartune - Chromatic Tuner	Music	4+	3.99	3.99	Varies with device	10.893829	Varies with device	2.1.3	4.5	4.0	Paid	Paid	The most critically acclaimed instrument tuner...	"My most indispensable app is my guitar tuner....	True
4	TouchOSC	TouchOSC	Music	4+	4.99	4.99	255k	4.066406	1.9.10-a2	1.9.8	4.5	4.0	Paid	Paid	TouchOSC is a modular OSC and MIDI control sur...	TouchOSC is a modular OSC and MIDI control sur...	True

Among those apps released in both market, we find an interesting phenomenon: in every 5 of them, there are one have different prices in these two market. Why this happens?

we draw a Sankey diagram for those same app but have different prices in two market:



From this diagram, we can find that there is a huge difference in number when one is free but another is paid. For those in Android free but iOS paid, there are 132 observations, but for those in iOS free but Android paid, there are only 5. we continue to study and find that for those 132 observations, $\frac{3}{4}$ of them have In-App-Purchase. We guess that maybe in android market.,.

Conclusion & Future Plan

- App developers are more willing to set free download of their app at first in Android market than in iOS market to attract customers, and use other methods like in-app-purchase to make profits.
- More exploration on both Markets
- Relationships between Reviews and Rating
- Classify the review into positive/negative based text itself (Sentiment Analysis)

Sentiment Analysis Plan:

	appid	title	rating	review	Title Sentiment	Review Sentiment
0	281796108.0	Come back in - the water is fine	5.0	Having been a long time and premium user of Ev...	0.2023	0.7690
1	281796108.0	Customer Service	1.0	If you are looking for a company that has acce...	0.0000	0.9892
2	281796108.0	Lost photos I uploaded for recipes	1.0	I used to love Evernote until today. I bragged...	-0.3182	-0.9513
3	281796108.0	No customer service	2.0	So I've been using Evernote for over three yea...	-0.2960	0.7014
4	281796108.0	It's a great tool	5.0	It's a great tool for various different reason...	0.6249	0.9223

We try to do the sentiment analysis (close to -1 negative, close to 1 positive) for those review and title of apple apps. It works well for title, but for review..... not good.

We find the problem is that:

It works well for a single meaning, but when it faced with some turning point, it cannot catch the more important part of the sentence:

```
In [171]: sid.polarity_scores('I love it.')['compound']
Out[171]: 0.6369
```

```
In [172]: sid.polarity_scores('I dislike it recently')['compound']
Out[172]: -0.3818
```

```
In [173]: sid.polarity_scores('I love it.However,I dislike it recently')['compound']
Out[173]: 0.3818
```

```
In [174]: sid.polarity_scores('I dislike it.However,I like it recently')['compound']
...
Out[174]: -0.0258
```

Thus, in the remaining time, we plan to analyze more about why some of our sentiment does not match the rating. Whether the user give wrong rating or wrong review? Whether our sentiment analysis have some constraints? Or whether that sentence cannot be decided without context?