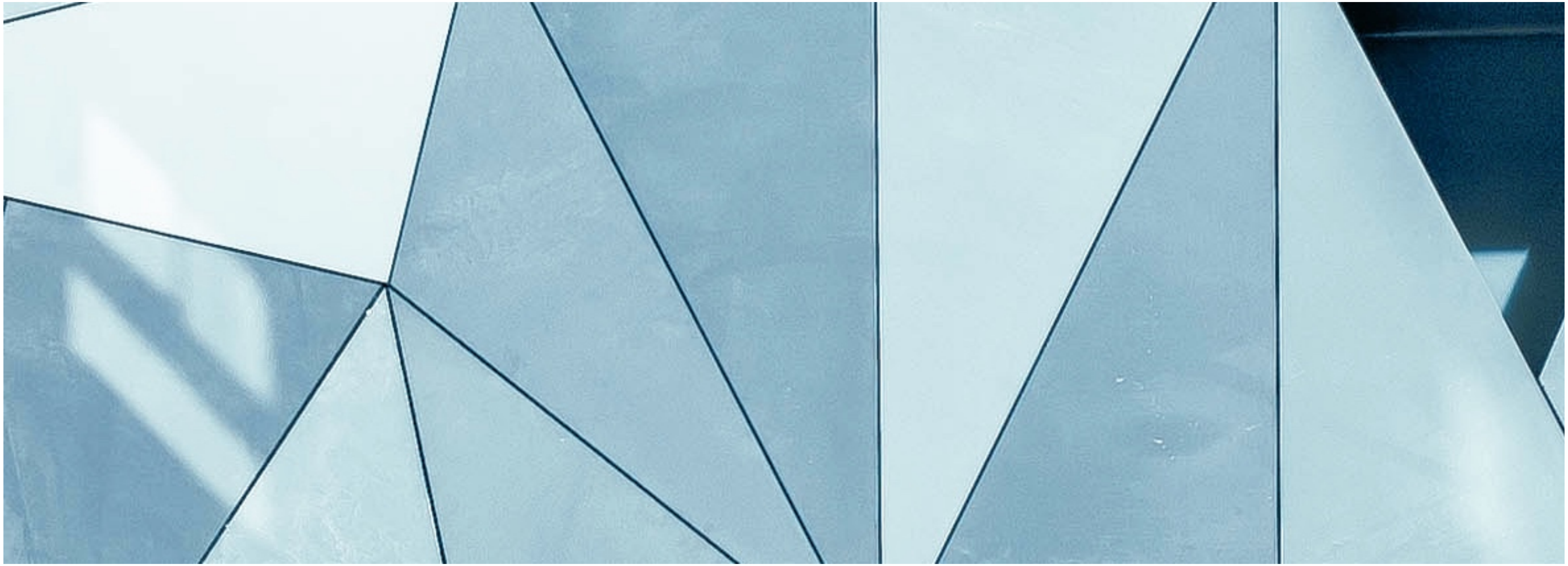


TARGET AUDIENCE FOR DIRECT MARKETING IN STARBUCKS REWARDS MOBILE APP

Katerina Bosko, PhD

- September 2019 -

CONTENTS

PROBLEM STATEMENT

GOAL:

Target Audience For A Marketing Campaign

QUESTIONS:

- Which Customers Love Coupons?

- Which Don't?

- What Types of Offers Send to Whom?

SOLUTION:

Customer Segmentation

Using PCA & K-means Clustering

METRICS:

Response Rate (RR)

Conversion Rate (CVR)

DATASET OVERVIEW

portfolio.json (10 offers x 6 fields)
- offer types sent during 30-day test period (bogo, discounts, informational offers)

profile.json (17000 users x 5 fields)
- demographic profile of app users (age, income, gender, membership duration)

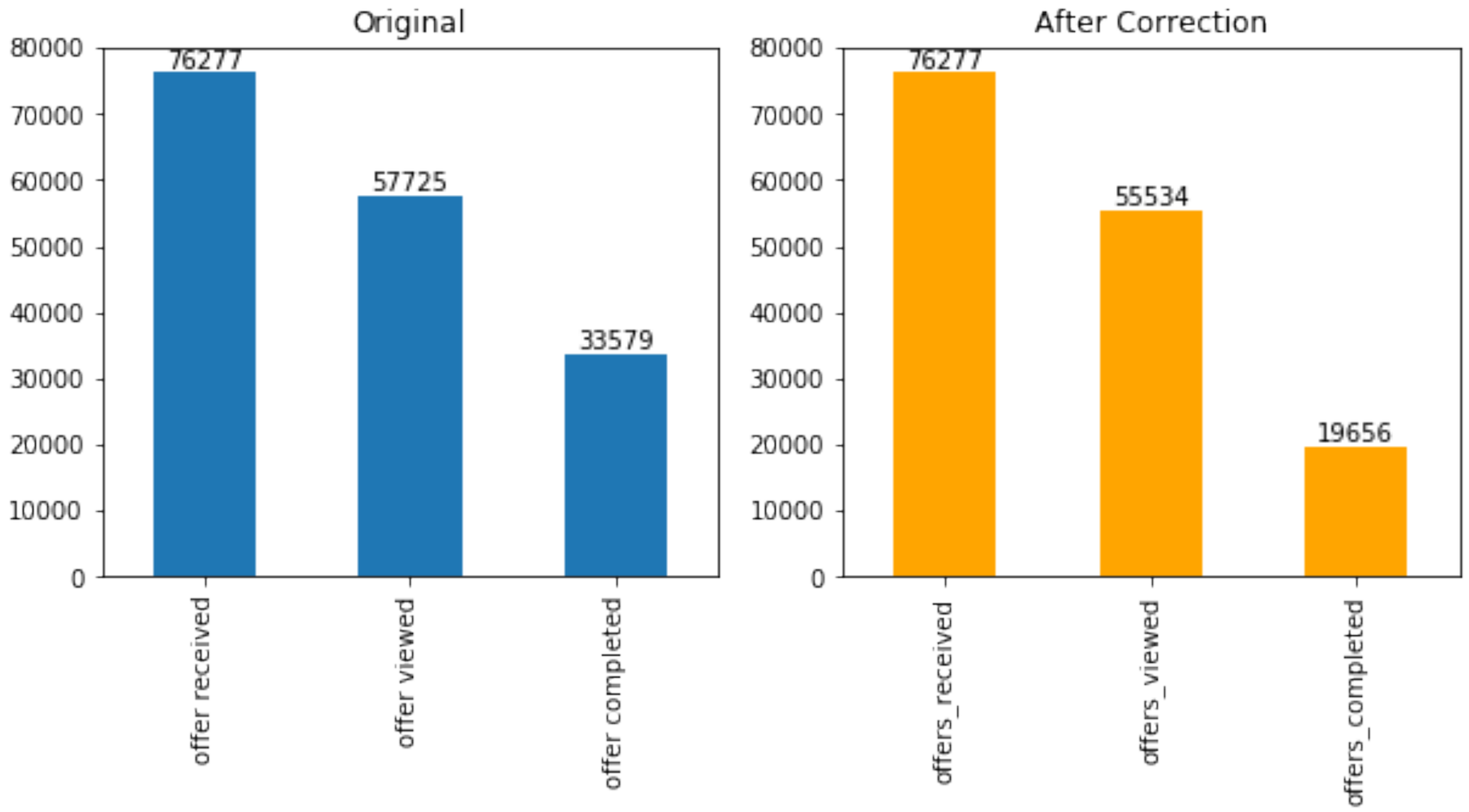
transcript.json (306648 events x 4 fields)
- event log on transactions, tracking of offers received, viewed, completed

DATA CLEANING

Challenge: improper tracking of offers viewed & offers completed

Impact: correct tracking could save $70,000 per month

DATA CLEANING RESULTS



FEATURE ENGINEERING

Calculated 27 new Features for Each Customer

total_amount

offers_received

offers_viewed

offers_completed

transactions_num

bogo_received

bogo_viewed

bogo_completed

avg_order_size

discount_received

discount_viewed

discount_completed

avg_reward_size

informational_received

informational_viewed

avg_bogo_size

avg_discount_size

total_rewarded

total_bogo

total_discount

offers_rr

bogo_rr

discount_rr

informational_rr

offers_cvr

bogo_cvr

discount_cvr

EXPLORATIVE DATA ANALYSIS

Average Starbucks Rewards App Customer:

- middle-aged (median - 55 years)
- $64000 income
- spent $104.44 in total
- got $5.6 rewarded
- made 8 transactions spending $13.34 per order
- received 4-5 offers, viewed 3 offers, completed 1 offer

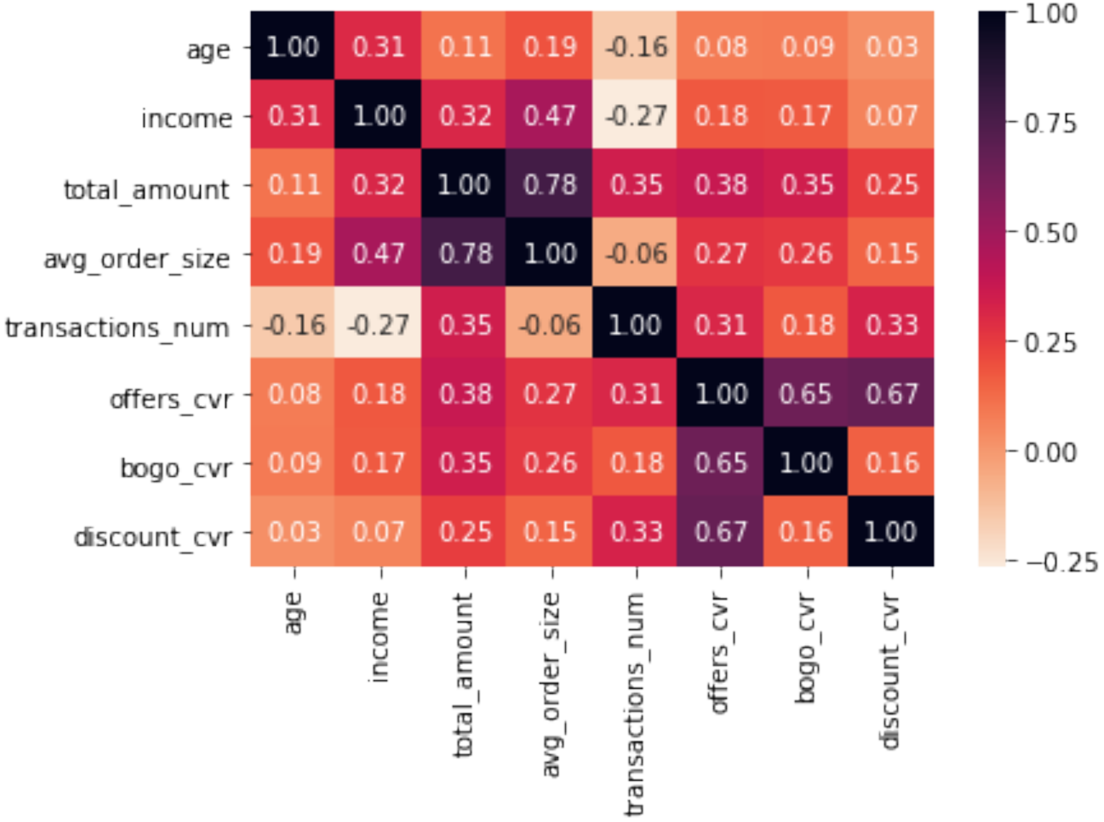
Metrics:

Response Rate (RR) - 73% of offers received

Conversion Rate (CVR) - 34% of offers viewed

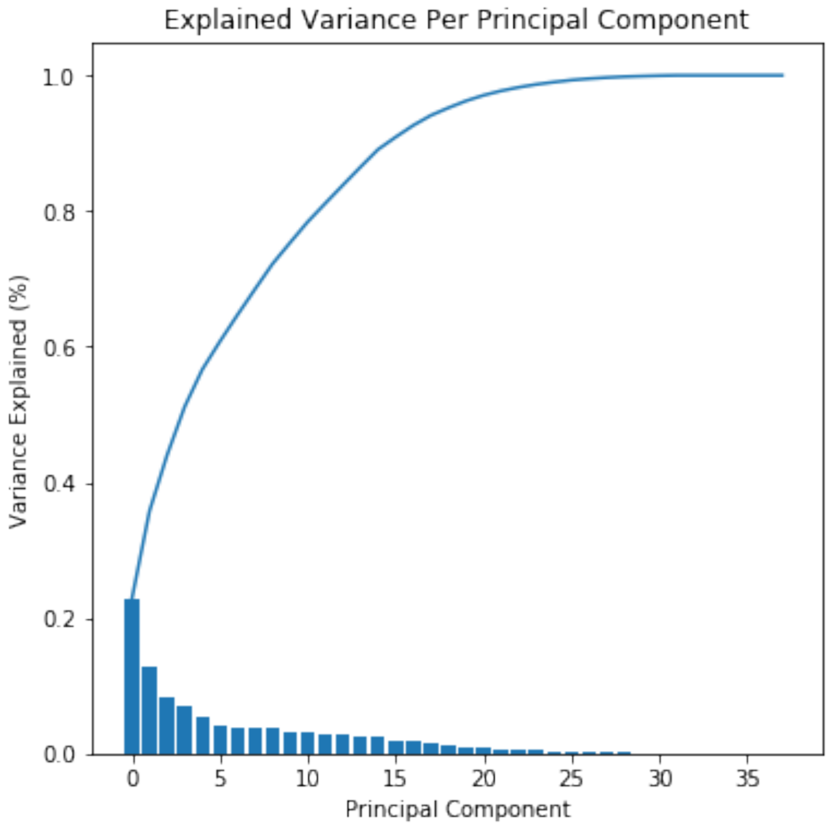
EXPLORATIVE DATA ANALYSIS

Correlation Matrix:



Not demographics,
but spending habits correlate more with Conversion Rates.

DATA PRE-PROCESSING



Imputed Missing Values
(12.8% of data in the profile dataset)

One Hot Encoded Categorical Features

Scaled Features with Standard Scaler

Reduced Dimensionality with PCA:

kept the first 10 components that in total captured almost 80% of the variance in the data.

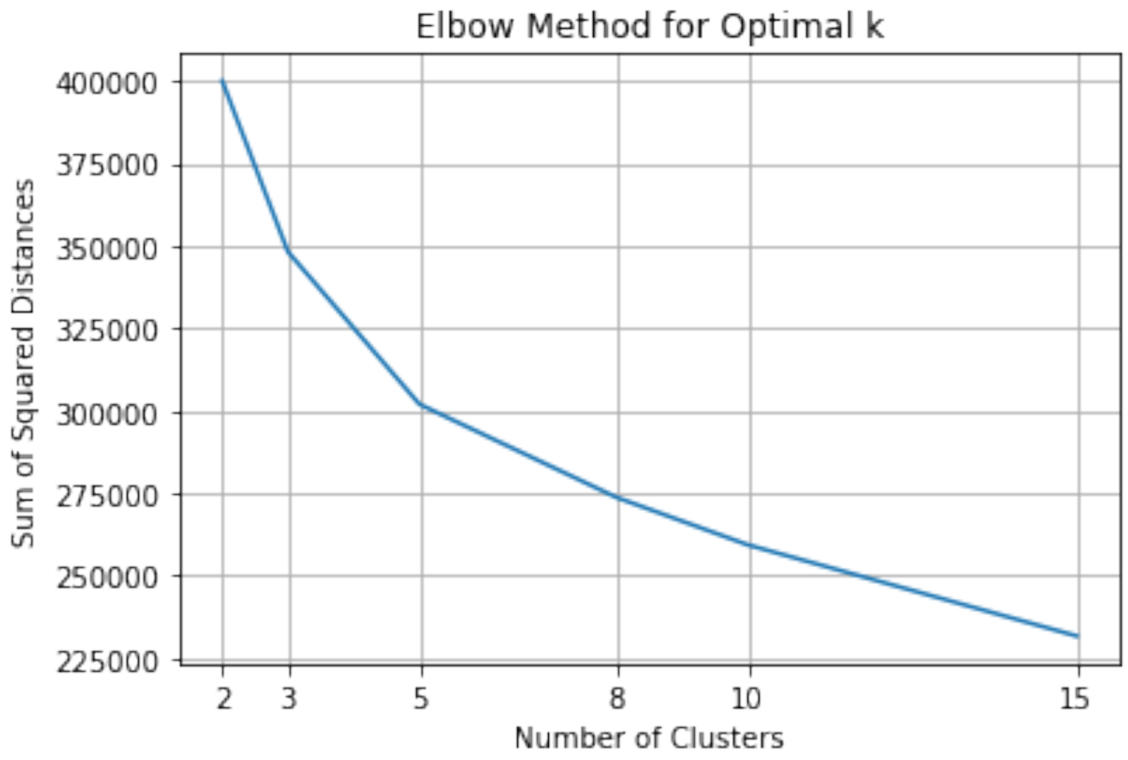
MODELLING

Run K-Means Clustering on Pre-Processed Data

Decided upon Optimal Cluster Number:

- Elbow Method – 3 or 5 clusters
- Silhouette Score – for 3 higher than for 5
- Visual Validation – 3 better than 5

Decision: 3 clusters



MODELLING RESULTS

Cluster 1 - “Disinterested”:

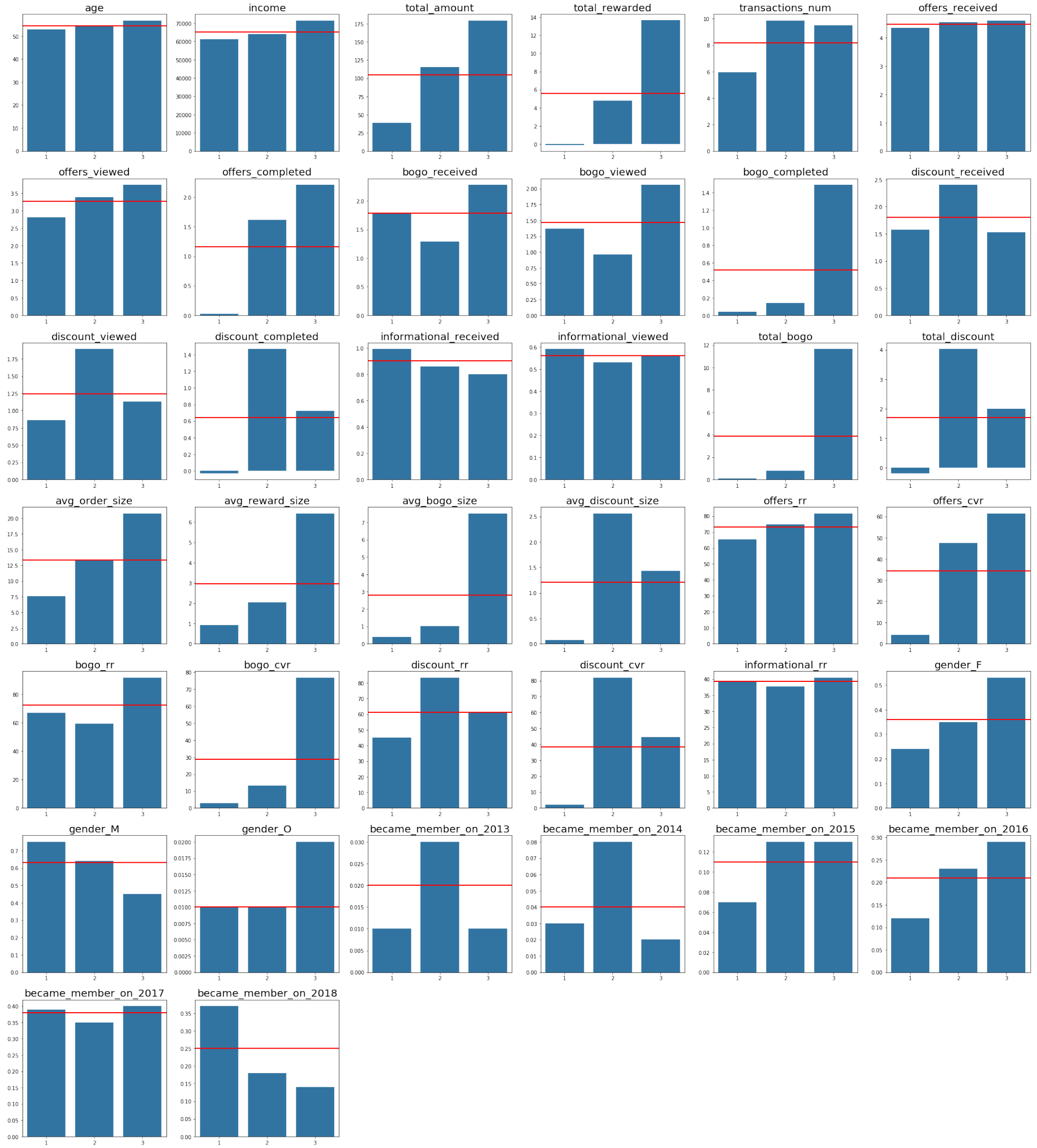
This group of customers are predominantly male that just recently became members. They tend to spend not much with below average number of transactions and small average order size. Although slightly more than 60% in this group view offers, they don’t complete them.

Cluster 2 - “Discount-Type”:

This group of customers are also mostly male but with the longest membership status (since 2013/2014). They tend to receive more discounts, which they love and actively complete. Their spending habits are slightly above average - they make small orders but buy frequently.

Cluster 3 - “Bogo-Type”:

This is the only segment where female dominate over male. The customers in this group tend to be older and have higher income. They are loyal customers for few years already. They spend a lot - make huge orders and buy frequently. With such spending habits, no wonder that they are interested in bogo and get rewarded the most. They complete bogo offers way beyond average, but also react to discounts from time to time.



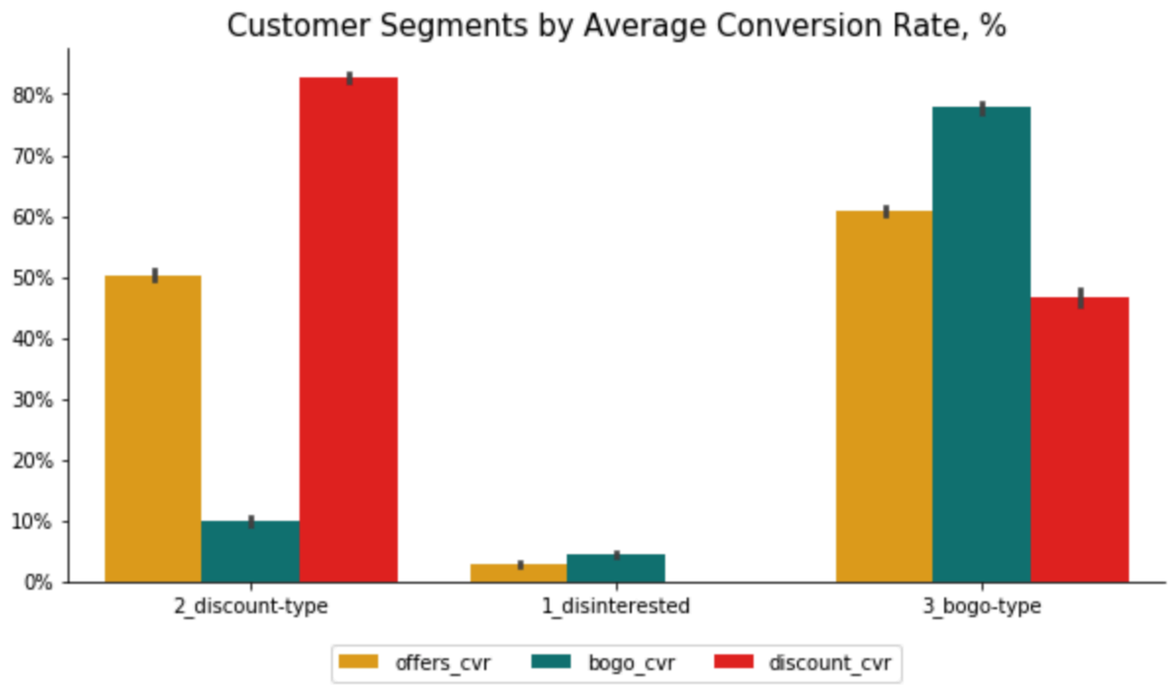
BUSINESS CASE CONCLUSION

3 Customer Segments:

- Disinterested (~40%)
- Discount-Type (~31%)
- Bogo-Type (~29%)

Next Step:

A/B testing
with push notifications



TECHNICAL CONCLUSION

Performed full Analysis Cycle

- cleaning & preprocessing the data
- dealing with missing values
- feature engineering
- feature scaling
- one hot encoding
- dimensionality reduction
- clustering

Automated Reporting by refactoring & writing functions