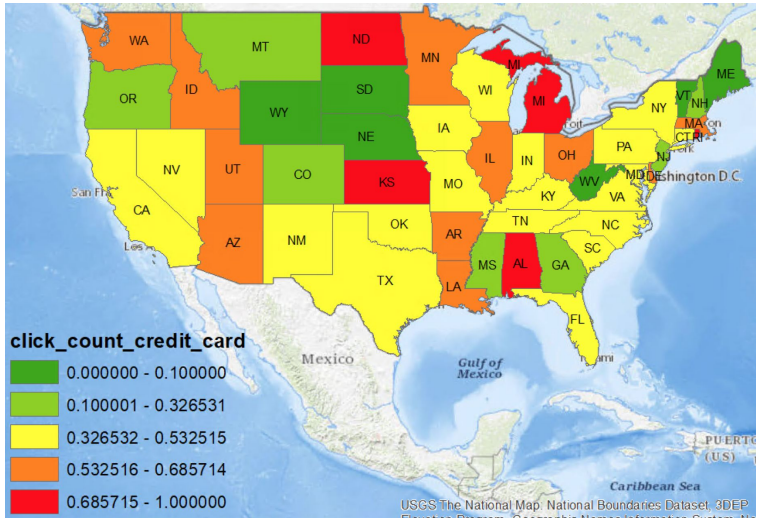
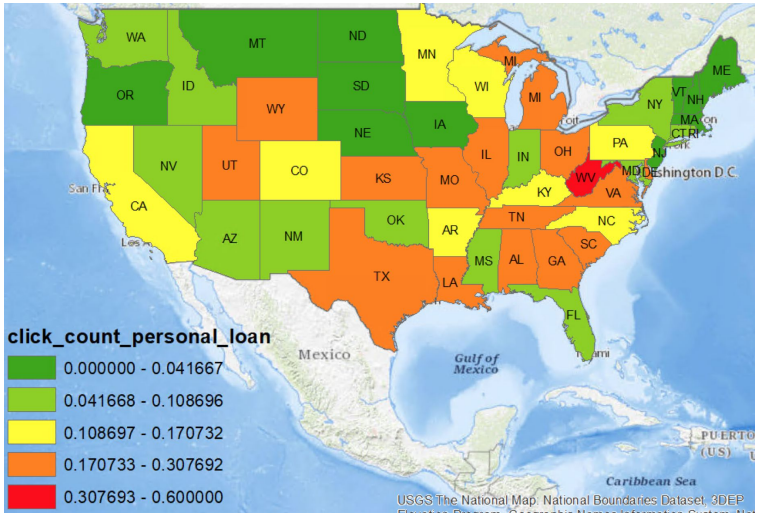
Credit Sesame User Clustering and Product Offering Predicting

Goal

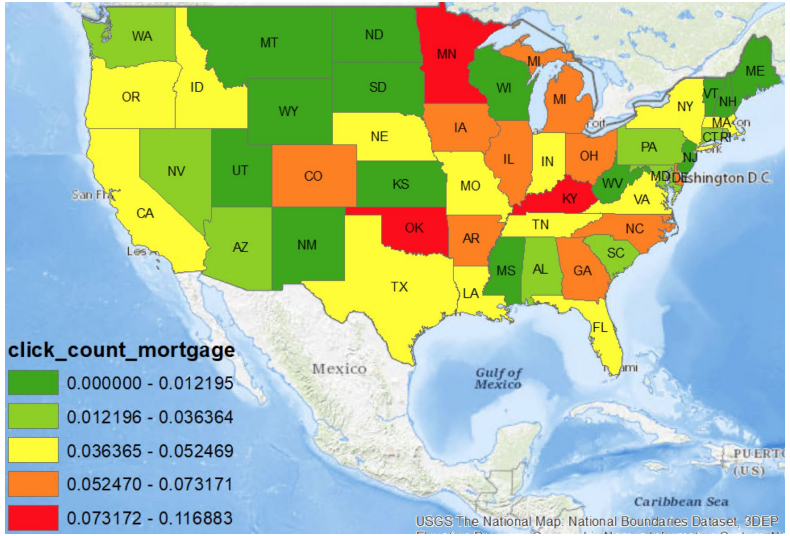
- Made clustering analysis on users
- Classify new user to most likely label
- Predict what kind of product may interest new user
- Visualize the data to better help company to do product recommendation

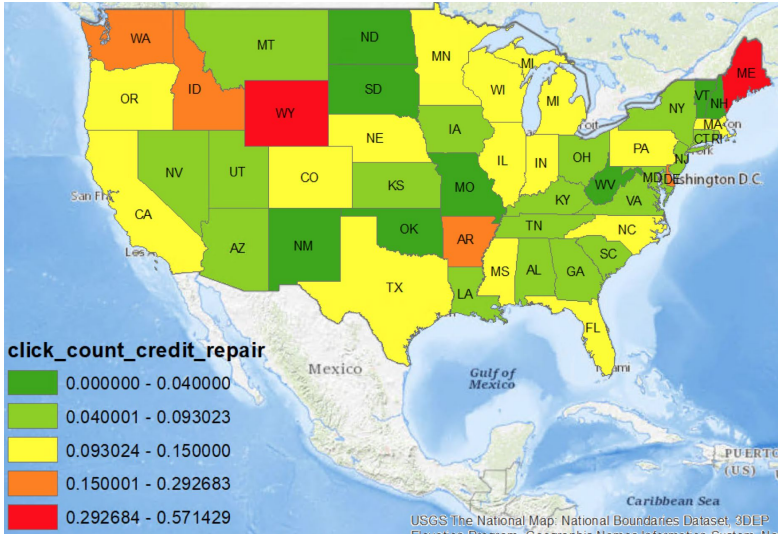
Exploration



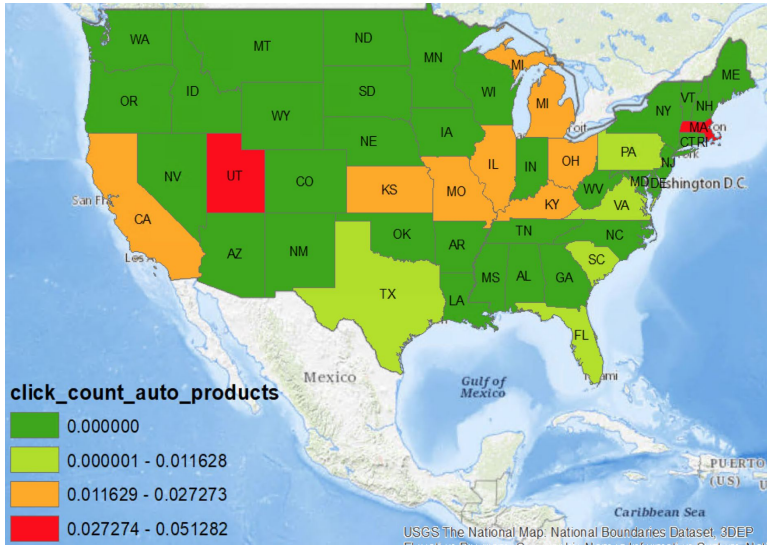


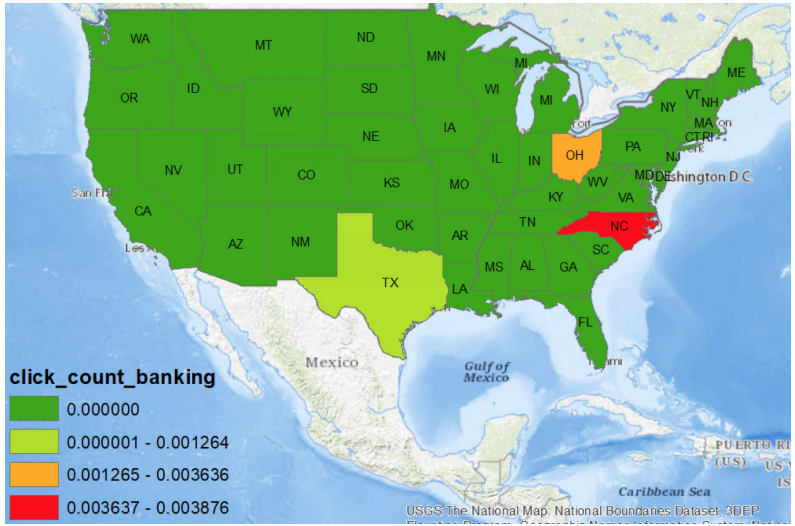
Exploration



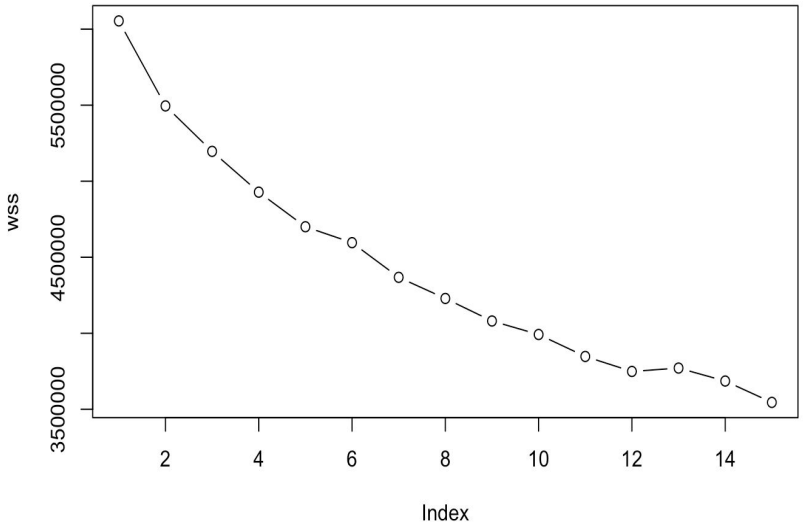


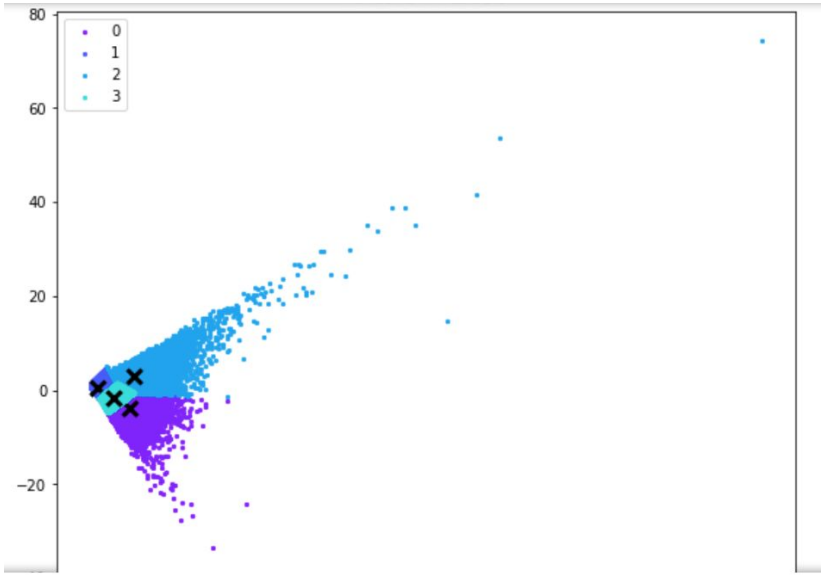
Exploration



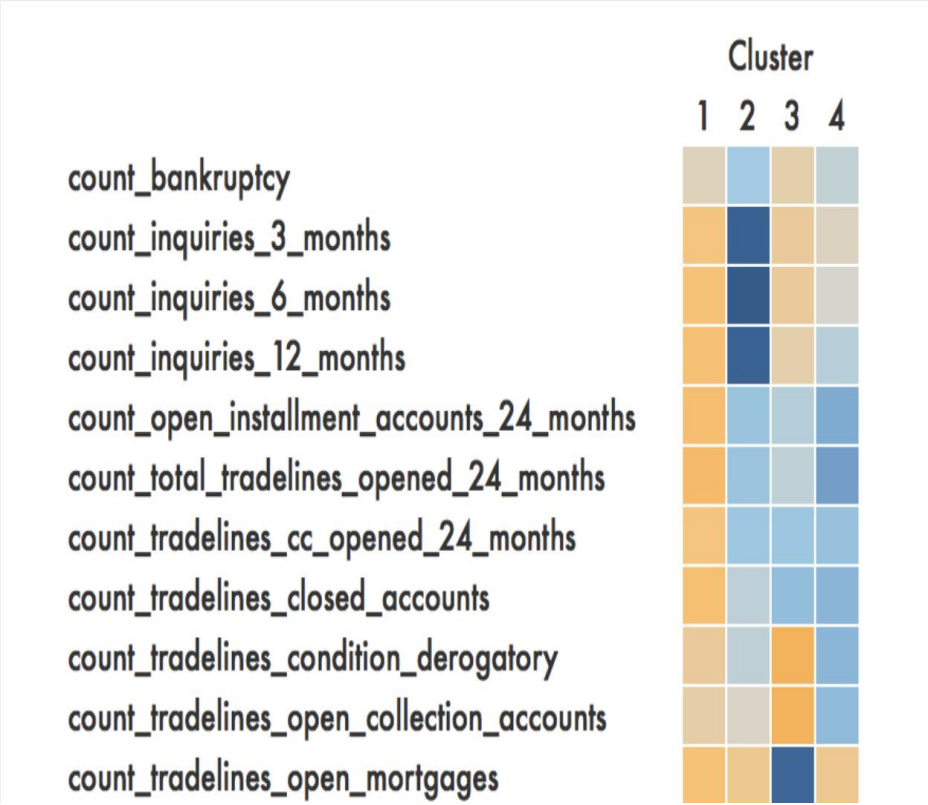


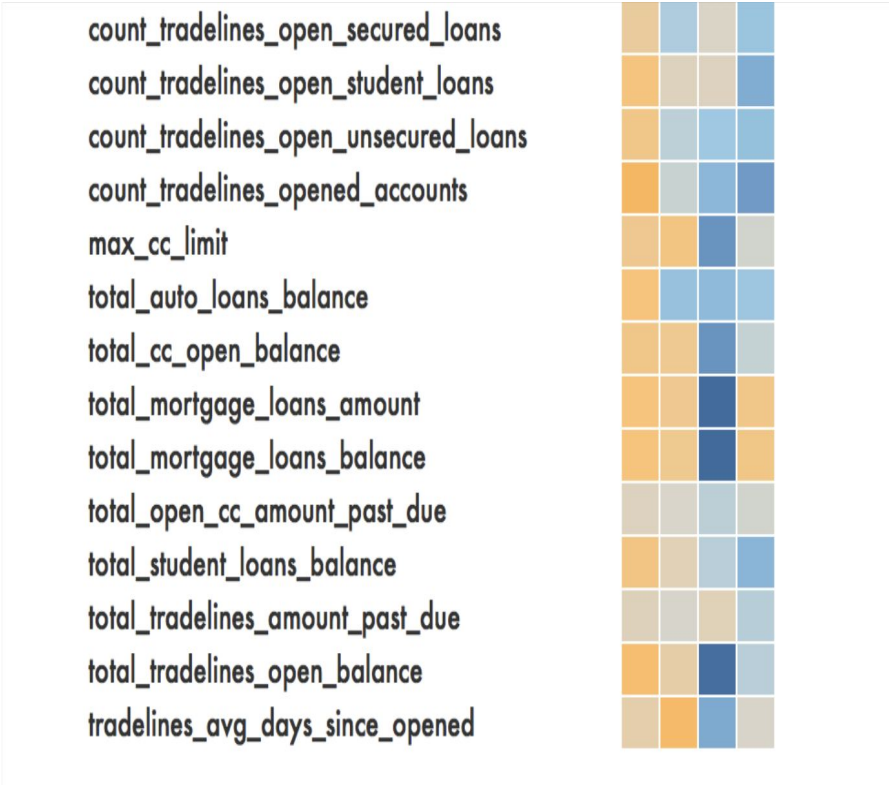
Kmeans clustering



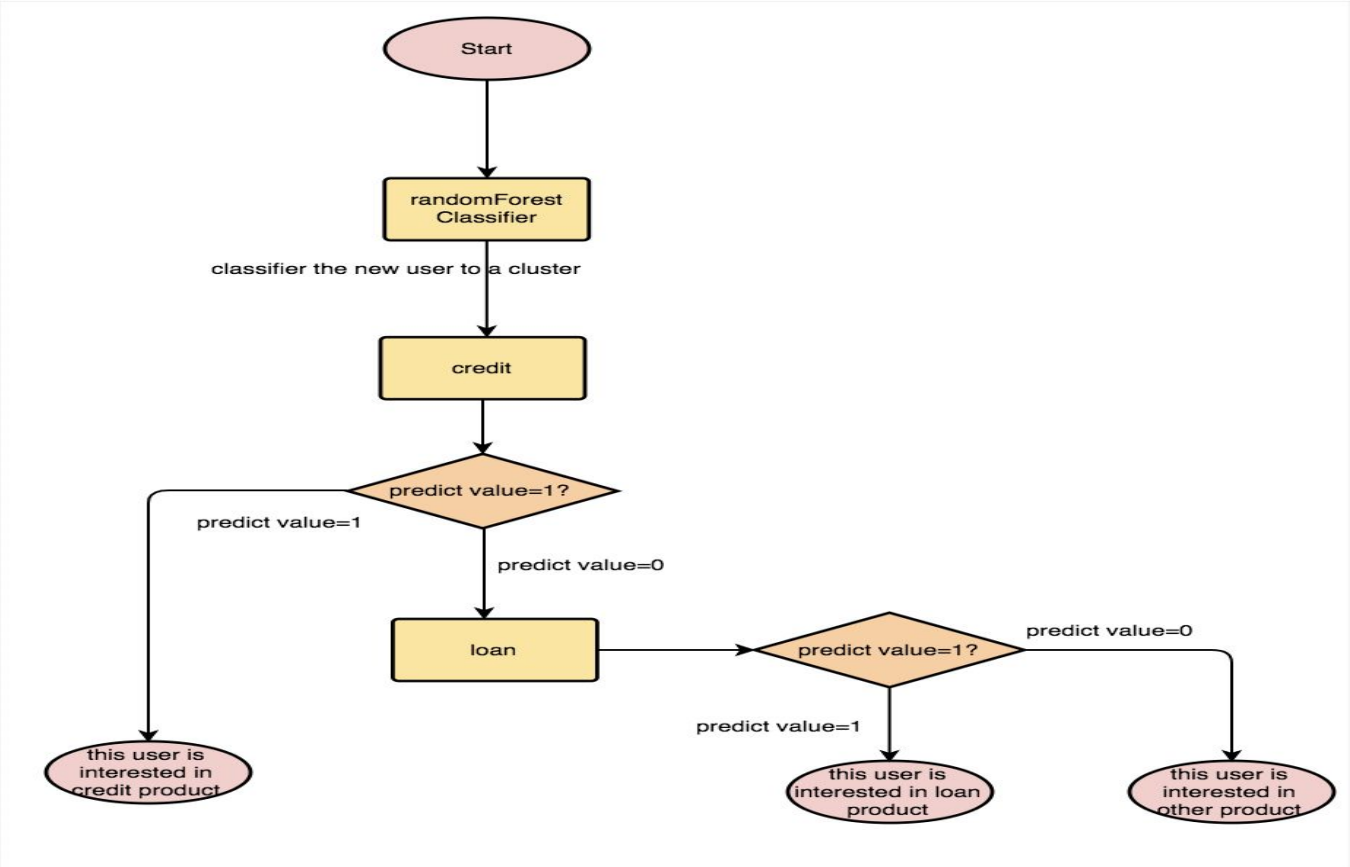


Kmeans clustering





Modeling


Start
randomForest Classifier
classifier the new user to a cluster
credit
predict value=1?
predict value=1
predict value=0
loan
predict value=1?
predict value=0
predict value=1
this user is interested in credit product
this user is interested in loan product
this user is interested in other product

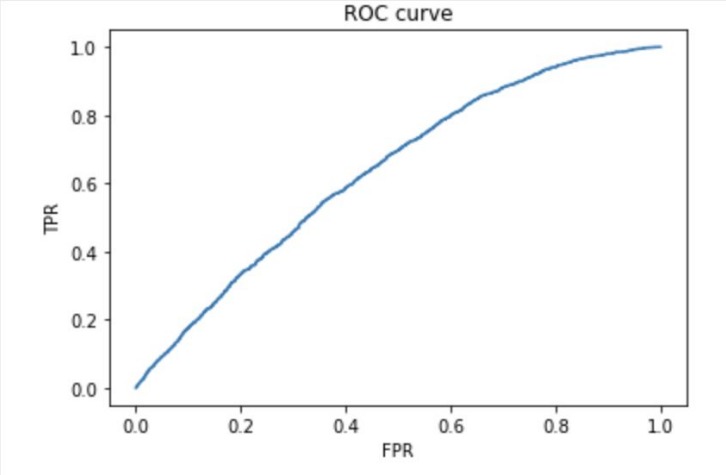
Classification

```
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)
```

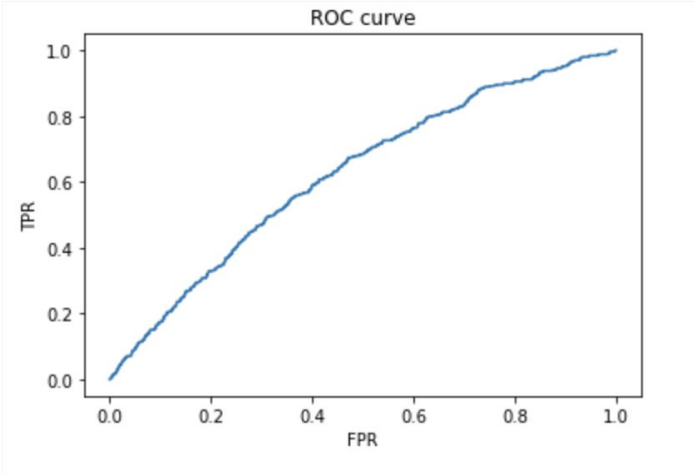
We want to use this classifier to classify the new user to the most similar cluster based on our clustering result.

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 achieve with a better understanding of how current webpages are organized.
3. Since the performance of our models still have potential to improvement, we can conduct further parameter tuning.