Earnin Assessment Test

Shicheng Chen

Questions:

1. Marketing Attribution

- a. What campaign was responsible for each user's finding our app?

2. Low Sales

- It looks like sales have been a bit low the last couple of days of the sales data set. Is this something we should be worried about?

Marketing Attribution Insight and Report

Data Preparation:

- After importing the data file into R, change the column names into same format. For example for different ‘create_on’ column to ‘attribution_created_on’,’device_created_on’ and using the same format on the user id and device id.

- Checking duplication rows and each data set. The only duplication shown is on ‘item’ data set, which is reasonable because same sale id can sell same items.

- Checking the connection between user_id and device_id. There are five users does not have device id, which means they did not use phone to sign on campaign, so drop of those five users.

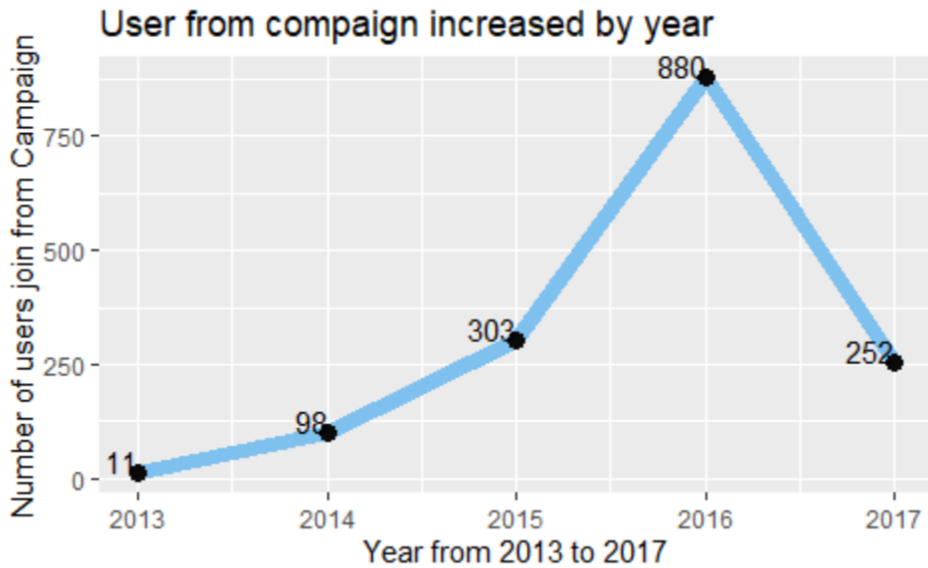
- Deleting the rows in attribution data set which not shown in the user_with_device data set, this part erases the people whom only have record on attribution and never become a user.

- Merging the data sets attribution and user_with_device, which shows if each user row are created after an attribution date, and if the date is less than the average register time on each campaign, and mark True on these rows, the rest mark Fasle.

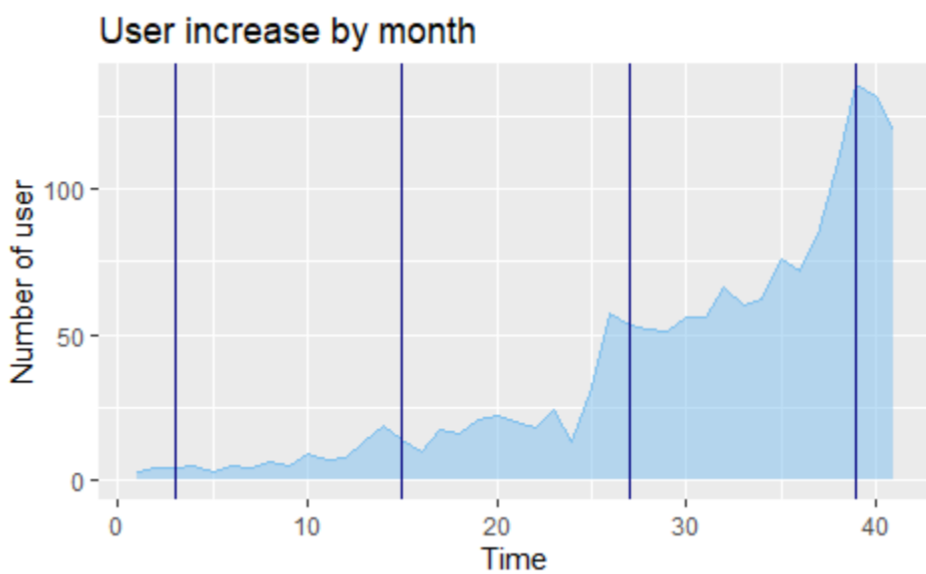
- Drop the 344 not valid rows, which became user too long after the campaign.

Report:

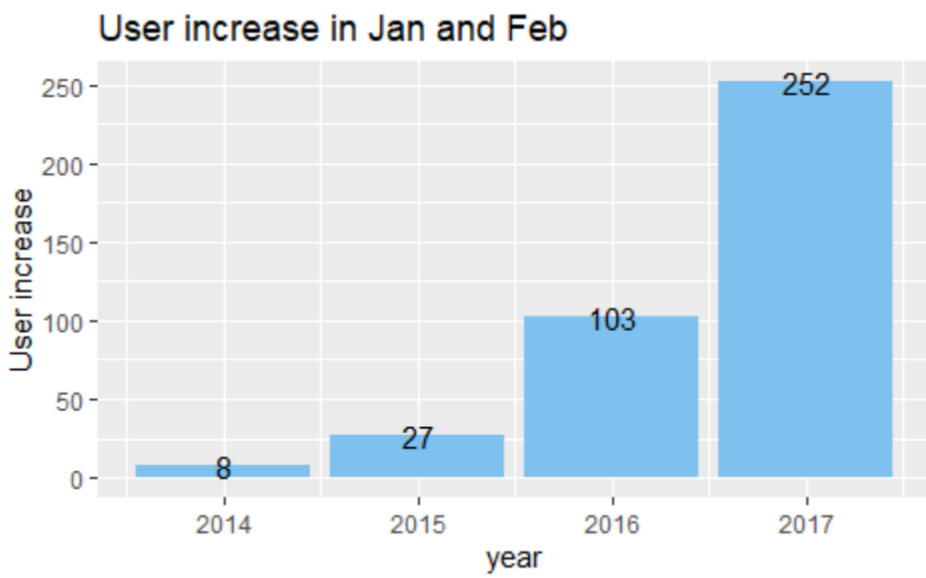
- From the plot below, we can see that the registration of user is increasing year by year, and the drop of 2017 because we only have two months' data in year of 2017.



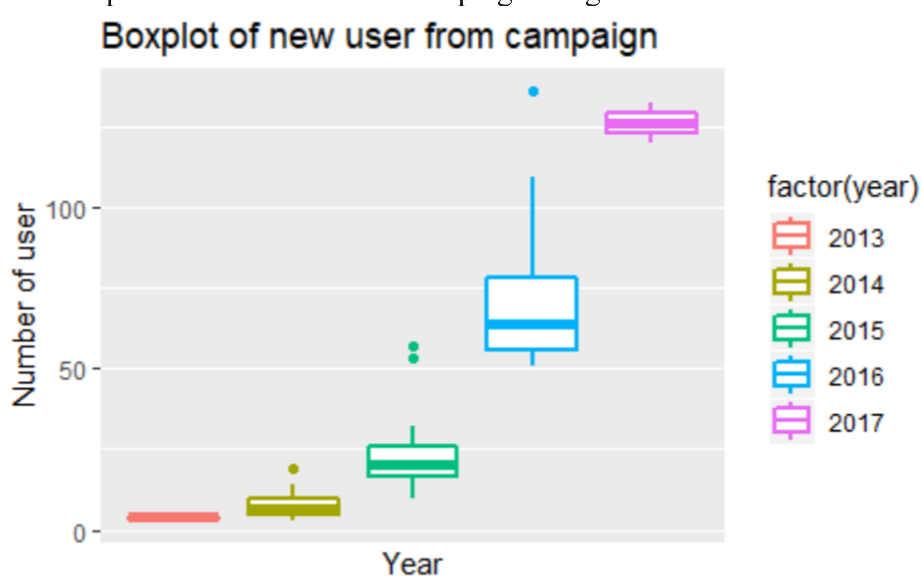
- The plot below shows the monthly register by the time from Octo 2013 to Feb 2017. The blue vertical line shows the January for each year, and we can see the increasing slope is higher at in the last quarter of each year and at the first quarter of each year it drop down a little.



- From the bar-plot below, we can see if we choose January and February's user registration number, we can see 2017's registration is the highest.



- The boxplot below shows each campaign's registration number is increasing by years.



- Overall there are 4031 records in attribution, and we have 1544 users register from the campaign, the registration rate is 38.3%

- I identify two main variables on the influence of the campaign, the number of users register through the campaign and the rate of becoming user after the campaign. I offered equal weight on those two with the score from 50 to 0.5, decreasing by 0.5. Finally, these are the top 10 most influence campaign I got:

```
6XC4I2 9VTGJV QSY7GR JFJFGZ QY02HA
Y59GNL FE91LN D8IC8B CPIAHO 5IHLGC
```

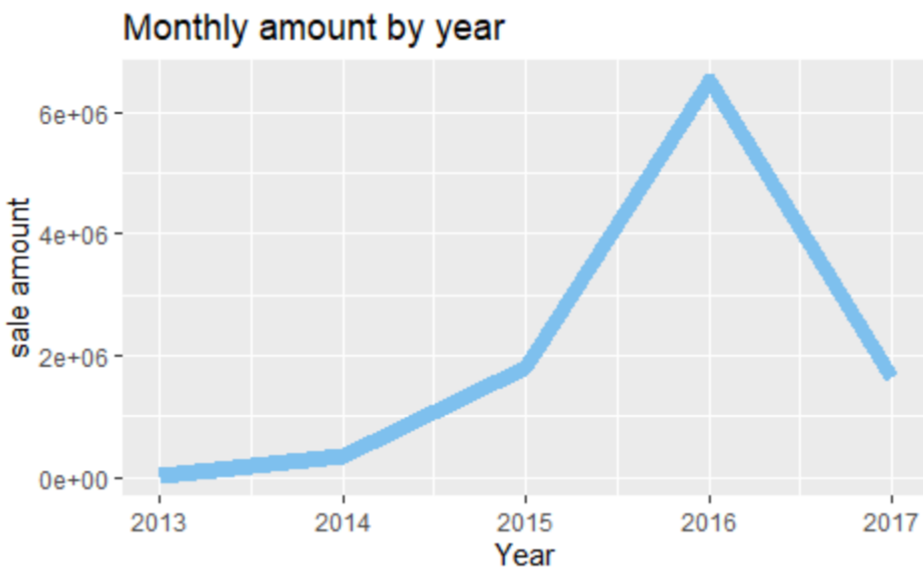
Low Sales Insight and Report

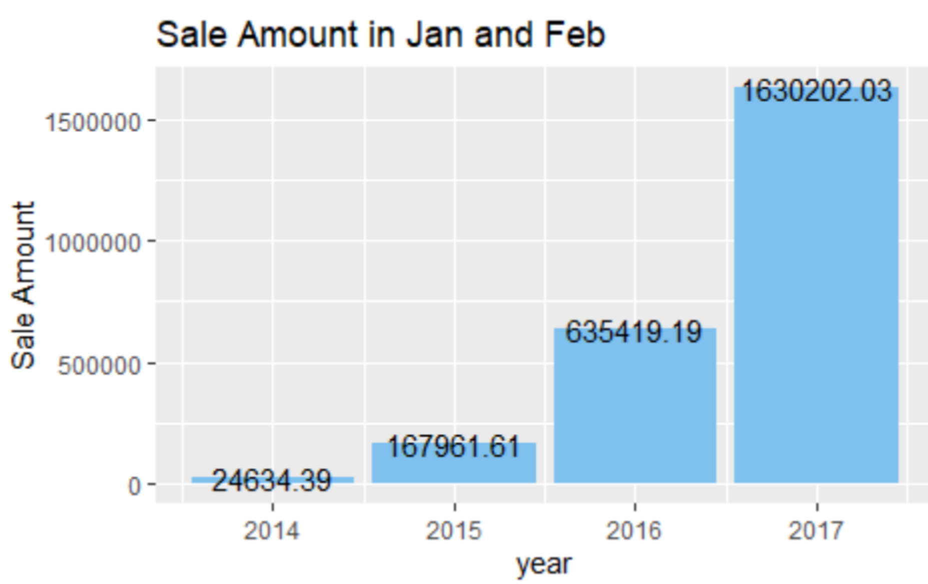
Data Preparation

- Changing weekday column from ‘0 to 6’ to ‘Sunday to Monday’.
- Adding ‘device_type’ column into sales data from the user ID use the phone for this app, and there are several user id whom did the sale but not have a phone.
- Adding ‘Validate’ column to identify if the sale from the user whom register after a valid campaign.
- Taking out the sale amount and date from data set ‘sale’, and change those two columns into a time series data set.

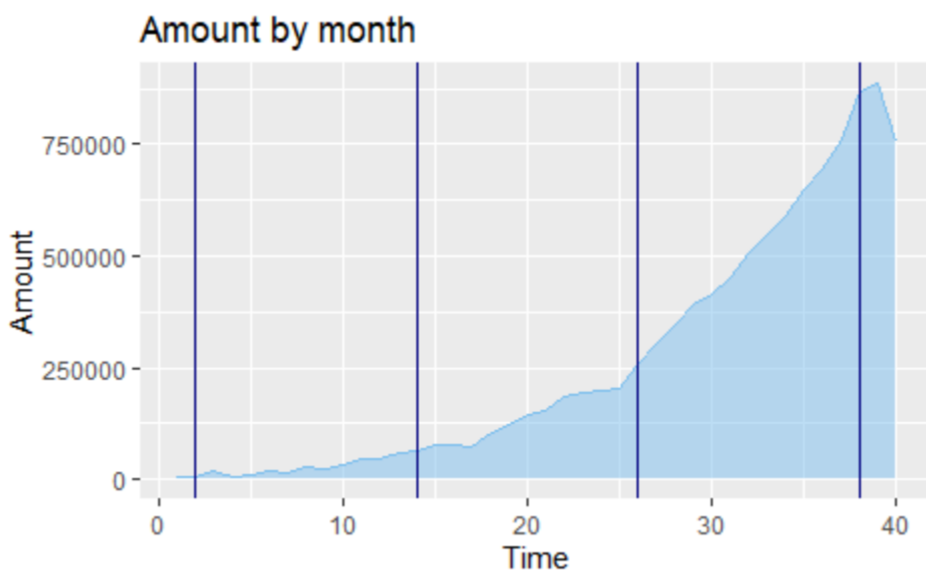
Report:

- From the two plots below, we can see there is a drop in the year of 2017 but we only have two month sale data in 2017. If we collect the month January and February out, we can see the amount of selling in 2017 has an exclusive increasing than the year before on the same time period.

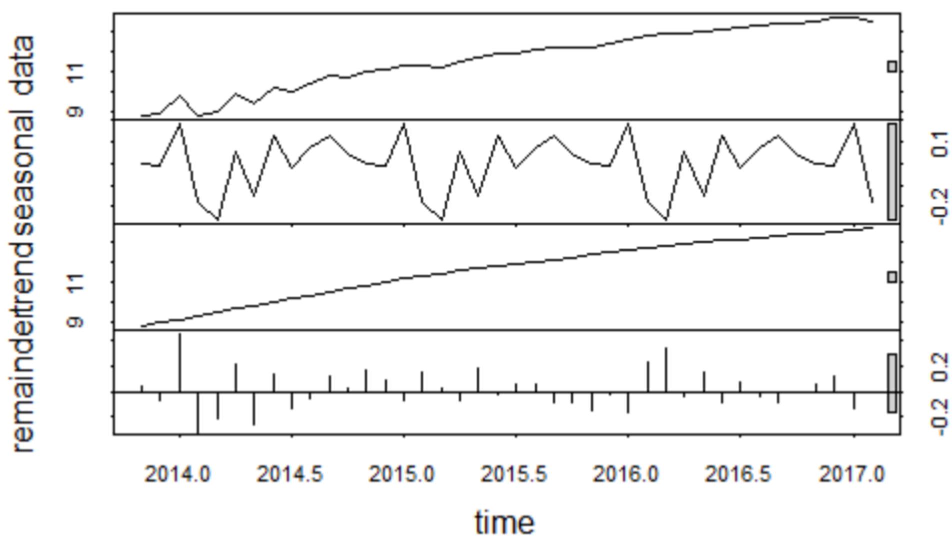




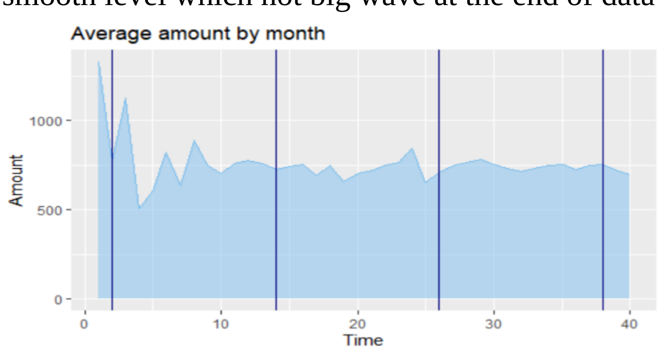
- For the plot below shown the sale amount trend through time, and we can see the sale amount has wave in the middle but overall it is increasing. As the question asked, at the end of days, the sale amount is decreasing, it might be cause by seasonal.



- I put the data as time series check the seasonal and trend. From the plot below, it shows the seasonal is very clear, and out end-time (January and February) are at the decreasing slope of the seasonal. From the trend, we can see that it continue increasing by time.



From the plot below we can see the average sale amount by each user is staying in a smooth level which not big wave at the end of data set.

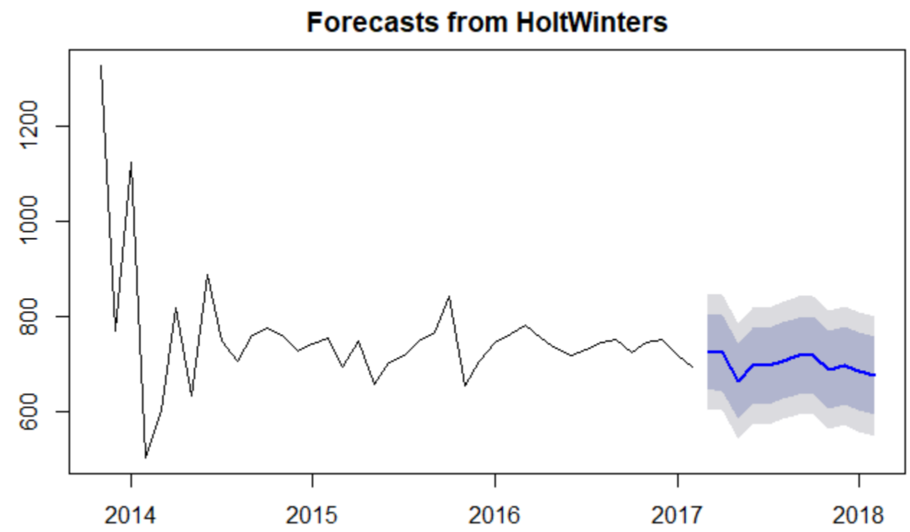


- Also I used hypothesis test to get that we have confidence to say the average sale in Feb 2017 has not difference than the time before.

```
        Welch Two Sample t-test

data:  history_amount and test_amount
t = 2.1891, df = 15.851, p-value = 0.04391
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
   1.616726 103.154853
sample estimates:
mean of x mean of y
 759.3058  706.9200
```

- Because of the seasonal is obvious in the above plot, so I use Holt-Winter method to forecast the next year average user sale amount. From the plot we can see the average sale amount by user stays in an even level which labels each user will use our app in an stable amount.

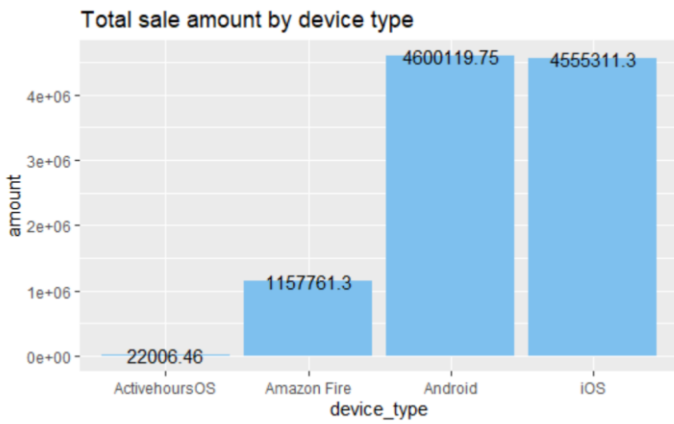


Conclusion:

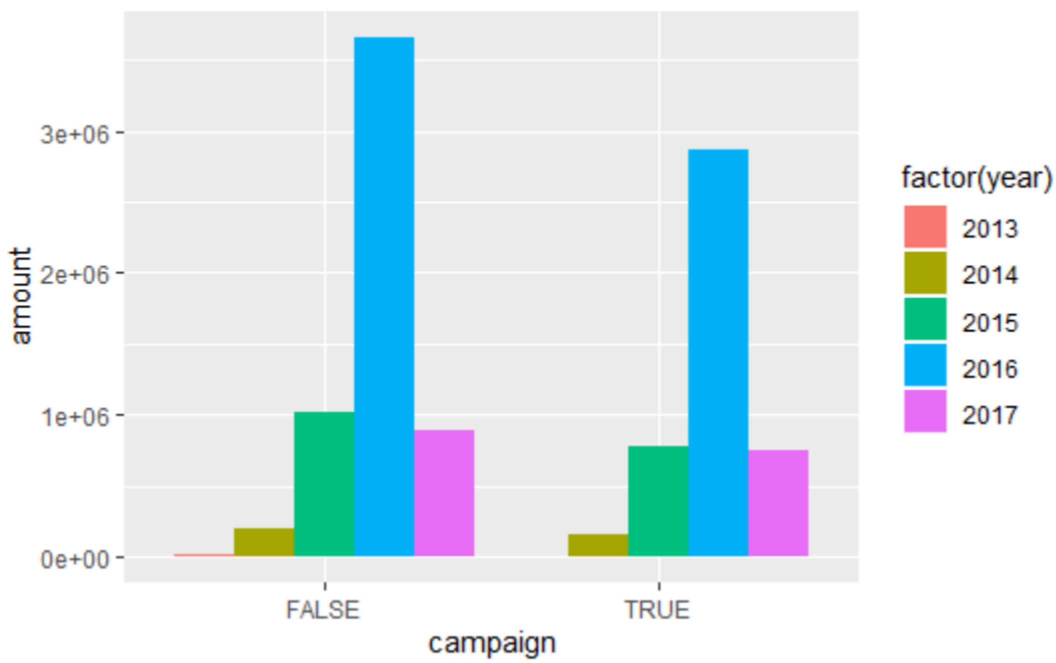
We do not need to worry about the sale amount is decreasing at the end of the data set. From the time series plot, this is the seasonal drop at the beginning of the year. Comparing to the previous year, it is the new highest sale amount of these two months than past years. The average sale amount does not have a deep change, which means our active user stays the same selling level as before. There is one thing we need to worry about is the number of the new users join, because our average user amount is stable, the only way we can keep increasing is having more users on our app. Which I will recommend doing more attribution on the campaigns which has high register rate and find a create activities to make people whom know our app not through campaign.

Further Thoughts

- To identify which variable are most influence the sale amount, I put weekday sand device type as my predictors and using linear regression to get the model. From the summary I can see that on Mondays and Thursday people love to do sale on out app, and for people who use phones are less than the people who use ActivehoursOS system. But because of the user number of using IOS and Android, the total sale amount is belong to Android and iOS.



- From the plot below we can see the people become users through campaign (True) has less sale amount compare to people who becomes user on other method. We might thing what other method from people know this app, or make a inviting bonus on people invite other users joining in to this app.



- From the table below we can see most people are Android and iOS user, and only 27 people are ActiveHoursOs user, but the ActivehoursOs user has the highest sale amount on average. We might think how to make people under this device type to join into our app.

```
      ActivehoursOS Amazon Fire Android  iOS
 FALSE             10         940    3367 3500
 TRUE              17         507    2941 2725
```