THE CAPSTONE PROJECT

SPRINGBOARD

Presented by Bolor Boldbaatar
Mentor Matt Fornito, Summit Analytics LLC

My project starts with this

I had no idea what I was getting myself into

OUTLINE

1. Introduction
2. Data Exploration
3. Recommendation Systems
4. Data Analysis
5. Conclusion

1.

Introduction

Background

- Santander bank is originated from Spain, but has headquarters in many countries

Problem

- Currently, their system doesn't recognize the right products to their customers, so that some customers get too many offers while others don't have any.
- Their wish for the solution is to have a recommendation system to predict which products their existing customers will use in the next month based on their past behavior and that of similar customers.

Purpose for the Project

- Find a recommendation system for predicting next bank products
- Try different methods of recommendation system

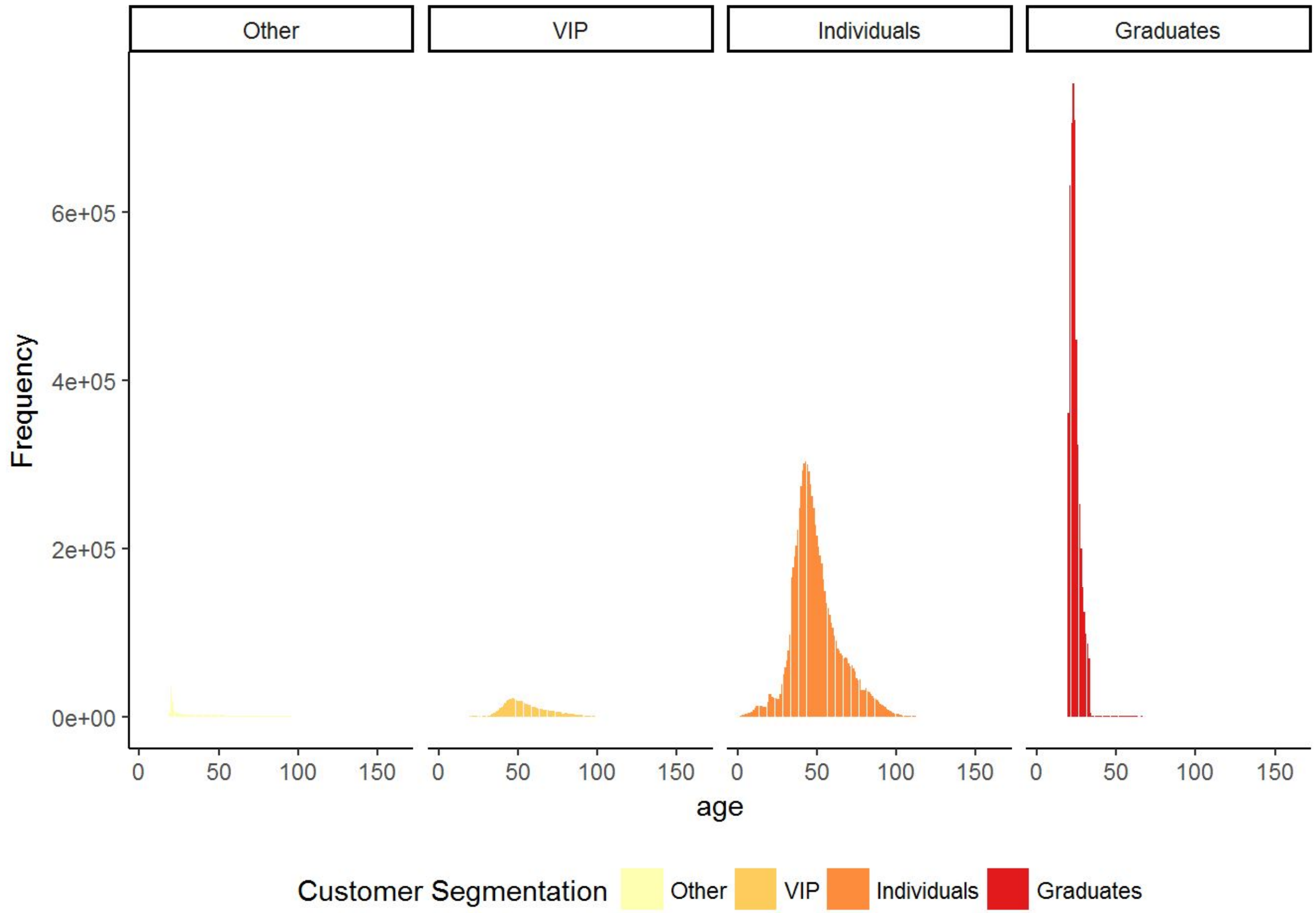
2.

Data Exploration

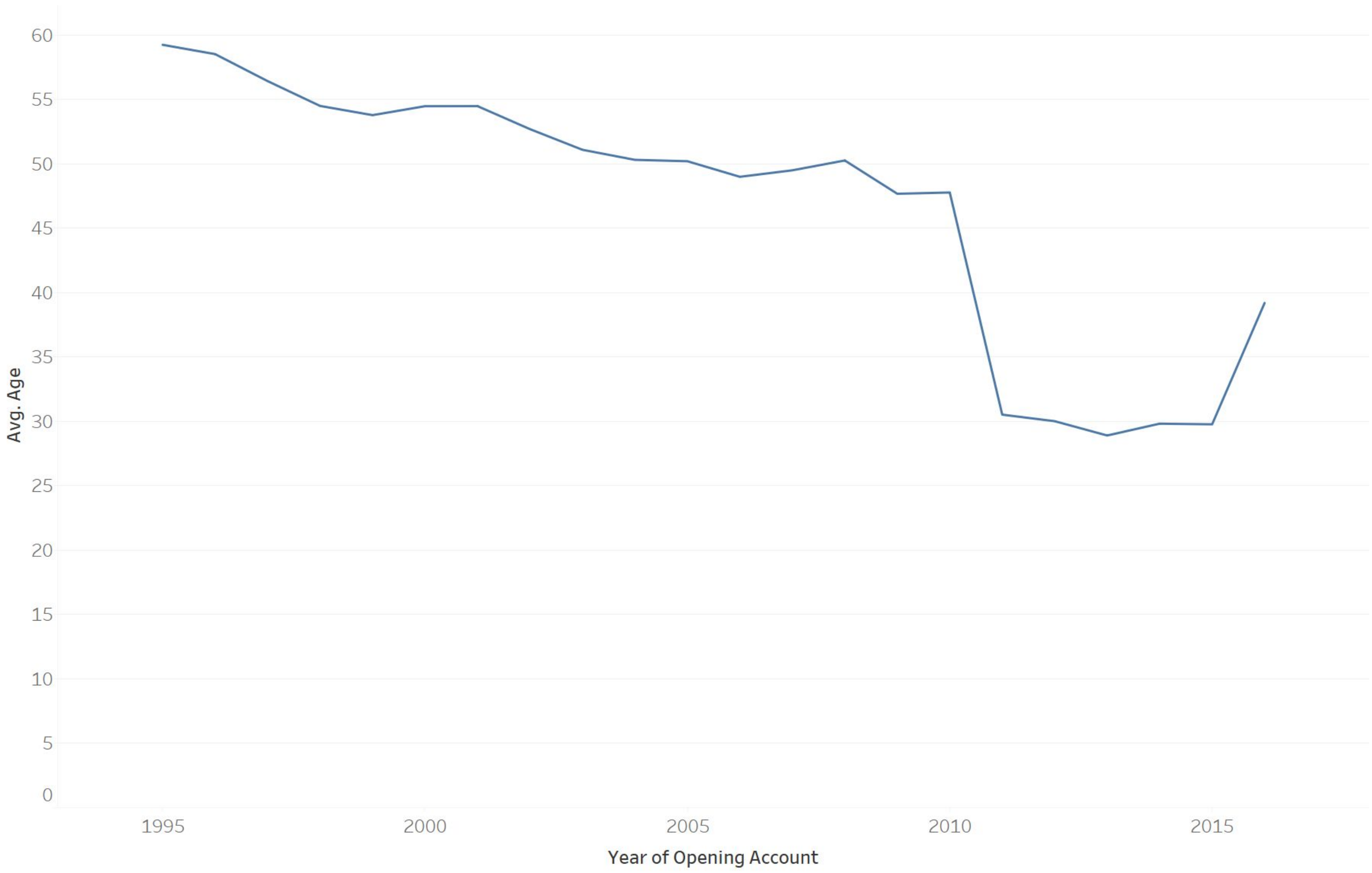
Main dataset insights

- There are more than 13 million user database and 48 different variables
- First 24 of them user demographic variables, and last 24 of them bank products
- User demographic info includes age, gender, income, location and etc.
- Bank products are all 0-1 data, 1 being user consumes that product

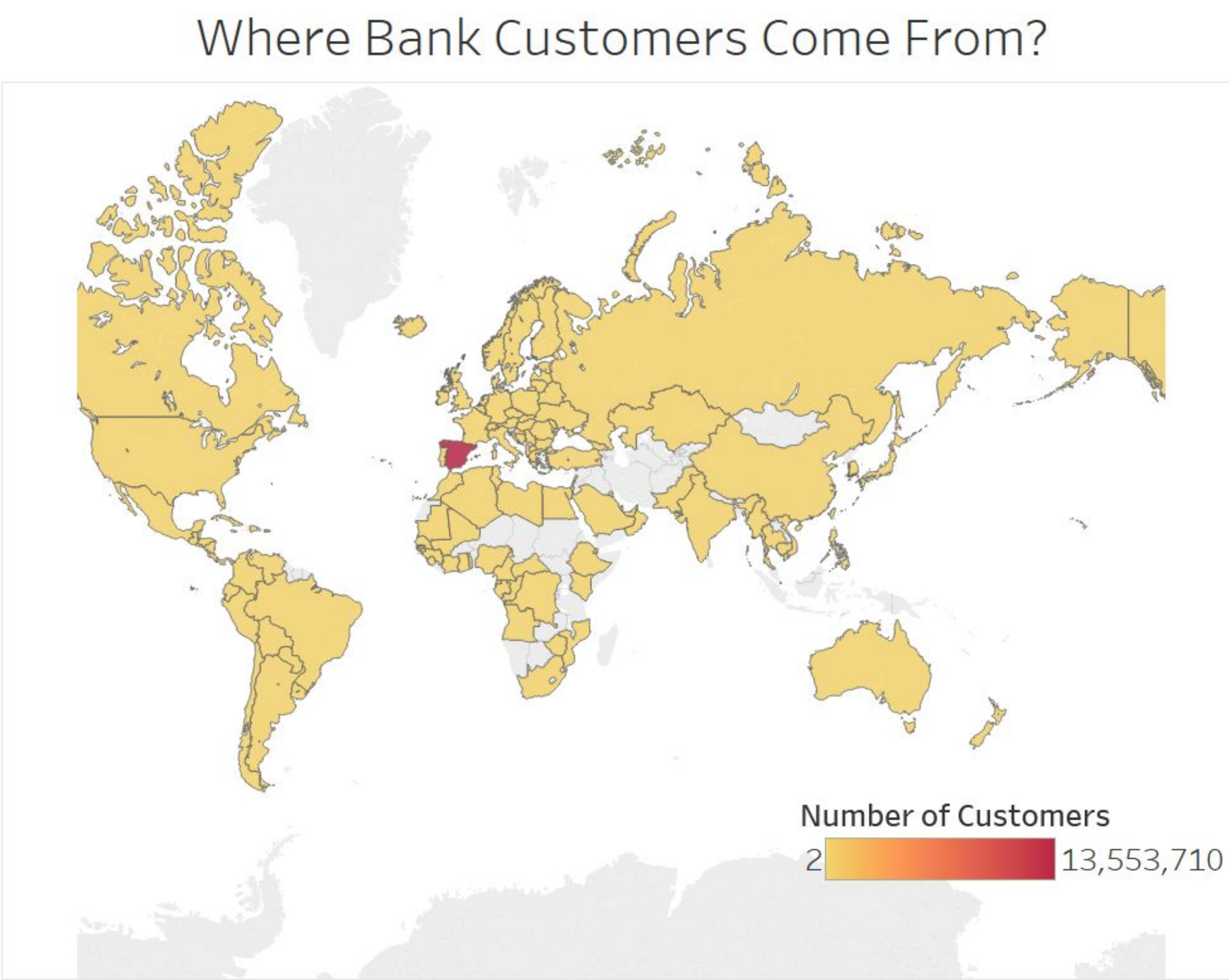
Exploring Customer Age



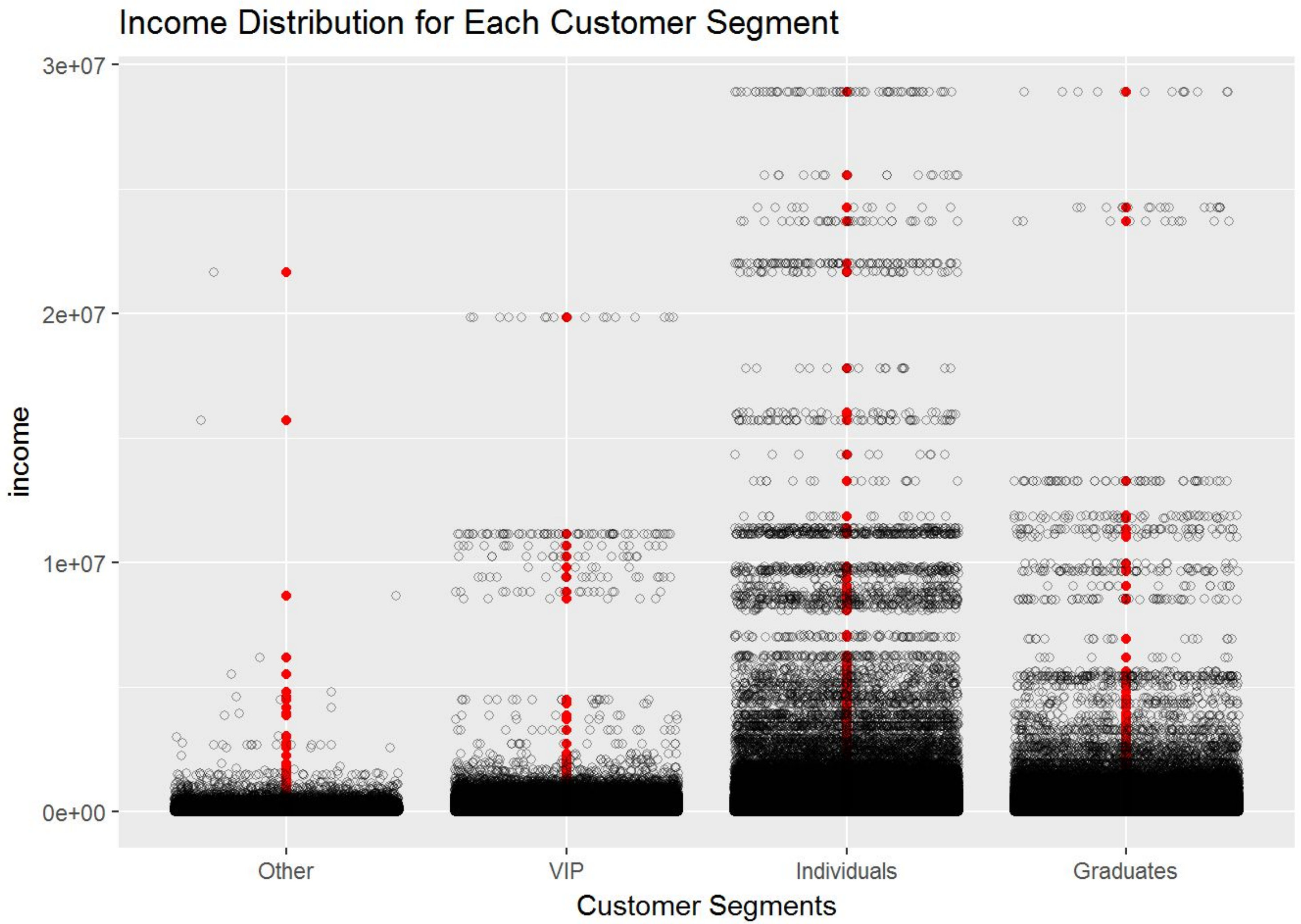
Changing Age Distributions Over the Years



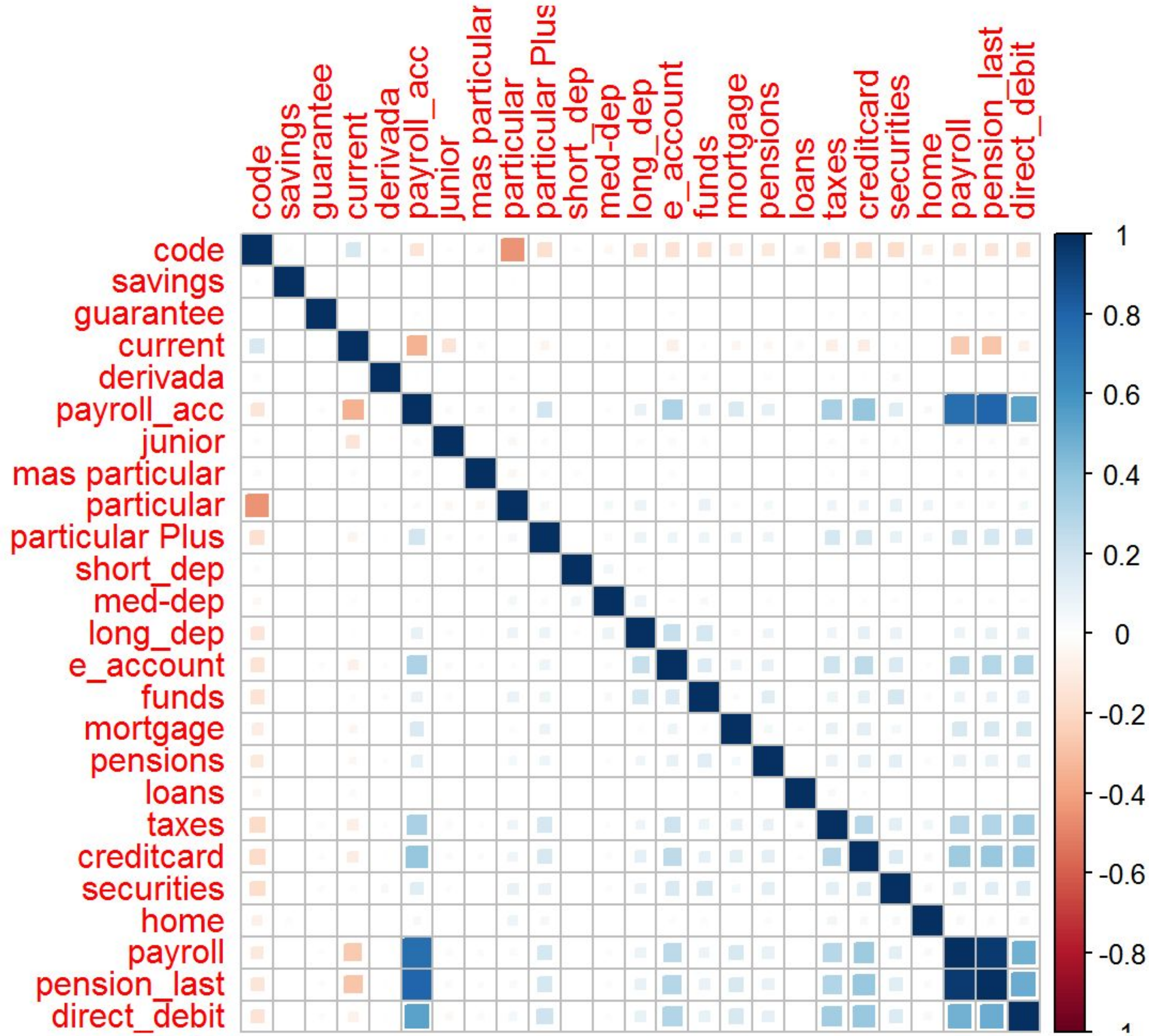
Exploring Customer Location



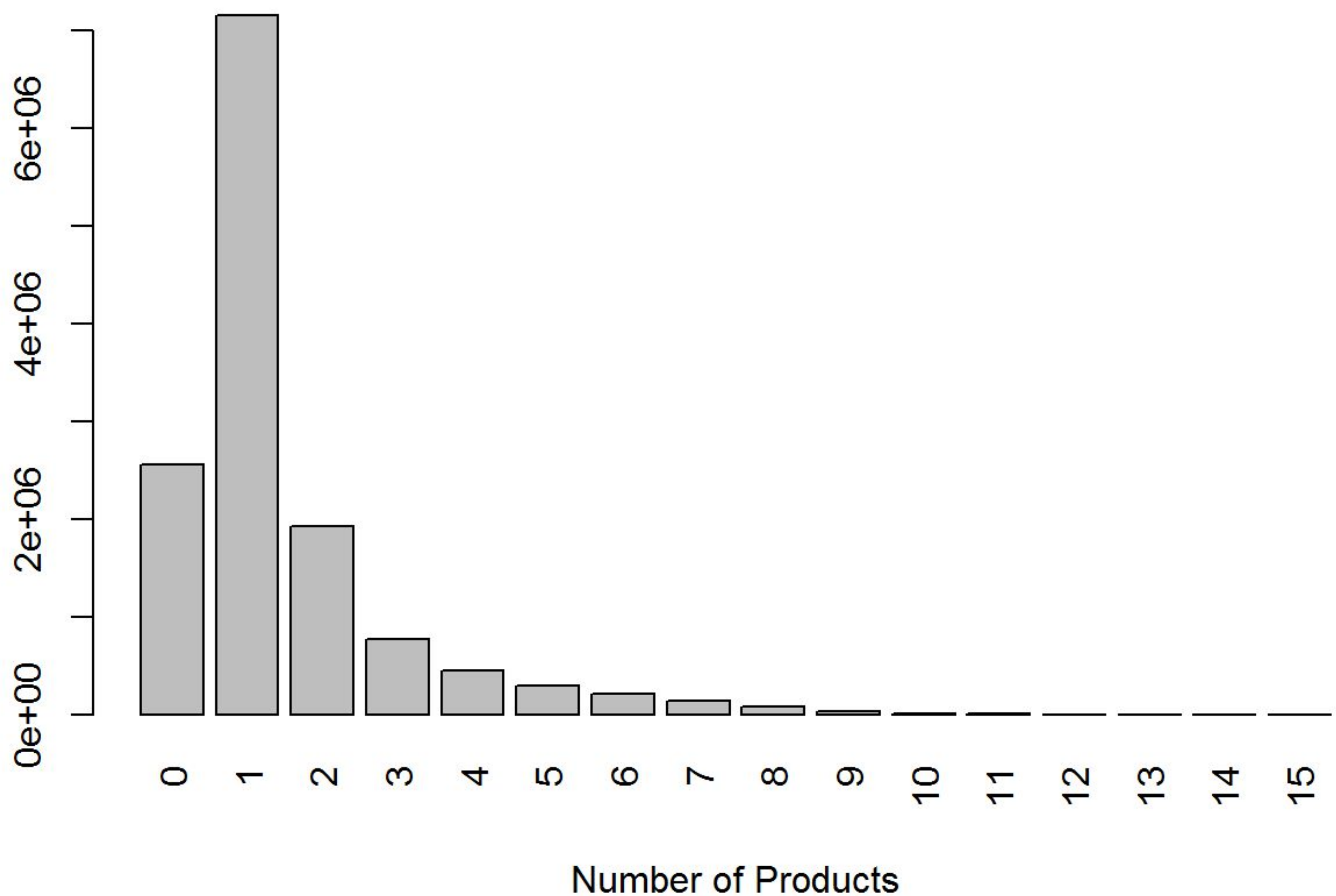
Exploring Customer Income



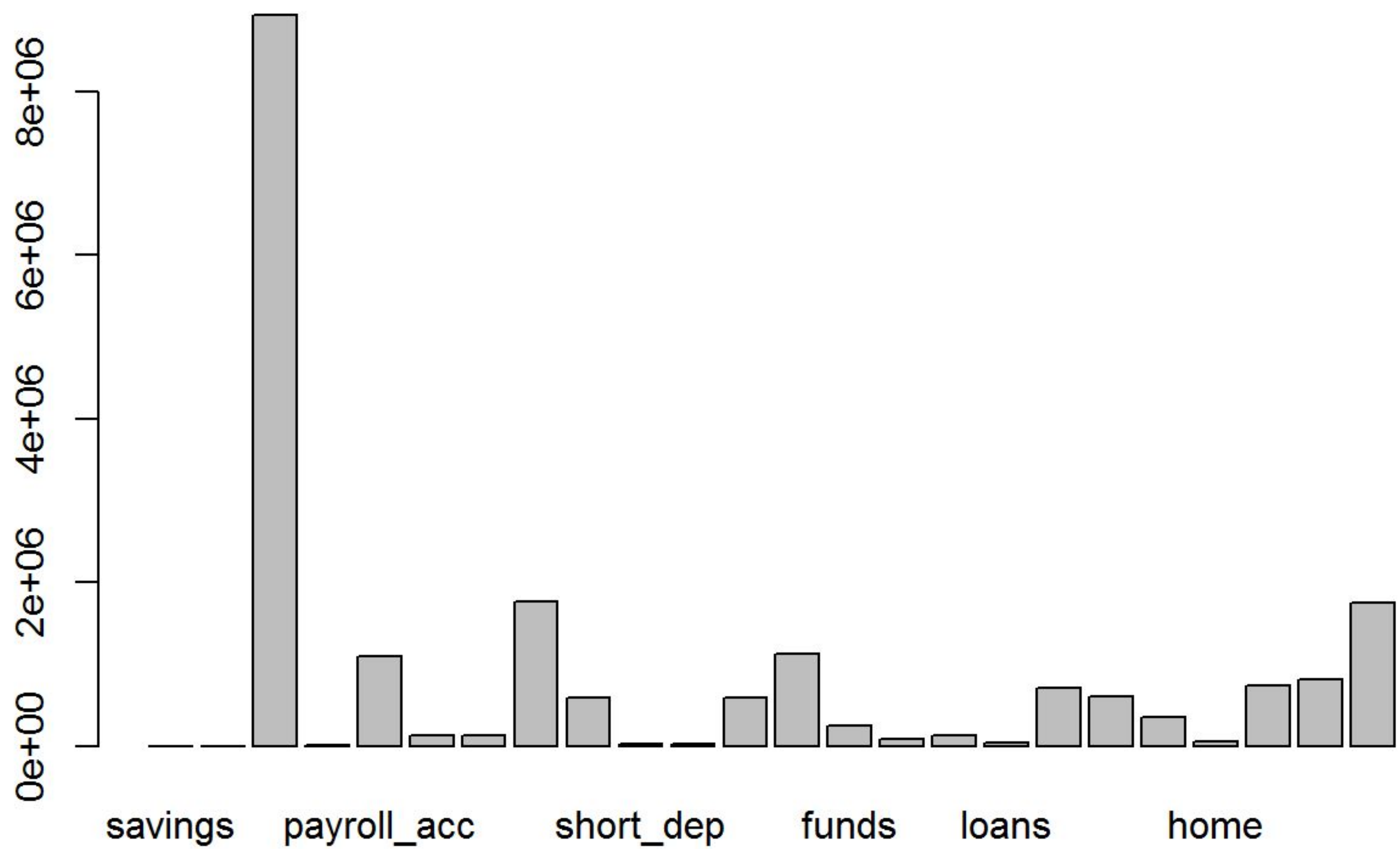
Exploring Bank Products



Exploring Bank Products



Exploring Bank Products



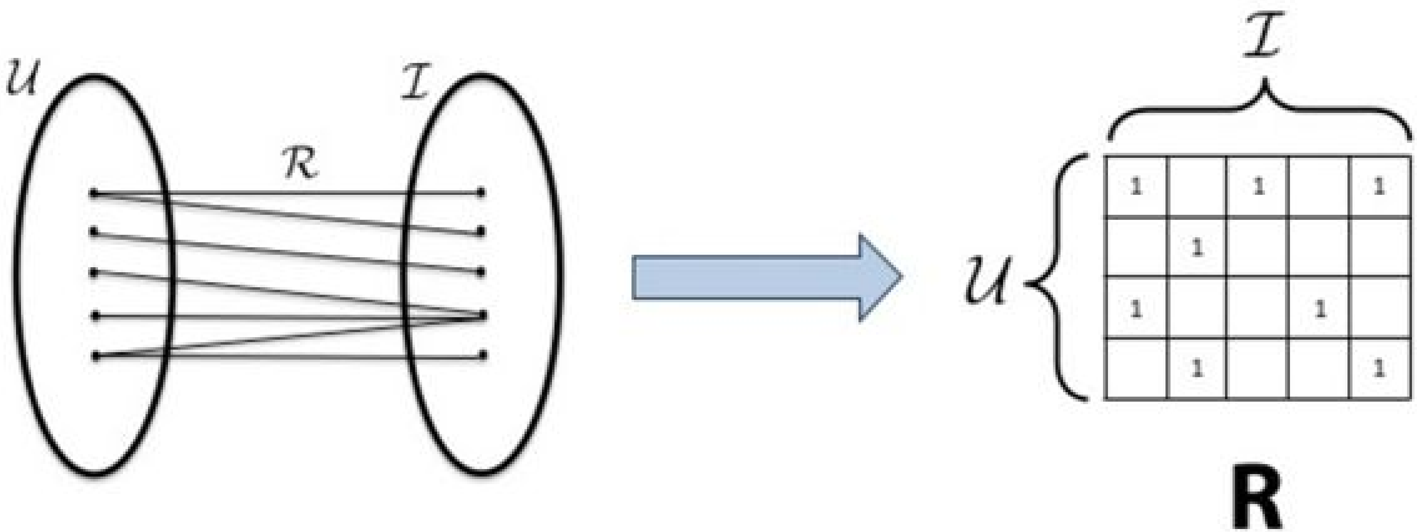
3.

Recommendation Systems

Binary Recommendation System

- Binary recommendation system is bit different than traditional recommendation system
- Rating is not based on 1-5 Likert Scale
- 0-1 data, another word Positive-only data, to show user preference
- 0 represents either user has no preference now, or user doesn't know about the product, or does not like the product

Binary Recommendation System



Collaborative Filtering

Item-based

Determine similar products

Give more weights to the most similar products

User-based

Determine similar users

Give more weights to the opinions of similar users

4.

Analysis: Item-based Collaborative Filtering

Process of Item-based Collaborative Filtering

Create matrix

Find similarities using cosine measure

Find similar items

#1 Step

- Create matrix that holds NA from the current dataframe

```
holder <- matrix(NA, nrow=ncol(bank.recom),ncol=ncol(bank.recom),dimnames=list(colnames(bank.recom),colnames(bank.recom)))

bank.recom.similarity <- as.data.frame(holder)


for(i in 1:ncol(bank.recom)) {
  for(j in 1:ncol(bank.recom)) {
    bank.recom.similarity[i,j]= getCosine(bank.recom[i],bank.recom[j])
  }
}
```

#2 Step

- Calculate similarity matrix between all items based on available users' ratings.
- Cosine Similarity

$$cos(A, B) = \frac{A \times B}{||A|| ||B||}$$

- Final similarity matrix:

```
##                current payroll_acc particular e_account     taxes
## current     1.00000000  0.04887032  0.3687681 0.2747616 0.1902221
## payroll_acc 0.04887032  1.00000000  0.1391024 0.3674849 0.3687115
## particular  0.36876805  0.13910241  1.0000000 0.1153420 0.1672927
## e_account   0.27476158  0.36748492  0.1153420 1.0000000 0.2688076
## taxes       0.19022212  0.36871153  0.1672927 0.2688076 1.0000000
## payroll     0.05957614  0.76568900  0.1245008 0.3153189 0.3153222
##                payroll pension_last direct_debit
## current     0.05957614   0.06127241    0.3545048
## payroll_acc 0.76568900   0.79755111    0.5742283
## particular  0.12450081   0.13023025    0.1897989
## e_account   0.31531889   0.33744902    0.3753149
## taxes       0.31532222   0.33478083    0.4053543
## payroll     1.00000000   0.95606375    0.5062049
```

#3 Step

- Store only n closest items to each item;

```
bank.neighbours <- matrix(NA,
nrow=ncol(bank.recom.similarity),ncol=2,dimnames=list(colnames(bank.recom.similarity)))
```

- Calculate predicted rating for each item based on available ratings of user u by weighting available ratings of users on similarities.

```
For (i in 1:ncol(bank.recom))
{
  bank.neighbours[i,] <-
(t(head(n=2,rownames(bank.recom.similarity[order(bank.recom.similarity[,i],decreasing=TRUE),][i]))))
}
```

Results: Recommended Products for First 5 Users

	Users	Additional Products
1	"657788"	particular
2	"657795"	pension_last
3	"657790"	current
4	"657794"	direct_debit
5	"657789"	direct_debit

Conclusion

- Limitations:
  - Limited Time
  - Little experience to deal with large dataset
  - User-based approach
- Recommendation:
  - Model Evaluation
  - Model expansion
  - Other models

Thanks!

Any questions?

You can find me at

bbplus2@gmail.com

Credits

Special thanks to all the people who made and released these awesome resources for free:

- Presentation template by SlidesCarnival
- Photographs by Unsplash