App Utilization Analysis

PayActiv Practicum

Fri, June 3rd

Michael Chiu, Tejaswini Kambhampati, Fanyu Kong, Hongjin Wu

Advisor: Prof. Sunghun Chung

Contents:

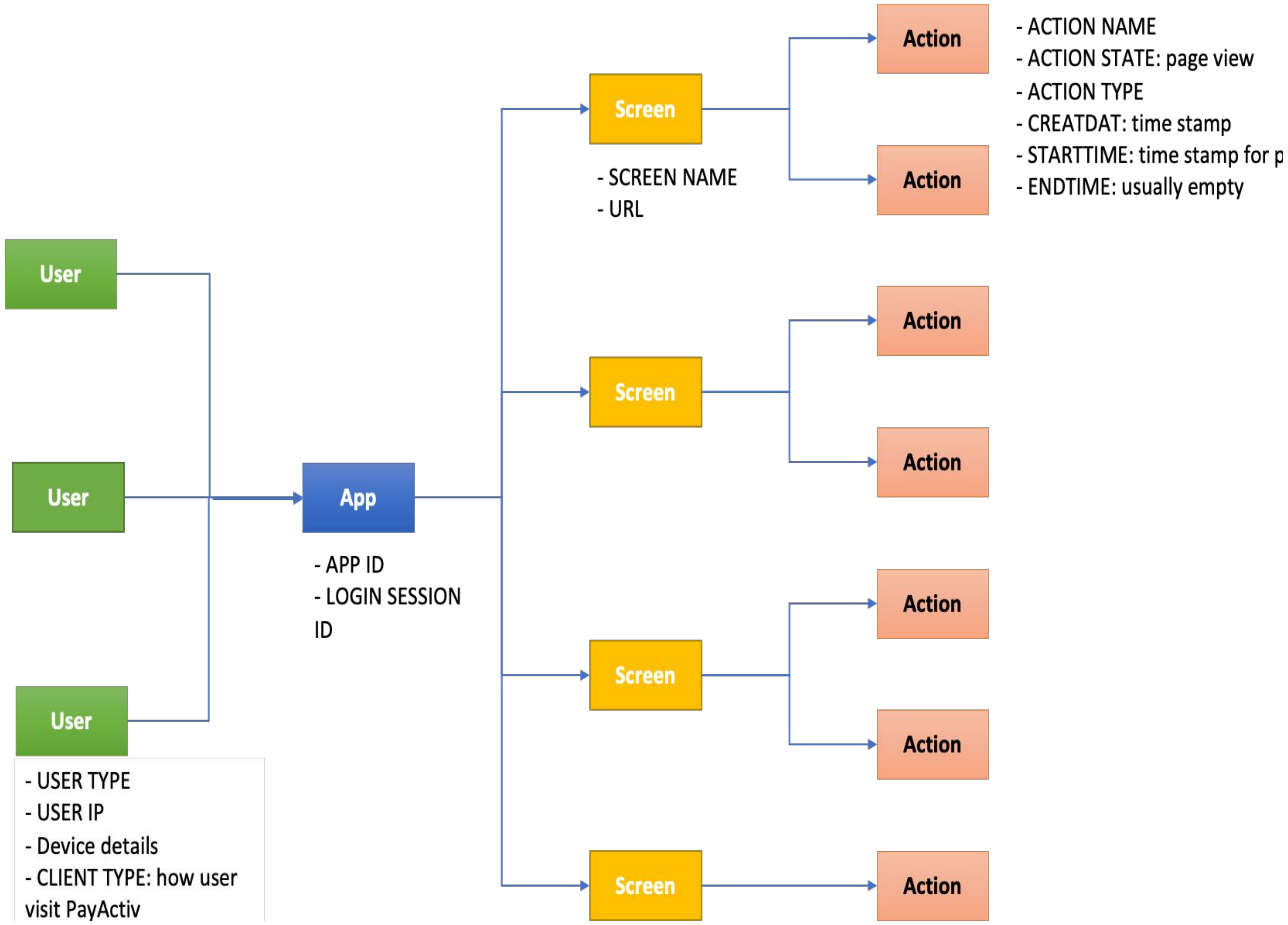
- Holistic financial wellness platform
- Direct to Business (Enterprise Users) - Amazon, Walmart, Visa, Uber etc., (EWA,Payactiv connect)
- Direct to consumer - Direct users (Employers not in PayActiv network)
- Payactiv card - discounts, Savings, Financial education
- Connect and engage deskless teams

Scope & Objectives

- Analyze ‘how’ direct users navigate through the app
- ‘What’ services are used

Data Dictionary

- Usertype: Direct,Enterprise
- Clienttype: Web, MobileApp,Mobileweb
- AppID: Appbits among different environments
- 91180 unique screens
- 52717 unique actions



Direct User Dataset Information

- 375,450 unique Direct User Ids
- 16,216,523 entries
- From 2021-04-24 to 2022-05-16 (based on CreatedAt timestamp)
- Direct User accounts for ~74% of total user (375,450 out of 507,884)

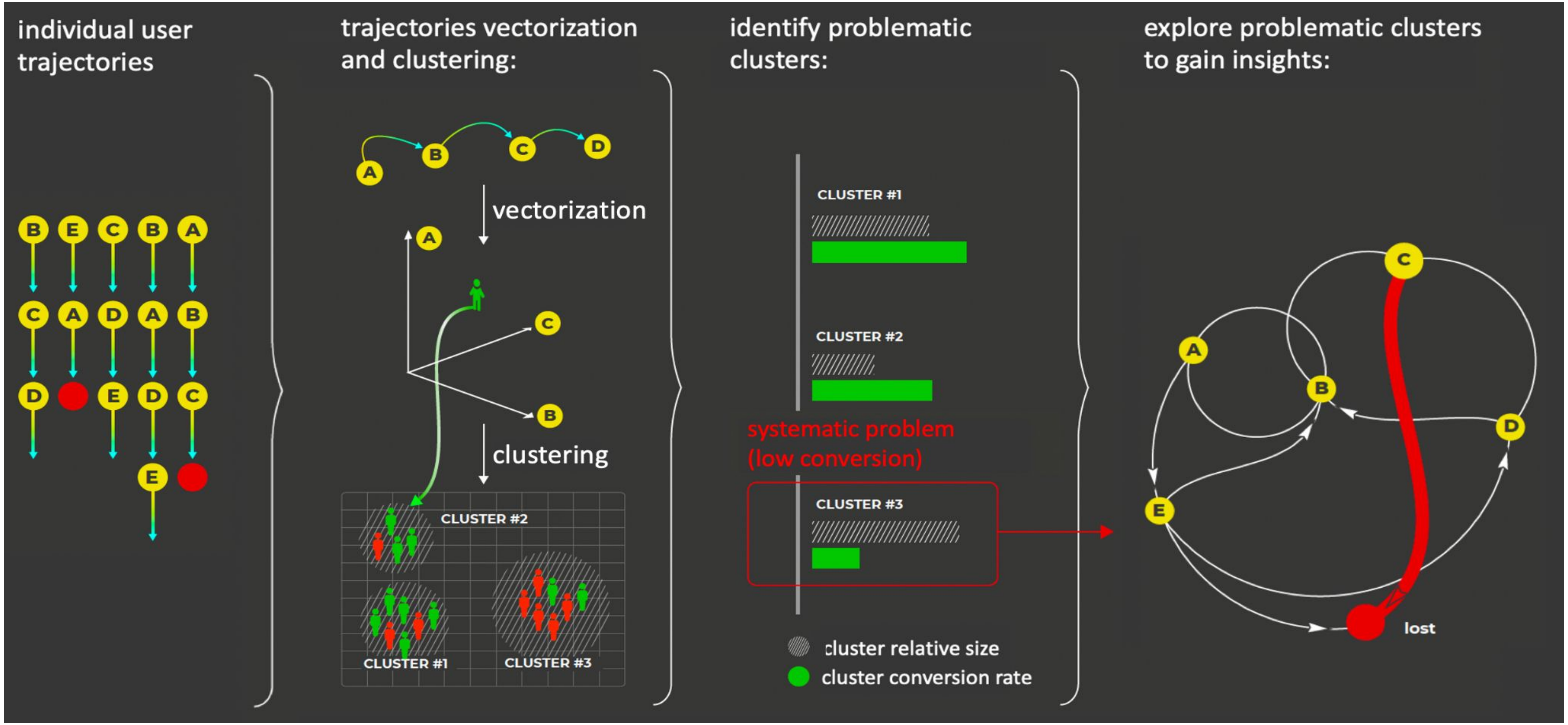
Key Challenges

S.No	CHALLENGES FACED	SOLUTION
1.	Volume of data	Querying relevant data
2.	Injecting datalakes from SNOWFLAKE into Python for analysis	Use SNOWFLAKE built-in package in python, connecting to snowflake to access the data lake directly
3.	Identifying and Eliminating Test Records	Identify Demo accounts
4.	Identifying Exclusive Direct Users	Eliminate users with EWA
5.	Multiple paths to Target screen	Ideal Screen flow focus
6.	Identify screen changes wrt to meaningful actions	Standardizding the action label (feature engineering)
7.	Tracking users trajectories	Retentioneering

Intro to Retentioneering

A Python library to analyze user behavior through graph visualizations, adjacency matrix and clusterization

Approach



individual user trajectories
B
E
C
B
A
C
A
D
A
B
D
E
D
C
E
trajectories vectorization and clustering:
A
B
C
D
vectorization
A
C
B
clustering
CLUSTER #2
CLUSTER #1
CLUSTER #3
identify problematic clusters:
CLUSTER #1
CLUSTER #2
systematic problem
(low conversion)
CLUSTER #3
cluster relative size
cluster conversion rate
explore problematic clusters to gain insights:
C
A
B
D
E
lost

Key Deliverables

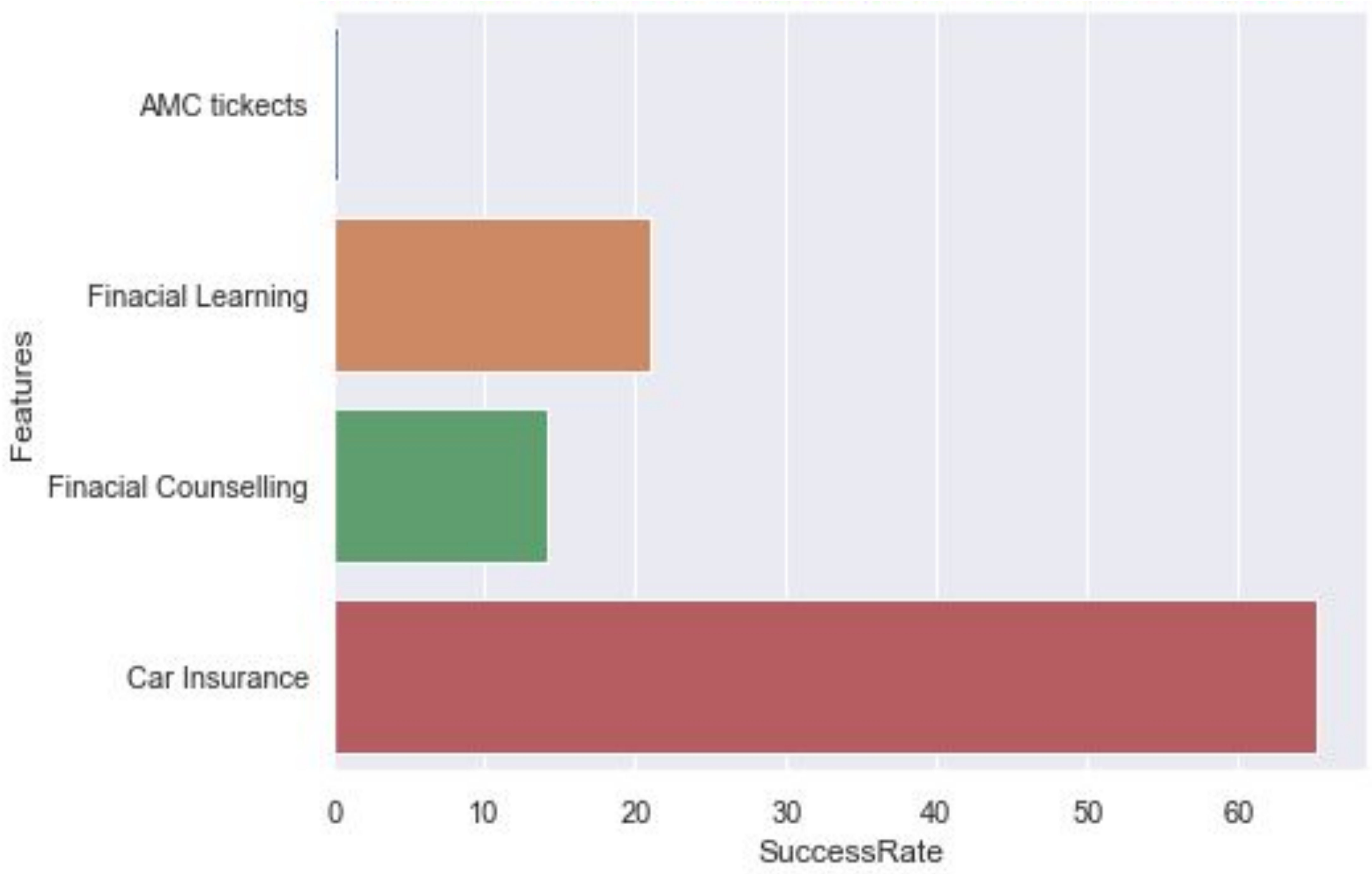
1. EDA on Direct Users

Feature Popularity

Most popular features

```
card                            3356965
payactiv-direct                 2427904
home                             822730
connect                          335766
welcome                          313362
savings                          284254
billpay                           81337
activate card                     54798
signup                            54489
card - Direct Deposit             46932
card - Activation                 39181
card - Pay Friend                 39134
sso                               15307
plaid transition_view             14416
plaid search_institution           9767
referral                           7941
Referral                           7819
gas-discount                       7664
rxdiscount                         6355
friend-indeed                      5815
access-plus                        4802
amc                                4635
card - Make Payment                4632
device-activity                    4303
plaid open                         3820
plaid exit                         3554
ewa                                3381
turbo-tax                          2917
smartspend                         2848
savings - Auto Insurance           2546
```

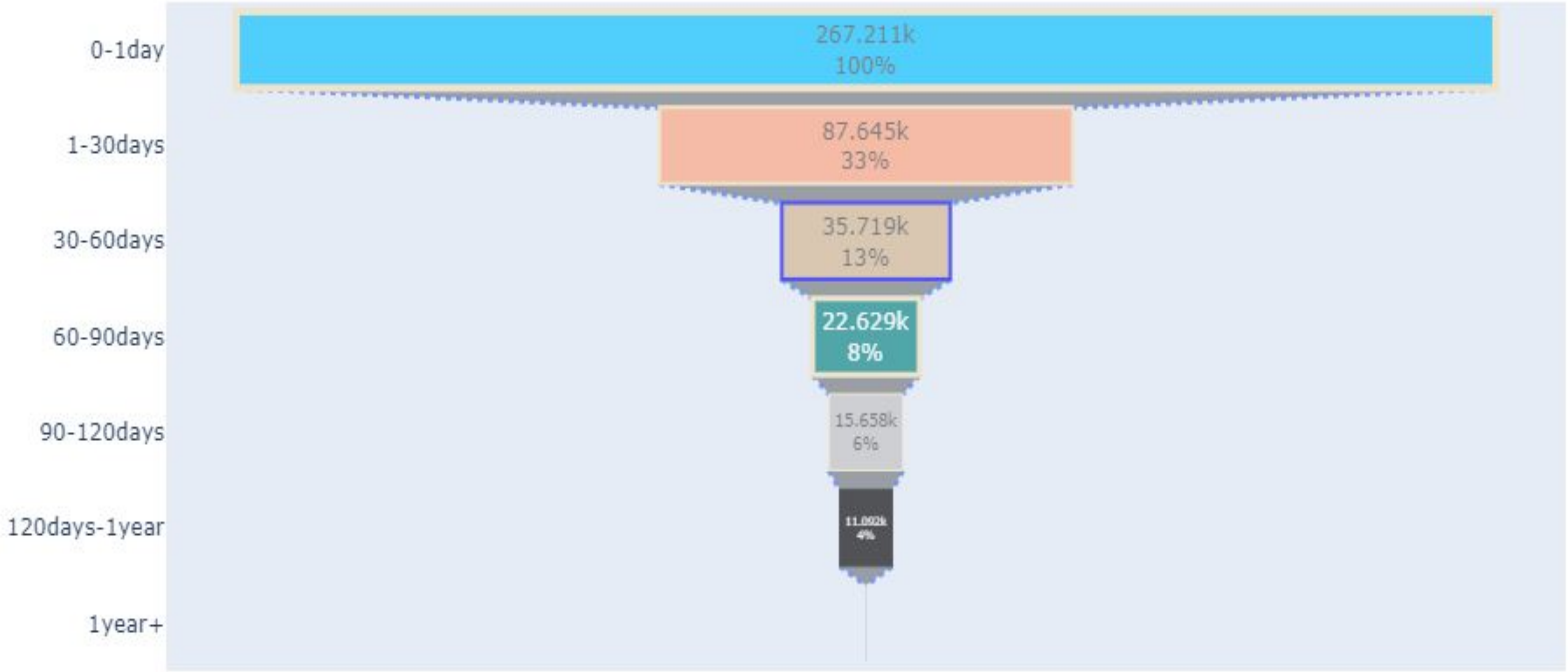
Features vs. Success Rate



Although AMC ticket feature is more visited than Car Insurance and Financial services, it has significant lower success rate. We also observed frequent visitors on Financial services. We would suggest:

- To provide more competitive offers on AMC tickets
- To create more quality Financial Learning articles
- To promo Car Insurance and Financial services features more to Direct Users

User Retention Rate



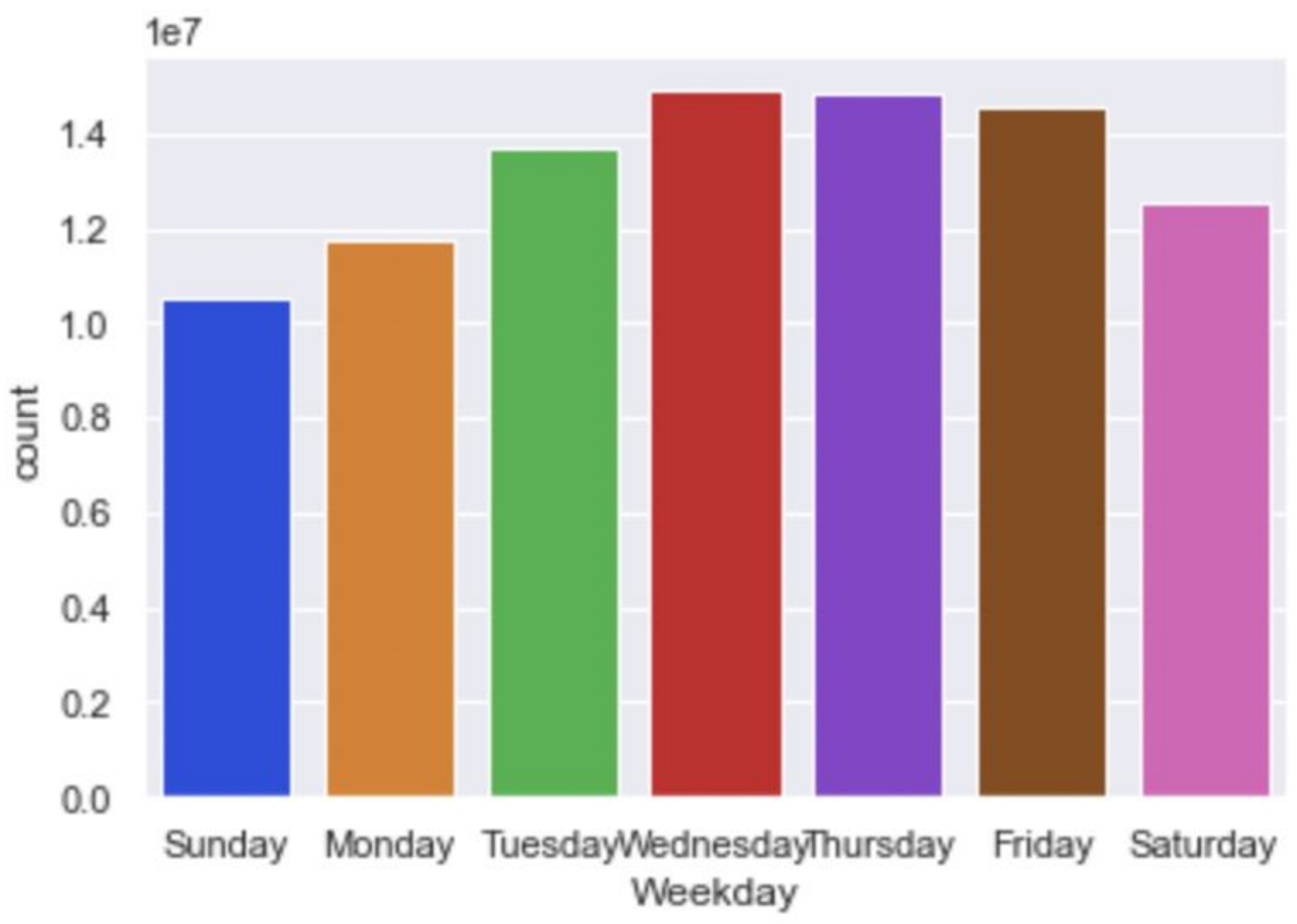
‘After one-time’ and ‘after one month’ are two major leak points for Direct Users. We would suggest:

- To push app tutorials or key feature introduction(such as car insurance, or financial services) to Direct Users after their first use.
- To pop out feature suggestions when users complete corresponding activities.
- Develop a Direct User loyalty program to keep users stay with the app

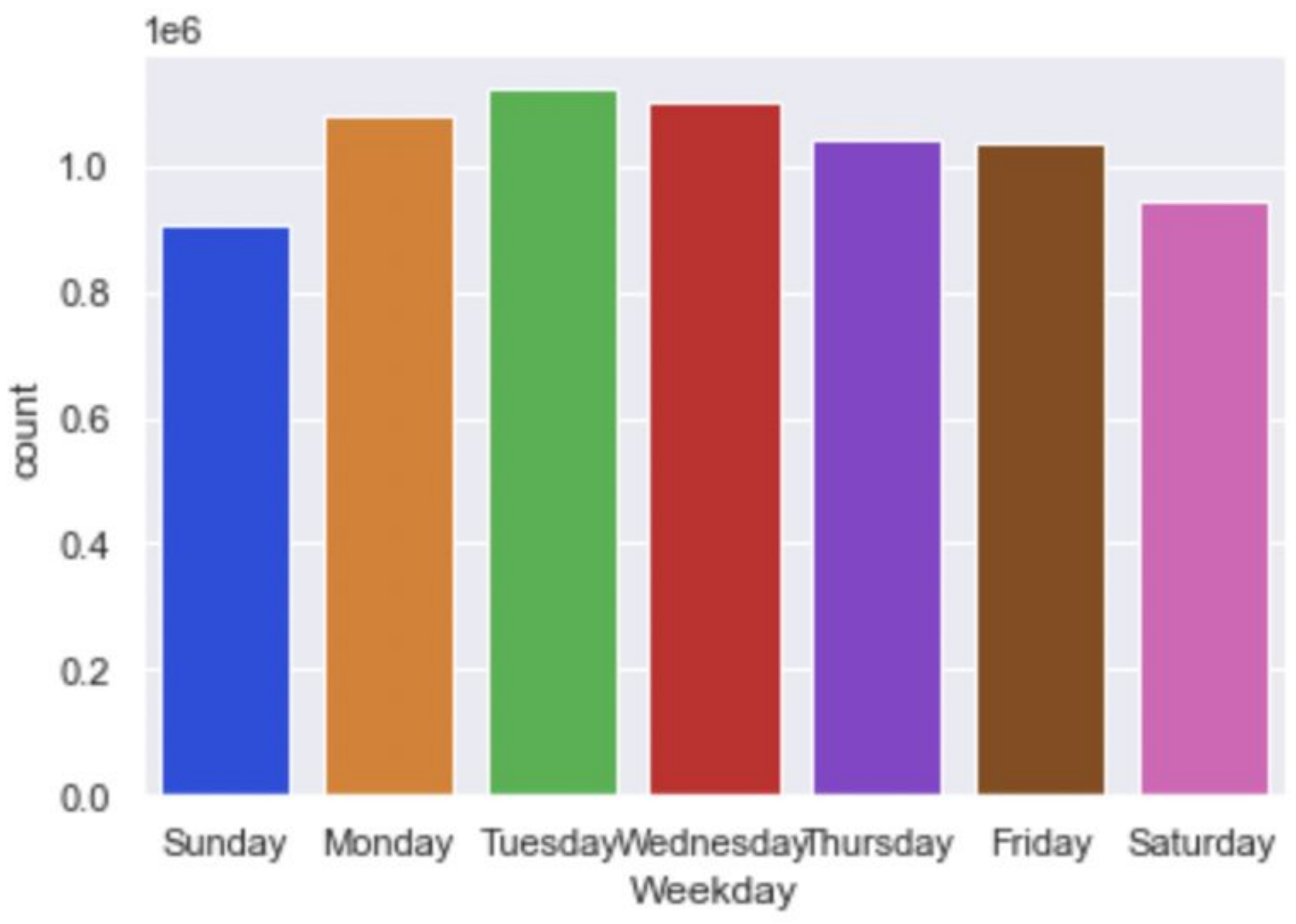
Findings on One-time Users

Among all direct users, 66% of the users logged in only once in the data time range (slightly longer than one year).

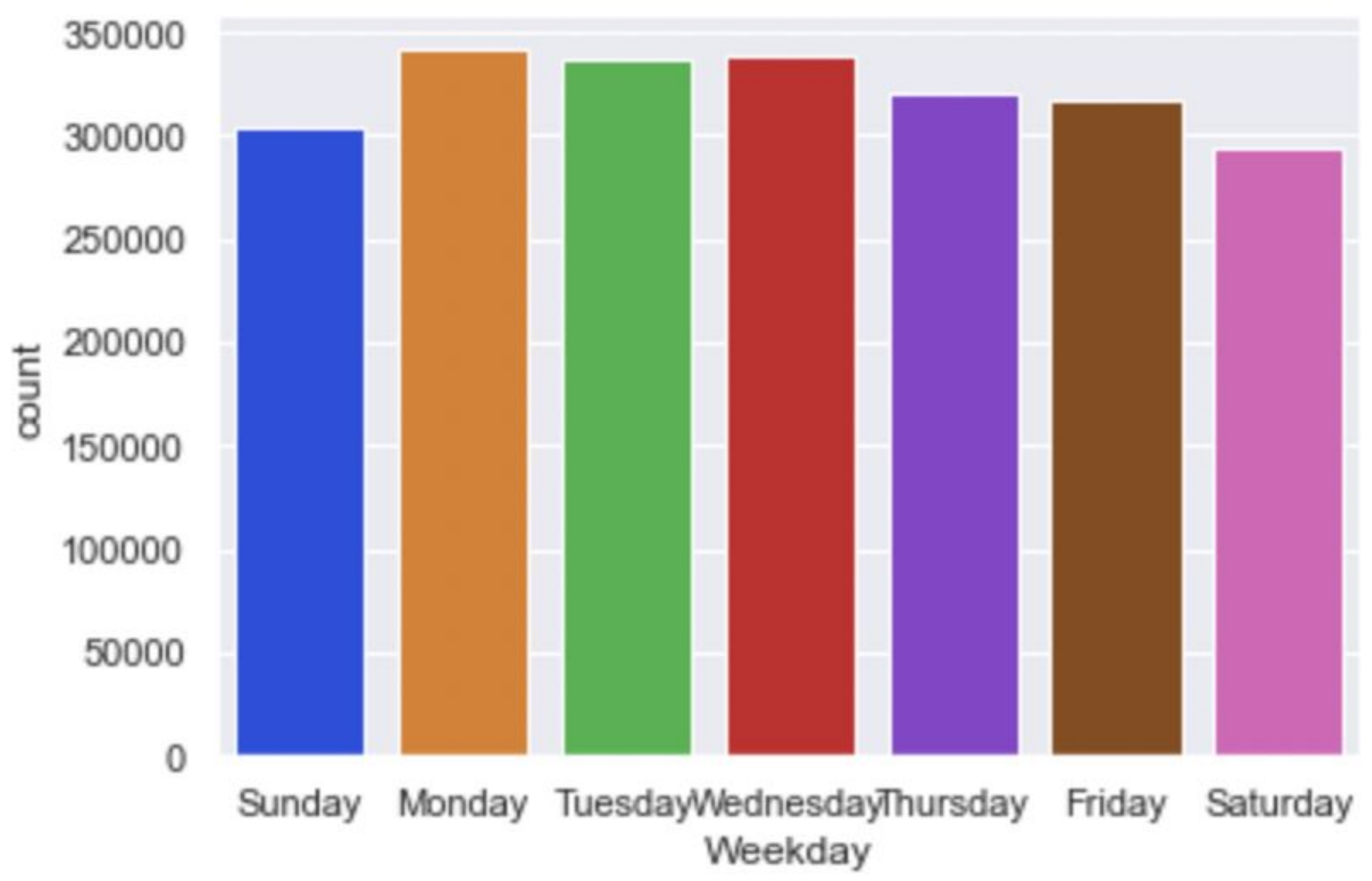
Day Time Usage



Enterprise User



Direct User

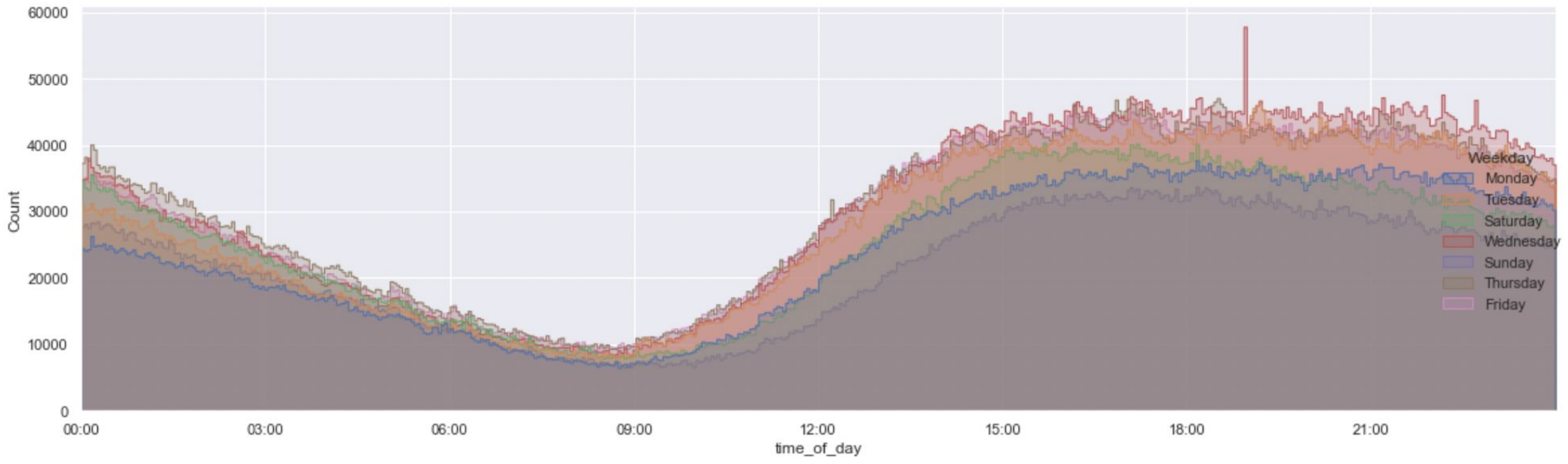


One-time User

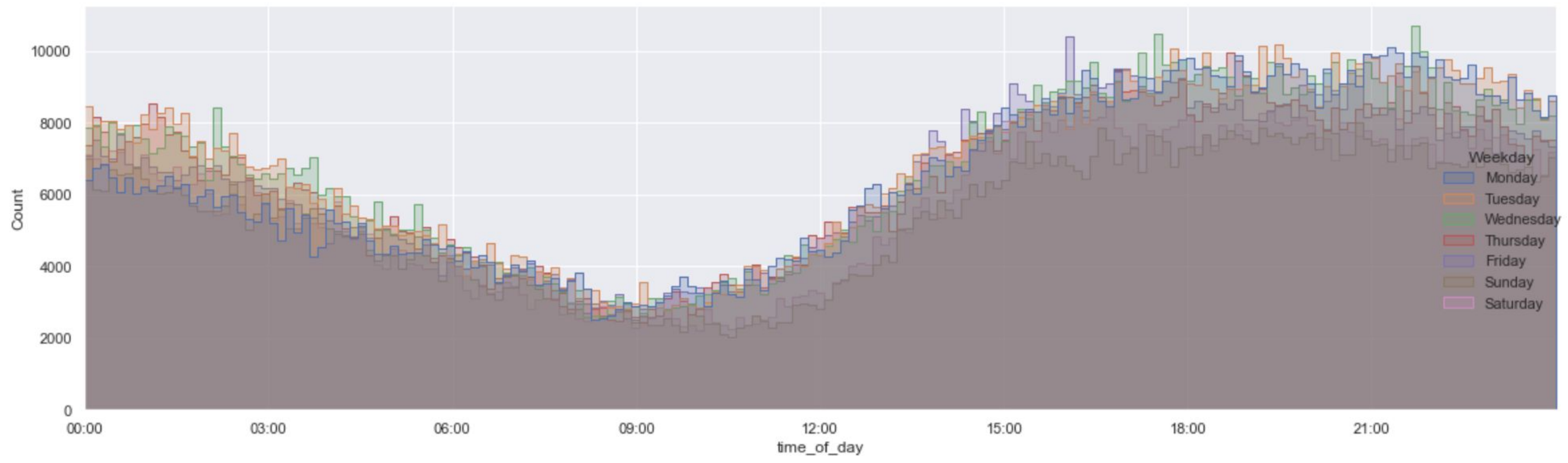
One-time users showed a heavily use trend on Monday

Day Time Usage

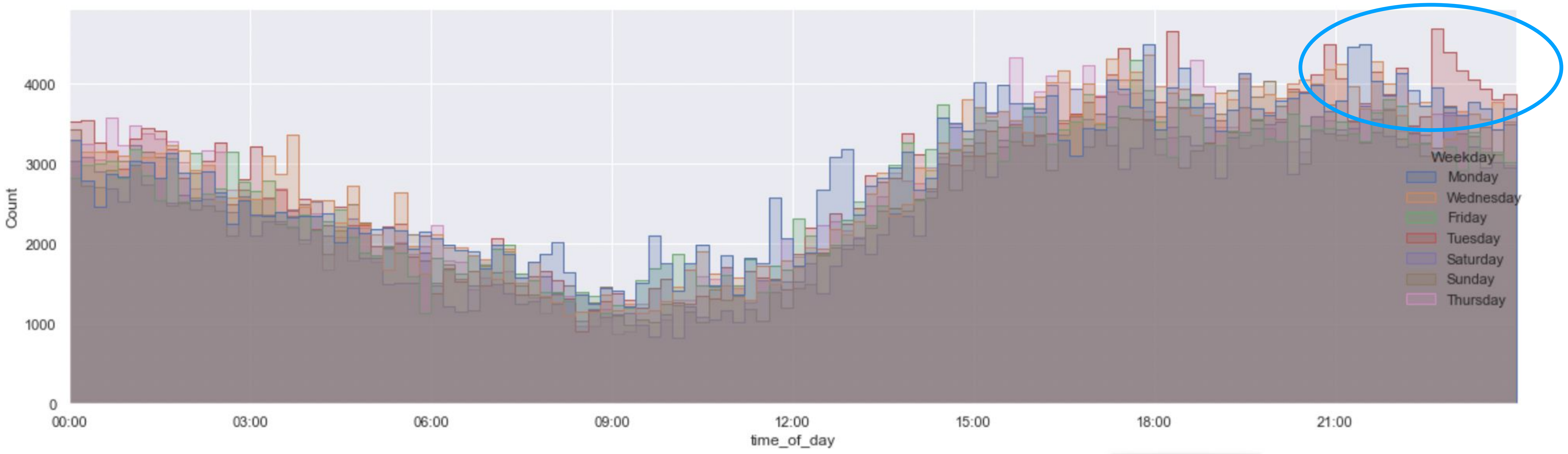
Enterprise User



Direct User



One-time User



One-time users showed usage peak on Sunday night

Start Screens

```
Card - Card Design & Terms                 81609
Payactiv-Direct - Employer                 69890
Payactiv-Direct - Home                     32566
Progress component                         31701
Signup - Employer Offer                     3413
Login                                        797
Card - Design                                332
Payactiv-Direct - Employee-ID                144
Signup - Create User                         113
Payactiv-Direct - Employee Details            82
Card - Personal Info                          58
Card - Pin Setup                              55
Card - Home                                   49
Card - Declined                               45
Signup - Createing Account                    44
Forgot Password                               37
Card - Services Dashboard                     34
Home - Profile                                30
Card - Questions                              23
accept-terms-conditions                       22
Name: SCREENNAME, dtype: int64
```

'Card Design & Terms' is the starting screen for PayAcitv Card application process.

Along with the previous graphs, we would suggest:

- A considerable amount of one-time users are new users who try to apply for PayAciv Card
- They could have potentially higher conversion possibilities on Mondays and Sunday nights
- The company can have new user acquisition campaigns targeting these two time frames

Session End Screens (drop-off)

Entire Dataset vs. One-Time User

Entire Dataset

SCREENNAME	
Payactiv-Direct - Home	333883
Home - Profile	132935
Card - Services Dashboard	97668
Payactiv-Direct - Employee-ID	67816
Payactiv-Direct - Employee Details	44414
Payactiv-Direct - Employer	42483
Card - Card Design & Terms	37172
Login	36814
Card - Home	28270
Home - Need Help	25817
Connect - Connect	24761
Card - Manage Card	22939
Activate Card - Success	20706
Savings - Home	16644
Connect - notifications	16514
Signup - Employer Offer	14293
Home - Edit Profile	14095
Connect	12988
Card - Pin Setup	12808
Card - Personal Info	11564

One-Time User

SCREENNAME	
Payactiv-Direct - Employee-ID	36008
Payactiv-Direct - Home	28553
Card - Card Design & Terms	26826
Home - Profile	17045
Payactiv-Direct - Employer	16774
Payactiv-Direct - Employee Details	15807
Signup - Employer Offer	9721
Card - Personal Info	6973
Card - Home	6336
Activate Card - Success	4974
Card - Declined	4283
Card - Services Dashboard	4196
Home - Need Help	4195
Card - Questions	3909
Card - Pin Setup	3886
Connect	3718
Card - Review Details	3045
Savings - Home	2453
Login	2196
Progress component	1865

2. Retentioneering Analysis

Segmentation

Based on number of sessions attended (4 Groups)

- Only 1 Session--- One time Visits
- Between 2 and 6 sessions
- Between 7 and 10 sessions
- Greater than 10 sessions

*number of logins session aggregated over the lifetime of datasets

Normalization over entire dataset and Feature Engineering

- weight: (USERID) means the fraction of users taking such transition route
- => by customer level
- edge weight, represented by the thickness of path, means the total number of given transitions in the dataset
- Standardize Actionname into 35 features from over 50k of unique actions

Identifying screen flows

Target events:

1. Financial Services (Savings),
2. Auto Insurance under Saving,
3. Bill Payment Management Service
4. **Direct Deposit Setup**

*Direct deposit is the most important feature as it brings cash advance to the users by allowing funds to be automatically credited to user’s account in PayActiv Apps

Findings: Most of the time not on Monetary Transactions

Get_Adjacency Matrix

	Referral	amc	billpay	card	card - Acti	card - Dire	card - Pay	connect	gas-discou	home	payactiv-c	referral	rxdiscoun	savings	savings - A	signup	turbo-tax	url
Referral	222	7	14	132	1	5	29	39	4	87	405	16	1	34	2	1	1	1
amc	5	190	16	86	0	4	14	22	11	163	277	2	11	67	9	0	1	2
billpay	12	5	1595	1932	9	46	345	725	7	1051	3157	4	4	768	4	1	6	39
card	117	65	1831	8264	2706	2626	1298	5828	147	5511	8657	55	39	5828	33	357	86	4461
card - Acti	9	1	39	1789	1842	82	12	199	5	103	1523	2	0	219	0	0	1	47
card - Dire	3	7	83	2248	26	2000	47	373	6	472	1766	1	5	474	5	2	3	191
card - Pay	31	15	328	1336	6	34	767	436	10	719	2295	11	1	233	0	0	1	21
connect	42	28	746	5416	77	253	580	3948	42	4383	7331	55	22	4077	20	2	13	231
gas-discou	6	13	16	192	0	3	4	52	193	143	364	1	7	61	5	0	0	4
home	105	179	1226	6606	76	767	863	4703	154	6940	7733	271	144	2551	211	30	55	424
payactiv-c	406	224	3256	8652	624	1588	2187	6982	379	8259	8632	63	40	5113	25	989	426	1095
referral	8	2	3	64	1	0	6	37	4	270	153	166	1	16	0	1	0	0
rxdiscoun	2	10	6	60	1	1	4	29	8	128	105	0	157	124	16	0	2	29
savings	21	77	562	4959	62	287	429	5159	67	1323	5368	9	180	3073	85	1	14	104
savings - A	0	9	5	48	0	1	1	27	5	201	112	0	15	19	73	0	0	1

- 2 centered starting point (card, PayActiv Direct)
- Most of the activities occur after users apply the card and register as direct user

Visualization of Behavior For Each of the Group

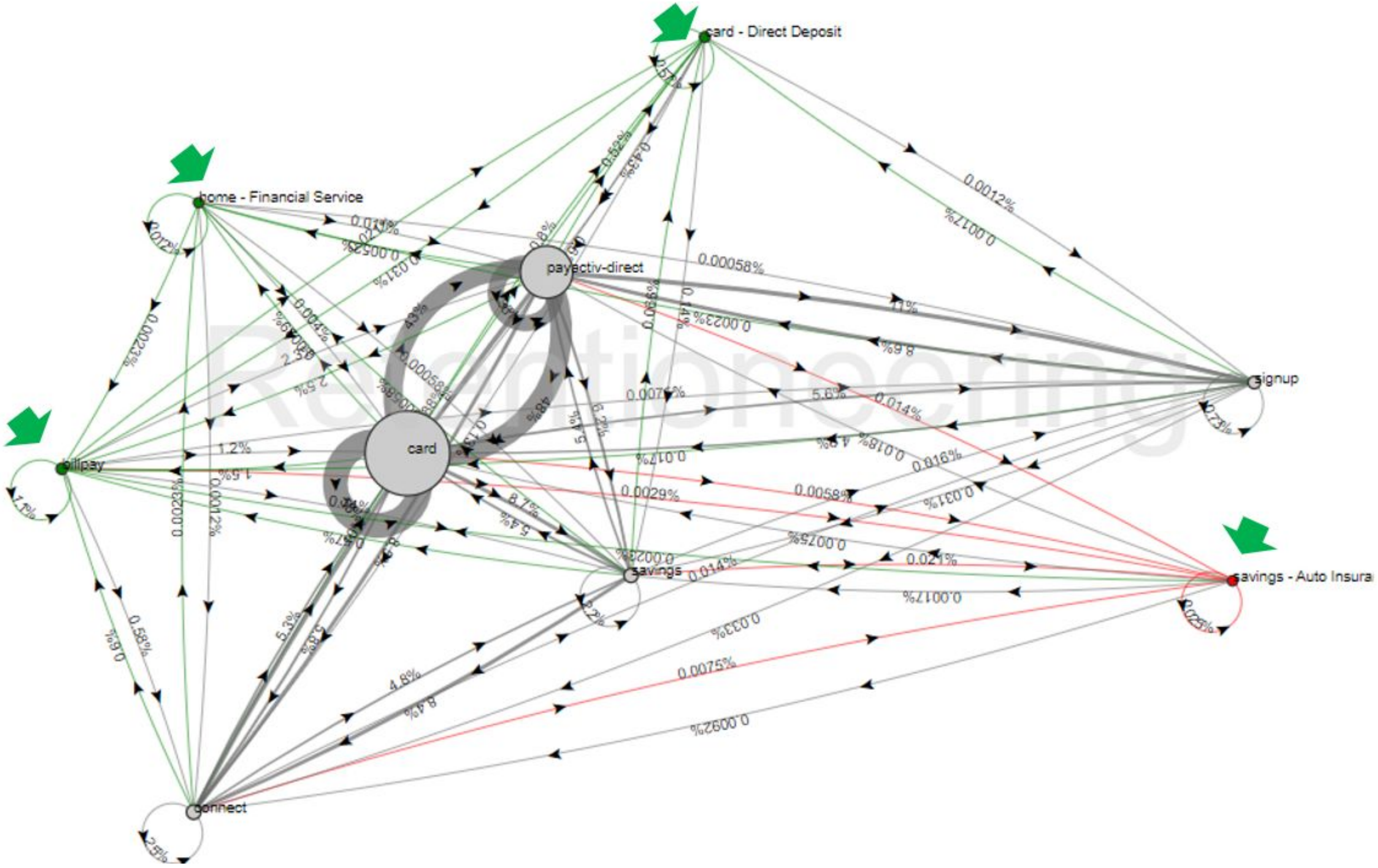
Group 1

1) Low conversion rate of Direct Deposit, 8.6% of group 1 after sign up would proceed to Payactiv Direct, 46% of which converse to direct deposit

2) Low conversion rate of the remaining targets

3) connection from bill pay to auto insurance is observed (Probably bill payment service is linked to insurance premium)

4) Strong weighted connection between Payactiv Direct and Connect (Heavy dependance on customer live agent)



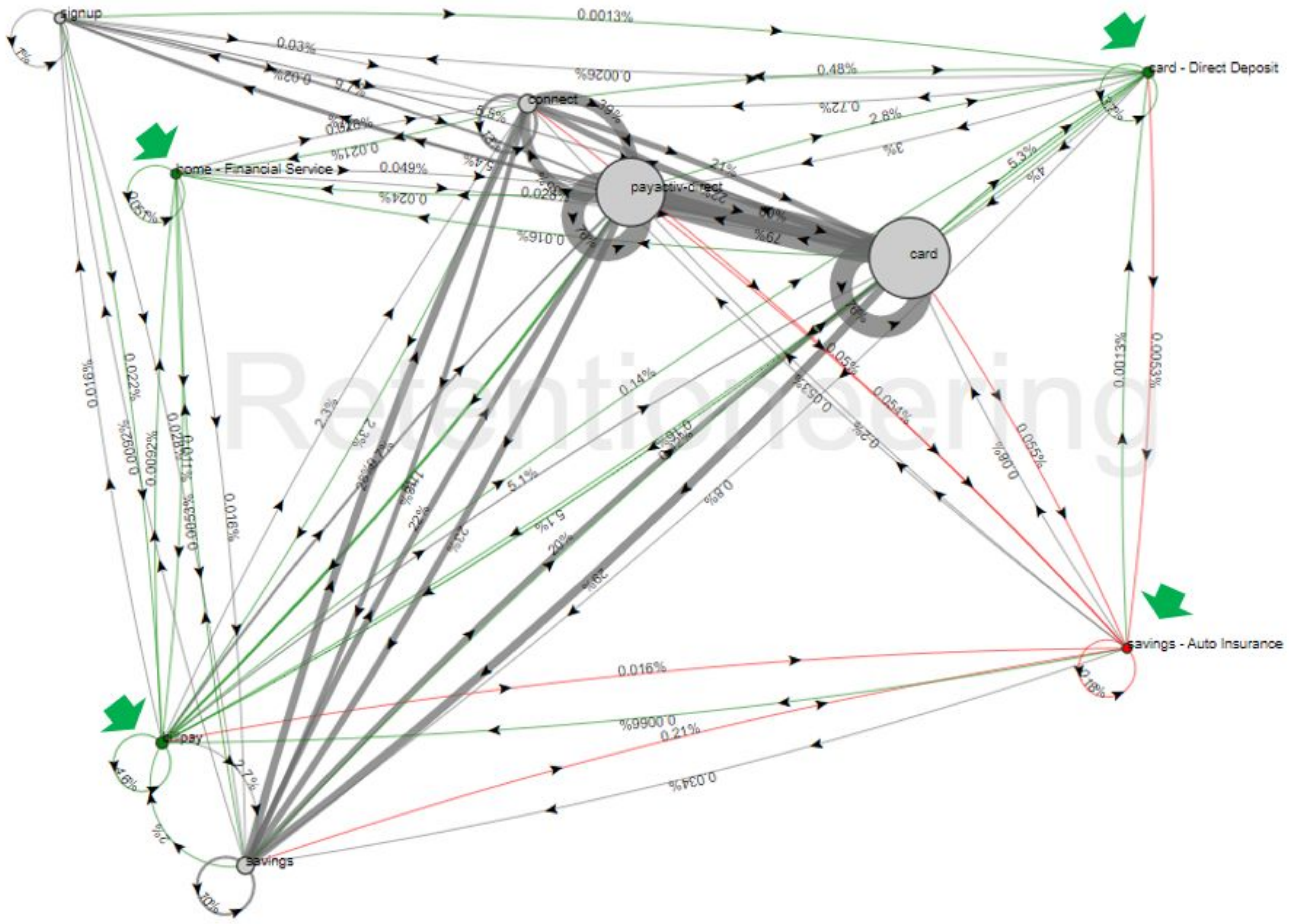
Visualization of Behavior For Each of the Group

Group 2

1) Low conversion rate of Direct Deposit, 9.7% of group 2 after sign up proceed to Payactiv Direct, 46% of which converse to applying the card

2) Low conversion rate of all the targets
3) connection from bill pay to auto insurance is observed

4) Strong weighted connection between Payactiv Direct and Connect (Heavy dependance on customer live agent)



Visualization of Behavior For Each of the Group

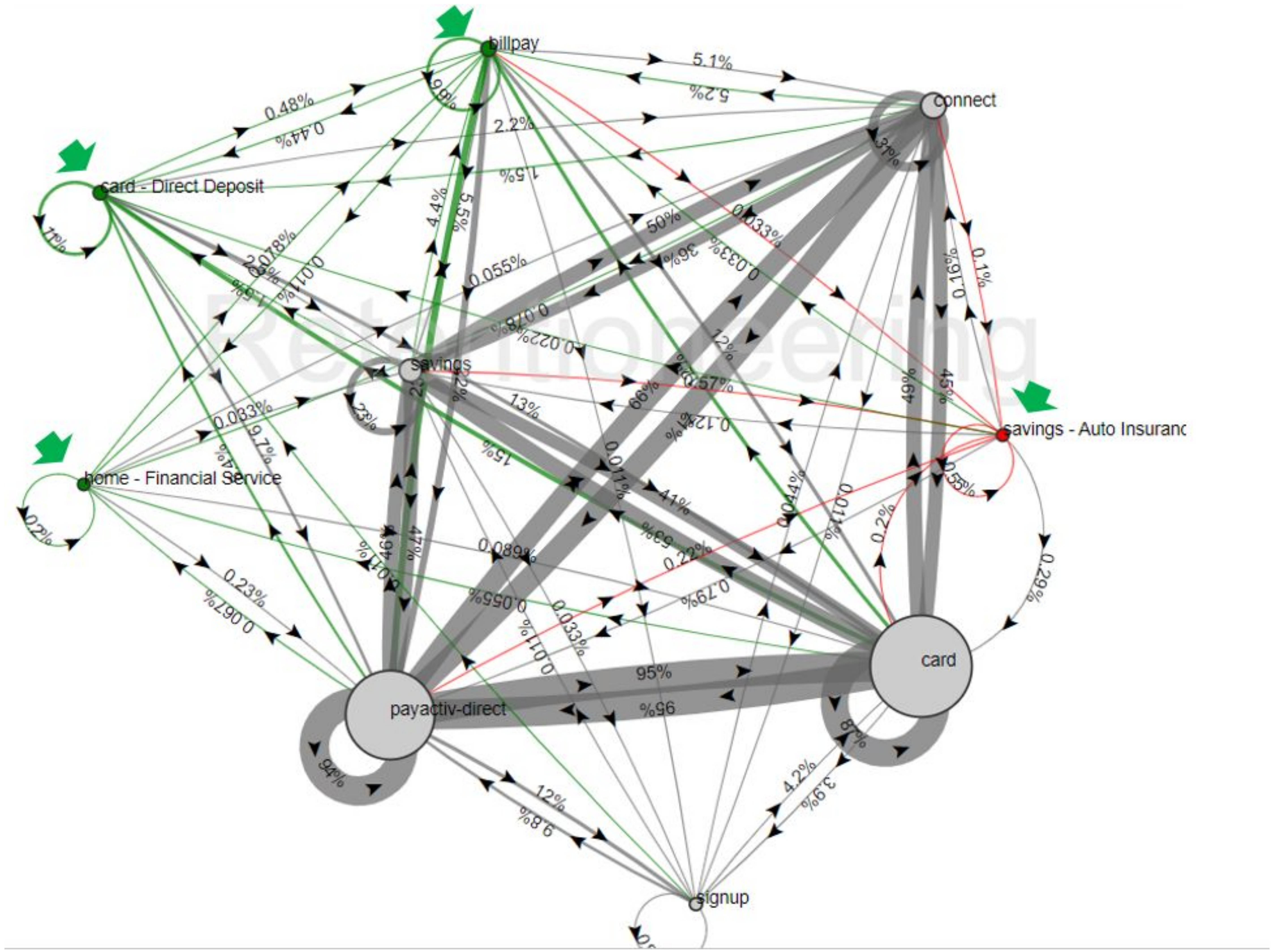
Group 3

1) Low conversion rate of Direct Deposit, 9.8% of group 3 after sign up proceed to Payactiv Direct, 95% of which converse to applying the card

2) Slightly higher conversion rate of all the targets

3) connection from bill pay to auto insurance is observed

4) Strong weighted connection between Payactiv Direct and Connect (Heavy dependance on customer live agent)



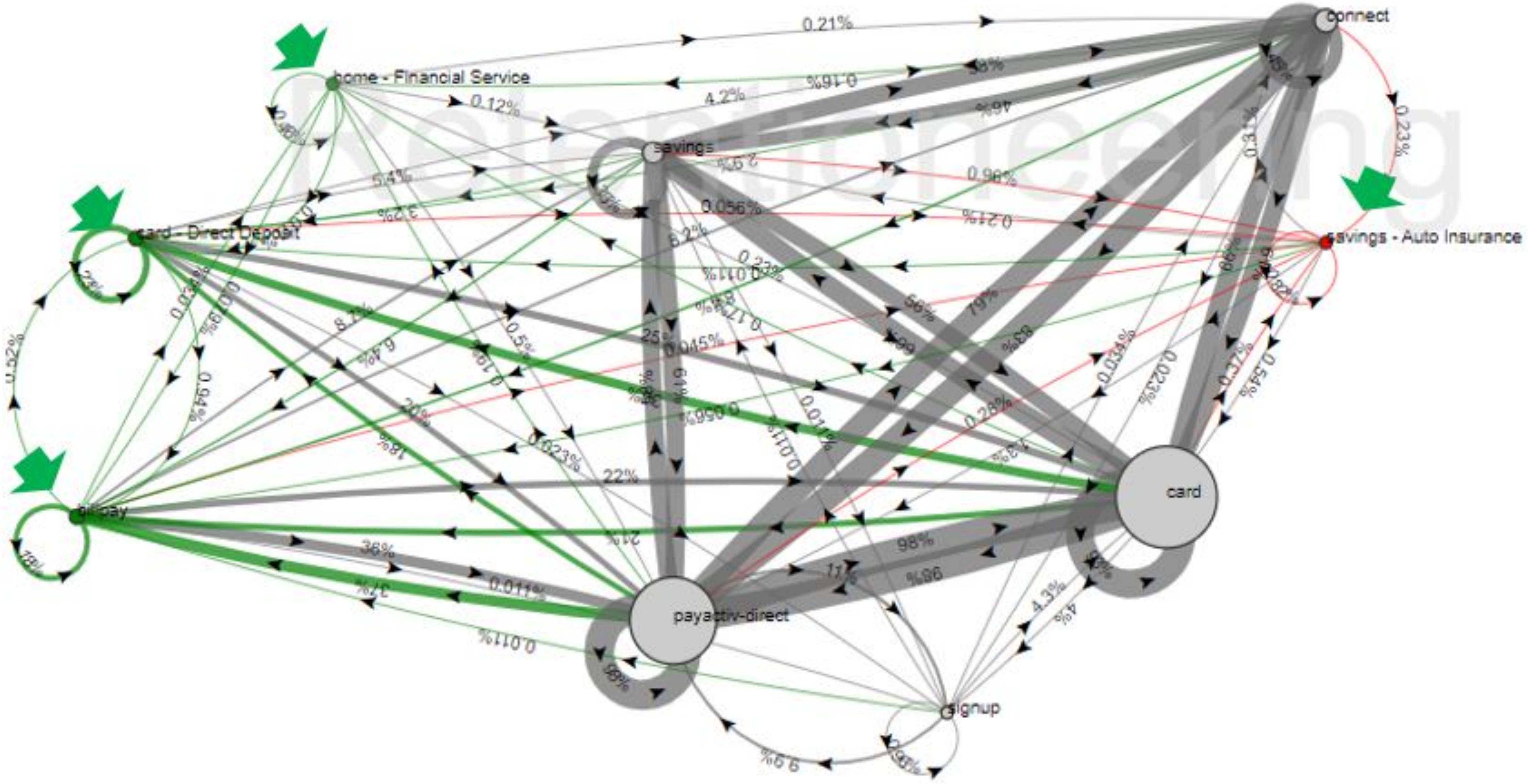
Visualization of Behavior For Each of the Group

Group 4

1) Low conversion rate of Direct Deposit, 9.9% of group 4 after sign up proceed to Payactiv Direct, 99% of which converse to applying the card

2) Highest conversion rate of all targets among the 4 groups (37% of people signing up will adopt Bill Payment Service, 19% of people signing up will enable direct deposit

3) However, low conversion rate is still observed for auto insurance



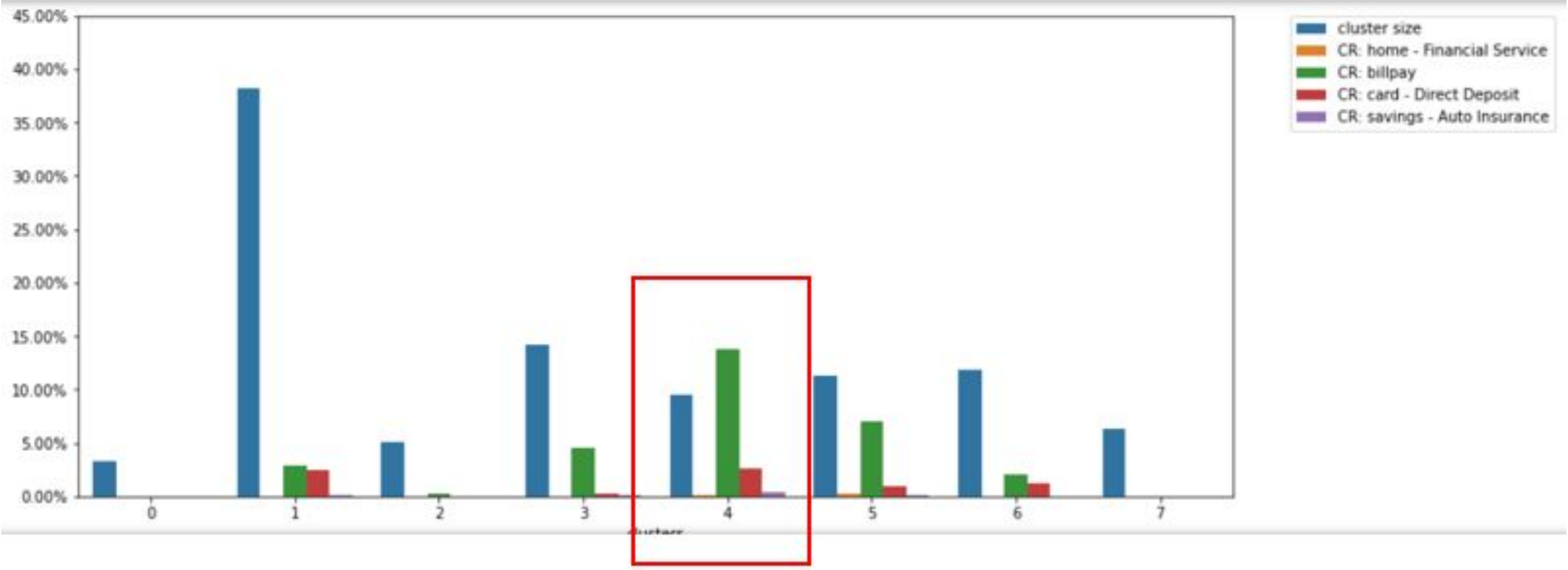
Behavioral Segmentation using Clusterization

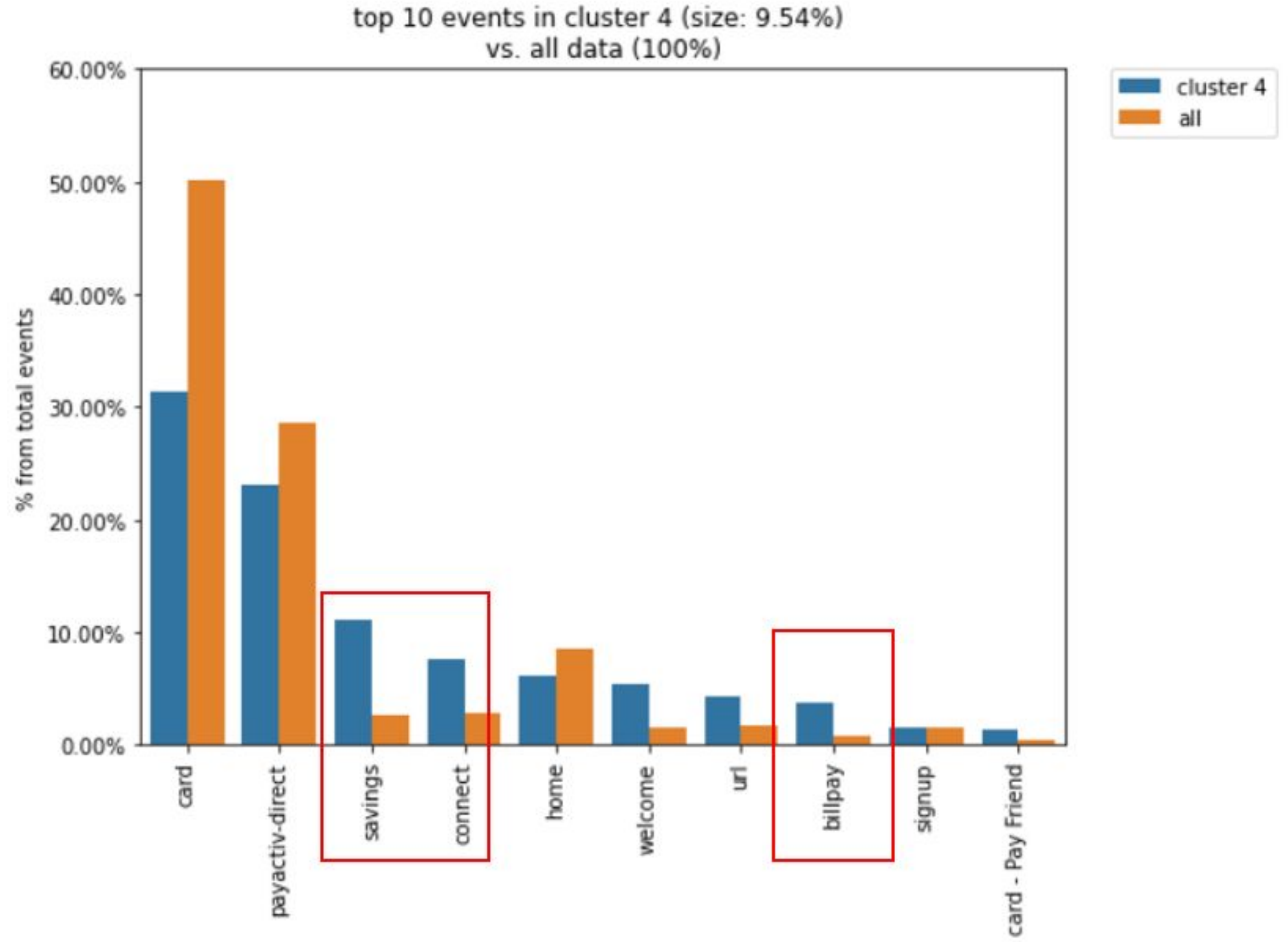
Within each group of users

- # of clusters: 8
- Clusterization methods: KMeans
- feature vectorizer: TFIDF , the significance of each user transition compared with whole population
- ngram: 1 to 2

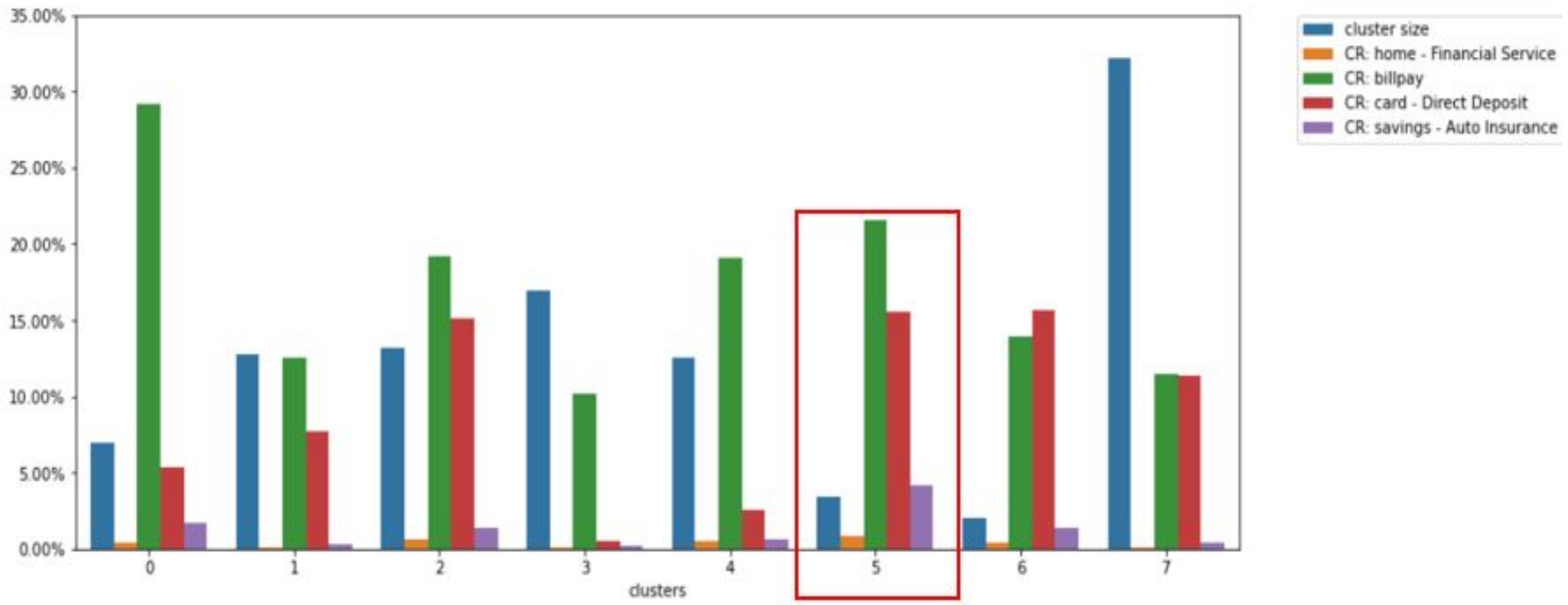
Behavioral Segmentation using Clusterization

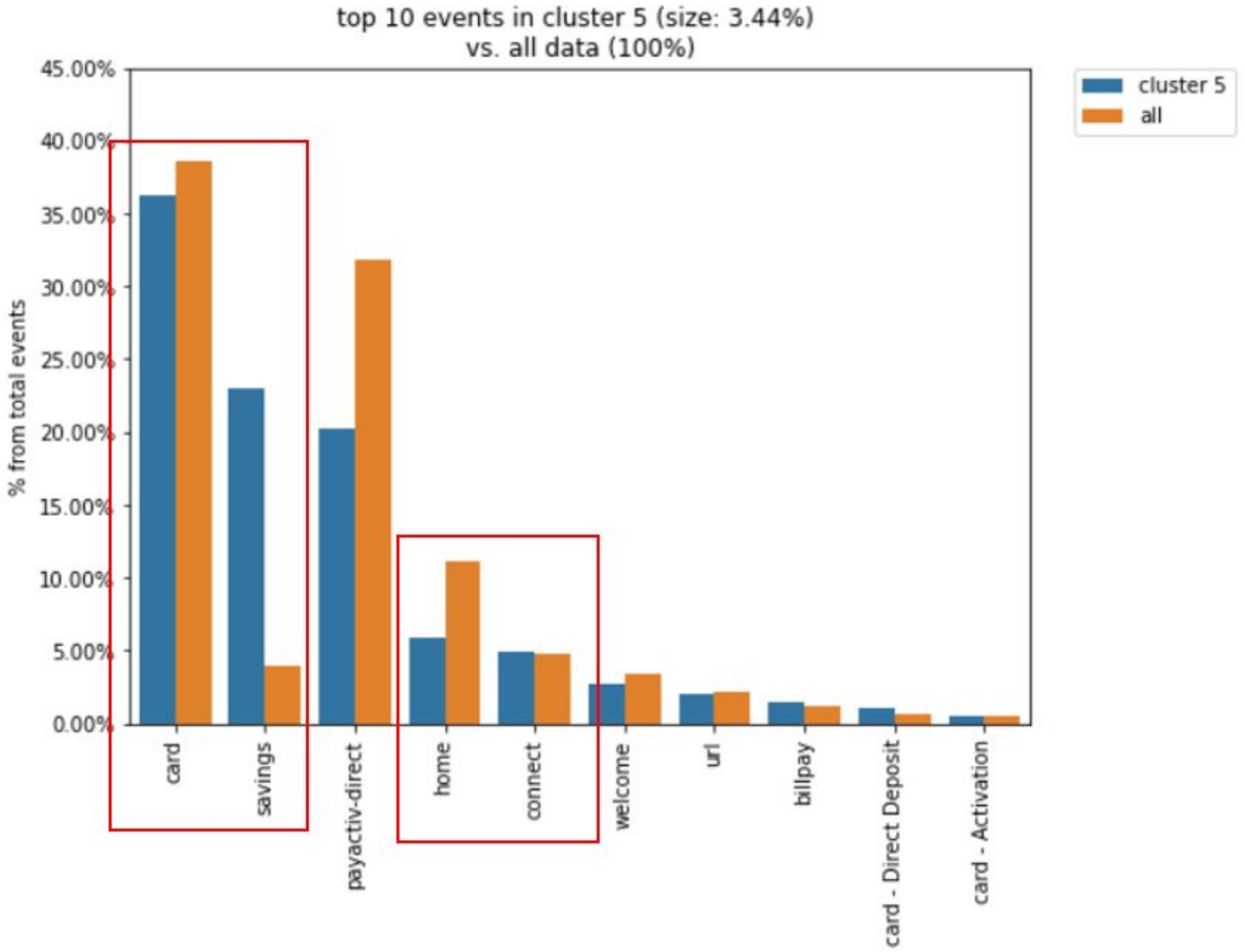
Group 1





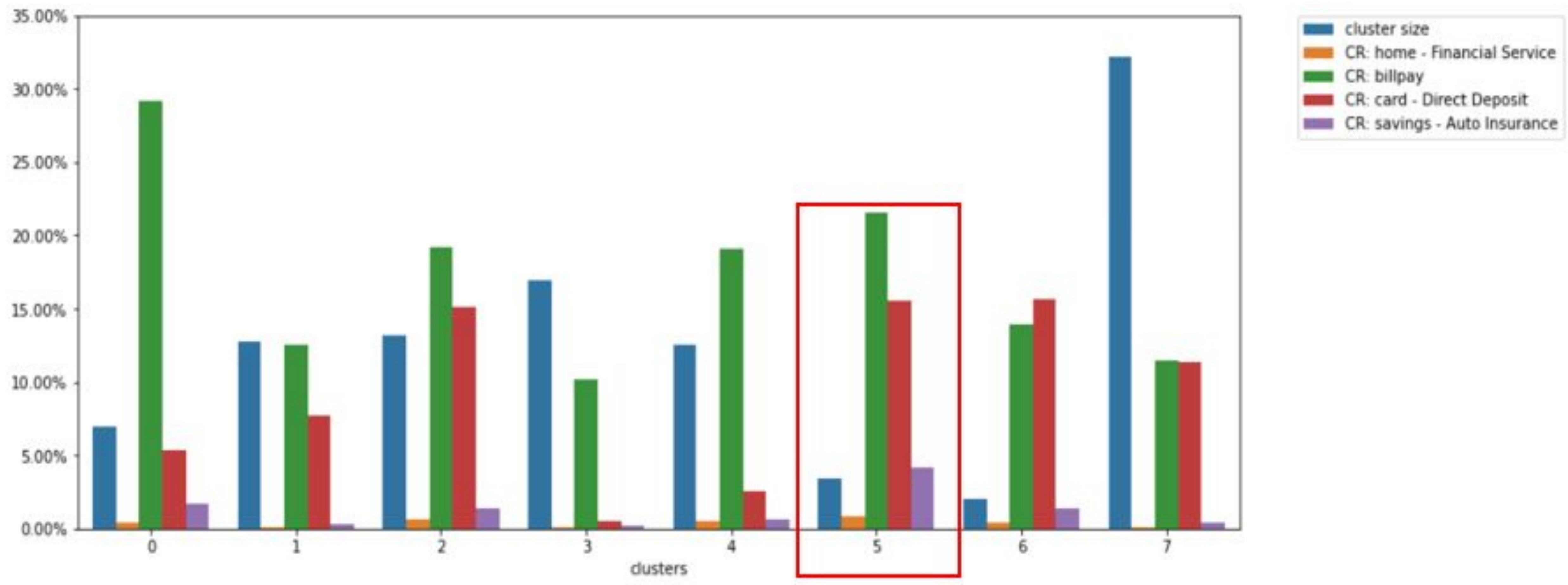
Group 2

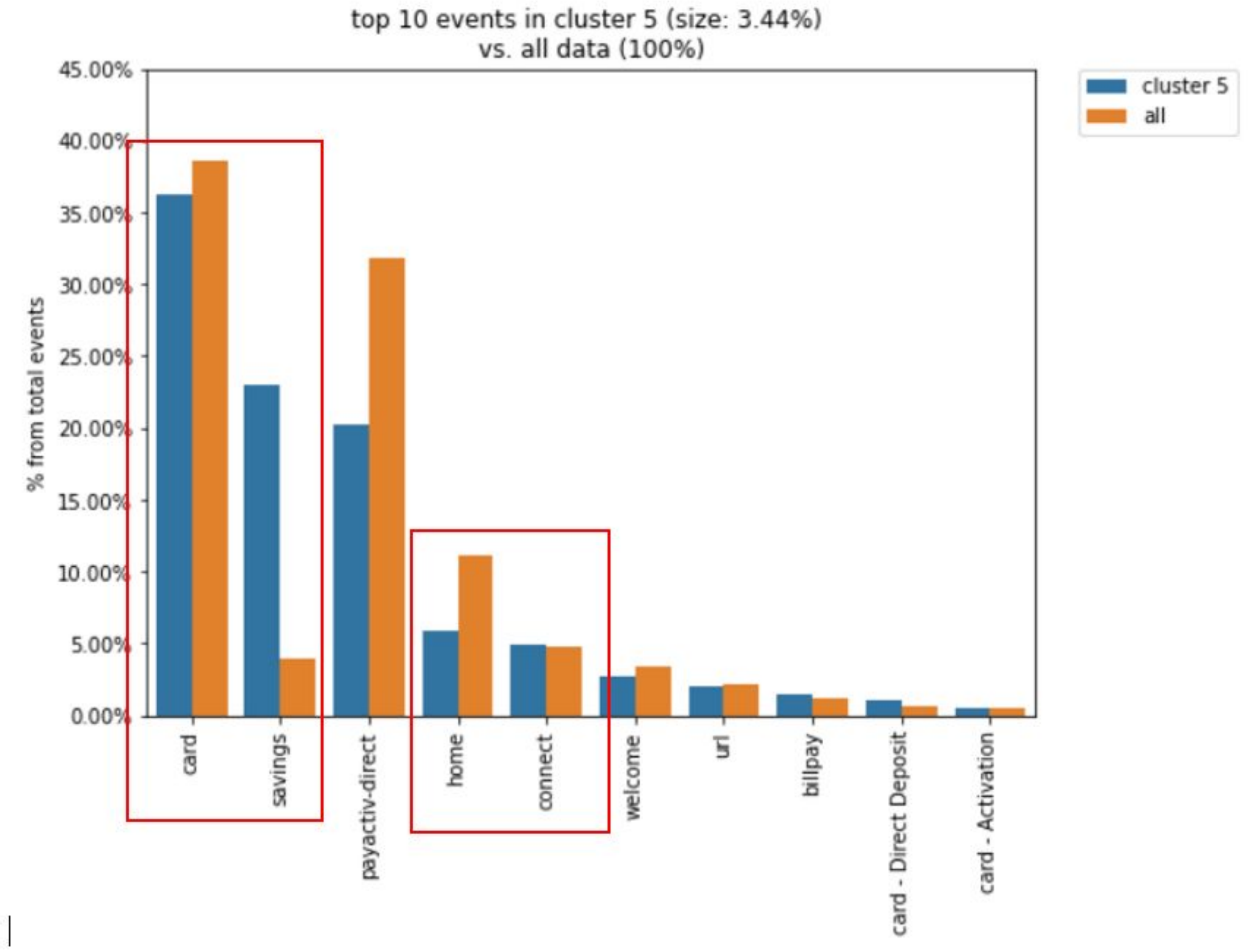




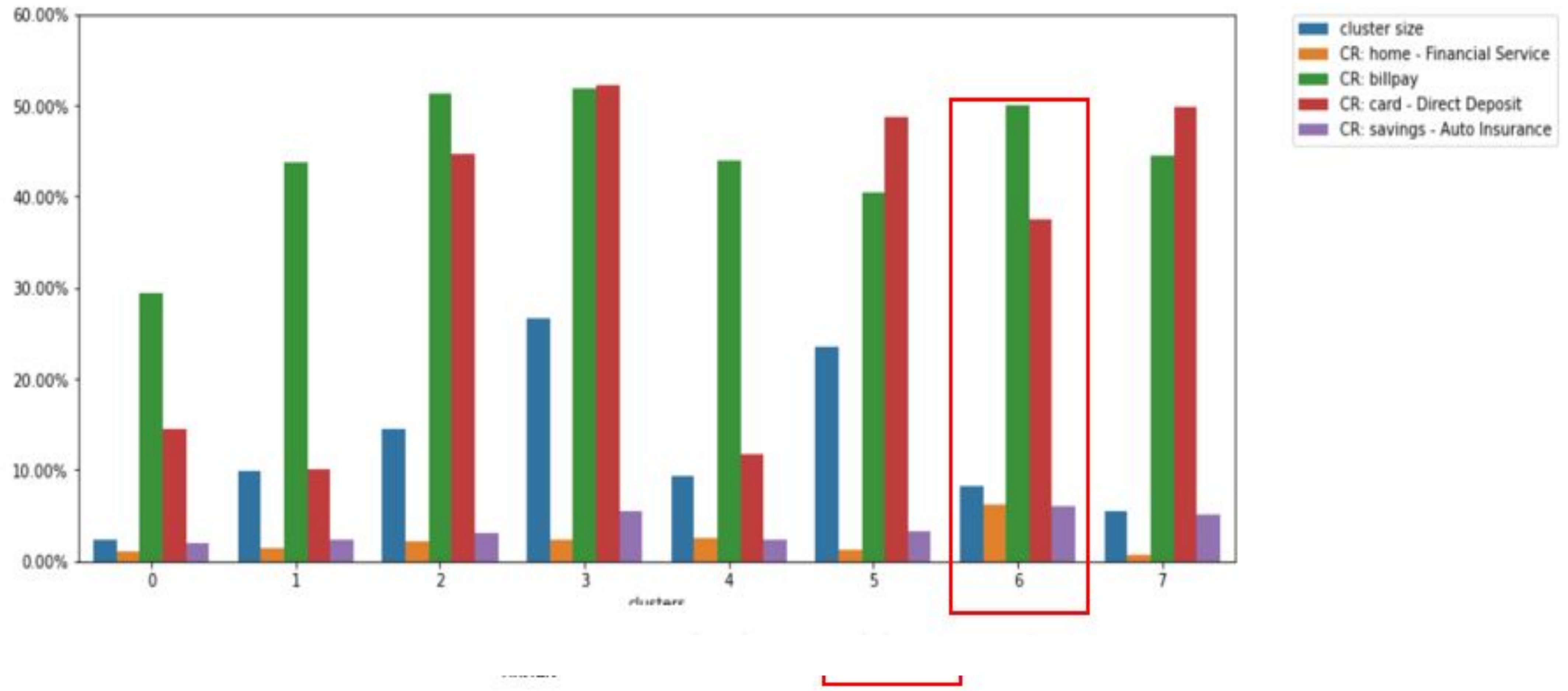
Behavioral Segmentation using Clusterization

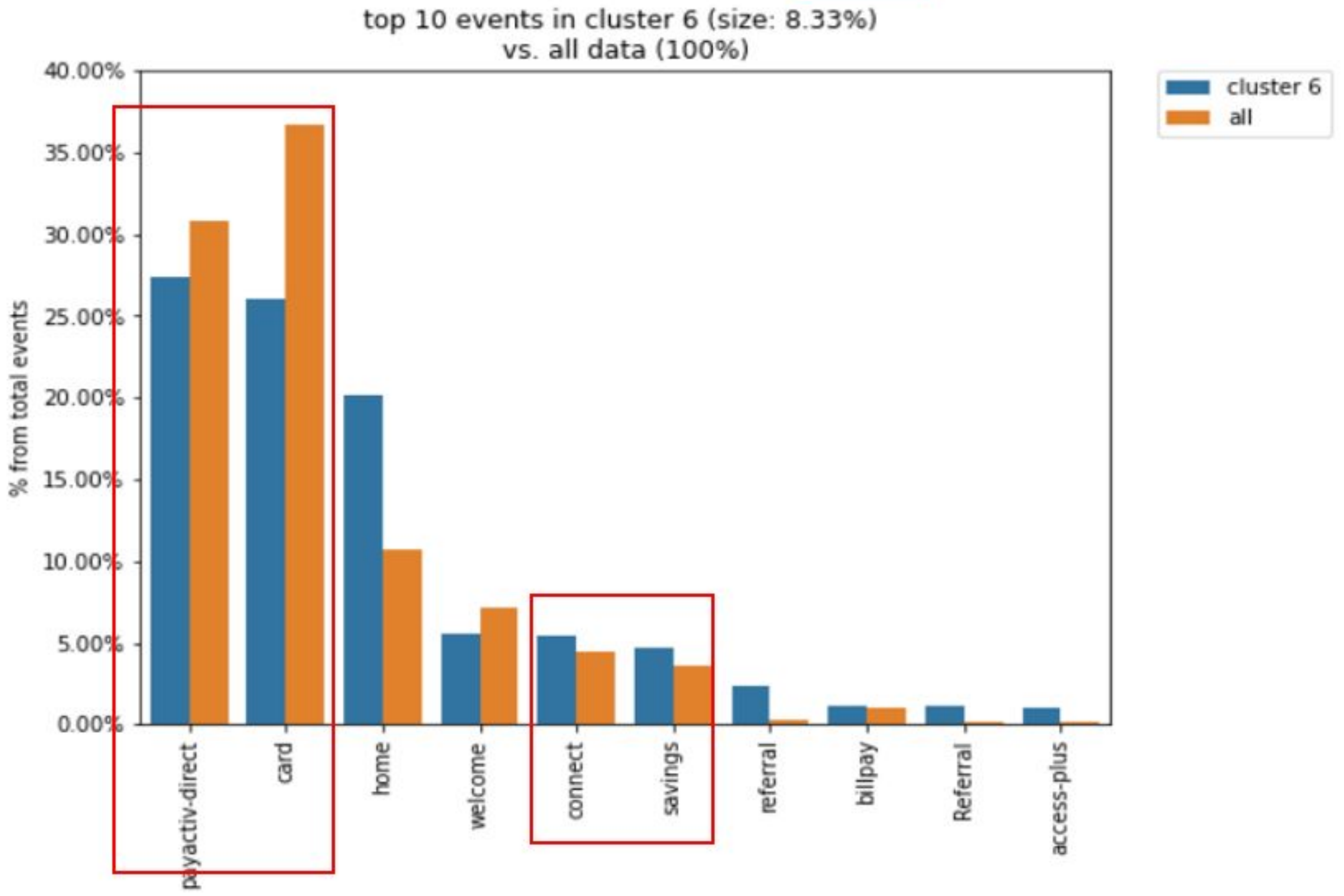
Group 3





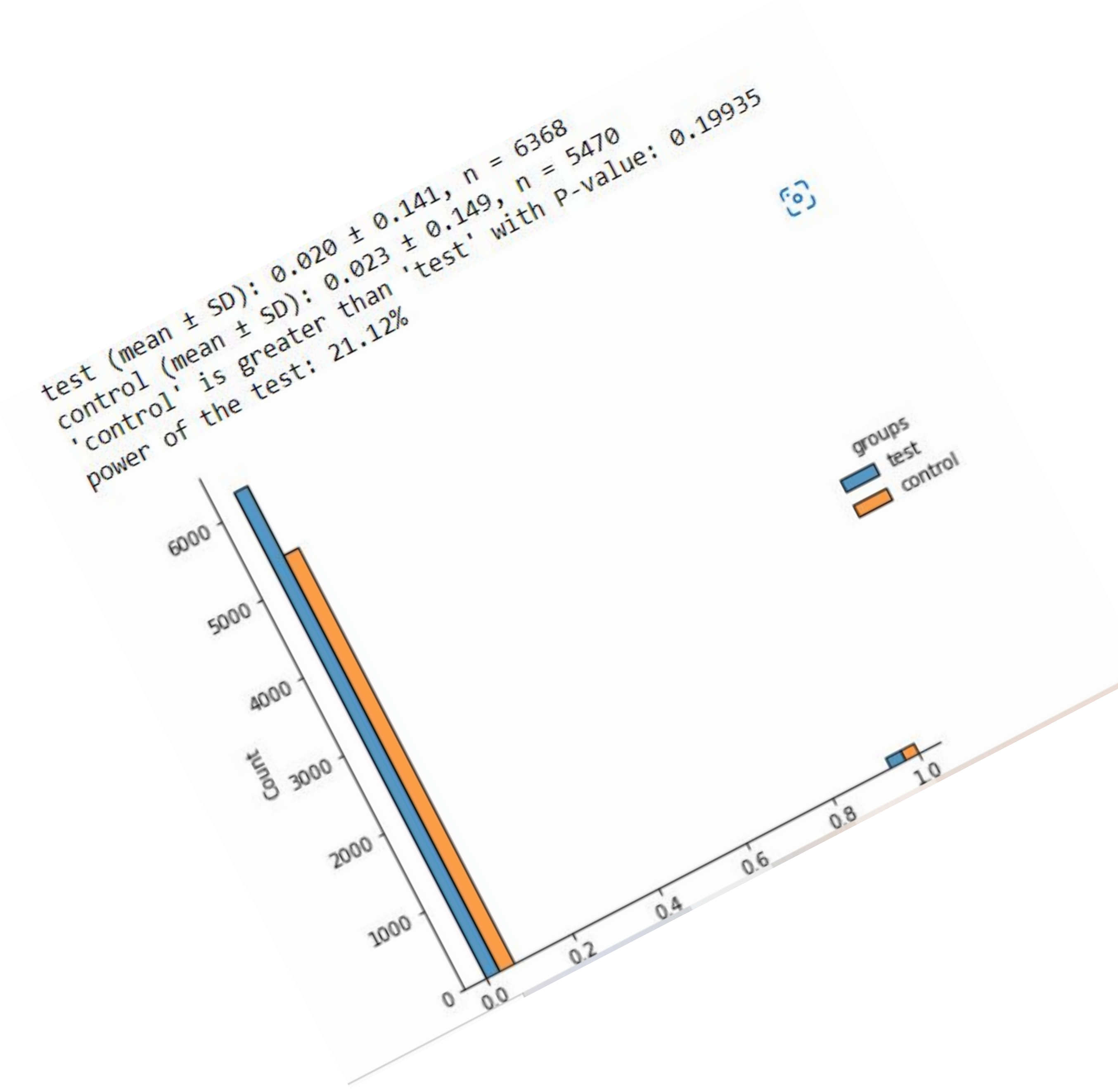
Group 4

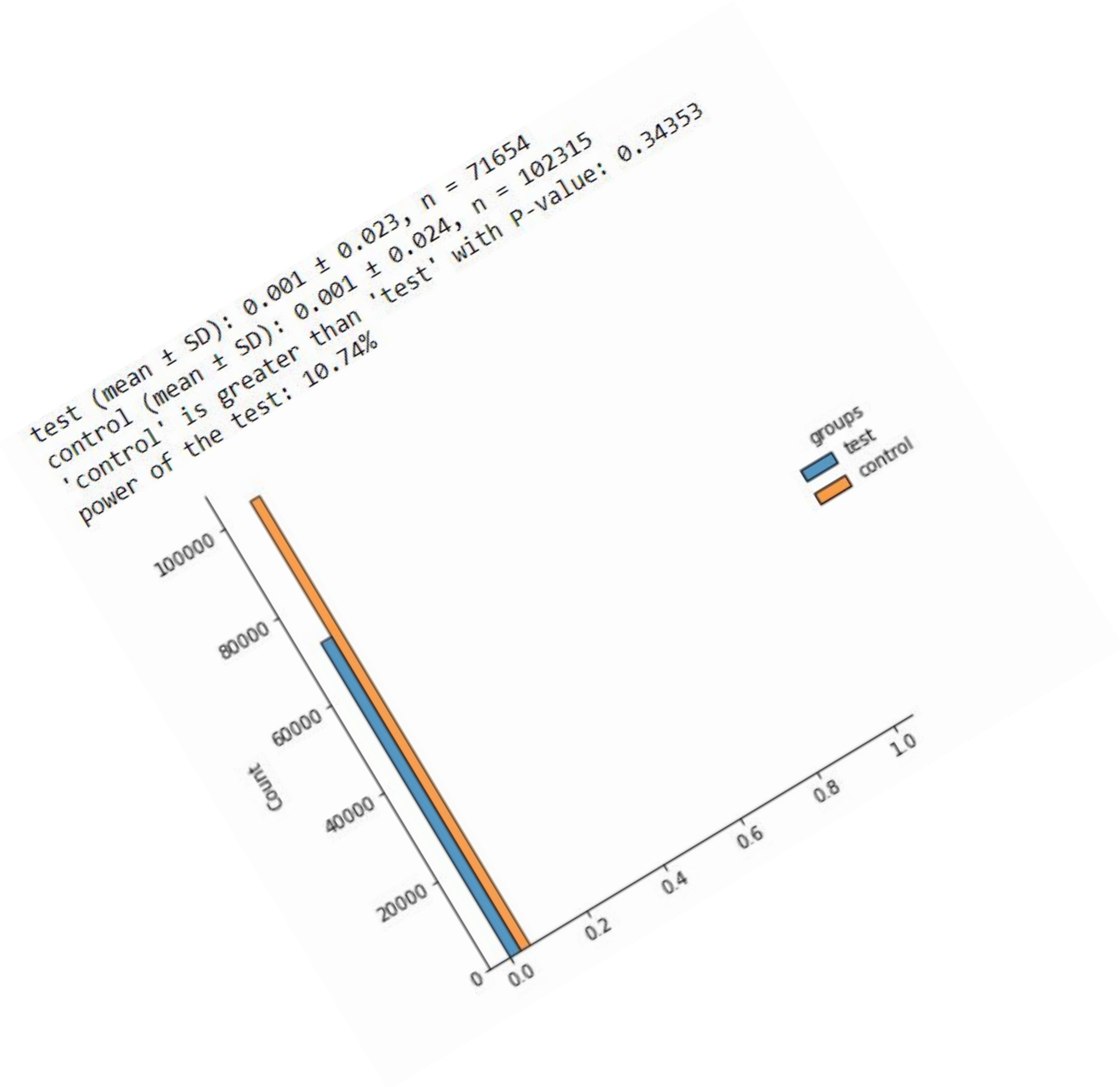


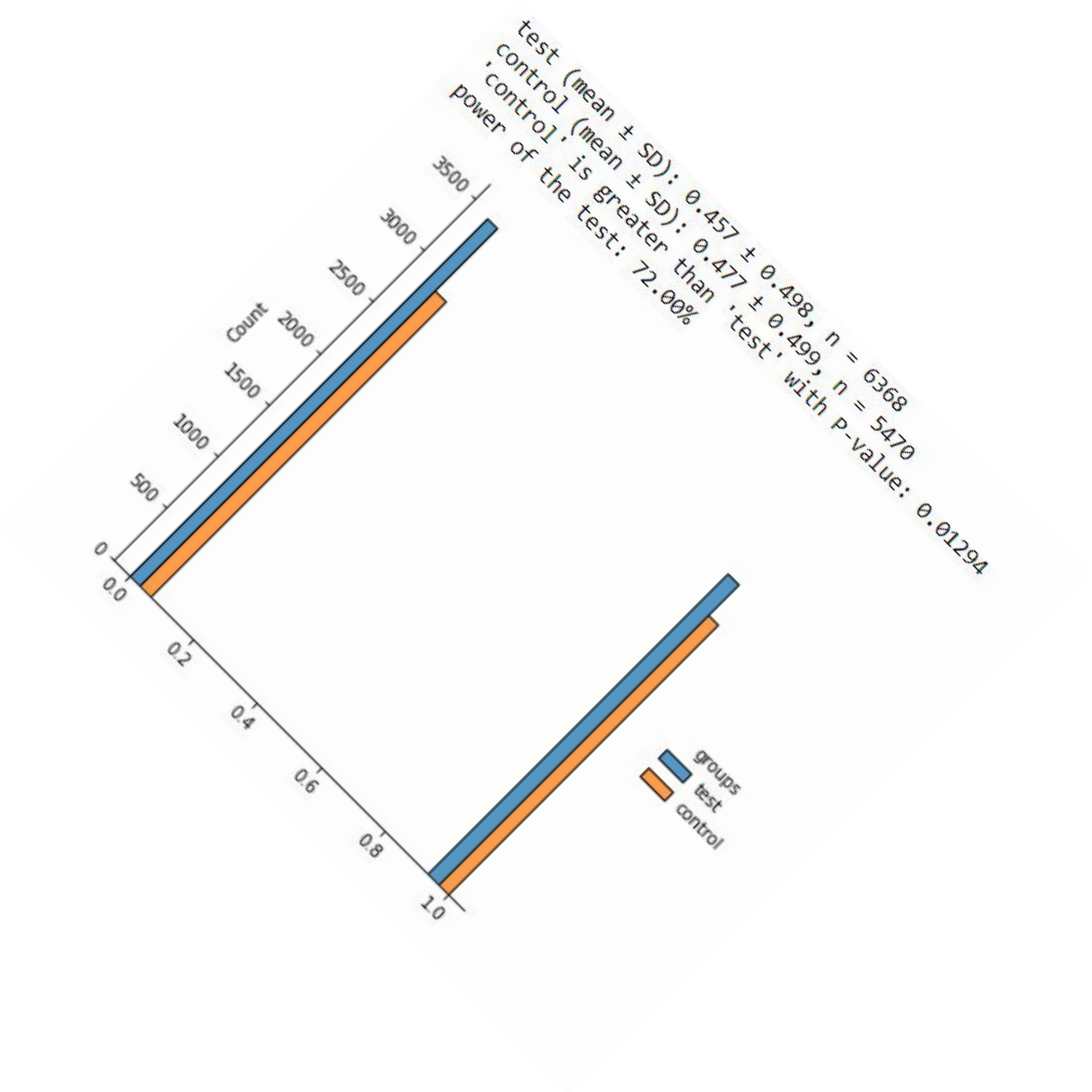


A/B Testing and Experimentation

- Using Jan 2022 as a timeline to split the data into test and control groups.
- Test: One sided T test with 95% testing level
- Testing variable: To test the conversion rate of the four target events







A/B Testing and Experimentation

- Direct Deposit conversion is more significant in 2022, indicating the demand of payroll advance access than the group in 2021

- Direct Deposit usually comes in effective later than the other target events, reflecting the difficulty to get the service effective

- For those who have control group being more significant than the test group, it means the features is not being used continuously in 2022, leak of users is a sign

Group	Target event	P Value	Conclusion
1	Home- Financial Service	control is greater than test: 0.33616	Fail to reject null hypothesis
1	Bill Management Service	control is greater than test: 0.0000	Control group has higher conversion rate
1	Direct Deposit Setup	test is greater than control: 0.0000	Test group has higher conversion rate
1	Savings - Auto Insurance	control is greater than test: 0.00125	Control group has higher conversion rate
2	Home- Financial Service	control is greater than test: 0.03591	Control group has higher conversion rate
2	Bill Management Service	test is greater than control: 0.13505	Fail to reject null hypothesis
2	Direct Deposit Setup	test is greater than control: 0.0000	Test group has higher conversion rate
2	Savings - Auto Insurance	test is greater than control: 0.06233	Fail to reject null hypothesis
3	Home- Financial Service	control is greater than test: 0.47157	Fail to reject null hypothesis
3	Bill Management Service	control is greater than test: 0.00556	Control group has higher conversion rate
3	Direct Deposit Setup	test is greater than control: 0.02266	Test group has higher conversion rate
3	Savings - Auto Insurance	control is greater than test: 0.087	Fail to reject null hypothesis
4	Home- Financial Service	control is greater than test: 0.22347	Fail to reject null hypothesis
4	Bill Management Service	control is greater than test: 0.0547	Fail to reject null hypothesis
4	Direct Deposit Setup	test is greater than control: 0.00008	Test group has higher conversion rate
4	Savings - Auto Insurance	control is greater than test: 0.22266	Fail to reject null hypothesis

Recommandation and insight

Live Agent

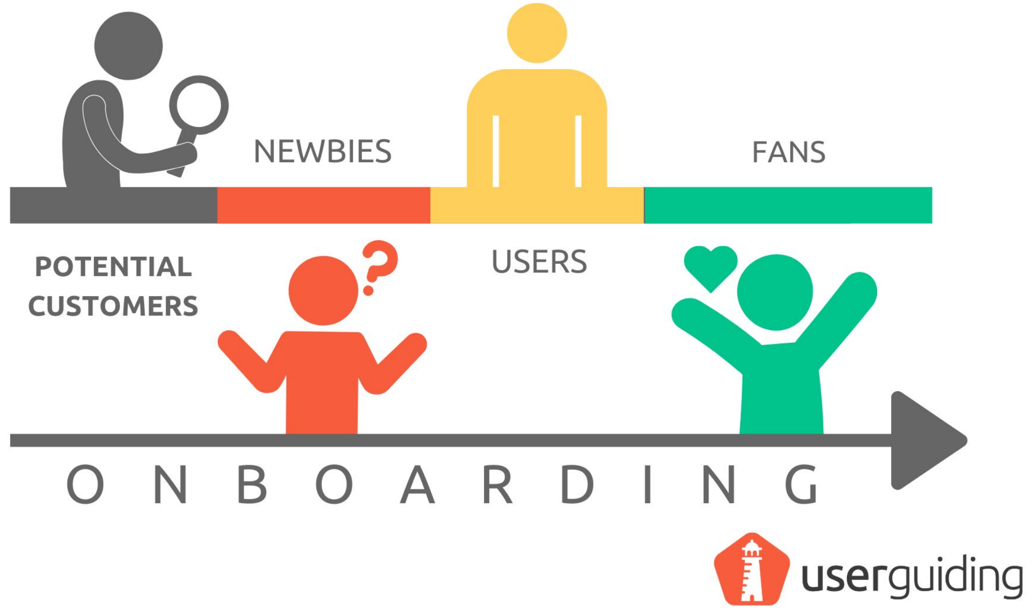
- Strong edge linkage between Connect and PayActiv Direct registration service
- Moderate to heavy dependance on customer live agent service
- Chatbot root cause analysis

Launch Car Financing Product

- given that prior to using PayActiv app, most of the users already possess a vehicle and no significant demand for purchasing new vehicle
- launch insurance bridging program, but need to know the current user insurance subscription expiry date, ads is needed to pop out to prompt users.

Apps Enhancement

- by default put the in-app product subscription service into the in-app bill payment service, or provide some pop-out to encourage users to converge to bill payment service after subscripting to in-app product service



PayActiv Direct Onboarding

- Getting user to successfully on-board PayActiv Direct service is the key to increased bank/card linked to PayActiv apps.
- Actively look into the Plaid search and study the companies (by names, segment) to expand customer outreach team to promote the apps and actively help user/company to onboard PayActiv apps,

Machine Learning Pipeline

1. Feature Engineering: Label user status as inactive after idling for two months
2. Goal: To predict user drop off (target var = status)
3. Vectorizer: TFIDF
4. Classifiers: Logistics Regression, Random Forest Classifiers

```
from sklearn.linear_model import LogisticRegression
from sklearn import metrics
clf = LogisticRegression(solver='lbfgs', multi_class='auto', max_iter=5000, random_state=42)
clf.fit(X_train, y_train)
accuracy = clf.score(X_test, y_test)
print('Accuracy score of the {} is {:.2f}'.format(clf.__class__.__name__, accuracy))
```

```
LogisticRegression(max_iter=5000, random_state=42)

Accuracy score of the LogisticRegression is 0.90
```

```
from sklearn.ensemble import RandomForestClassifier

clf = RandomForestClassifier(n_estimators=100, n_jobs=-1, random_state=42)
clf.fit(X_train, y_train)
accuracy = clf.score(X_test, y_test)
print('Accuracy score of the {} is {:.2f}'.format(clf.__class__.__name__, accuracy))
```

```
RandomForestClassifier(n_jobs=-1, random_state=42)

Accuracy score of the RandomForestClassifier is 0.94
```

Project Benefits to the Company

Project Benefits

1. Exploratory Analysis Findings: Visualize Patterns and trends across seasons
2. A/B Testing Findings: Predict customer value and optimize operations
3. Recommendations:

a. Launching auto financing programs to boost up auto-insurance subscription rate

b. Some pop-out to encourage users to converge to bill payment service

4. Dropout Screen Analysis: Enhance customer journey maps and reveal the attribution of initiatives launched towards the customer experience.

Future Scope

- When more User Data (like Demographics, Locations, Transactions) can be made available, environmental attributes that can influence an outcome can be identified
- A more promising approach to deep learning for Customer Behavior Analytics is to use Recurrent Neural Networks (RNN)

Thank you