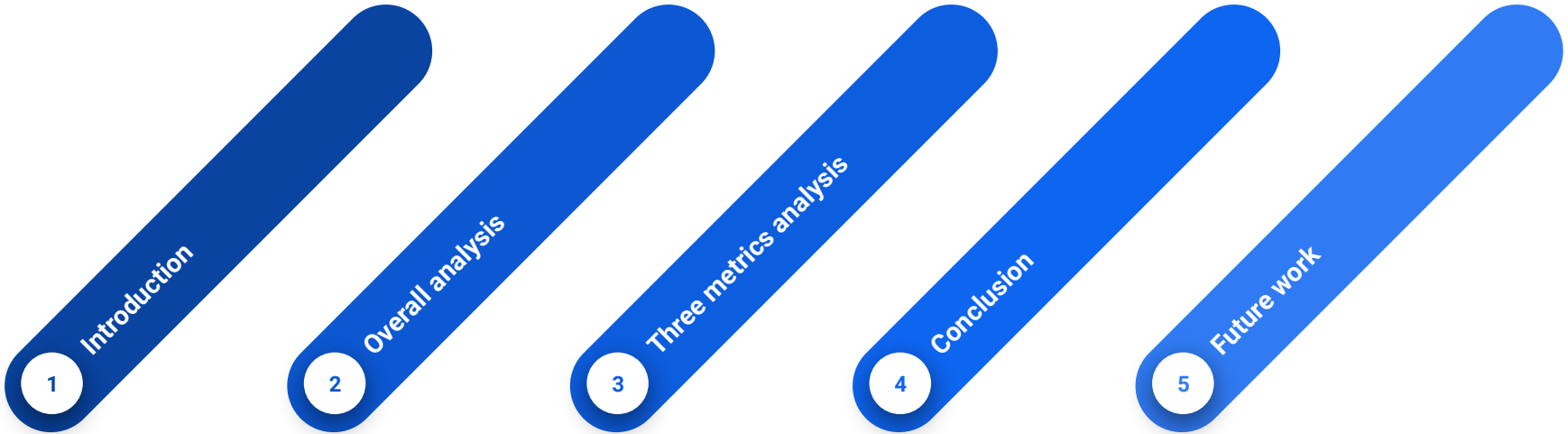
Cohort Analysis for

Jane Liu
May 22, 2019


1
Introduction
2
Overall analysis
3
Three metrics analysis
4
Conclusion
5
Future work

1 Introduction

2 Overall analysis

3 Three metrics analysis

4 Conclusion

5 Future work

What is cohort analysis?

- Belong to: behavioral analytics

- Don’t: look at all users as one unit
- Do: break them into related groups for analysis

(population to special sample: started from the same time)

Source: https://en.wikipedia.org/wiki/Cohort_analysis

- Start: 201840 (Oct. 1, 2018)
- End: 201851 (Dec. 23, 2018)
- Tool: Sql & Excel

1 Introduction

2 Overall analysis

3 Three metrics analysis

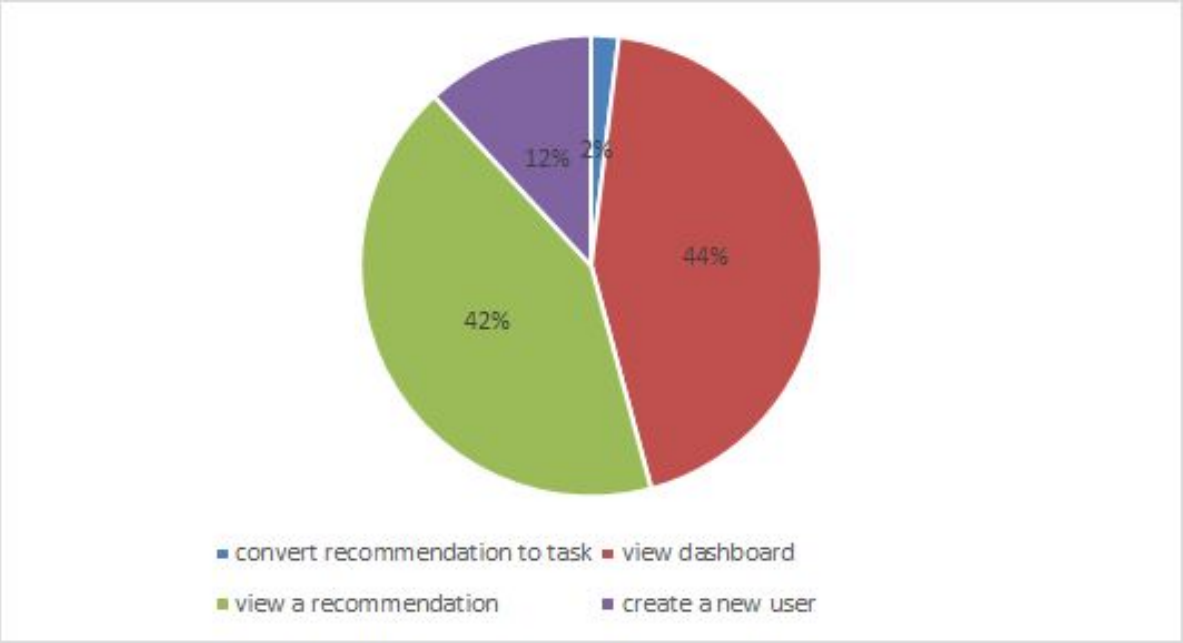
4 Conclusion

5 Future work

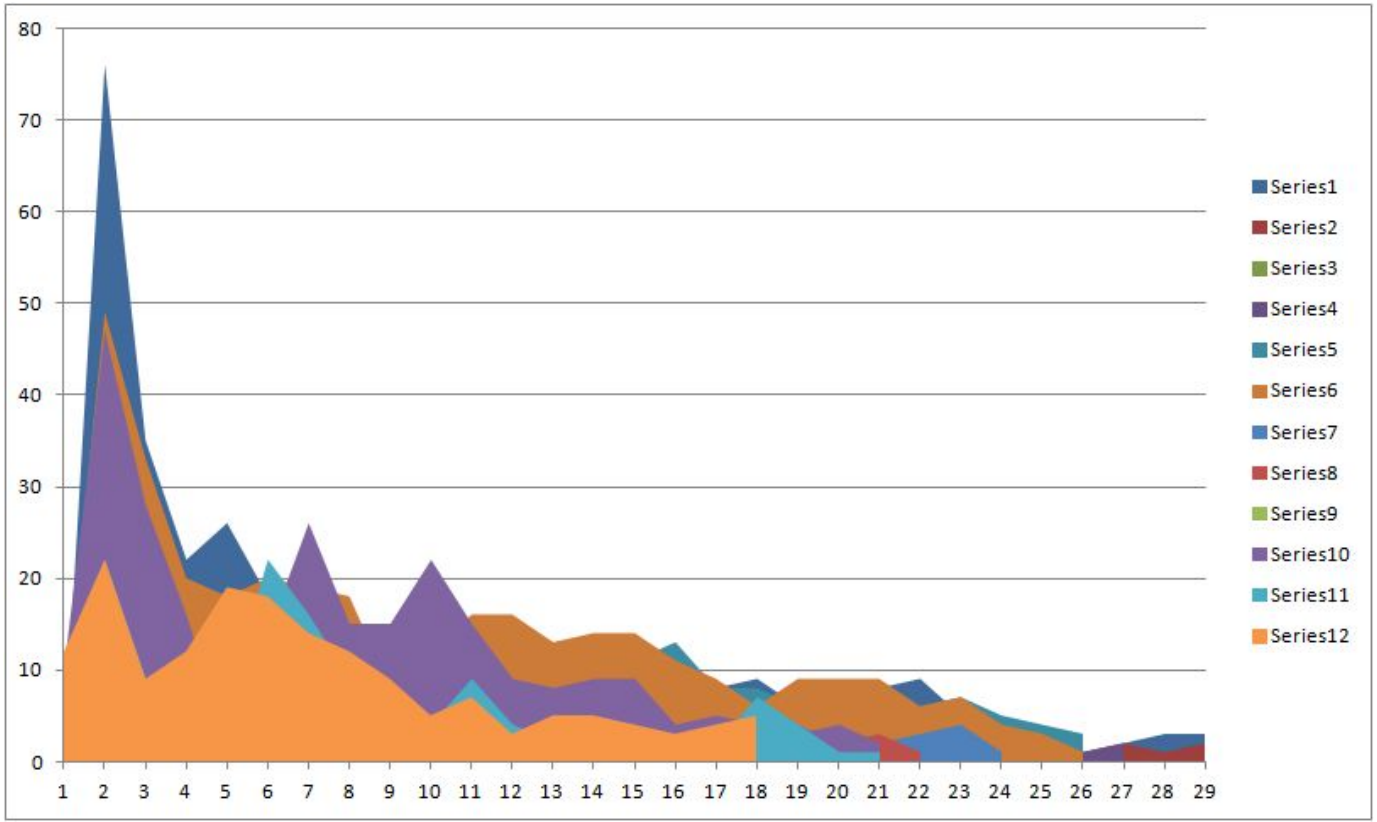
Goal

Analyze three user actions to determine which best correlates to **user retention** over a 12 week period.

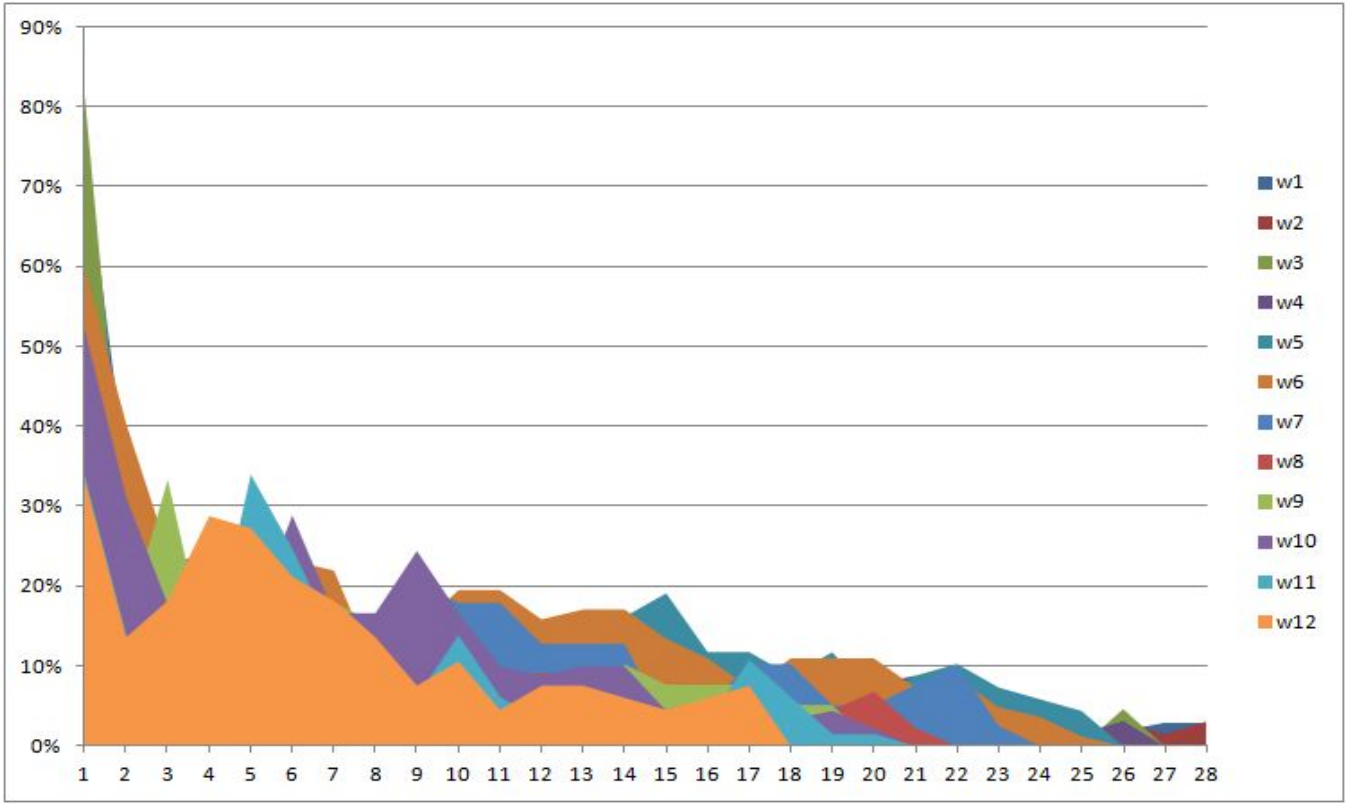
Distributions of types



Retention Number: Max Last Period 29 weeks



Max Retention Rate: 81% and drop quickly



1 Introduction

2 Overall analysis

3 Three metrics analysis

4 Conclusion

5 Future work

44%

View Dashboard

View Dashboard_number retention

Activity periods: from week j, j=1...27

week	join	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	102	74	28	13	15	10	7	5	4	4	4	2	1		4	4	2	3	2	1	4	2	2	3			1	2
2	65	47	12	7	3		1		1	1	2							1		1	1	2	1					
3	43	33	7	6	3	2	1	3	2	1			1			1	1											
4	63	31	12	5	6	4	2	3	1		1	3	2	5	1	2	2		1		1	1	2	1	1	1	1	
5	68	34	12	8	3	3	1	4	2	2	5	3	3	3	8	7	7	4	4	4	1	3	5	4	3	1		
6	82	36	15	10	12	11	9	10	5	6	4	4	4	6	5	5	4	2	3	3	4	3	3	1	2	1		
7	39	13	4	4	3	1	3	1	2	4	3	3	2	2	3		2	2	2	1	1	2	3					
8	44	12	6	5	2	4		3	2		1	1		2		1		1		1	2							
9	39	10	3	4	2	1	1		1	2		1	1	1	3	3	1	2	1	1								
10	90	34	18	8	1	6	12	6	6	10	8	6	6	5	5	2	1	2	1	3	2							
11	65	17	5		1	5	4	3	3	2	9	2	1				1	2										
12	66	18	7	5	9	6	3	6	5	3	2	2	2	2	3	1	2	2										

Join number for week i, i=1...12

View Dashboard_retention rate

Number of week j / join number of week i, i =1...12; j= 1...27

week	join	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	1	73%	27%	13%	15%	10%	7%	5%	4%	4%	4%	2%	1%	0%	4%	4%	2%	3%	2%	1%	4%	2%	2%	3%	0%	0%	1%	2%
2	1	72%	18%	11%	5%	0%	2%	0%	2%	2%	3%	0%	0%	0%	0%	0%	0%	2%	0%	2%	2%	3%	2%	0%	0%	0%	0%	0%
3	1	77%	16%	14%	7%	5%	2%	7%	5%	2%	0%	0%	2%	0%	0%	2%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
4	1	49%	19%	8%	10%	6%	3%	5%	2%	0%	2%	5%	3%	8%	2%	3%	3%	0%	2%	0%	2%	2%	3%	2%	2%	2%	2%	0%
5	1	50%	18%	12%	4%	4%	1%	6%	3%	3%	7%	4%	4%	4%	12%	10%	10%	6%	6%	6%	1%	4%	7%	6%	4%	1%	0%	0%
6	1	44%	18%	12%	15%	13%	11%	12%	6%	7%	5%	5%	5%	7%	6%	6%	5%	2%	4%	4%	5%	4%	4%	1%	2%	1%	0%	0%
7	1	33%	10%	10%	8%	3%	8%	3%	5%	10%	8%	8%	5%	5%	8%	0%	5%	5%	5%	3%	3%	5%	8%	0%	0%	0%	0%	0%
8	1	27%	14%	11%	5%	9%	0%	7%	5%	0%	2%	2%	0%	5%	0%	2%	0%	2%	0%	2%	5%	0%	0%	0%	0%	0%	0%	0%
9	1	26%	8%	10%	5%	3%	3%	0%	3%	5%	0%	3%	3%	3%	8%	8%	3%	5%	3%	3%	0%	0%	0%	0%	0%	0%	0%	0%
10	1	38%	20%	9%	1%	7%	13%	7%	7%	11%	9%	7%	7%	6%	6%	2%	1%	2%	1%	3%	2%	0%	0%	0%	0%	0%	0%	0%
11	1	26%	8%	0%	2%	8%	6%	5%	5%	3%	14%	3%	2%	0%	0%	0%	2%	3%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
12	1	27%	11%	8%	14%	9%	5%	9%	8%	5%	3%	3%	3%	3%	5%	2%	3%	3%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

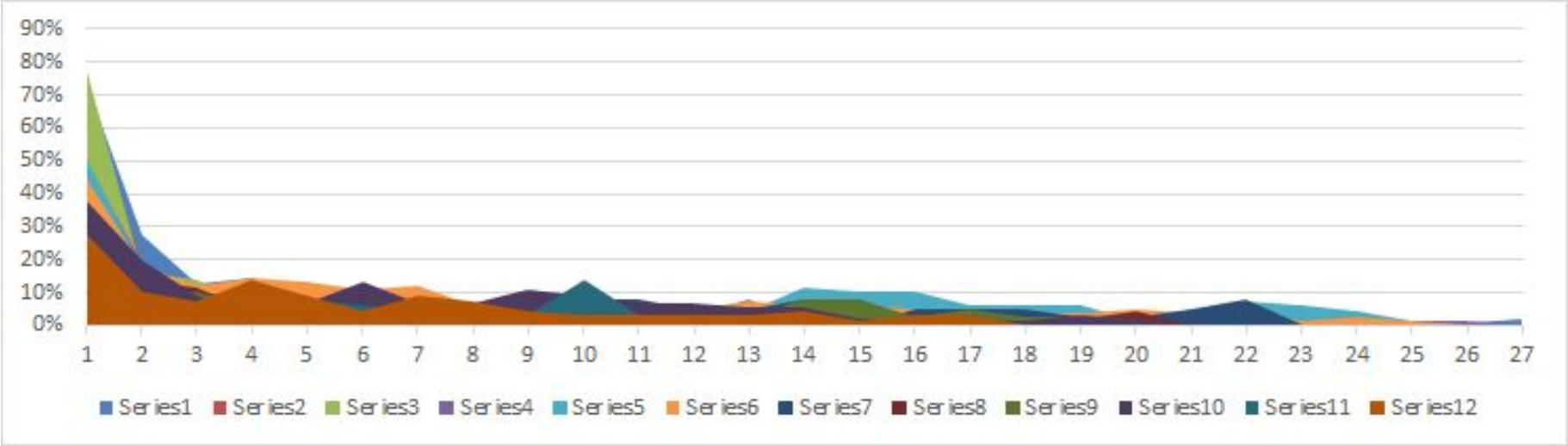
View Dashboard_AVG retention rate= 5%

Original rate -5%

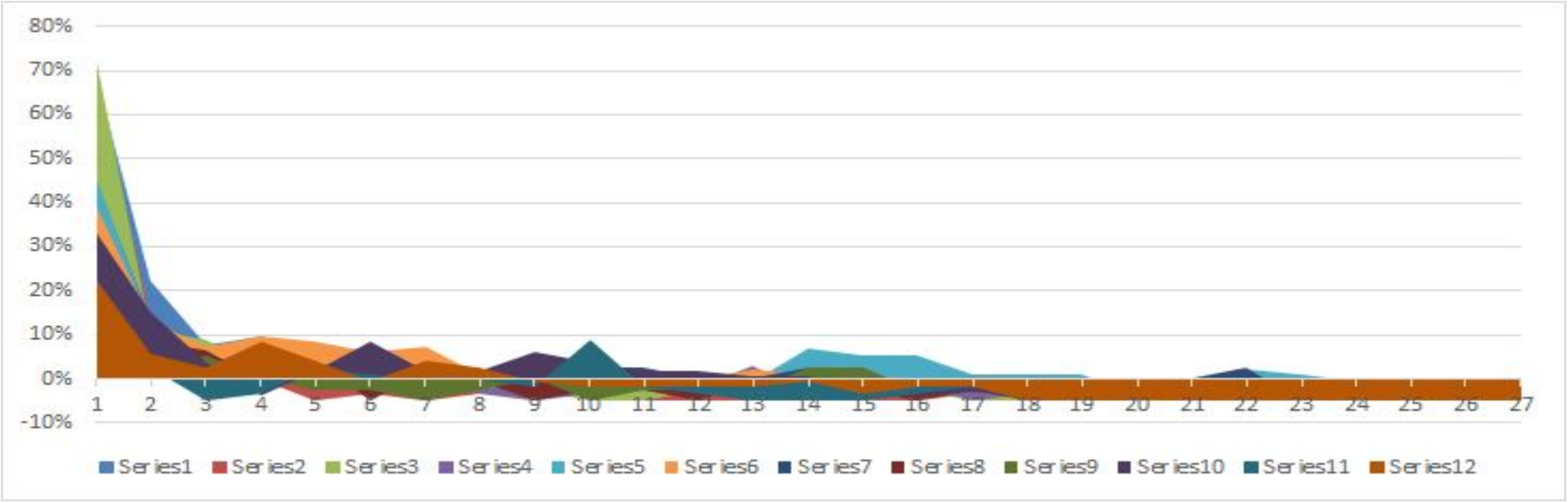
week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	68%	22%	8%	10%	5%	2%	0%	-1%	-1%	-1%	-3%	-4%	-5%	-1%	-1%	-3%	-2%	-3%	-4%	-1%	-3%	-3%	-2%	-5%	-5%	-4%	-3%
2	67%	13%	6%	0%	-5%	-3%	-5%	-3%	-3%	-2%	-5%	-5%	-5%	-5%	-5%	-5%	-3%	-5%	-3%	-3%	-2%	-3%	-5%	-5%	-5%	-5%	-5%
3	72%	11%	9%	2%	0%	-3%	2%	0%	-3%	-5%	-5%	-3%	-5%	-5%	-3%	-3%	-5%	-5%	-5%	-5%	-5%	-5%	-5%	-5%	-5%	-5%	-5%
4	44%	14%	3%	5%	1%	-2%	0%	-3%	-5%	-3%	0%	-2%	3%	-3%	-2%	-2%	-5%	-3%	-5%	-3%	-3%	-2%	-3%	-3%	-3%	-3%	-5%
5	45%	13%	7%	-1%	-1%	-4%	1%	-2%	-2%	2%	-1%	-1%	-1%	7%	5%	5%	1%	1%	1%	-4%	-1%	2%	1%	-1%	-4%	-5%	-5%
6	39%	13%	7%	10%	8%	6%	7%	1%	2%	0%	0%	0%	2%	1%	1%	0%	-3%	-1%	-1%	0%	-1%	-1%	-4%	-3%	-4%	-5%	-5%
7	28%	5%	5%	3%	-2%	3%	-2%	0%	5%	3%	3%	0%	0%	3%	-5%	0%	0%	0%	-2%	-2%	0%	3%	-5%	-5%	-5%	-5%	-5%
8	22%	9%	6%	0%	4%	-5%	2%	0%	-5%	-3%	-3%	-5%	0%	-5%	-3%	-5%	-3%	-5%	-3%	0%	-5%	-5%	-5%	-5%	-5%	-5%	-5%
9	21%	3%	5%	0%	-2%	-2%	-5%	-2%	0%	-5%	-2%	-2%	-2%	3%	3%	-2%	0%	-2%	-2%	-5%	-5%	-5%	-5%	-5%	-5%	-5%	-5%
10	33%	15%	4%	-4%	2%	8%	2%	2%	6%	4%	2%	2%	1%	1%	-3%	-4%	-3%	-4%	-2%	-3%	-5%	-5%	-5%	-5%	-5%	-5%	-5%
11	21%	3%	-5%	-3%	3%	1%	0%	0%	-2%	9%	-2%	-3%	-5%	-5%	-5%	-3%	-2%	-5%	-5%	-5%	-5%	-5%	-5%	-5%	-5%	-5%	-5%
12	22%	6%	3%	9%	4%	0%	4%	3%	0%	-2%	-2%	-2%	-2%	0%	-3%	-2%	-2%	-5%	-5%	-5%	-5%	-5%	-5%	-5%	-5%	-5%	-5%

View Dashboard

Max Retention Rate: 77%, last 27 weeks



AVG Retention Rate: most [-10%, 10%]



12%

Create a new user

Create a new user_number retention

week	join	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	102	2	3	1	4	3					2					1	1	2	2	1		1	1				1	1
2	65	1		1	2	2	1	1	4			2					1	2									1	
3	43		1								1																	
4	63	1	2	1	1		1											1							1			
5	68		1	1	1	4	1							3	1			2	2	4	1		2	1	1	1		
6	82	3	1			1	2	1	1	2	3	3	2	1		2	2	1	2		2	1	1	1				
7	39			1		1						1							1									
8	44	1	1		1	1		1	1	2						1		1		1	1							
9	39		2							1		1			1		1	1										
10	90	5	5	1			3	1	1	1	3			1	1	1	2		1	1								
11	65	3			3	2	2					1	1	1		2		1										
12	66			4	1	3	1	1	2	1	1		2				1	1										

Create a new user_retention rate

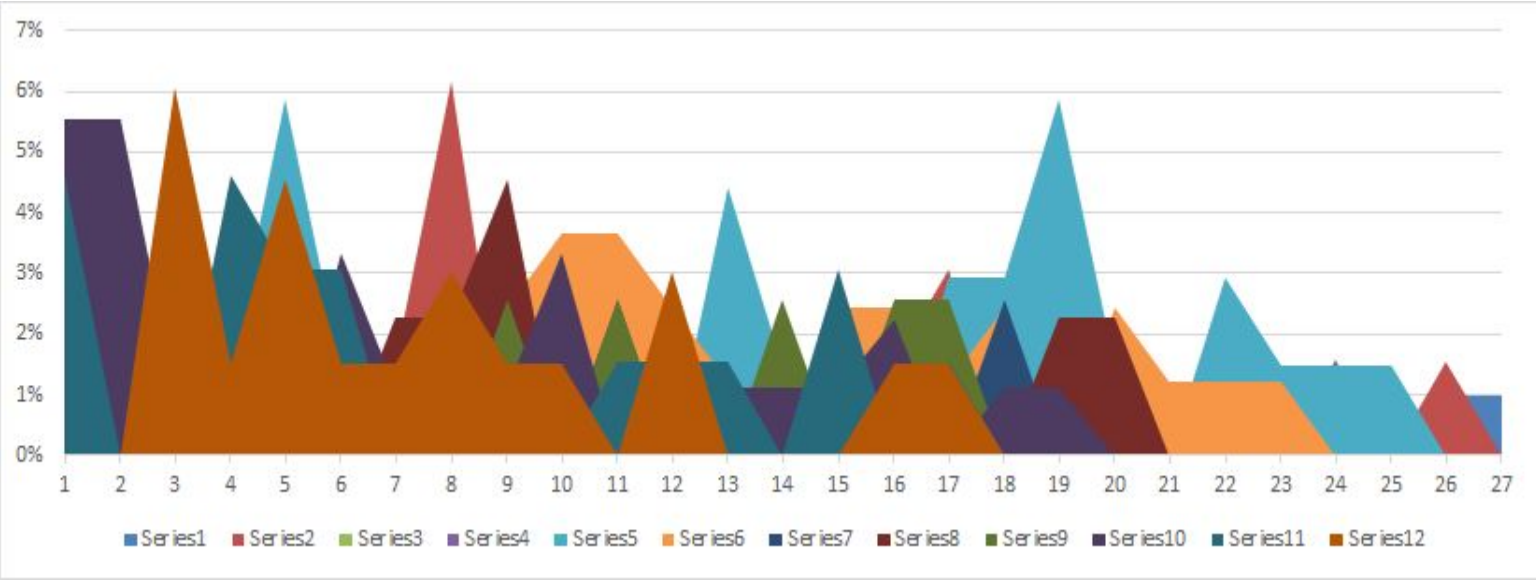
week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	2%	3%	1%	4%	3%	0%	0%	0%	0%	2%	0%	0%	0%	0%	1%	1%	2%	2%	1%	0%	1%	1%	0%	0%	0%	1%	1%
2	2%	0%	2%	3%	3%	2%	2%	6%	0%	0%	3%	0%	0%	0%	0%	2%	3%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%
3	0%	2%	0%	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
4	2%	3%	2%	2%	0%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%
5	0%	1%	1%	1%	6%	1%	0%	0%	0%	0%	0%	0%	4%	1%	0%	0%	3%	3%	6%	1%	0%	3%	1%	1%	1%	0%	0%
6	4%	1%	0%	0%	1%	2%	1%	1%	2%	4%	4%	2%	1%	0%	2%	2%	1%	2%	0%	2%	1%	1%	1%	0%	0%	0%	0%
7	0%	0%	3%	0%	3%	0%	0%	0%	0%	0%	3%	0%	0%	0%	0%	0%	0%	3%	0%	0%	0%	0%	0%	0%	0%	0%	0%
8	2%	2%	0%	2%	2%	0%	2%	2%	5%	0%	0%	0%	0%	0%	2%	0%	2%	0%	2%	2%	0%	0%	0%	0%	0%	0%	0%
9	0%	5%	0%	0%	0%	0%	0%	0%	3%	0%	3%	0%	0%	3%	0%	3%	3%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
10	6%	6%	1%	0%	0%	3%	1%	1%	1%	3%	0%	0%	1%	1%	1%	2%	0%	1%	1%	0%	0%	0%	0%	0%	0%	0%	0%
11	5%	0%	0%	5%	3%	3%	0%	0%	0%	0%	2%	2%	2%	0%	3%	0%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
12	0%	0%	6%	2%	5%	2%	2%	3%	2%	2%	0%	3%	0%	0%	0%	2%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Create a new user_AVG retention rate= 1%

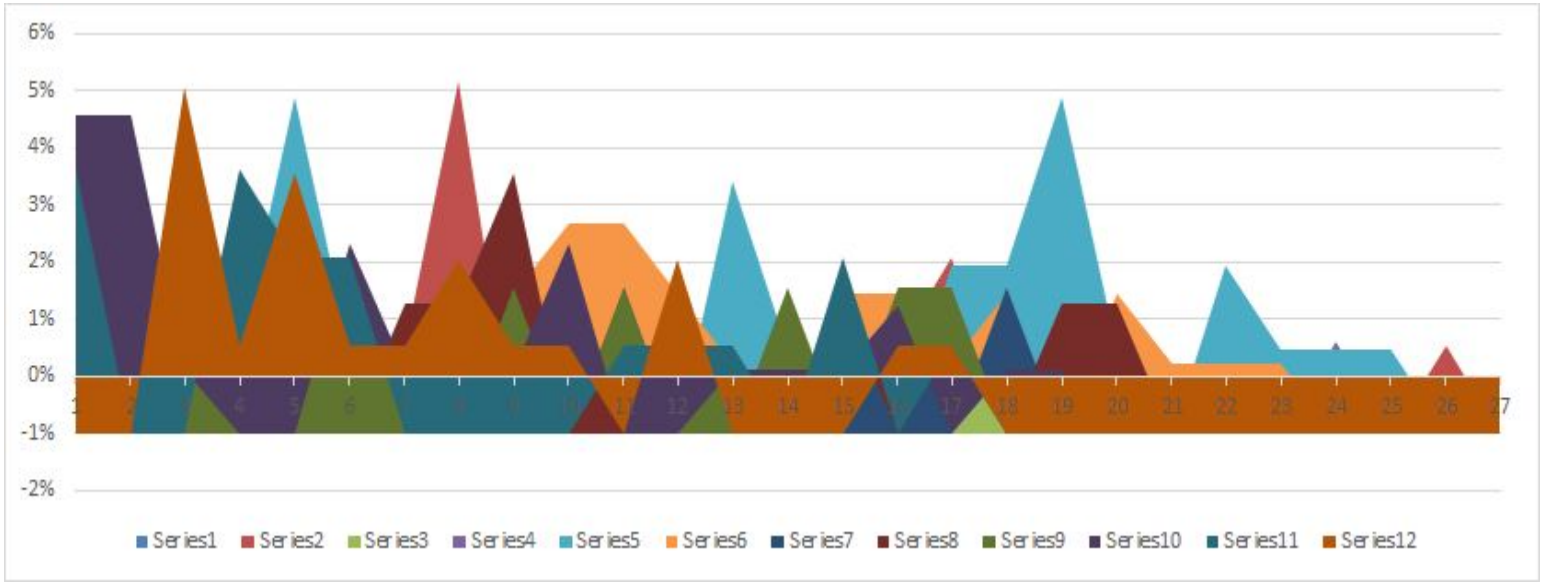
week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	1%	2%	0%	3%	2%	-1%	-1%	-1%	-1%	1%	-1%	-1%	-1%	-1%	0%	0%	1%	1%	0%	-1%	0%	0%	-1%	-1%	-1%	0%	0%
2	1%	-1%	1%	2%	2%	1%	1%	5%	-1%	-1%	2%	-1%	-1%	-1%	-1%	1%	2%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	1%	-1%
3	-1%	1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%
4	1%	2%	1%	1%	-1%	1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	1%	-1%	-1%	-1%	-1%	-1%	-1%	1%	-1%	-1%	-1%
5	-1%	0%	0%	0%	5%	0%	-1%	-1%	-1%	-1%	-1%	-1%	3%	0%	-1%	-1%	2%	2%	5%	0%	-1%	2%	0%	0%	0%	-1%	-1%
6	3%	0%	-1%	-1%	0%	1%	0%	0%	1%	3%	3%	1%	0%	-1%	1%	1%	0%	1%	-1%	1%	0%	0%	0%	-1%	-1%	-1%	-1%
7	-1%	-1%	2%	-1%	2%	-1%	-1%	-1%	-1%	-1%	2%	-1%	-1%	-1%	-1%	-1%	-1%	2%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%
8	1%	1%	-1%	1%	1%	-1%	1%	1%	4%	-1%	-1%	-1%	-1%	-1%	1%	-1%	1%	-1%	1%	1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%
9	-1%	4%	-1%	-1%	-1%	-1%	-1%	-1%	2%	-1%	2%	-1%	-1%	2%	-1%	2%	2%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%
10	5%	5%	0%	-1%	-1%	2%	0%	0%	0%	2%	-1%	-1%	0%	0%	0%	1%	-1%	0%	0%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%
11	4%	-1%	-1%	4%	2%	2%	-1%	-1%	-1%	-1%	1%	1%	1%	-1%	2%	-1%	1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%
12	-1%	-1%	5%	1%	4%	1%	1%	2%	1%	1%	-1%	2%	-1%	-1%	-1%	1%	1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%	-1%

Create a new user

Retention Rate: [0, 3%]



AVG Retention Rate:
most fall into [0, 2%] more equal distribution



2%

Convert recommendation to task

Retention Number: only sparse value 1

week	join	2	3	4	5	6	7	8	10	11	12	13	14	15	26
1	102	1												1	
2	65		1	1											1
3	43														
4	63														
5	68							1		1					
6	82				1				1	1			1		
7	39											1			
8	44										1		1		
9	39														
10	90	1				1									
11	65														
12	66			1			1			1		1			

Retention Rate range: only 1% and 3%

week	join	2	3	4	5	6	7	8	10	11	12	13	14	15	26
1	102	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%
2	65	0%	2%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%
3	43	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
4	63	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
5	68	0%	0%	0%	0%	0%	0%	1%	0%	1%	0%	0%	0%	0%	0%
6	82	0%	0%	0%	1%	0%	0%	0%	1%	1%	0%	0%	1%	0%	0%
7	39	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%	0%	0%	0%
8	44	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	2%	0%	0%
9	39	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
10	90	1%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%
11	65	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
12	66	0%	0%	2%	0%	0%	2%	0%	0%	2%	0%	2%	0%	0%	0%

Convert recommendation to task

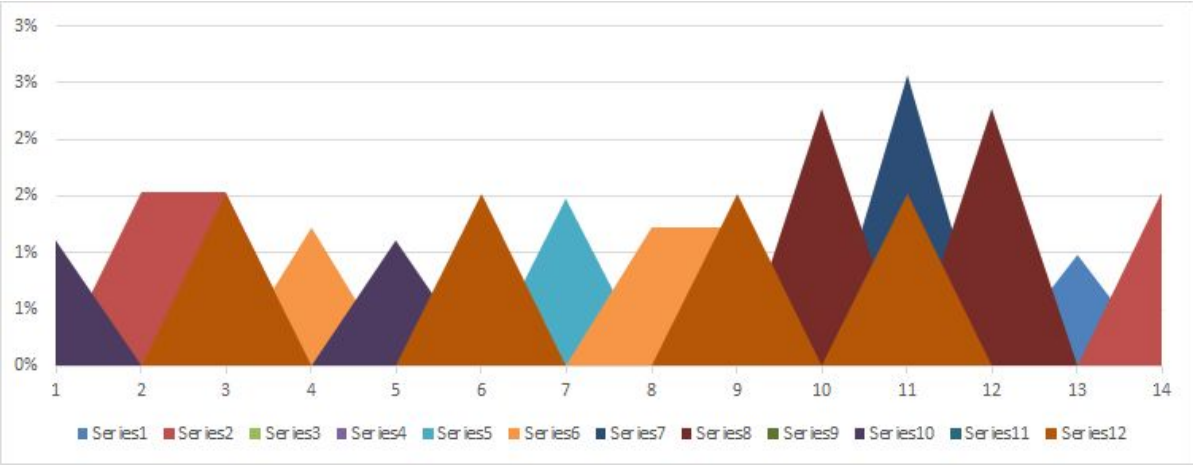
AVG Rate: 0.1773%

week	2	3	4	5	6	7	8	10	11	12	13	14	15	26
1	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%
2	0%	1%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%
3	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
4	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
5	0%	0%	0%	0%	0%	0%	1%	0%	1%	0%	0%	0%	0%	0%
6	0%	0%	0%	1%	0%	0%	0%	1%	1%	0%	0%	1%	0%	0%
7	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%
8	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	2%	0%	0%
9	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
10	1%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%
11	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
12	0%	0%	1%	0%	0%	1%	0%	0%	1%	0%	1%	0%	0%	0%

Retention Rate

More equal distribution, range is [0, 3%] than other two metrics visualization.

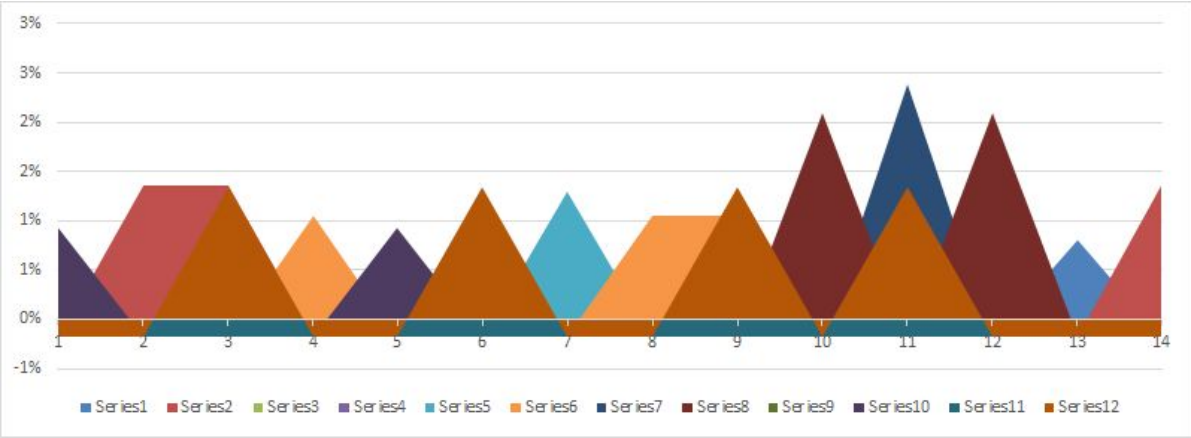
Specially, the highest value occurred in 10th week and 12th week.



Convert recommendation to task

AVG Retention Rate:

There is not big difference without considering zero value.



1 Introduction

2 Overall analysis

3 Three metrics analysis

4 Conclusion

5 Future work

Goal

Analyze three user actions to determine which best correlates to **user retention** over a 12 week period.

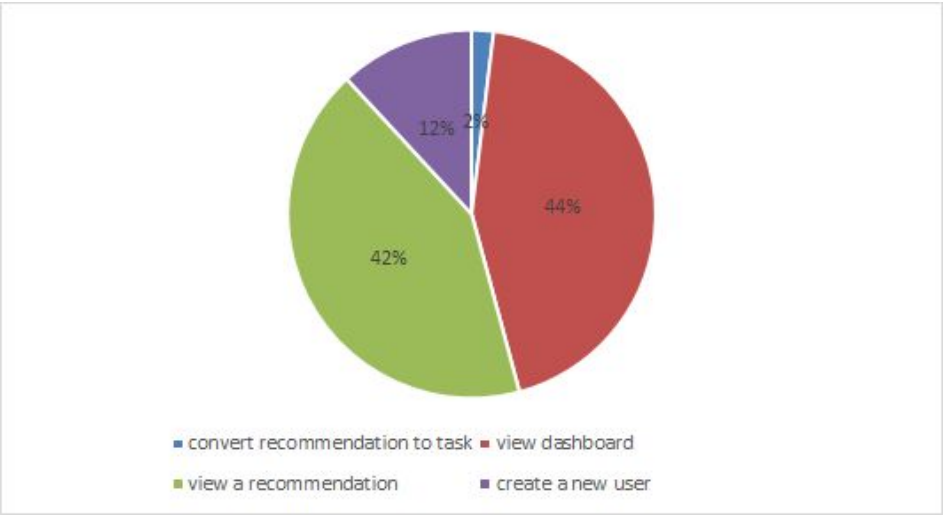
Which one?

View Dashboard

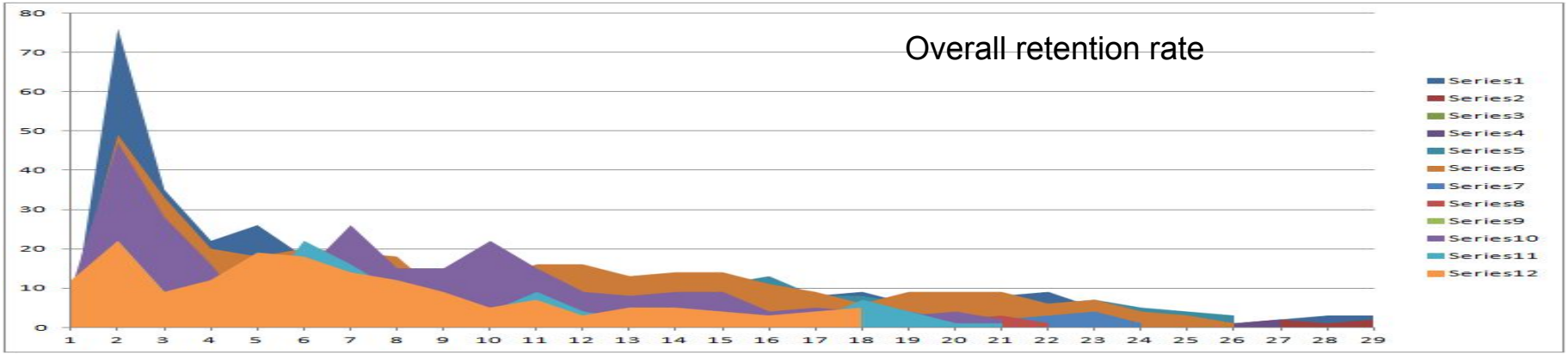
Reasons

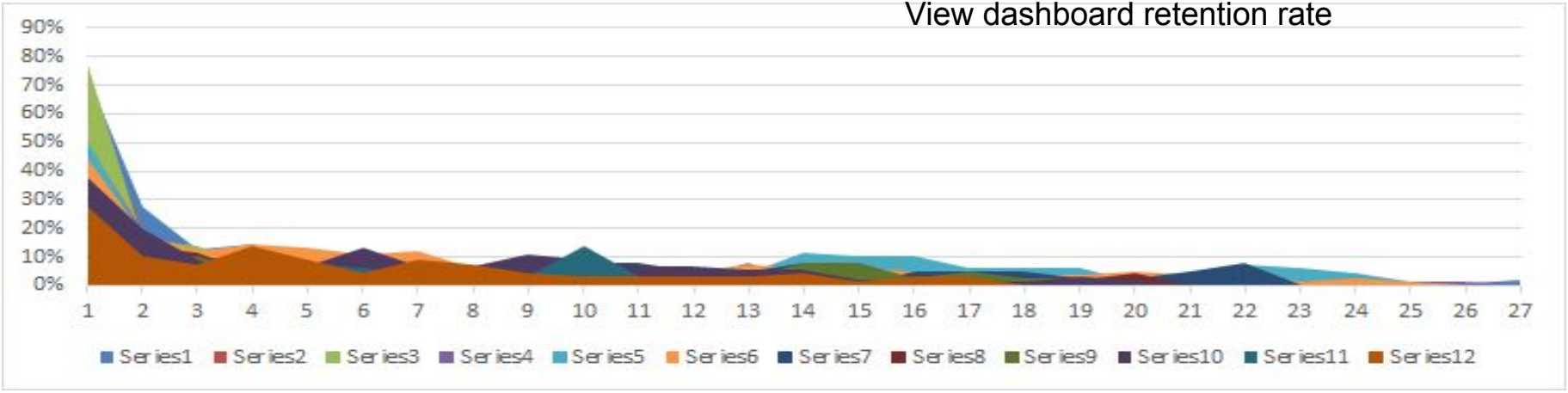
44%

convert recommendation to task	19
view dashboard	436
view a recommendation	420
create a new user	118



Reasons


Overall retention rate
80
70
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
Series1
Series2
Series3
Series4
Series5
Series6
Series7
Series8
Series9
Series10
Series11
Series12


View dashboard retention rate
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
Series1 Series2 Series3 Series4 Series5 Series6 Series7 Series8 Series9 Series10 Series11 Series12

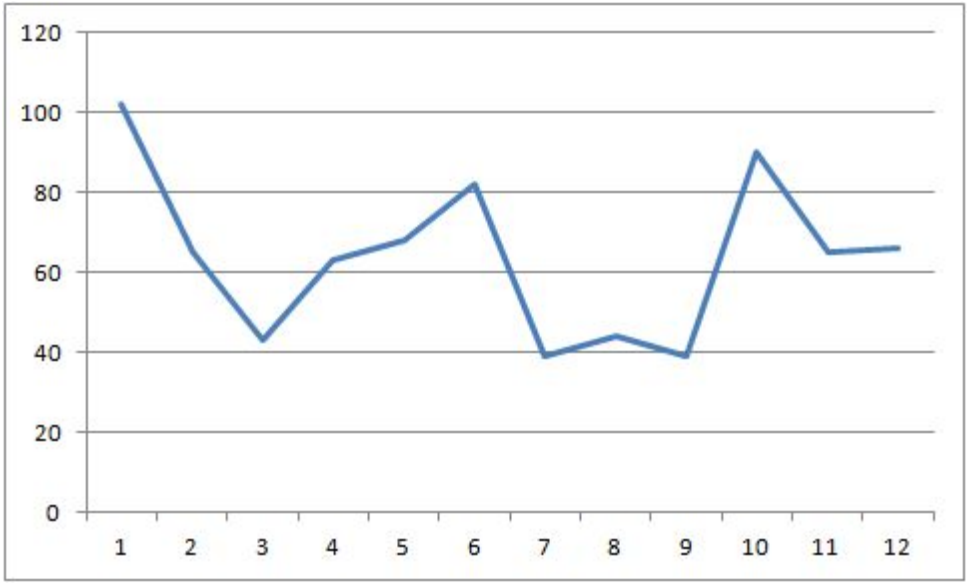
Reasons

- Active behavior
- Large enough data
- High percentage of total actions’ number
- Corresponding to the overall trends

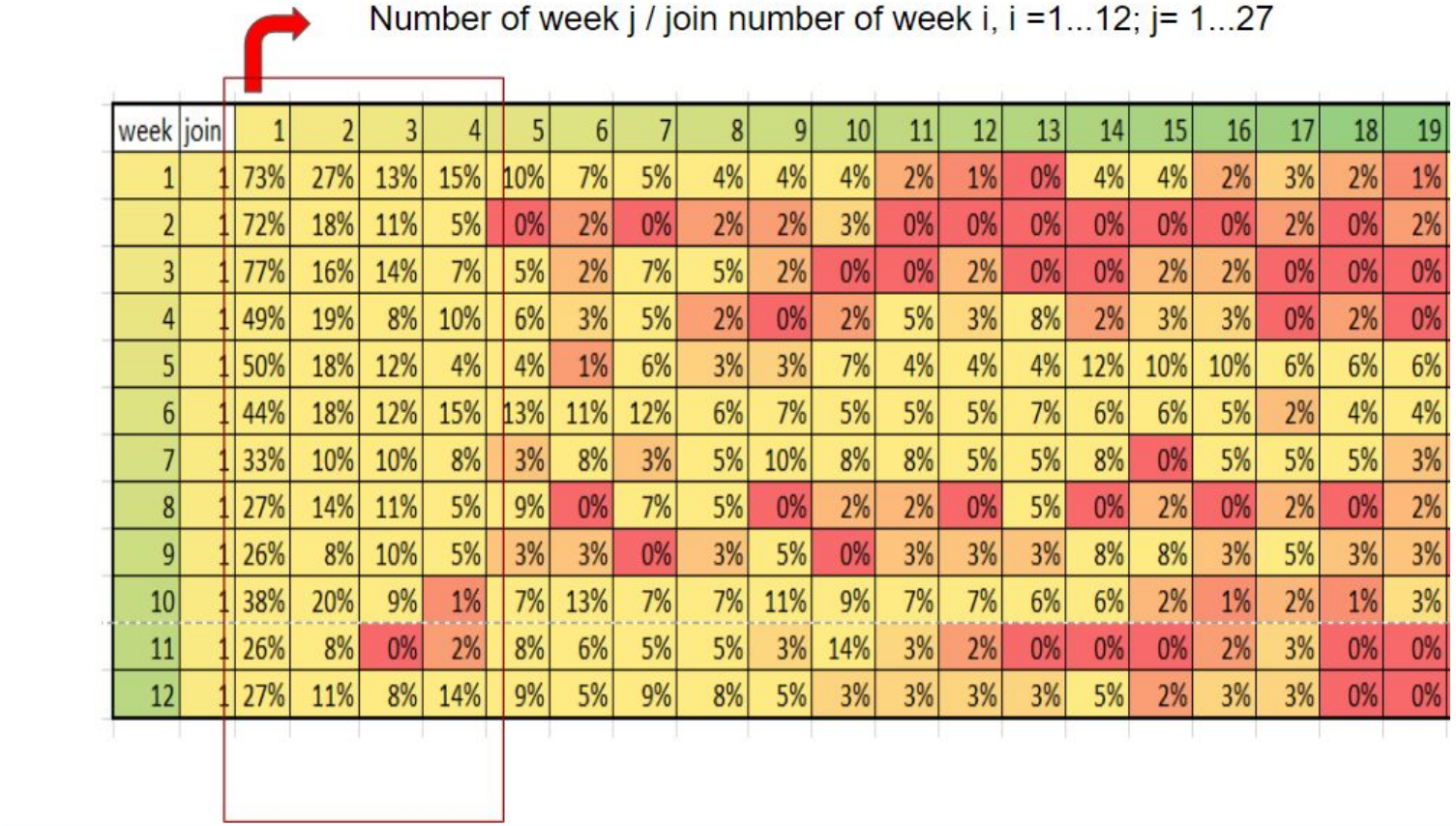
Number of Joined users weekly

Period pattern

Mean:64/month
Median: 65/month

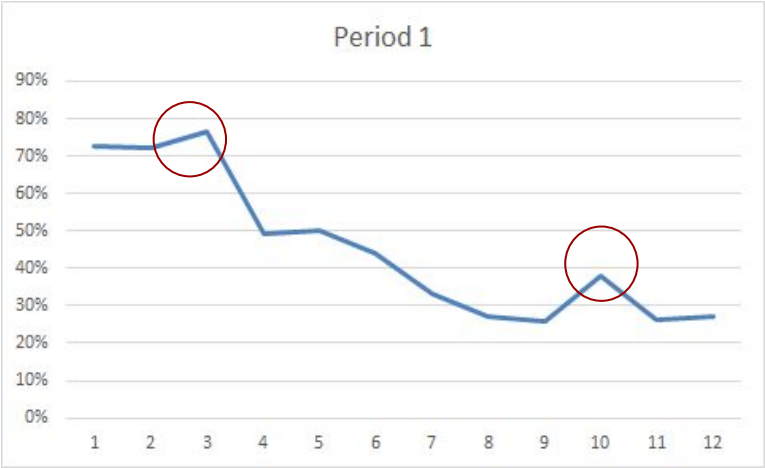


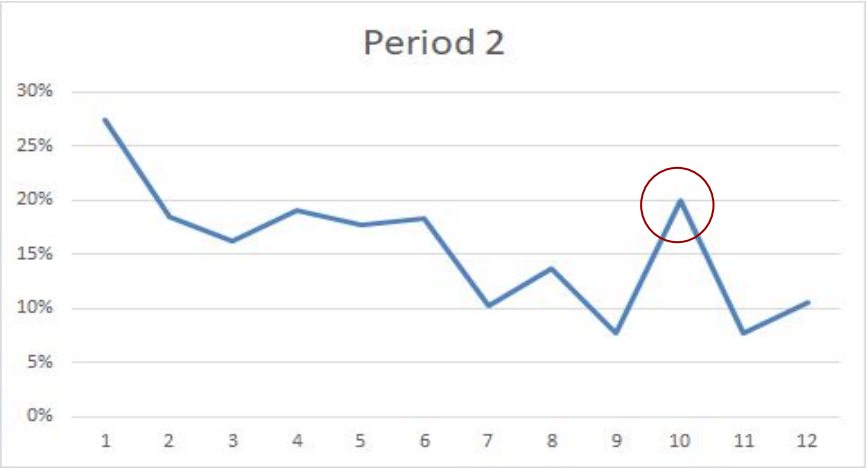
View Dashboard_retention rate



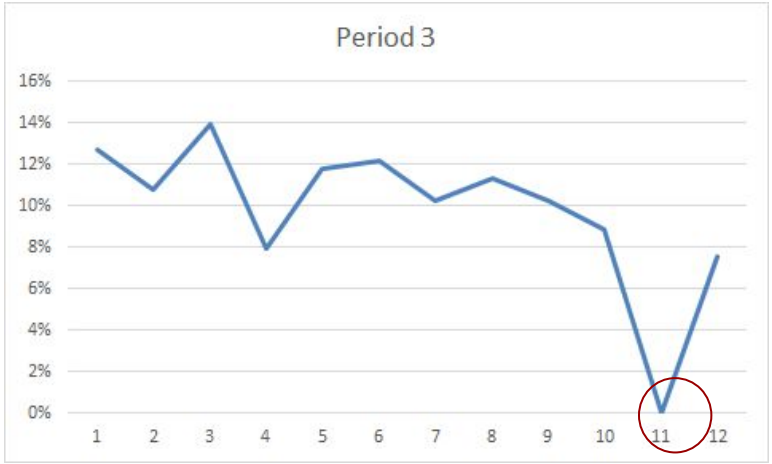
week	join	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1	73%	27%	13%	15%	10%	7%	5%	4%	4%	4%	2%	1%	0%	4%	4%	2%	3%	2%	1%
2	1	72%	18%	11%	5%	0%	2%	0%	2%	2%	3%	0%	0%	0%	0%	0%	0%	2%	0%	2%
3	1	77%	16%	14%	7%	5%	2%	7%	5%	2%	0%	0%	2%	0%	0%	2%	2%	0%	0%	0%
4	1	49%	19%	8%	10%	6%	3%	5%	2%	0%	2%	5%	3%	8%	2%	3%	3%	0%	2%	0%
5	1	50%	18%	12%	4%	4%	1%	6%	3%	3%	7%	4%	4%	4%	12%	10%	10%	6%	6%	6%
6	1	44%	18%	12%	15%	13%	11%	12%	6%	7%	5%	5%	5%	7%	6%	6%	5%	2%	4%	4%
7	1	33%	10%	10%	8%	3%	8%	3%	5%	10%	8%	8%	5%	5%	8%	0%	5%	5%	5%	3%
8	1	27%	14%	11%	5%	9%	0%	7%	5%	0%	2%	2%	0%	5%	0%	2%	0%	2%	0%	2%
9	1	26%	8%	10%	5%	3%	3%	0%	3%	5%	0%	3%	3%	3%	8%	8%	3%	5%	3%	3%
10	1	38%	20%	9%	1%	7%	13%	7%	7%	11%	9%	7%	7%	6%	6%	2%	1%	2%	1%	3%
11	1	26%	8%	0%	2%	8%	6%	5%	5%	3%	14%	3%	2%	0%	0%	0%	2%	3%	0%	0%
12	1	27%	11%	8%	14%	9%	5%	9%	8%	5%	3%	3%	3%	3%	5%	2%	3%	3%	0%	0%

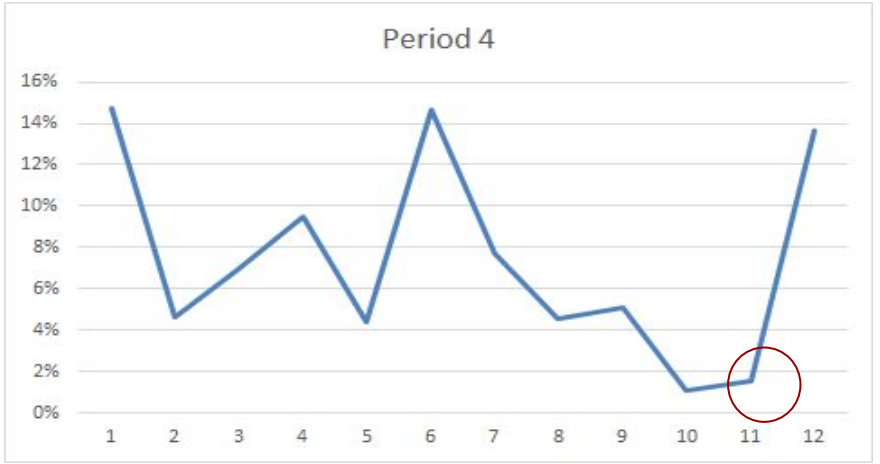
view dashboard--period retention rate for different join week





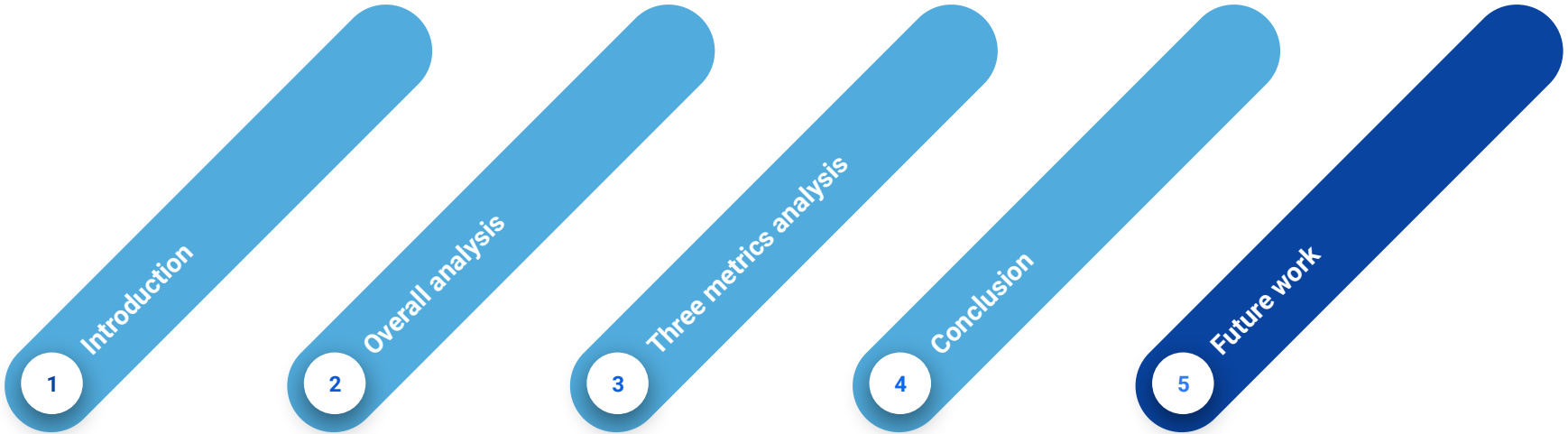
view dashboard--period retention rate for different join week





view dashboard--imply

- There is continuous decrease trends during 12 weeks period
- There is an increase in week 10 for lag1 and lag2
- There is a serious drop in week 11 but increase in week 12 for lag3 and lag4


1
Introduction
2
Overall analysis
3
Three metrics analysis
4
Conclusion
5
Future work

Future guess

- Diagnostic period pattern and try month retention
- Compare year’s retention rate and analysis the difference
- Compare our retention rate with competitors
- **Measure the engagement, eg, time spending on our products or frequency visit**
- Try advance model to predict purchase, eg, logistic regression

Thanks and best wishes!