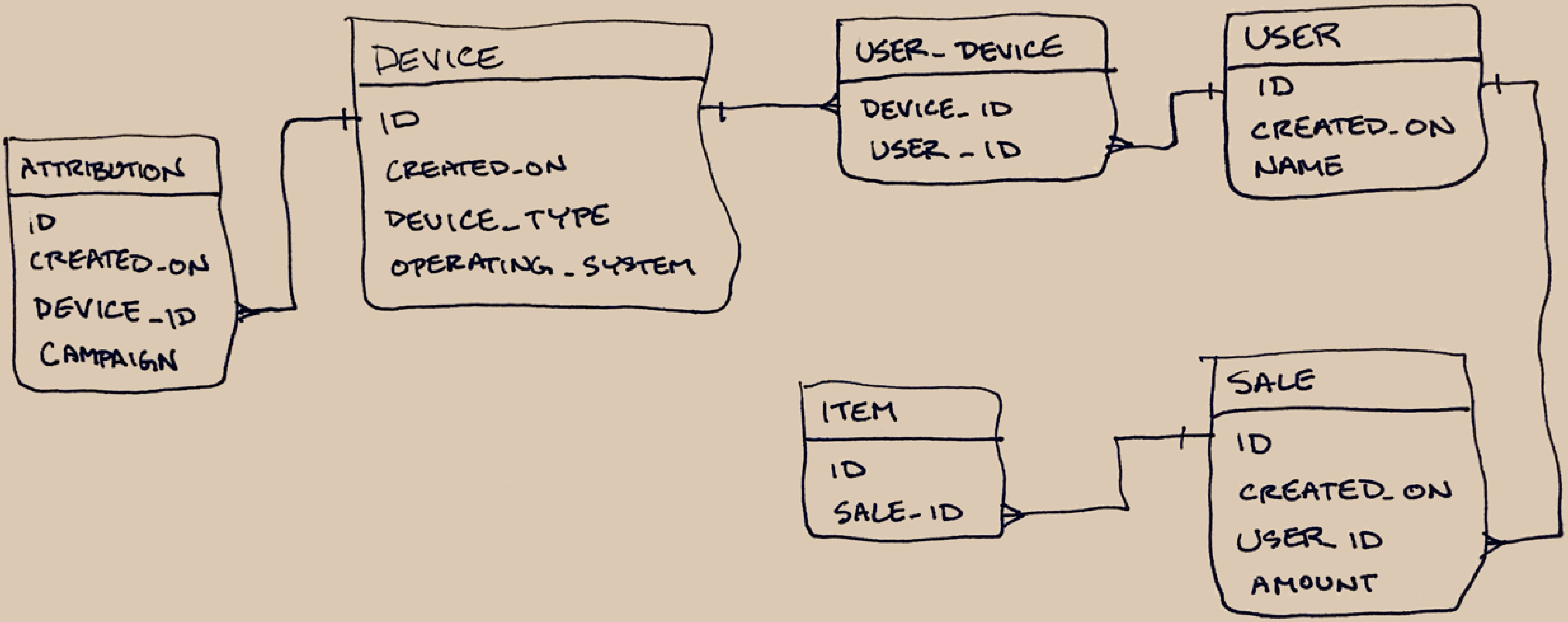
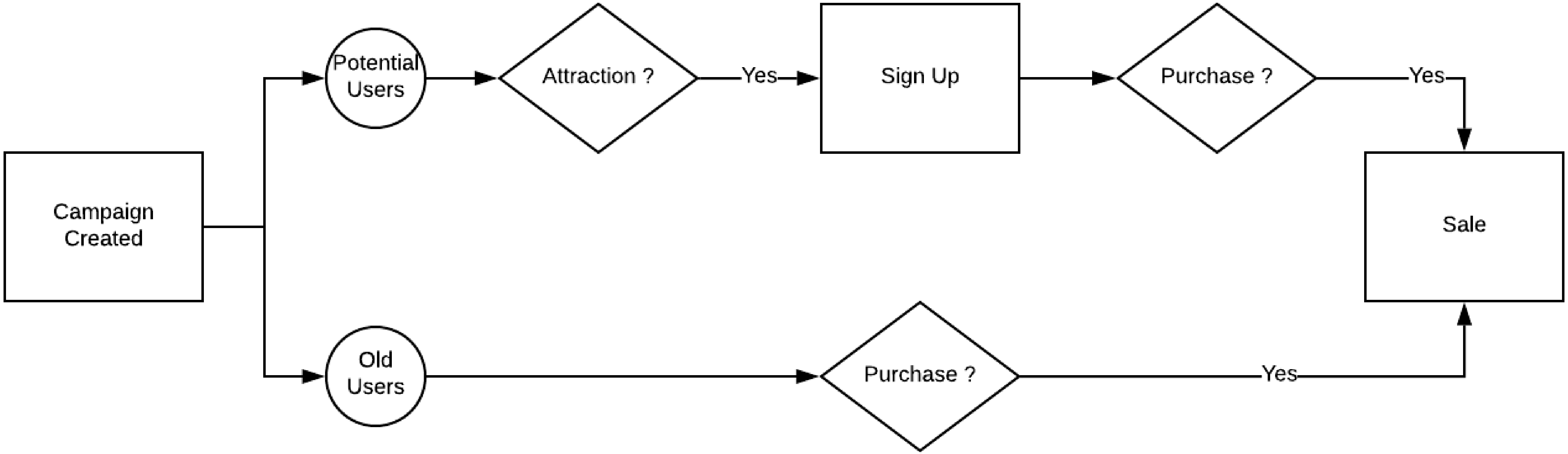
Earnin Take Home Test Presentation

Chicheng Zhang

Tables Relationship



How campaigns contribute to sale ?



Note: Sale is not only contributed by campaign’s conversion.

Question 1: What campaign was responsible for each user's finding our app?

Main Step:

1. Brute join USER, USER_DEVICE, ATTRIBUTION -> user_activity.
2. Apply filter, get valid new users activity records.
3. Count group by campaign (TOP 5)

Note:

- How to design filter for new users enrolled by campaigns.
- New users activities list in time series, should choose the earliest campaign for each user, while there're duplicated user_id in USER table, use created_on & name to identify.
- TOP 5: df_after_filter.groupby(['user_created_on', 'name']).apply(

  lambda x: x.sort_values('attribution_created_on').head(1)).campaign.value_counts().head()

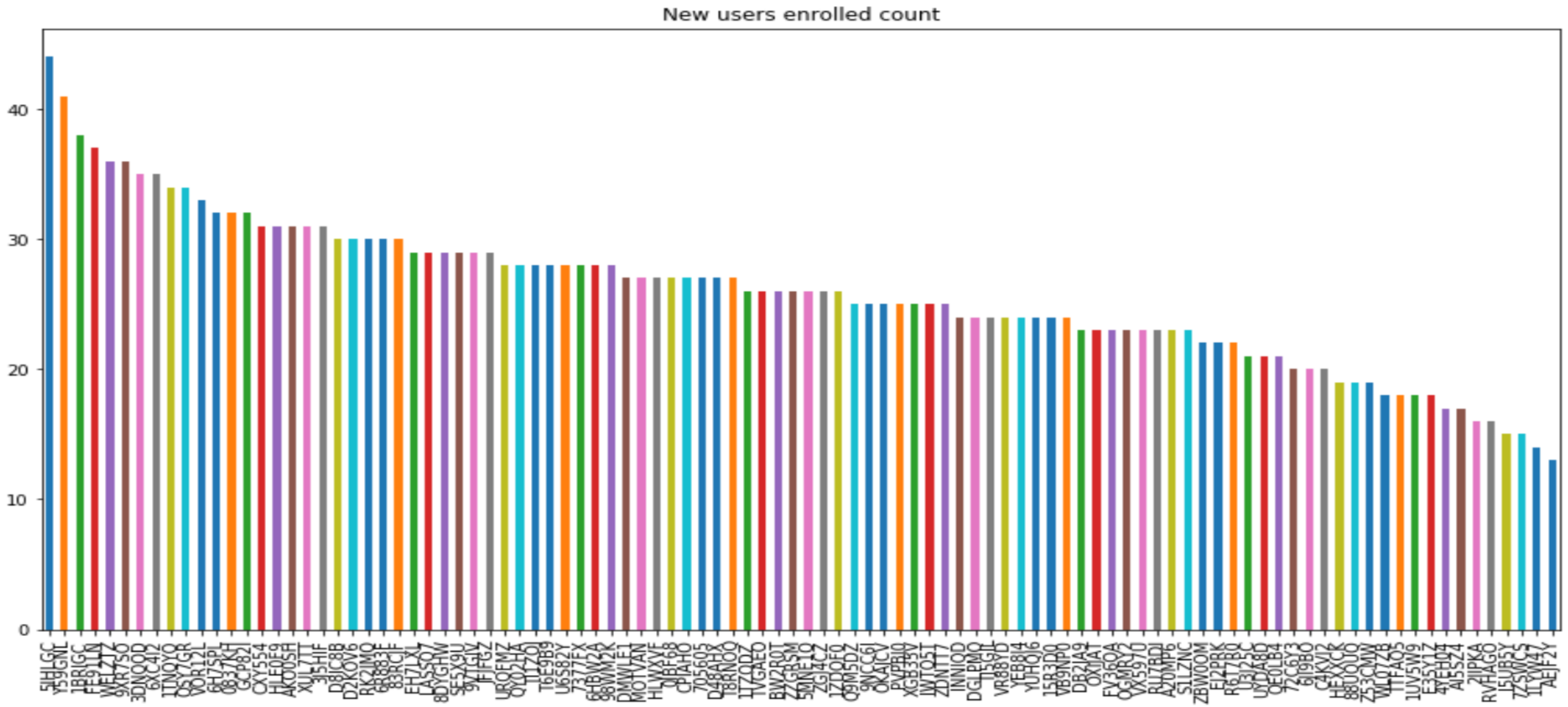
Simple Method

Filter: user_activity['user_created_on'] > user_activity['attribute_created_on']

Logic: Find all records that user created after than this user's interaction with campaign.

TOP 5 Campaigns

5IHLGC	44
Y59GNL	41
1BRIGC	38
FE91LN	37
WELZTZ	36



Assumption

- Workflow: campaign id created by merchant -> campaign info set to users -> device id created by users clicks -> user id created by users sign up.

- Since we are looking into the new users enrolled by campaigns, [attribution_created_on, device_created_on, user_created_on] happends in order and within a specific time period.

- For each stage, the available time period is 7 days.

Complex method

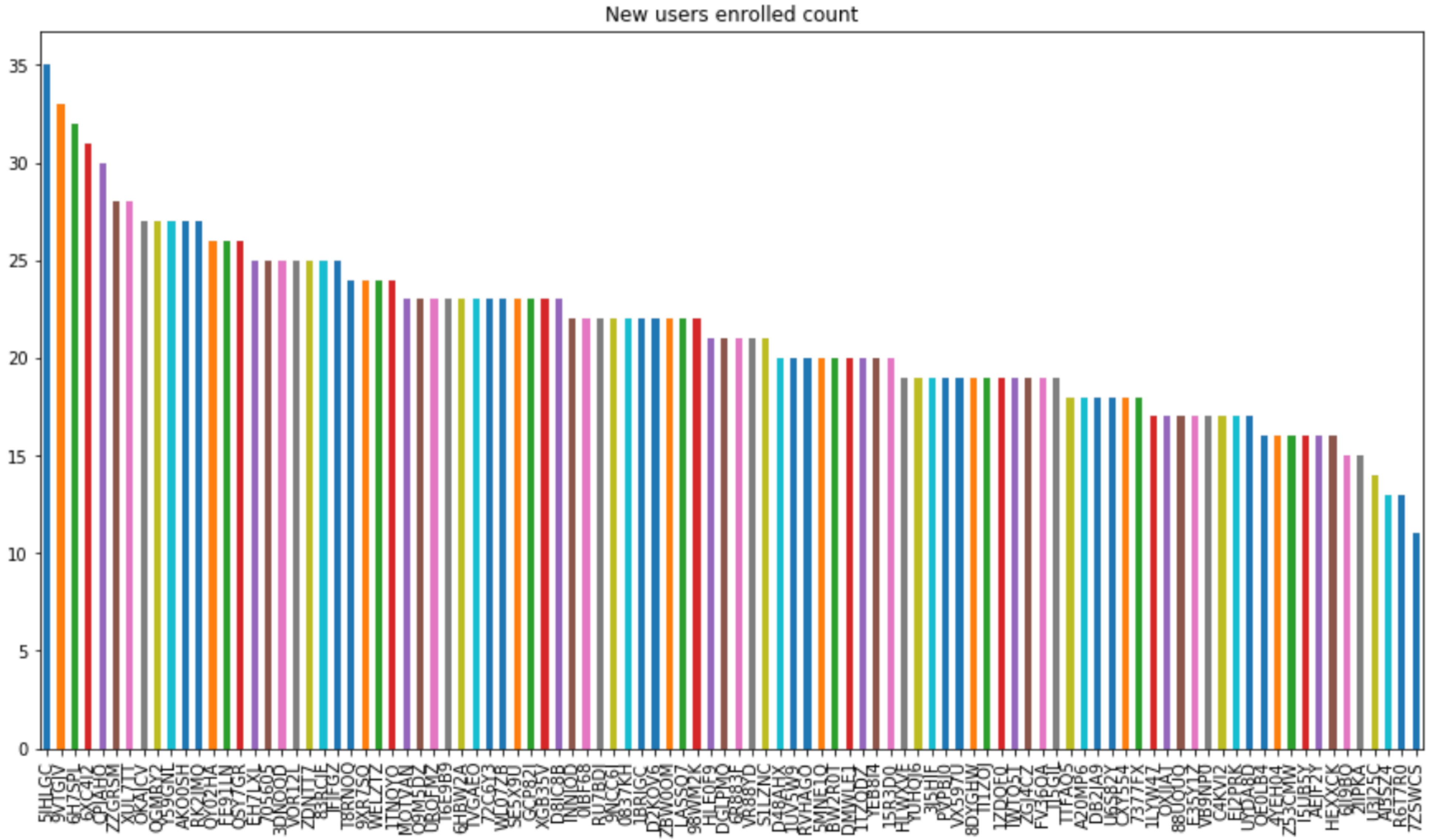
- User Device Filter: (user_activity['user_created_on’] > user_activity[‘device_created_on’]) & (user_activity['user_created_on'] - user_activity['device_created_on']) < timedelta(7))

- Device Campaign Filter: ((user_activity['device_created_on’] > user_activity['attribution_created_on’])) & ((user_activity['device_created_on'] - user_activity['attribution_created_on']) < timedelta(7))

Logic: [attribution_created_on, device_created_on, user_created_on] appends in order, also each stage has 7 days limit.

TOP 5

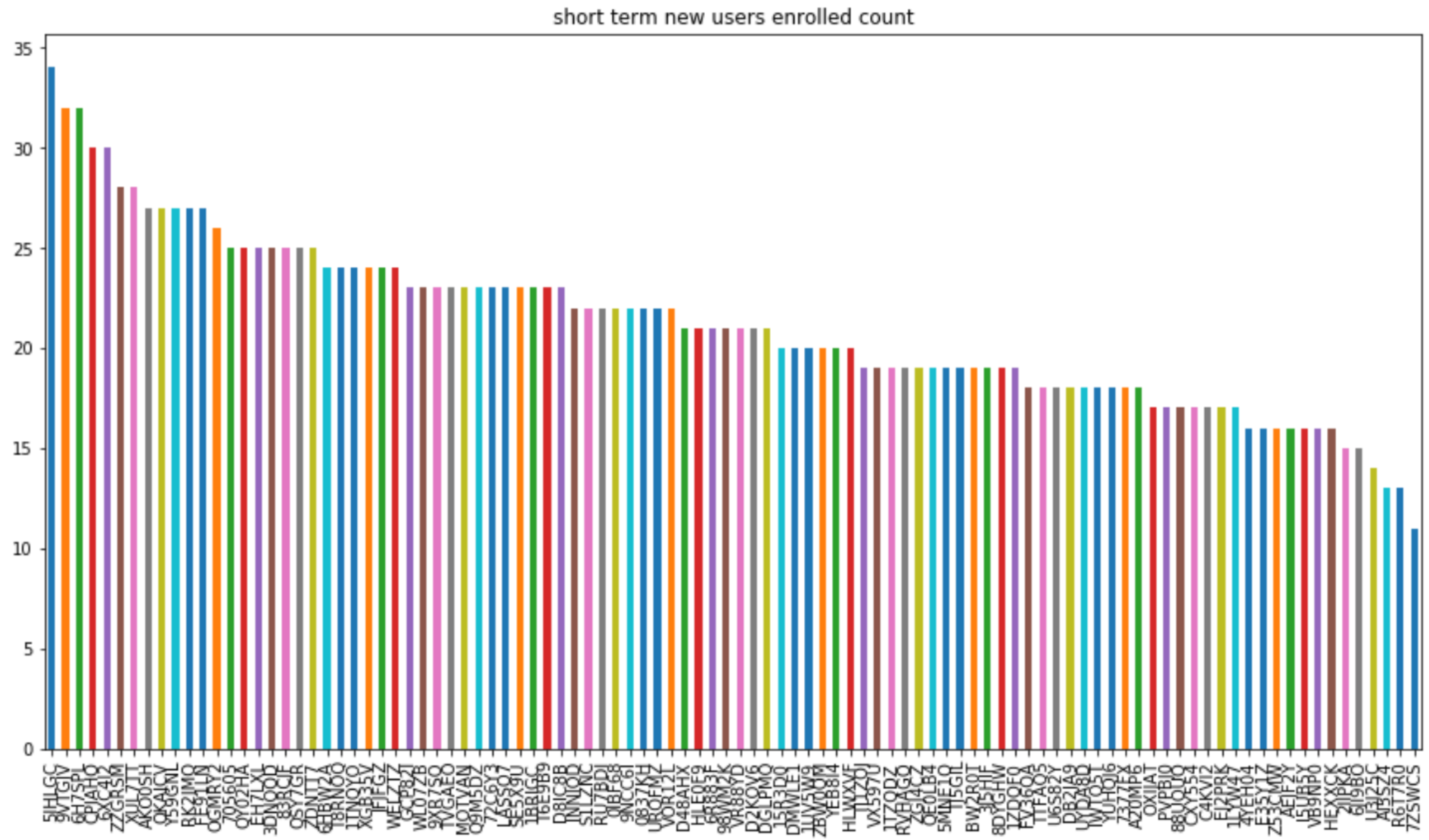
5IHLGC	35
9VTGJV	33
6H7SPL	32
6XC4I2	31
CPIAHO	30



- Try time period as 1 day(short term)

TOP 5

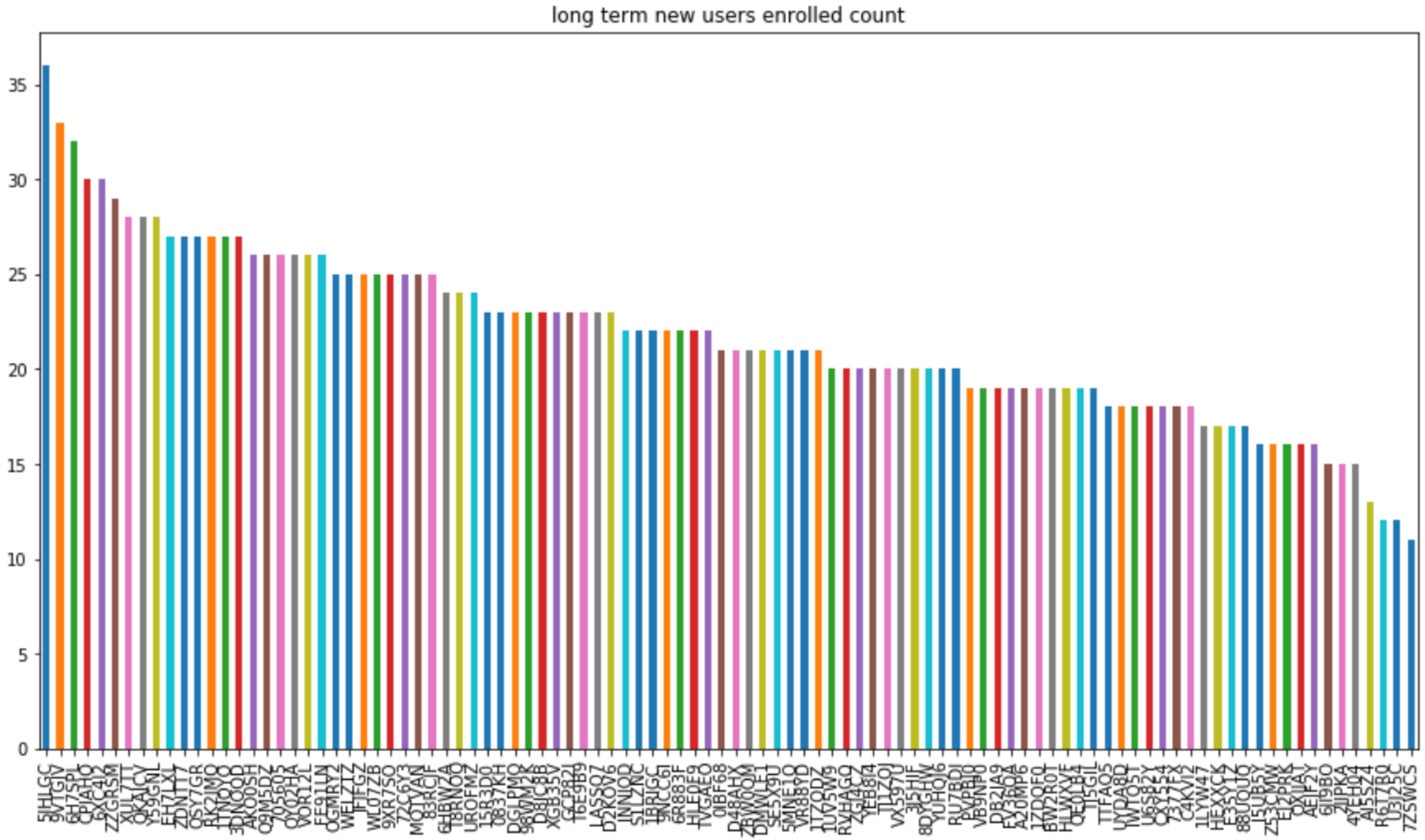
5IHLGC	34
9VTGJV	32
6H7SPL	32
CPIAHO	30
6XC4I2	30



- Try time period as 30 day(long term)

TOP 5

5IHLGC	36
9VTGJV	33
6H7SPL	32
CPIAHO	30
6XC4I2	30



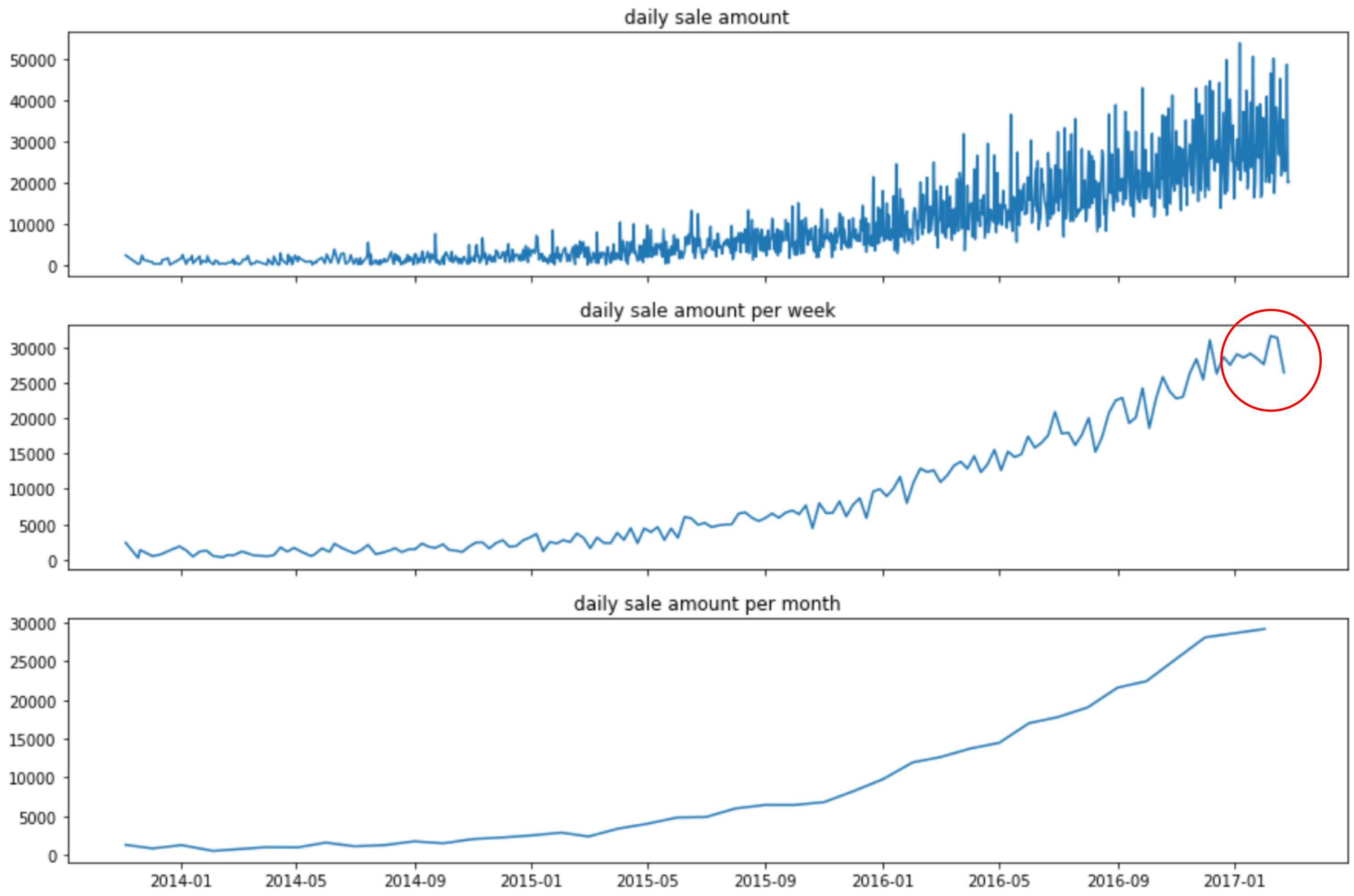
Discussion

- How to set the available time period?
  - Short term might lose some user_activitiey records that the potential users sign on delay.
  - Long term might incorrectly include user_activitiey records that the potential users attracted by another campaign or directly sign on without campaigns.
- Campaign Clustering

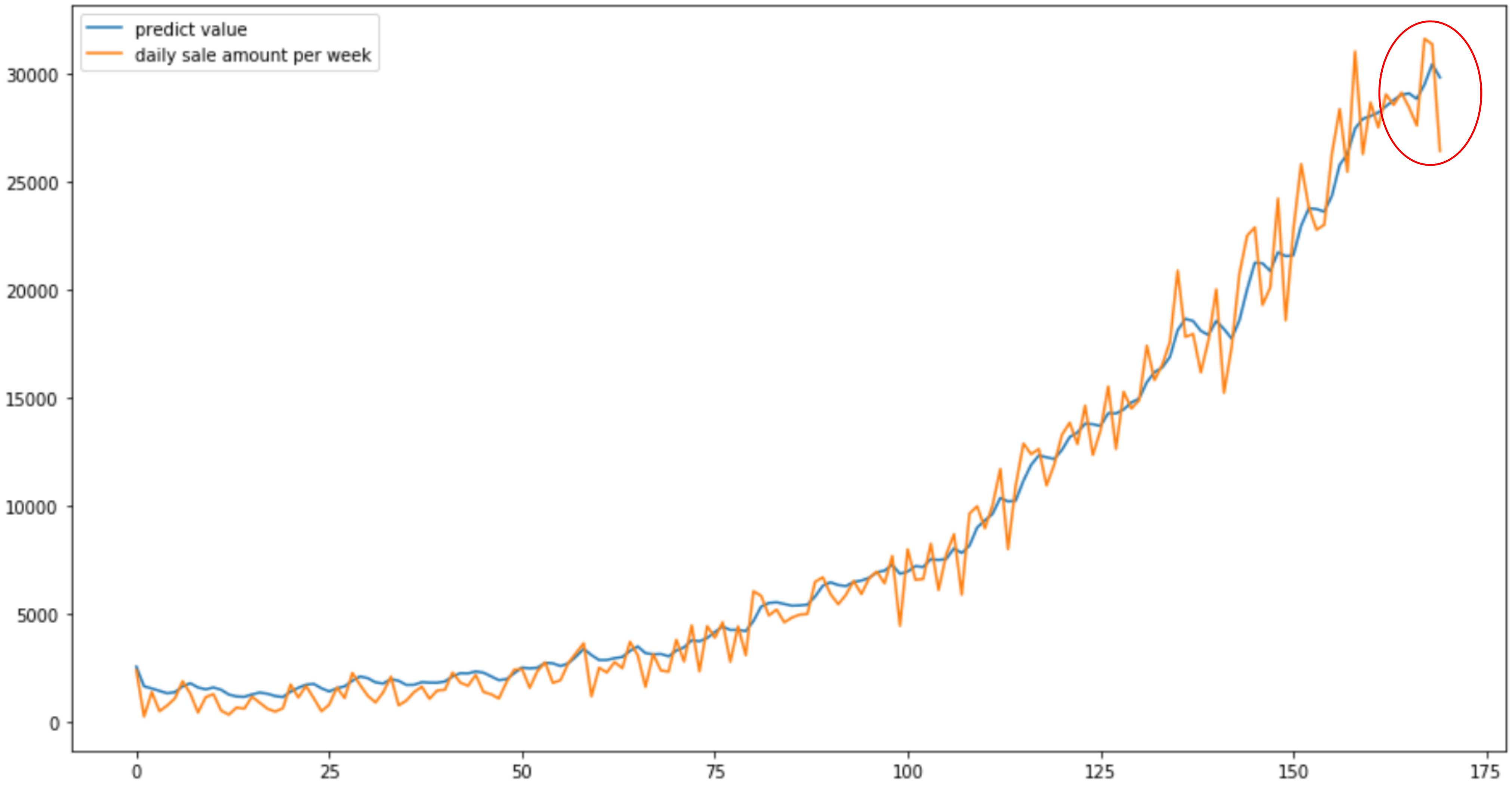
Question 2: Check out recent sale drop

- First step: sum up sale amount group by date -> total sale

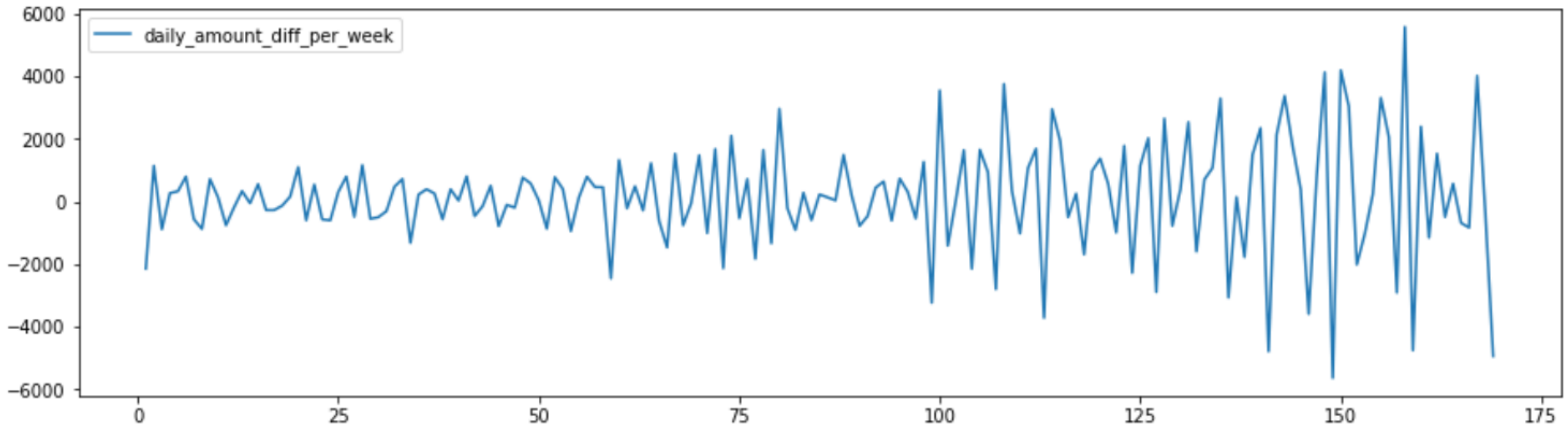
- Second step: calculate daily average sale amount group by week & month

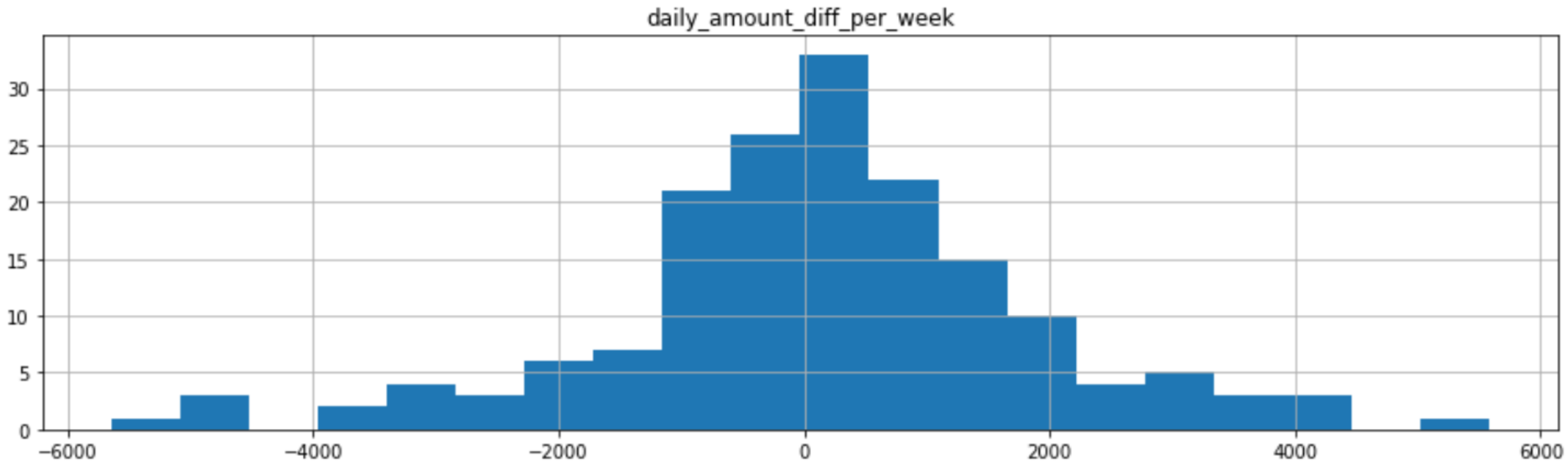


According to the weekly & monthly plots, we can easily find there's drop recently. Here we use weekly sale amount data to build ARIMA model to check if the drop is expected.



How to quantitatively measure this drop?





Use Normal distribution to fit the difference of daily sale amount per week.

Normal(mu, sigma)
where mu = mean(diff),
Sigma = std(diff)

95% confidence interval ≈ [mu± 2 × sigma]

Result:

- The 95% confidence interval is [-3348.190508525964, 3632.8102143132674], while last drop value is -4942.9995453531665.
- Comparing to previous week, the last week's daily sale amount dropped about 15.76%.

Drop Detection

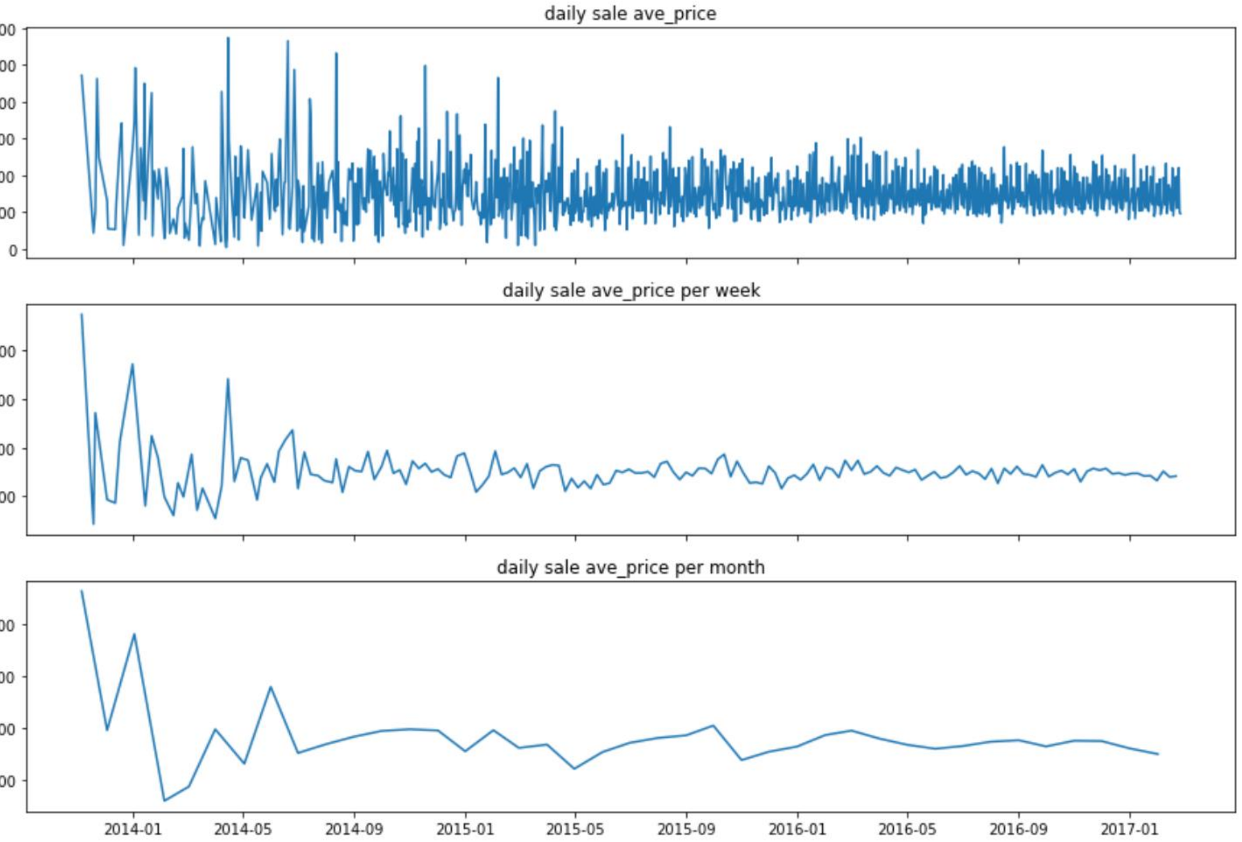
- Industry
- Sale Amount location distribution
- Special Event

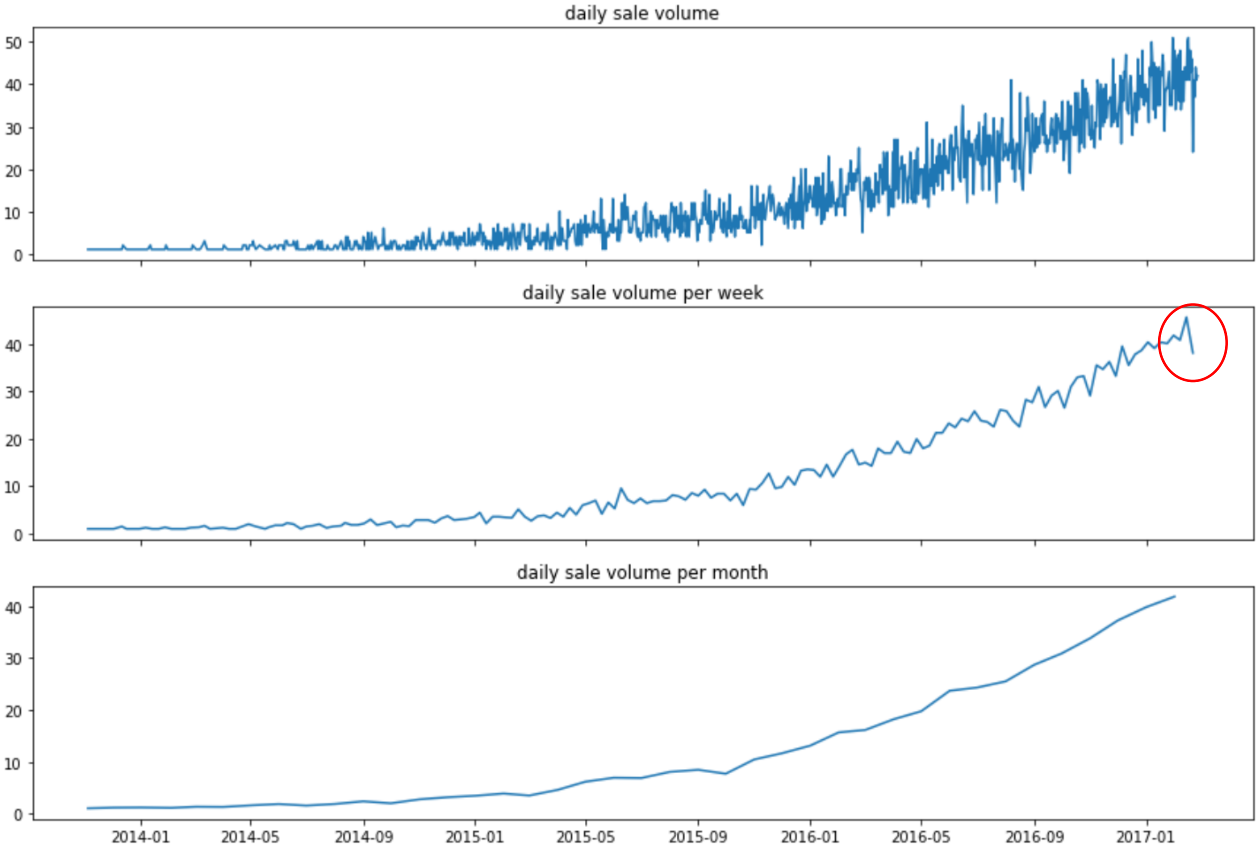
Given by limited information

- Sale Amount = Average Price * Sale Volume
- Group by week & month average respectively for Average Price & Sale Volume.

Average Price

Sale Volume

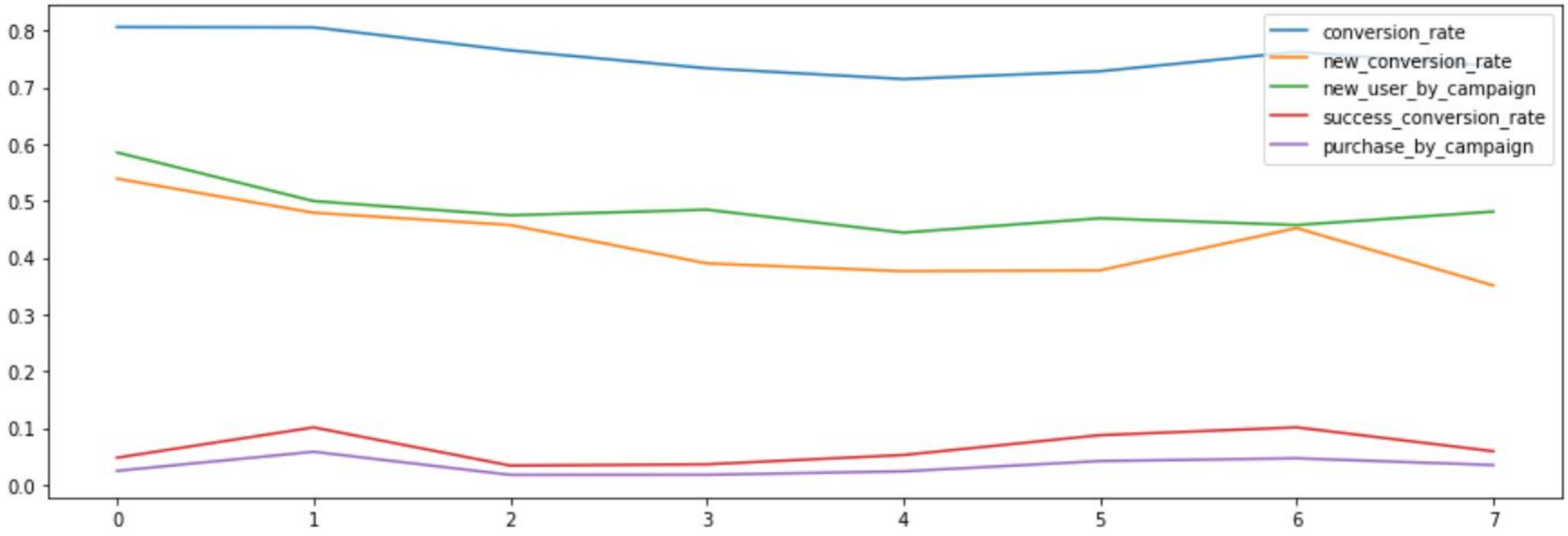




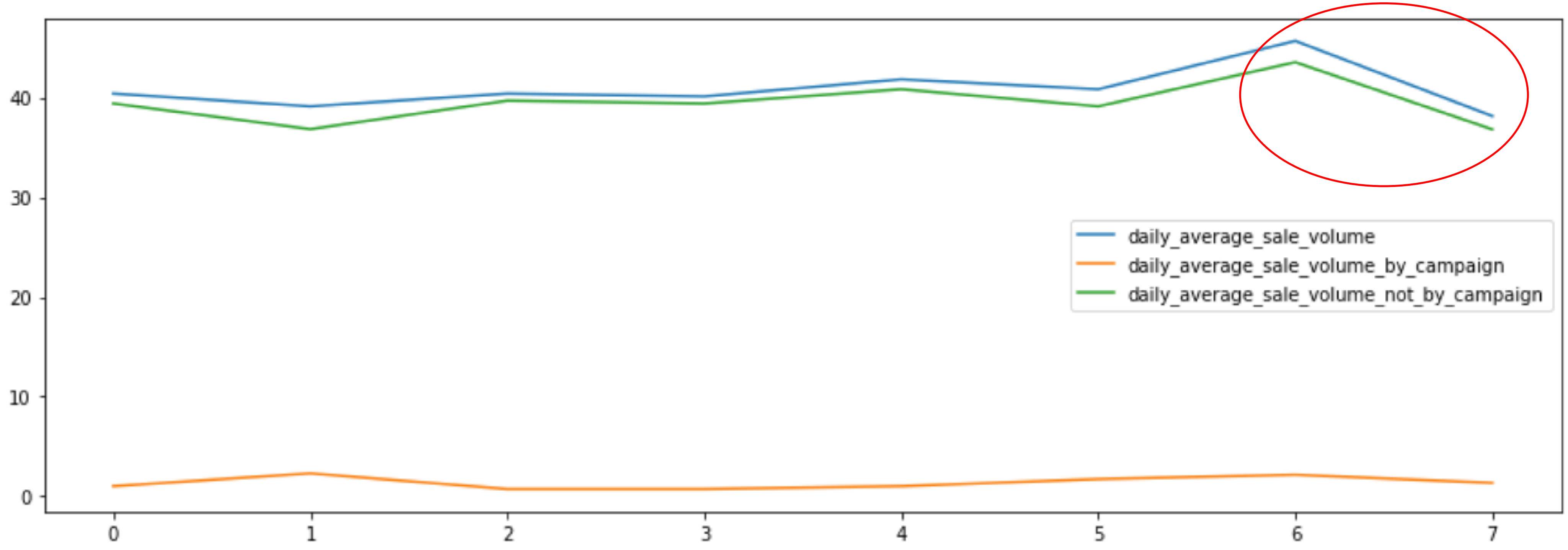
Some Metrics

- Conversion Rate: # campaign_user_id / # campaign_device_id
- New Conversion Rate: # new user enrolled by campaign / # clicks by device
- Success Conversion Rate: # sale by campaign / # campaign_user_id
- new_user_by_campaign: # new user enrolled by campaign / # total new user enrolled
- purchase_by_campaign: # sale volume by campaign / # total sale volume

	conversion_rate	new_conversion_rate	new_user_by_campaign	success_conversion_rate	purchase_by_campaign
0	0.806630	0.539474	0.585714	0.047945	0.024735
1	0.806122	0.479452	0.500000	0.101266	0.058394
2	0.765625	0.457831	0.475000	0.034014	0.017668
3	0.734043	0.390244	0.484848	0.036232	0.017794
4	0.715054	0.376471	0.444444	0.052632	0.023891
5	0.728723	0.378049	0.469697	0.087591	0.041958
6	0.762887	0.452381	0.457831	0.101351	0.046875
7	0.737705	0.351351	0.481481	0.059259	0.034934



	daily_average_sale_volume	daily_average_sale_volume_by_campaign	daily_average_sale_volume_not_by_campaign
0	40.428571	1.000000	39.428571
1	39.142857	2.285714	36.857143
2	40.428571	0.714286	39.714286
3	40.142857	0.714286	39.428571
4	41.857143	1.000000	40.857143
5	40.857143	1.714286	39.142857
6	45.714286	2.142857	43.571429
7	38.166667	1.333333	36.833333



Comparing to previous week, the last week's daily sale volume dropped about 16.51%.

Conclusion

- According to the analysis above, the root reason should be the drop of sale volume, especially sale volume not created by campaigns - churn rate.
- Campaigns play important role in new user enrollment (50%), while not in purchase.
- TODO: break down the sale volume into different items, find out which item's sale dropped most.
- Advice: price strategy, ads strategy, targeting strategy