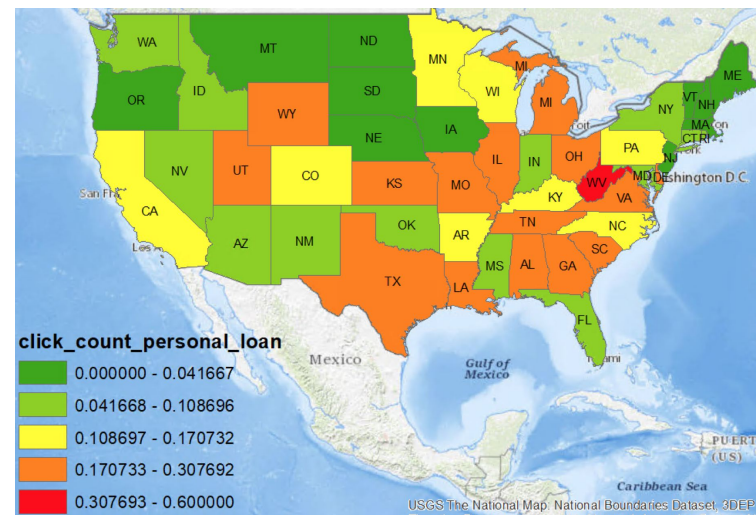
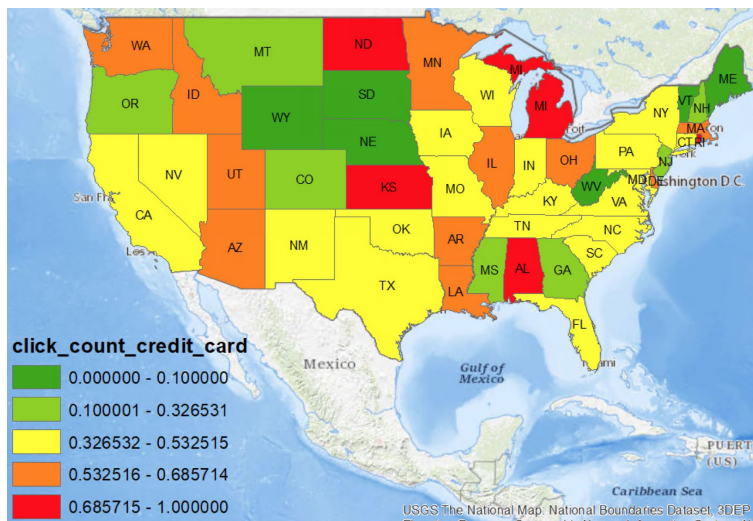


# Credit Sesame User Clustering and Product Offering Predicting

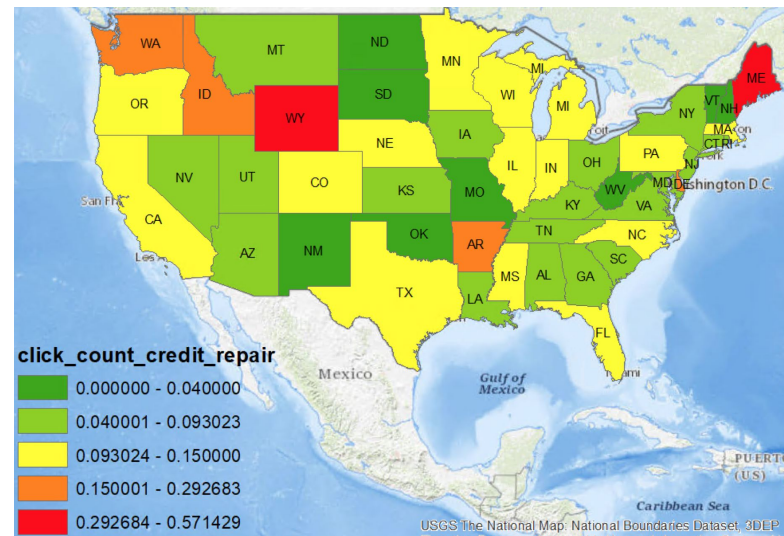
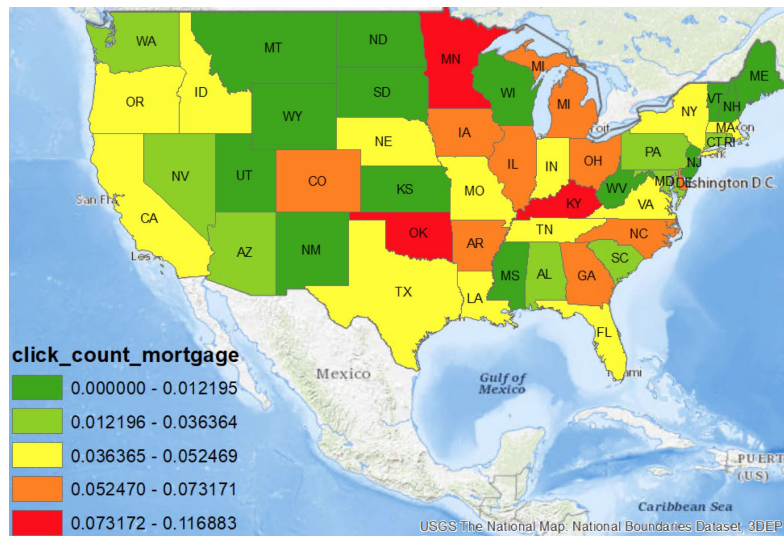
# Goal

- Made clustering analysis on users
- Classify new user to most likely label
- Predict what kind of product may interest new user
- Visualize the data to better help company to do product recommendation

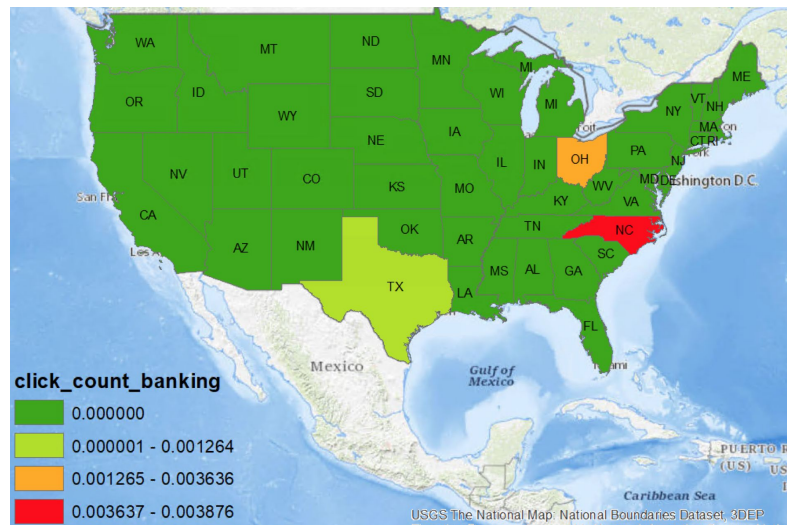
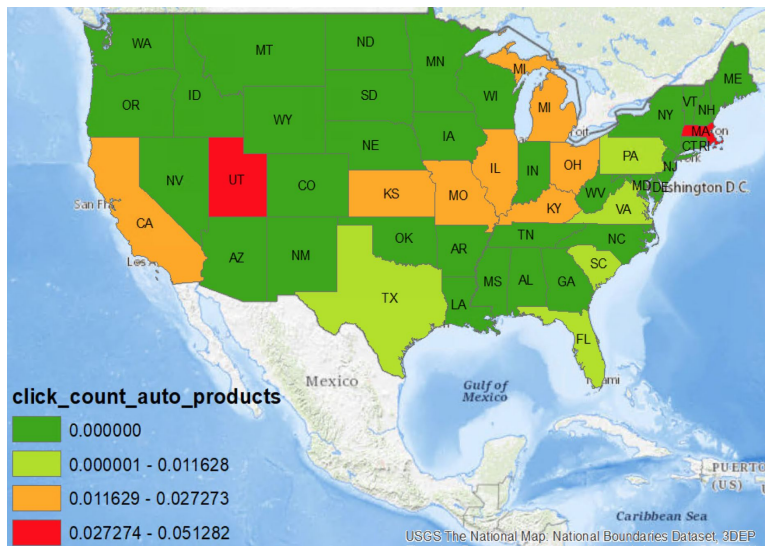
# Exploration



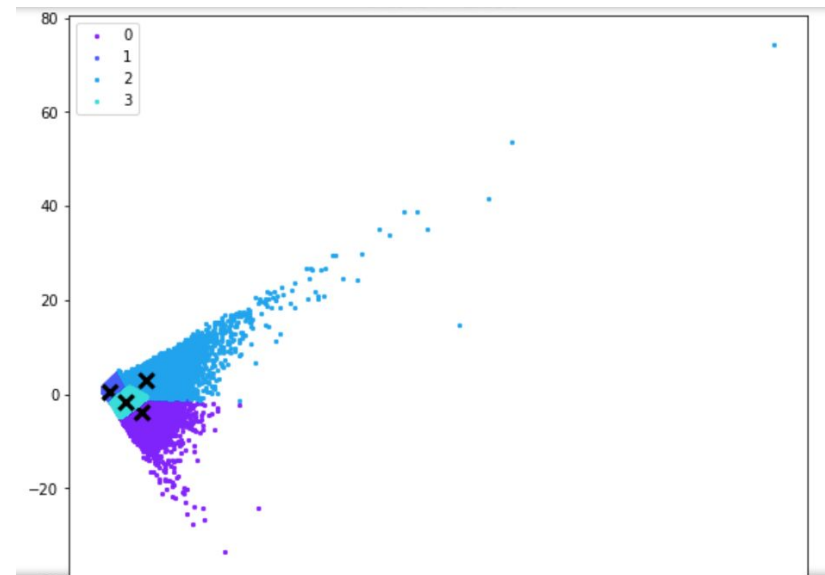
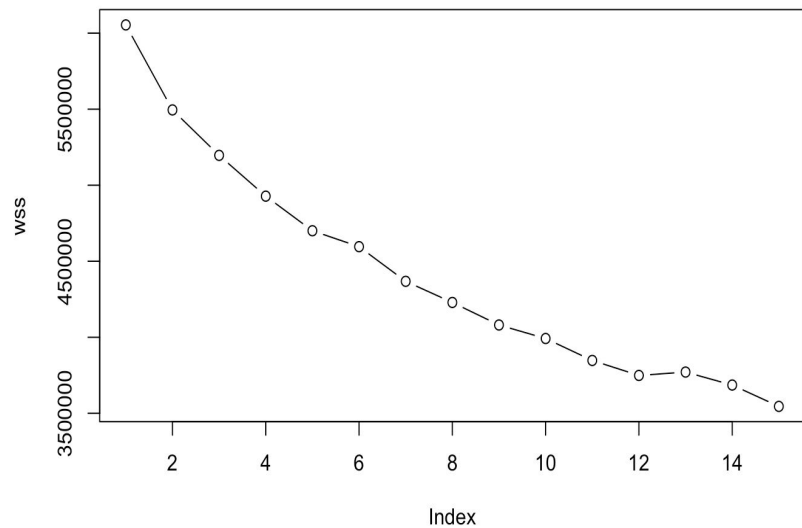
# Exploration



# Exploration



# Kmeans clustering

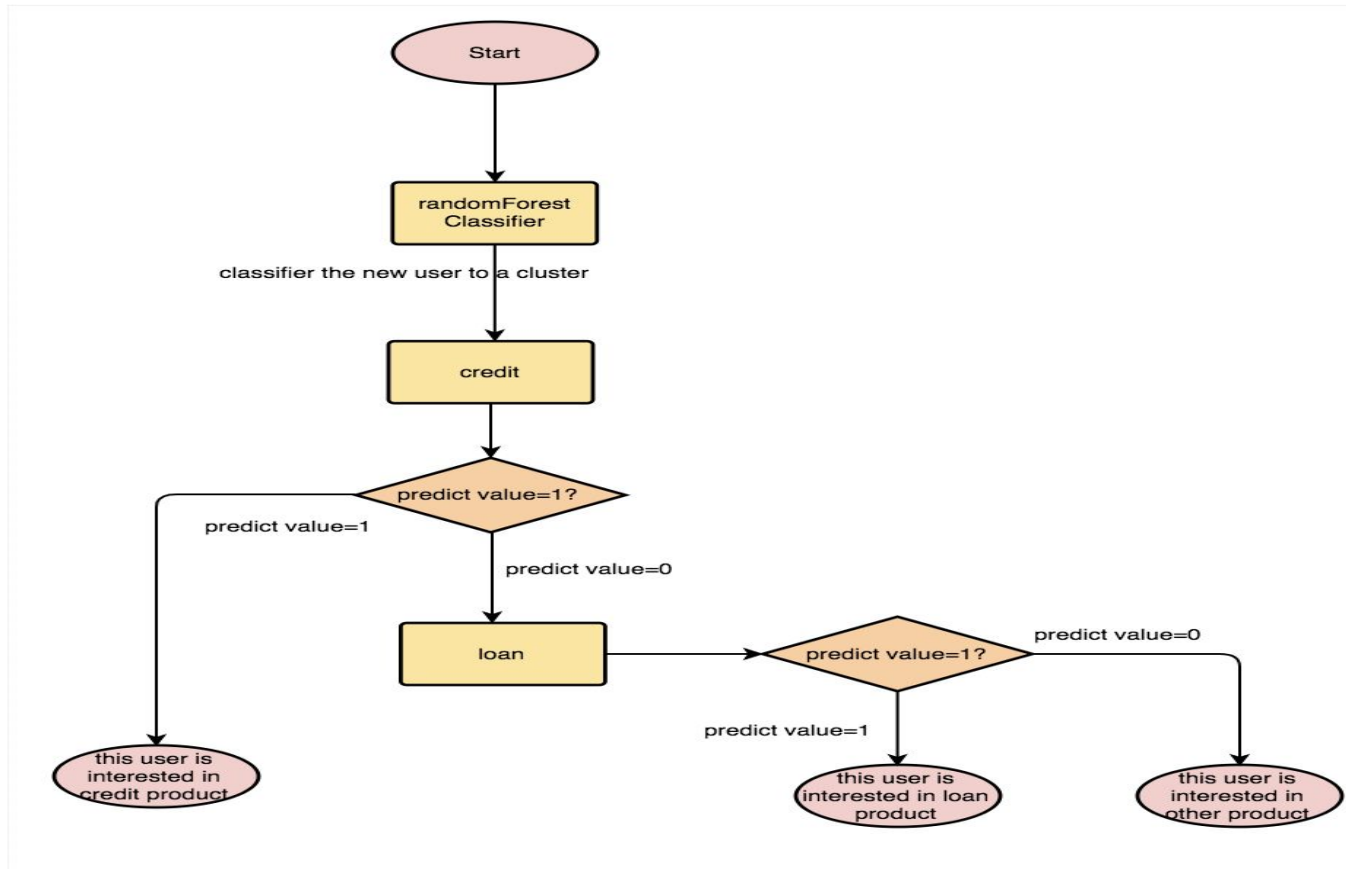


# Kmeans clustering

	Cluster			
	1	2	3	4
count_bankruptcy	1	2	3	4
count_inquiries_3_months	1	2	3	4
count_inquiries_6_months	1	2	3	4
count_inquiries_12_months	1	2	3	4
count_open_installment_accounts_24_months	1	2	3	4
count_total_tradelines_opened_24_months	1	2	3	4
count_tradelines_cc_opened_24_months	1	2	3	4
count_tradelines_closed_accounts	1	2	3	4
count_tradelines_condition_derogatory	1	2	3	4
count_tradelines_open_collection_accounts	1	2	3	4
count_tradelines_open_mortgages	1	2	3	4

count_tradelines_open_secured_loans	1	2	3	4
count_tradelines_open_student_loans	1	2	3	4
count_tradelines_open_unsecured_loans	1	2	3	4
count_tradelines_opened_accounts	1	2	3	4
max_cc_limit	1	2	3	4
total_auto_loans_balance	1	2	3	4
total_cc_open_balance	1	2	3	4
total_mortgage_loans_amount	1	2	3	4
total_mortgage_loans_balance	1	2	3	4
total_open_cc_amount_past_due	1	2	3	4
total_student_loans_balance	1	2	3	4
total_tradelines_amount_past_due	1	2	3	4
total_tradelines_open_balance	1	2	3	4
tradelines_avg_days_since_opened	1	2	3	4

# Modeling





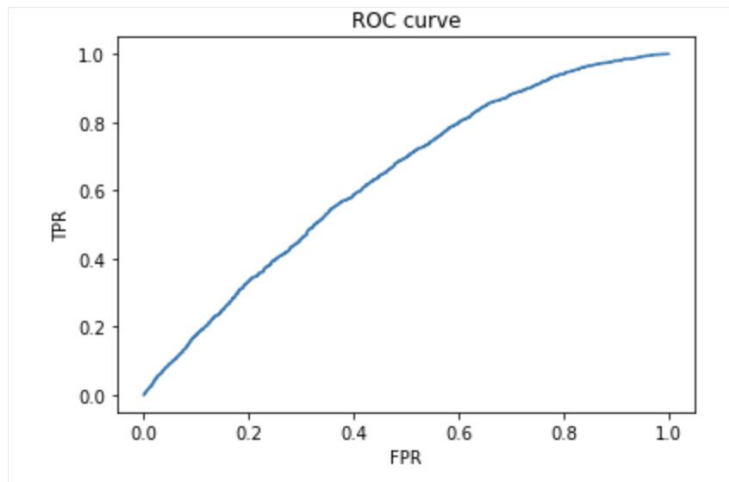
# Classification

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import GridSearchCV

para = {'n_estimators': [50, 100, 150, 200]}
gsearchRandom = GridSearchCV(estimator = RandomForestClassifier(random_state=42), param_grid = para, cv=5)
gsearchRandom.fit(x_train, y_train)
pred = gsearchRandom.predict(x_test)
```

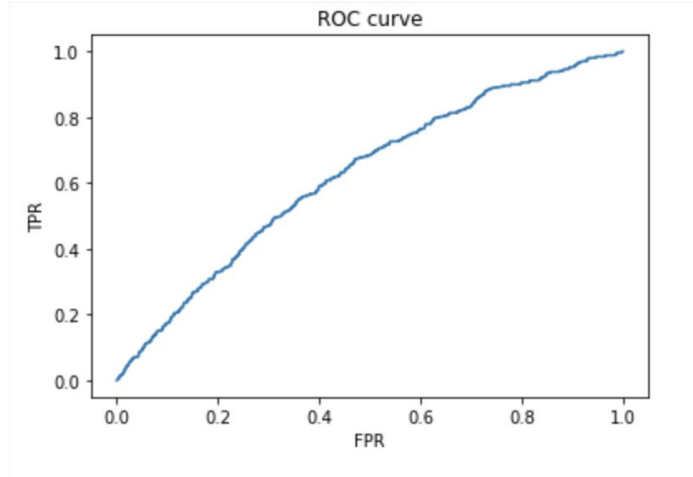
We want to use this classifier to classify the new user to the most similar cluster based on our clustering result.

# Credit



We want to use this classifier to predict whether the new user is interested in credit product.

# Loans



We want to use this classifier to predict whether the new user is interested in loan product.

# Next steps

As a next step we can:

1. Build a real-time clickstream visualization system to monitor the evolution of user behaviors
2. Look into user bounce rate by page which could be achieved with a better understanding of how current webpages are organized.
3. Since the performance of our models still has potential for improvement, we can conduct further parameter tuning.