Santander Product Recommendation

TEAM KWT Wen Li Yisong Tao Lydia Kan

01 Introduction

02 Data Cleaning and EDA

03 Feature Engineering

04 Training Model

05 Result and Finding

06 Future Steps

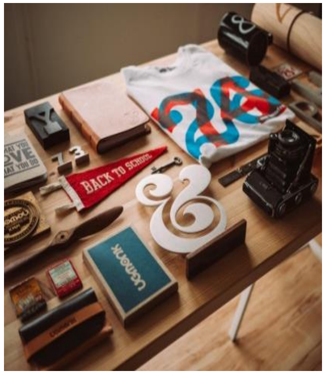
Introduction

Project Description

Santander Bank offers their customers personalized product recommendations time to time, in order to meet the individuals needs and satisfaction.
This challenge seeks to improve the recommendation system by predicting which products their existing customers will use in the next month based on their past behavior.

Goal

Achieve top 5% ranking and MAP@7 score on Kaggle leader board



Introduction

01

Data Size

Training Set:

13,647,409

Test Set:

02

Input Features

Categorical: 45

Continuous: 3

Customer Info. :

1: 24

Product Purchased Info:

25:48

03

Output Features

Multi-Classifier

Recommended

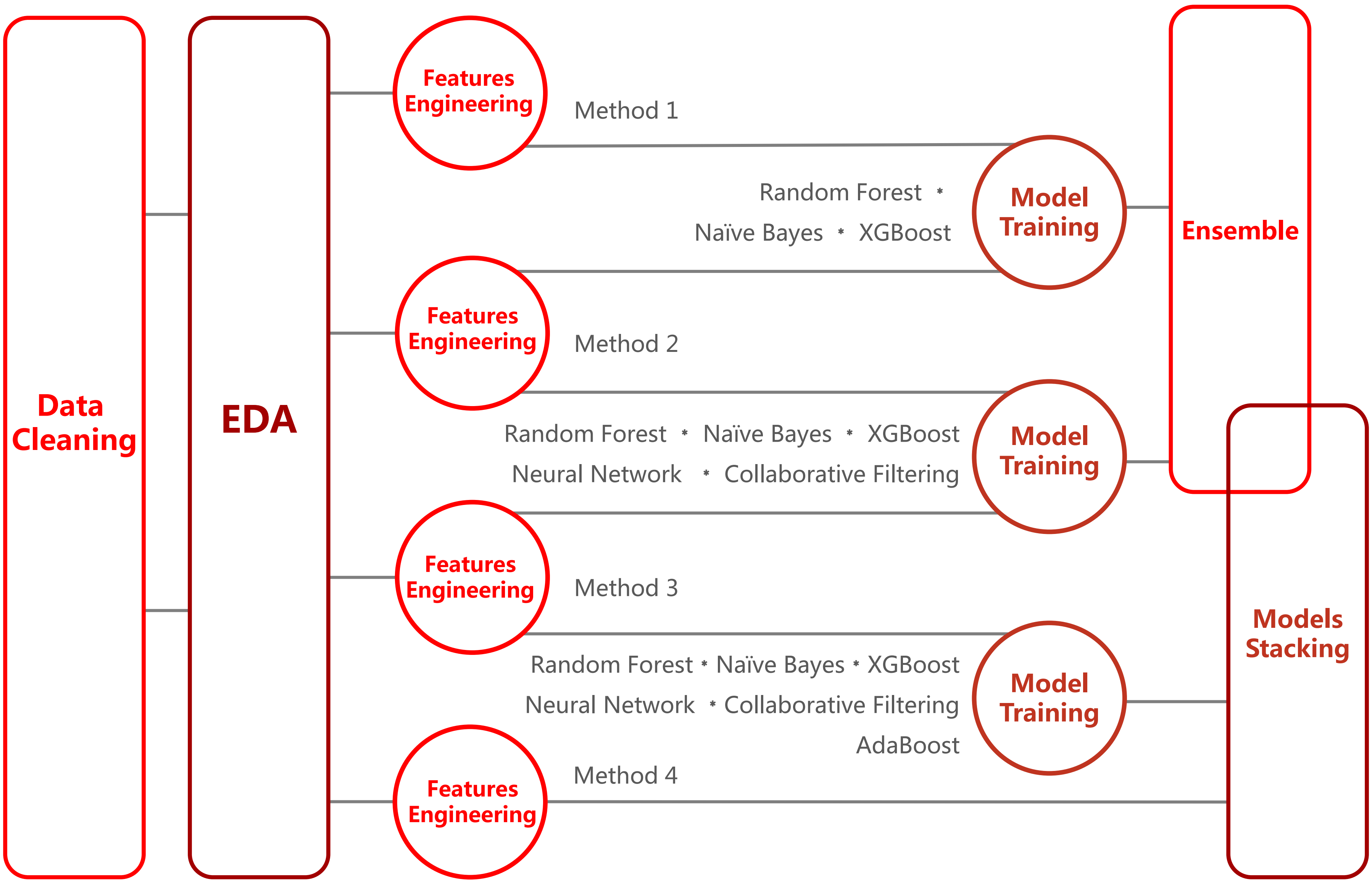
Products : 7

04

Evaluation

MAP@7

Workflow


Data Cleaning
EDA
Features Engineering
Method 1
Random Forest • Naïve Bayes • XGBoost
Model Training
Ensemble
Features Engineering
Method 2
Random Forest • Naïve Bayes • XGBoost
Neural Network • Collaborative Filtering
Model Training
Features Engineering
Method 3
Random Forest • Naïve Bayes • XGBoost
Neural Network • Collaborative Filtering
AdaBoost
Model Training
Models Stacking
Features Engineering
Method 4

Data Cleaning

Imputation

Contain Missing Values:

24 Features

Dropping Features

Drop 5 Features:

- Having over 95% missing value
- Repetitive of other features

Imputation

Unknown

- Sex
- Employee Index
- Country Residency
- Segmentation
- Residence Index
- Foreigner Index
- Channel to Join
- Primary
- Province Name

Common Type

- Customer Type
- Activity Index
- Rent

Others

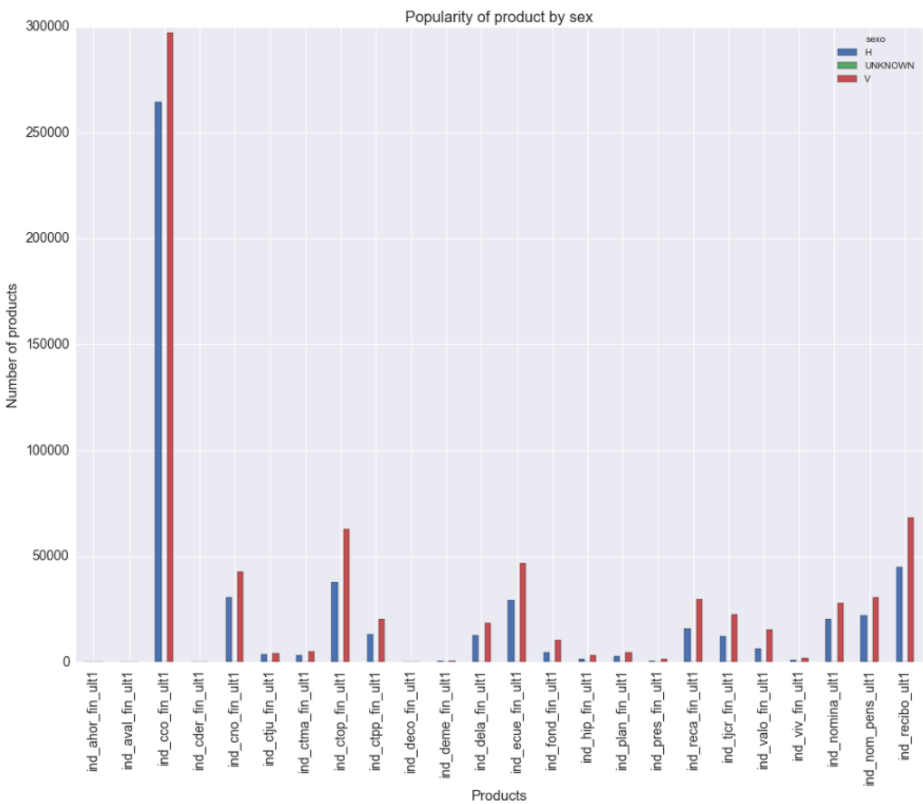
- New Customer – New
- Seniority – Min
- Age – Scale, Mean
- Relationship Type – 'A'
- Deceased Index – 'N'

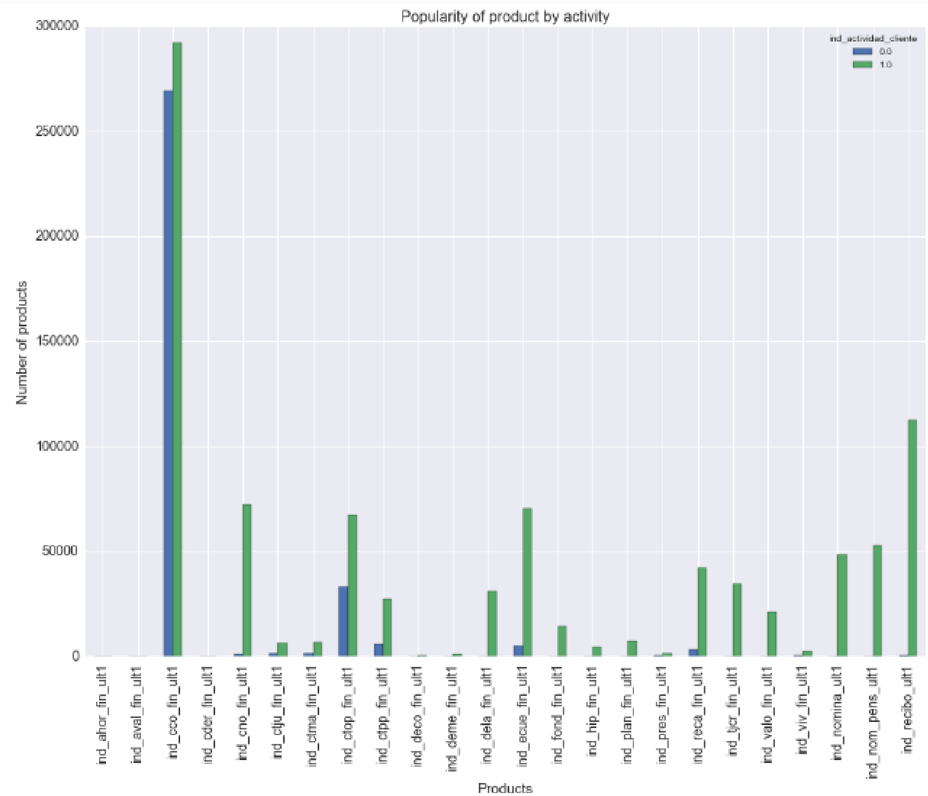
Products

- Payroll - 0
- Pensions - 0

EDA

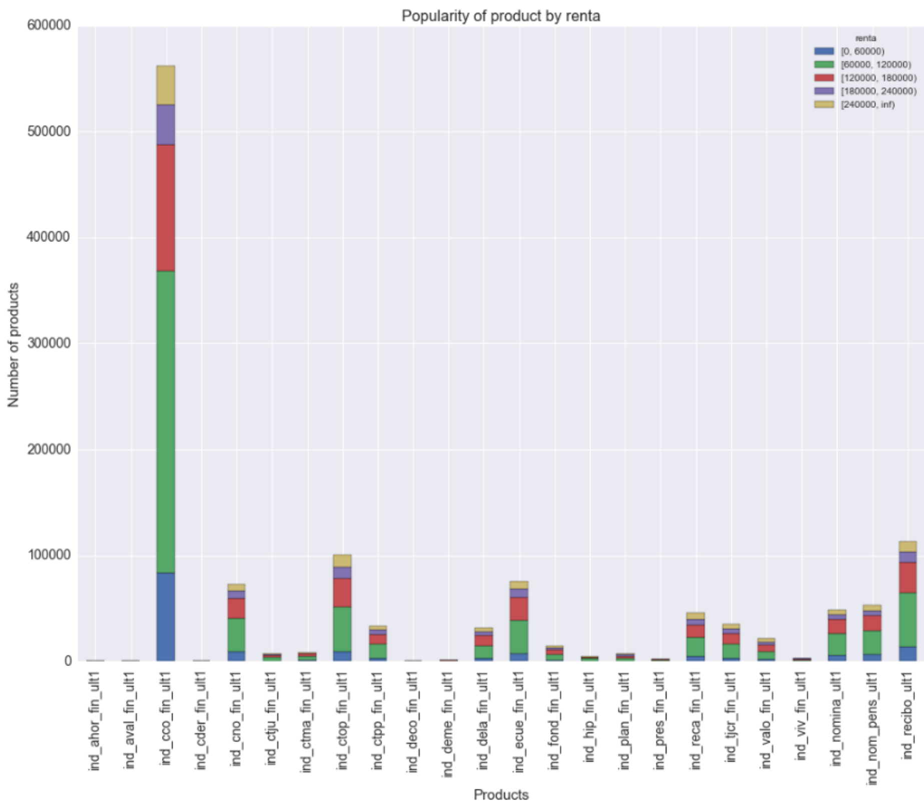
Product Sales Related to Customer's Info - 2016.5

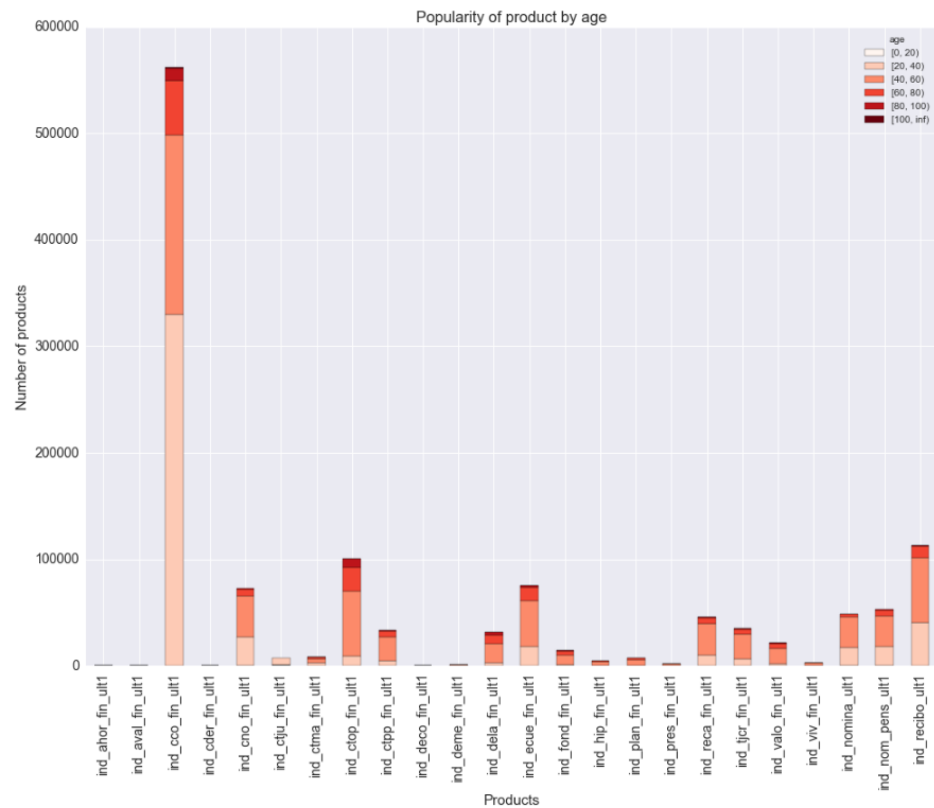




EDA

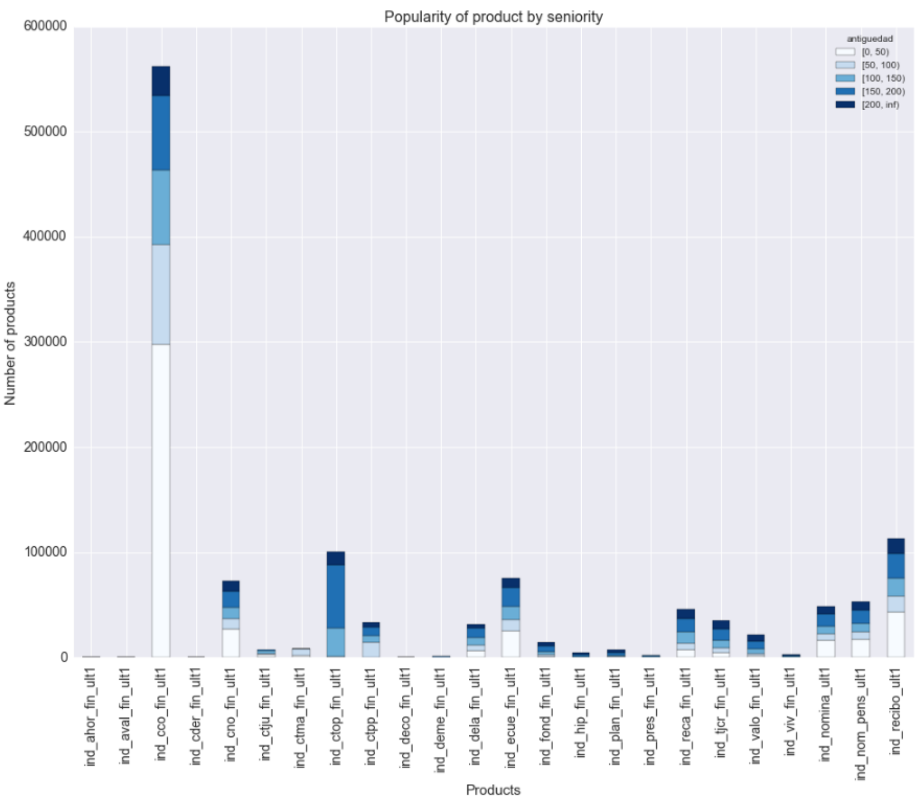
Product Sales Related to Customer’s Info - 2016.5

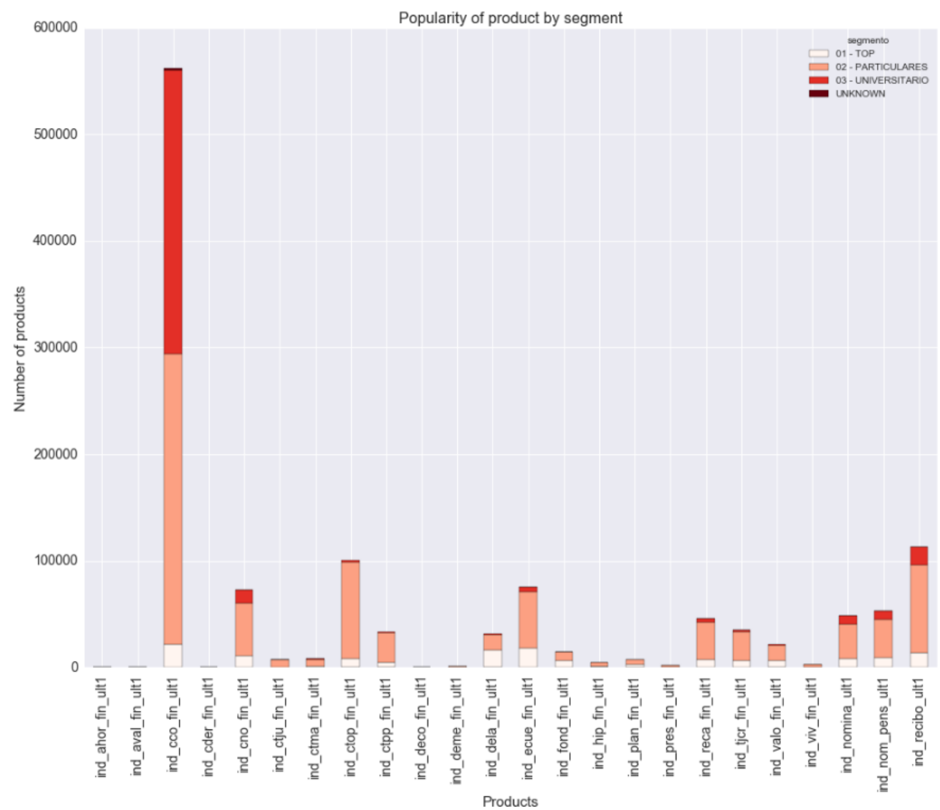




EDA

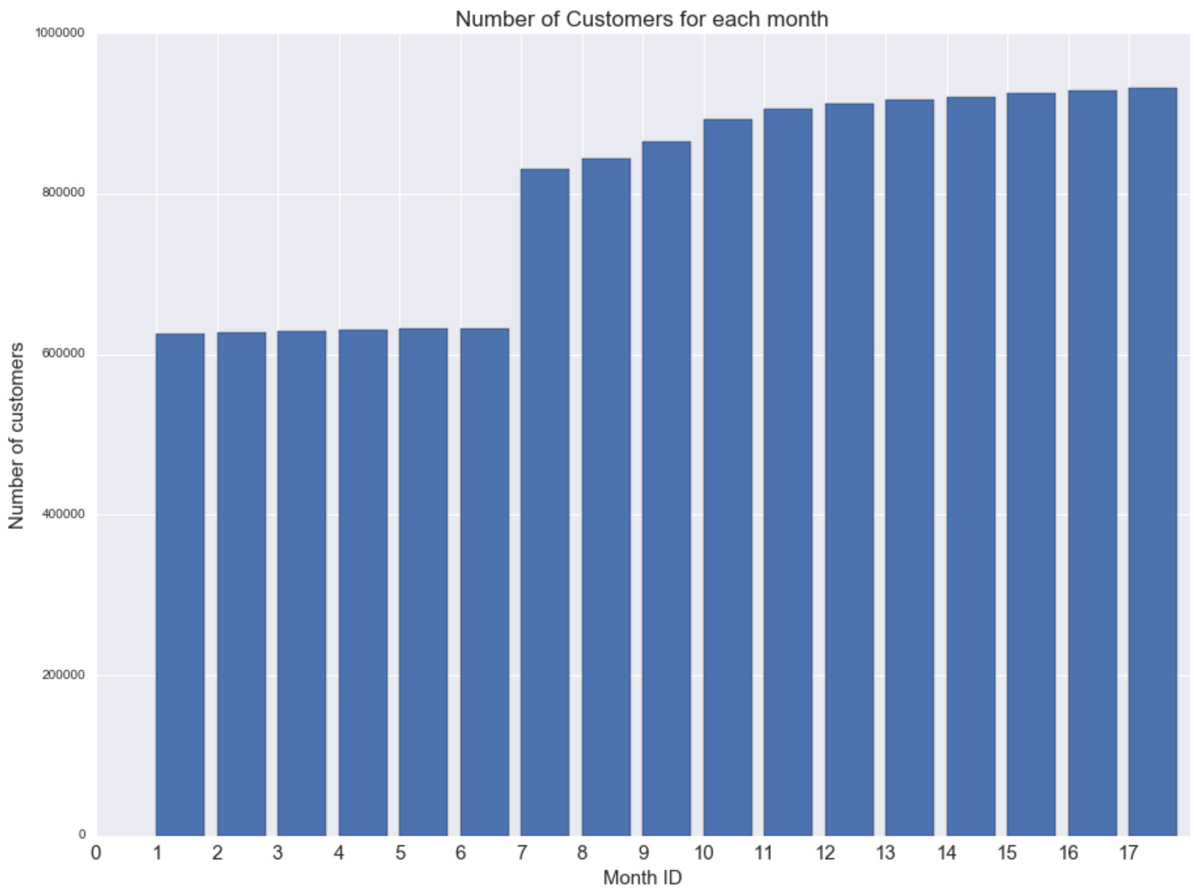
Product Sales Related to Customer's Info - 2016.5





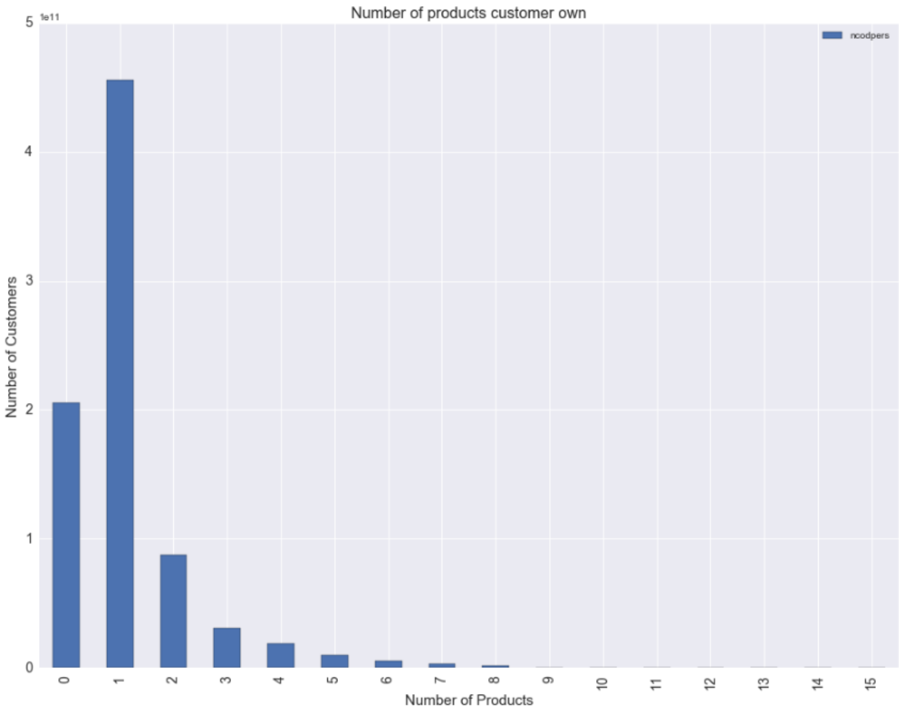
EDA

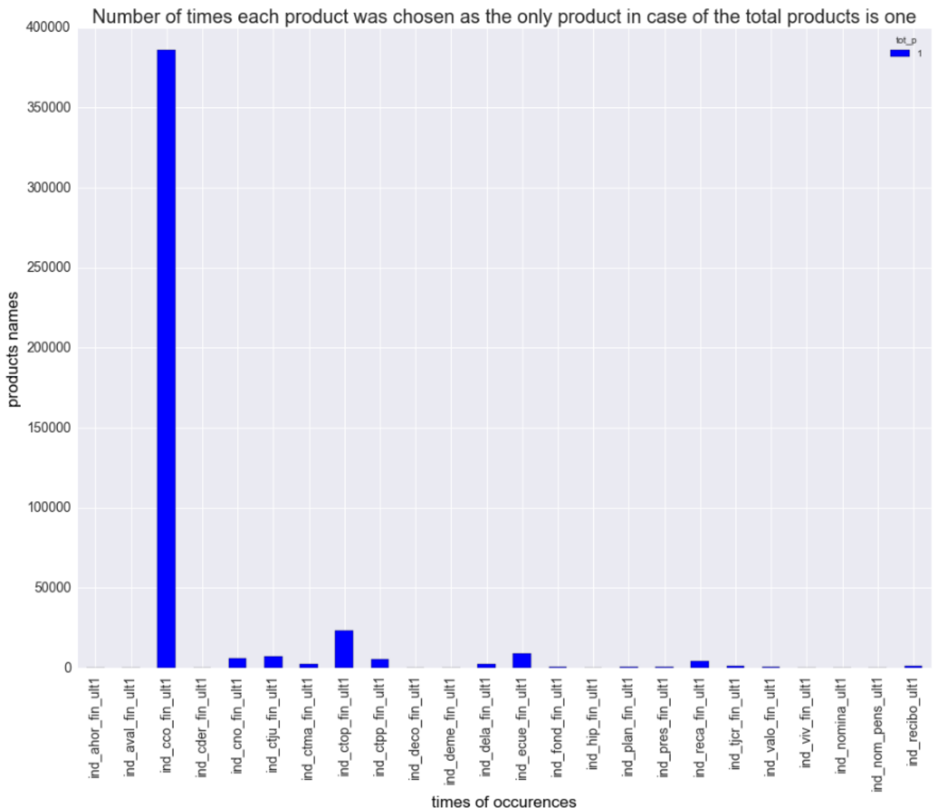
Number of Customers by Time



EDA

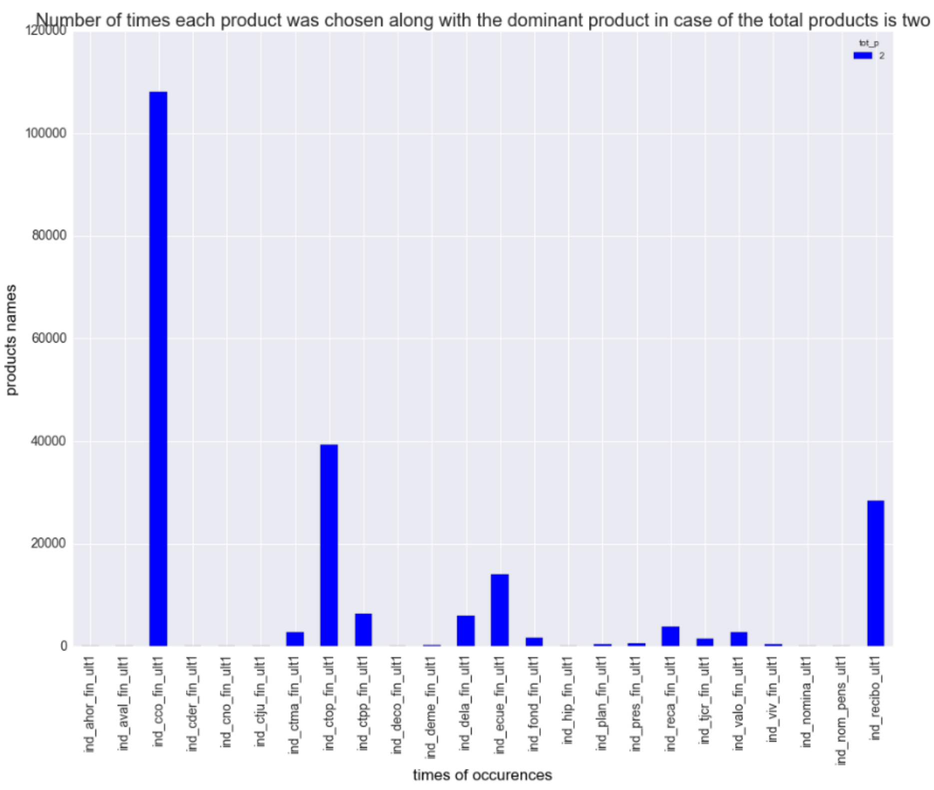
Number of Product Own - 2016.5

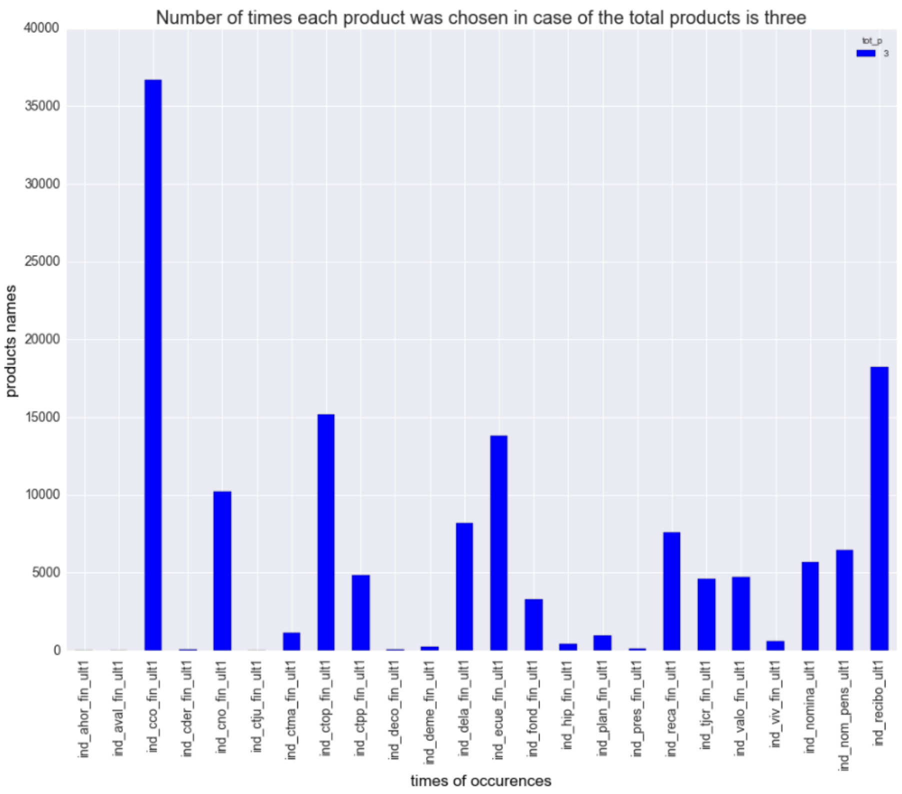




EDA

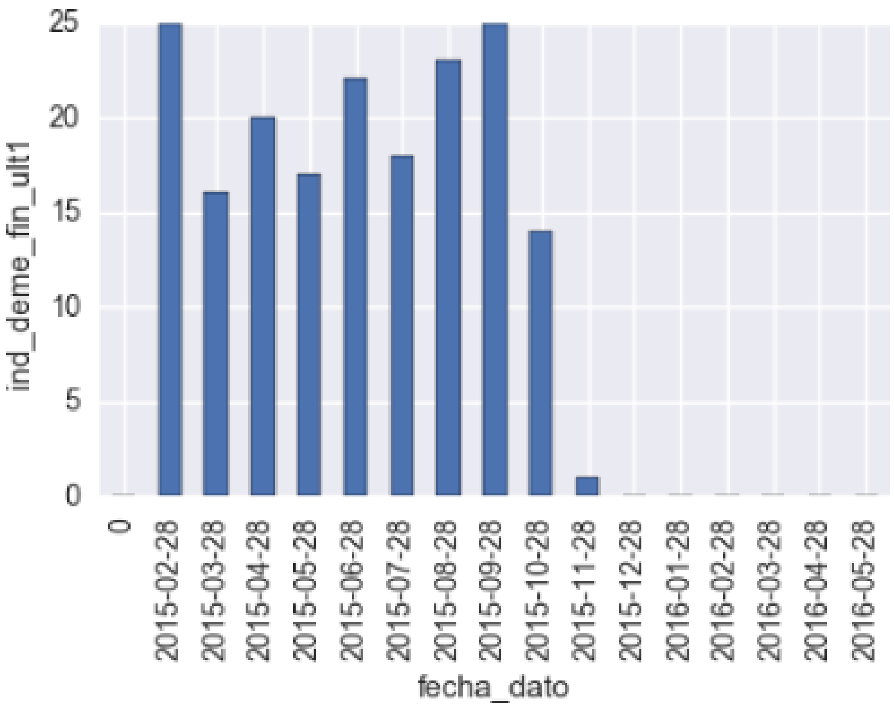
Number of Product Own - 2016.5

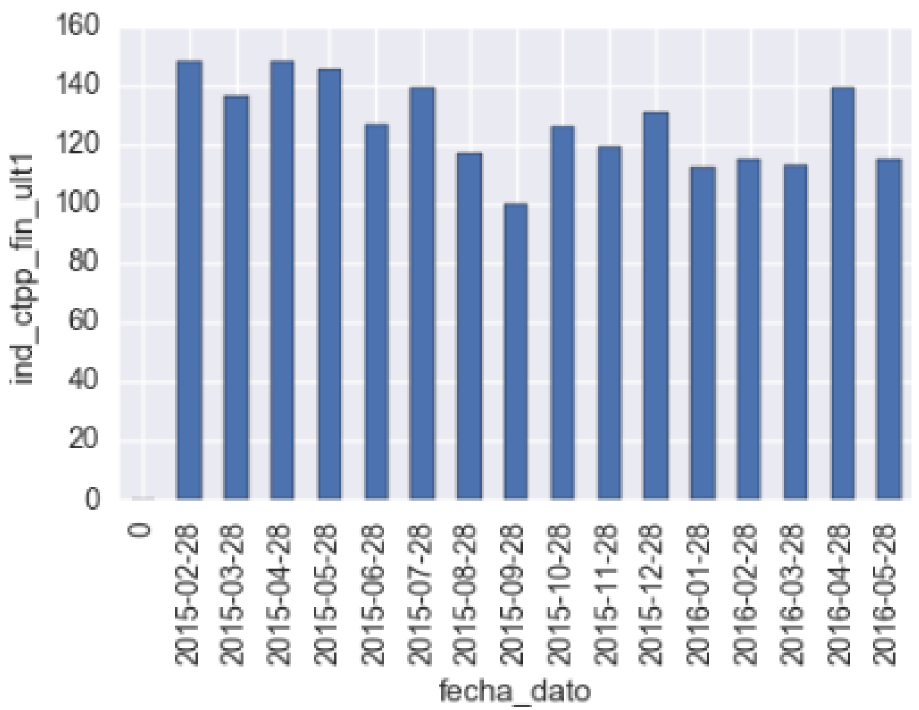




EDA

Number of Product Sales by Time





Feature Engineering

Input Features

Encoding

Output Features

Encoding

Use adjacent month
i.e. 2016.1-2016.2

Use the same month
i.e. 2015.5 – 2016.5

Use the seasonal month
i.e. 2016.3 – 2016.6

Input Features

Previous Month Products

Use adjacent month
i.e. 2016.1-2016.2

Use the same month
i.e. 2015.5 – 2016.5

Use the seasonal month
i.e. 2016.3 – 2016.6

Input Features

Create Change Features

i.e. Previous - Current

Time Series
Pick significant pattern
Level = 0 , 1
&
Create as new input features

Input Features

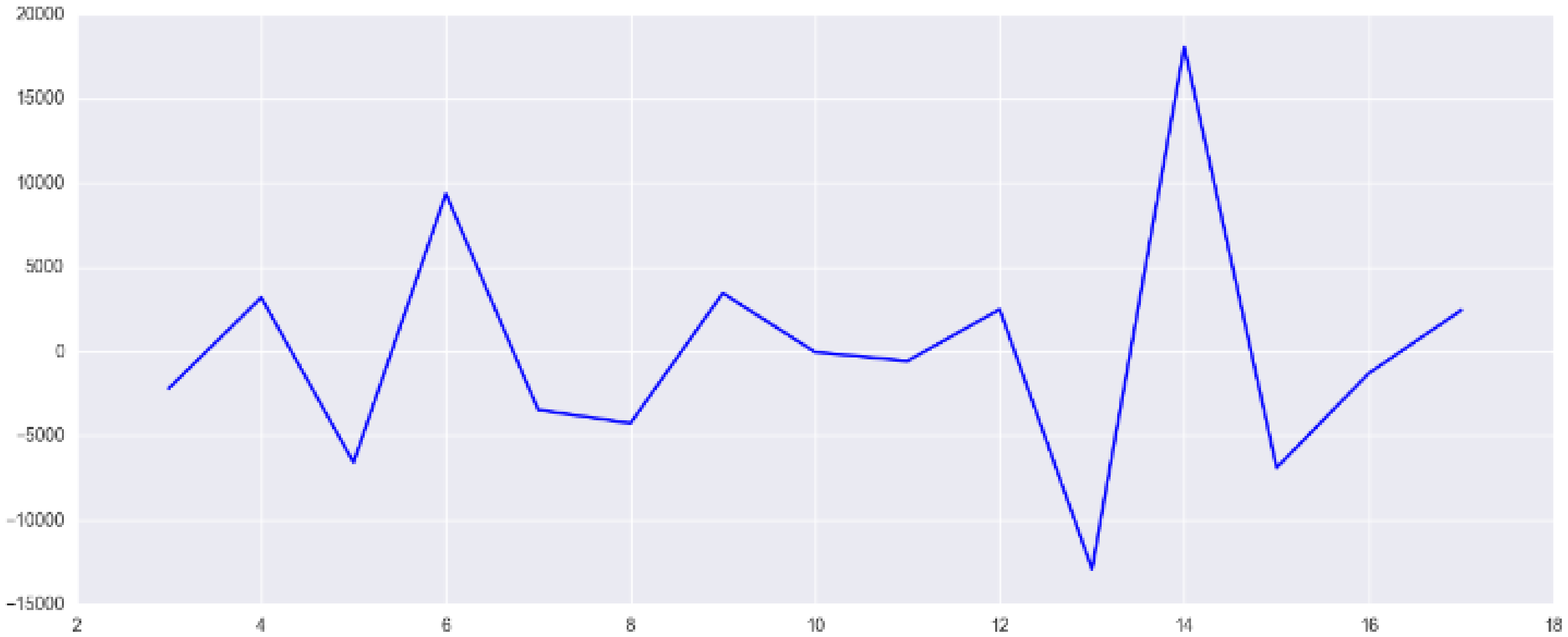
Time Series
Level = -1, 0, 1

Output Features

Drop features & add weight
Based on popularity of the products

Make recommendation based on products' popularity

Time Series



Results of Dickey - Fuller Test

Product Pension

Test Statistic	-3.163039
p-value	0.022226
No. Lags Used	4.000000
Critical Value (5%)	-3.232950
Critical Value (1%)	-4.331573
Critical Value (10%)	-2.748700

Models Training

0.0225

Random Forest

Naïve Bayes & Genetic Algorithm

0.0272

0.0269

XGBoost

Neural Network

0.01995

0.0

AdaBoost

Collaborative Filtering

0.02364

Ensemble - Voting

01 Popularity

02 Collaborative Filter

03 Naïve Bayes

0.0276

01 Popularity

02 Collaborative Filter

03 Naïve Bayes

04 XGBoost

05 XGBoost (new Features)

06 Random Forest (multiclass)

Result & Findings

Using the month from the previous year has better prediction than the previous month from the current year

Single Model, Naïve Bayes has the best performance

From the evaluation, it only penalized the false negative

Multiclass has better performance than multi-labels

Future Steps

- Using best models for model stacking
- Trying more features engineering
- More Ensemble and Stacking