Coupon Redemption Prediction

A report by Monil Gudhka

About

Title:
Coupon Redemption Prediction

Contest:
AmExpert 2019 - Machine Learning Hackathon

Repository:
https://github.com/monilgudhka/coupon_redemption_prediction

Problem Statement

- Client: retailer
- Marketing:
  - Attract new customers
  - Retain & reinforce loyalty of existing customers
  - Discount marketing: widely used promotional techniques
  - Prediction of the redemption behaviour are crucial
- Solution will help in
  - Accurately design coupon construct
  - Develop a more precise and targeted marketing strategies

Evaluation Metric

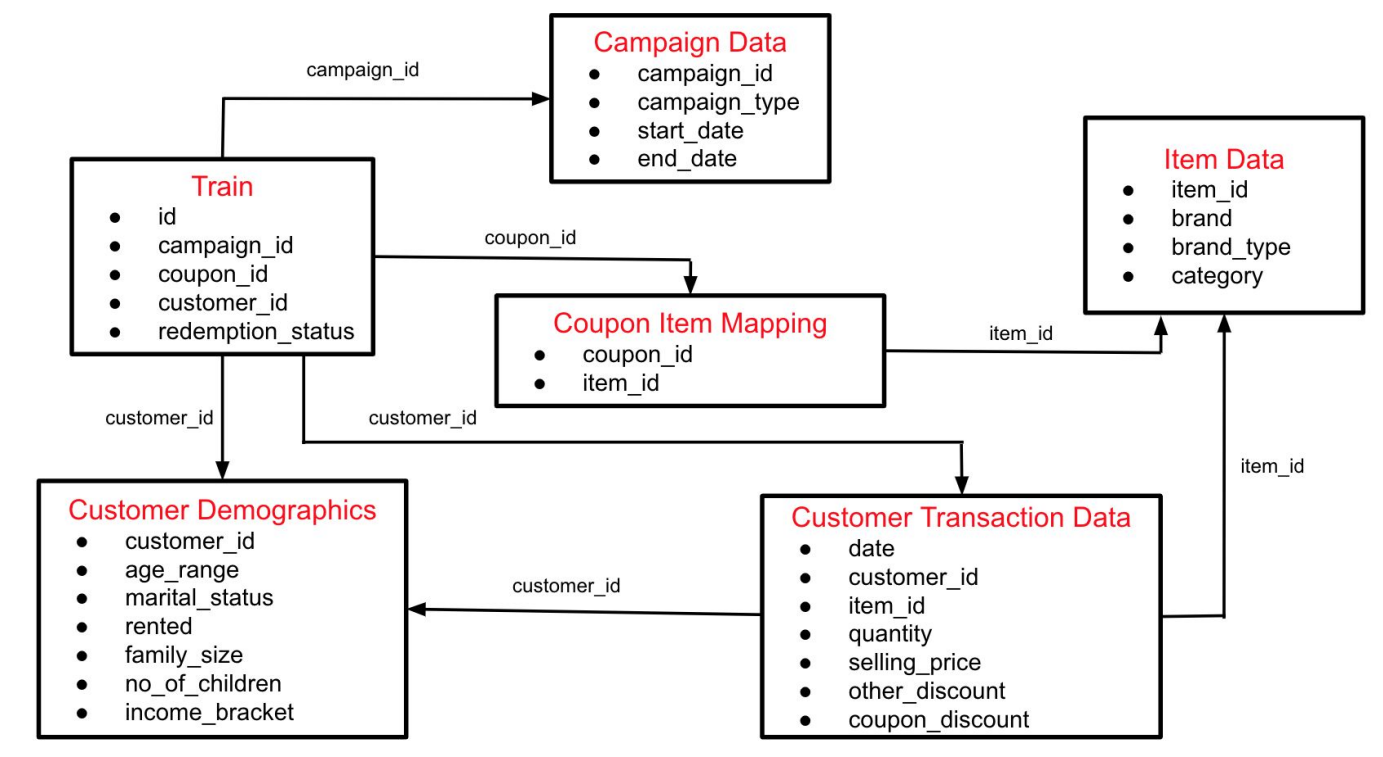
Area under the ROC curve

between the predicted probability and the observed target

(AUC-ROC)

across all entries in the test set

Dataset


Campaign Data
campaign_id
campaign_type
start_date
end_date
campaign_id
Train
id
campaign_id
coupon_id
customer_id
redemption_status
coupon_id
Coupon Item Mapping
coupon_id
item_id
item_id
Item Data
item_id
brand
brand_type
category
customer_id
customer_id
item_id
Customer Demographics
customer_id
age_range
marital_status
rented
family_size
no_of_children
income_bracket
customer_id
Customer Transaction Data
date
customer_id
item_id
quantity
selling_price
other_discount
coupon_discount

Data Cleaning

Missing data:

no_of_children:	Assuming it to be Zero
marital_status:	If family_size - no_of_children > 1 then Married else Single
Customers without Information:	Handle them with algorithm

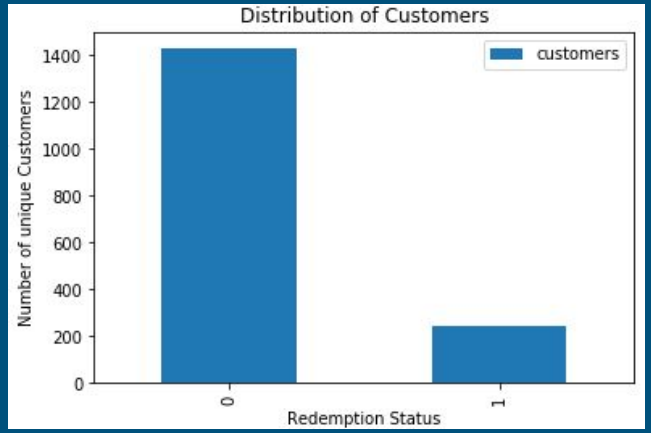
Outliers:

- Many outliers in the customer’s transactions
- Prediction algorithm has to be trained with outliers

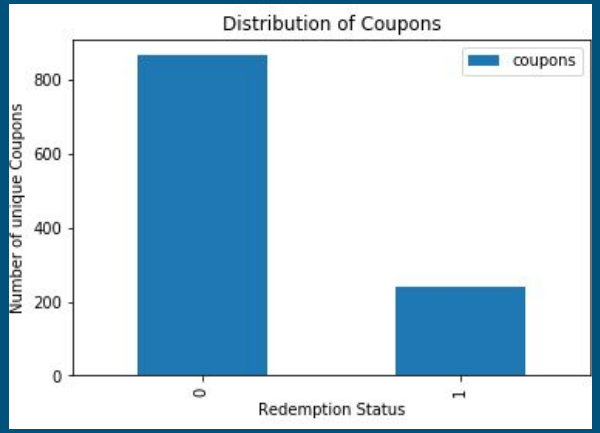
Data Merging

- Coupon Information
  - Extracting summary variables from coupon information
- Customer Behaviour
  - Extracting summary variables from customer’s transactions
- Campaign and Customer Information
  - Left Join of both data on customer_id
- Campaign and Coupon specific Customer Behaviour
  - Transaction summary from 223 days prior to 110 days prior of campaign’s start date
  - Transaction summary within the same date range but for items covered in the coupon
- Deriving Features
  - Represents change in customer behaviour
  - Represents match of customer behaviour with that of coupon

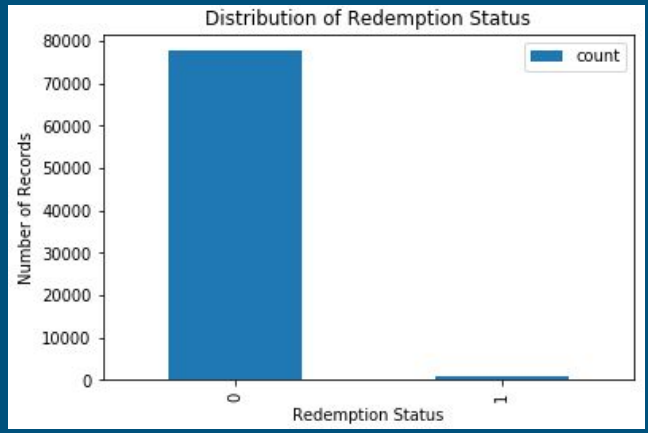
Analysis: Redemption



78% coupons were never redeemed by any customers

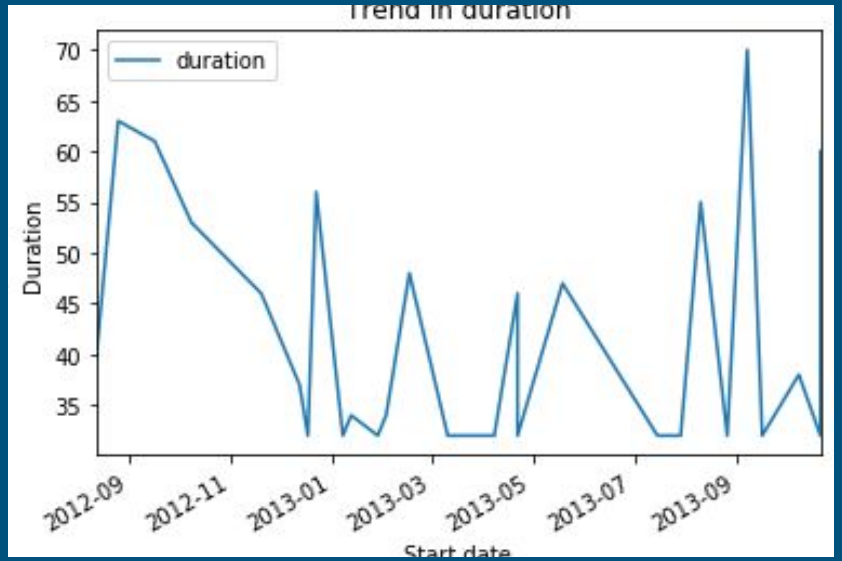


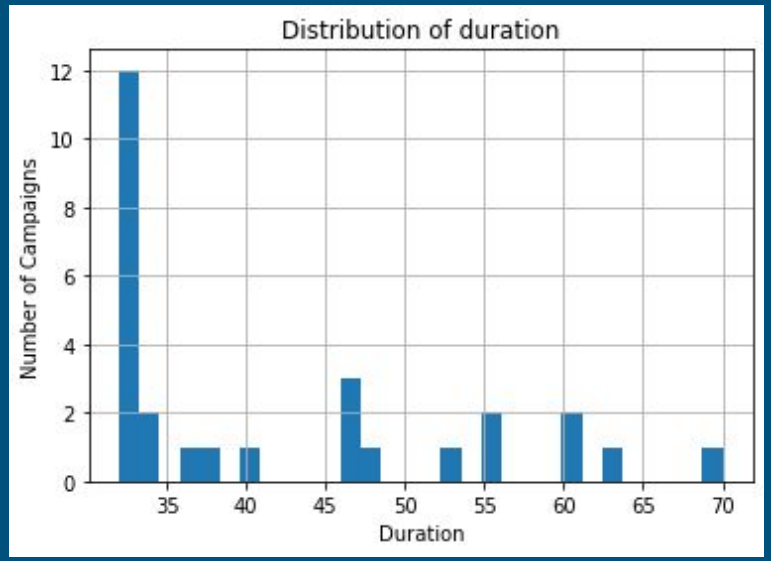
85% customers has never redeemed any coupons



0.93% campaign, coupon and customer combination has positive redemption status

Analysis: Campaigns

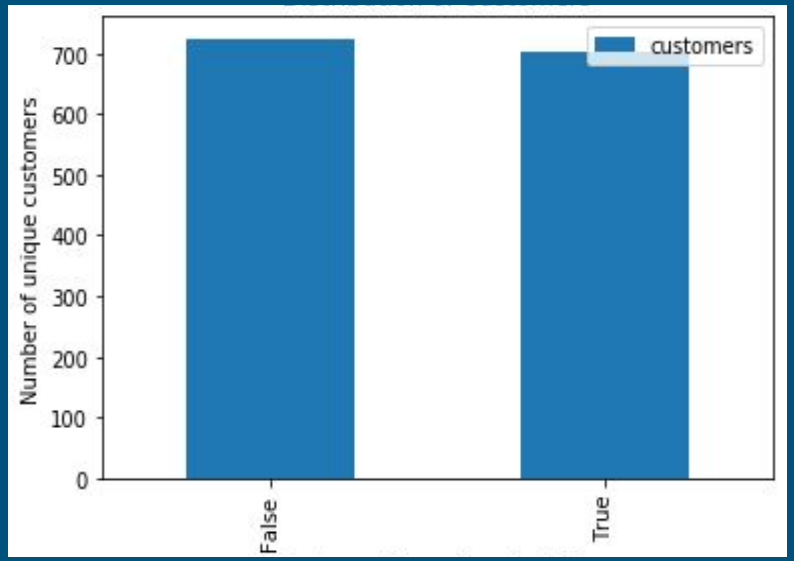


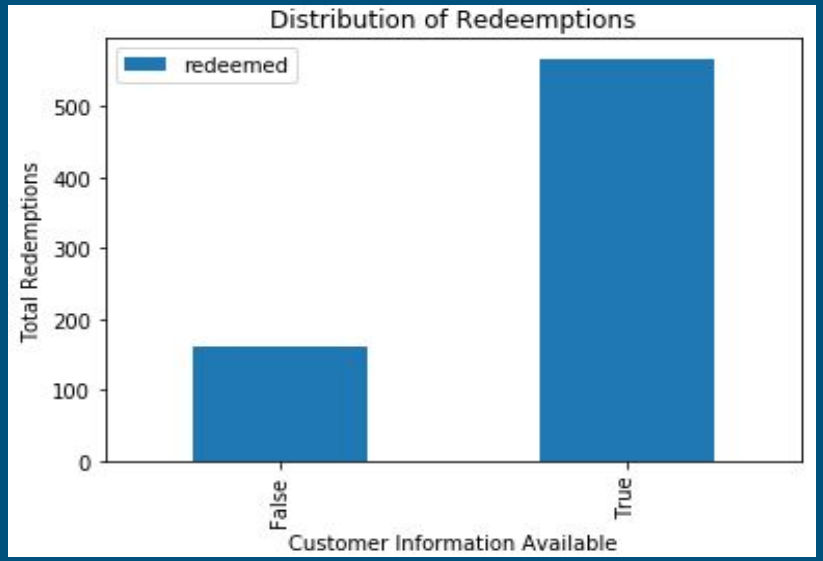


- Longer Campaigns in start
- Later on introduced campaigns with 35 or less days

- One Campaign of 70 days

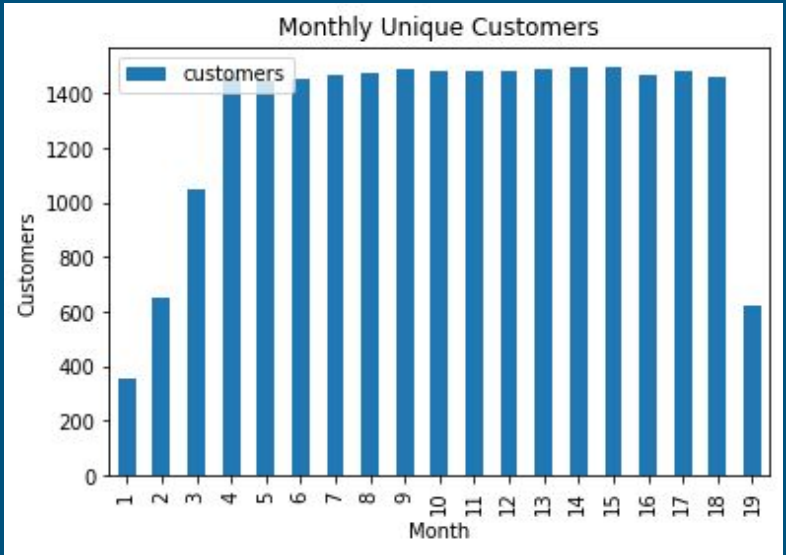
Analysis: Customer’s Information

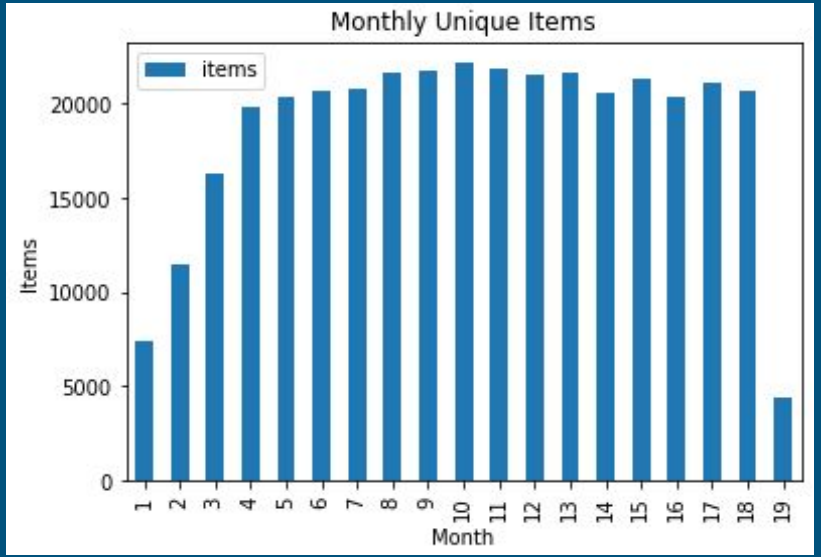




- 50% Customer's information is not available
- Customers with information redeems more coupons than customers without any information

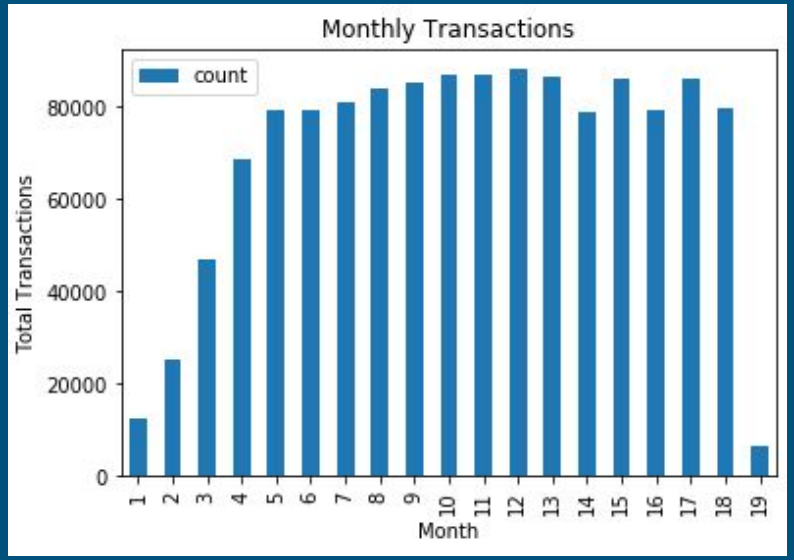
Analysis: Customer Transactions





- Not much transaction in the initial 3 months
- Thereafter, Customers and Items remains almost stable

Analysis: Customer Transactions *(continued)*



- Total transactions also remained stable after initial 3 months
- Transactions with coupon discount has fluctuations

Modelling

Defaults in missing values	0.9016	>	0.8973	NaN in missing values
With Customer Behaviour	0.9016	>	0.8690	Without Customer Behaviour
With Percentage and Count	0.9016	>	0.8803	Only Percentage
Tuned LightGBM	0.9060	>	0.9016	Default LightGBM
Combining models	0.9132	>	0.9060	One model

Combining models gave the private score of 0.8979.

Further Improvements

1. Merging the existing features
2. Different summary features
3. Parameter tuning
4. Fixing FPs and FNs from confusion matrix

Thank You