Qi-Wen Ng | DSI 14

Customer Segmentation Analysis with STARBUCKS

Outline

The Problem

Data Cleaning & Wrangling

Clustering & Analysis

Conclusion & Recommendations

STARBUCKS REWARDS™

16,000,000

active members

40%

of total sales from rewards program

Available on a mobile app to **market products**, send out **offers**, collect **payments**, and give **rewards** for freebies.

The Problem

Problem statement

~~AVERAGE CUSTOMER~~

Given the **diversity** in customer behavior, how can we identify and create experiences for **effective acquisition and retention**?

Problem statement

Customer Segmentation:

process of dividing customers into groups based on common characteristics

Problem statement

How can we use unsupervised machine learning to segment our customers?

Stakeholders

Marketing

- What offers should be sent to each customer ?

Sales

- Which customers we should focus for retention?
- Who are our most profitable customers?
- Which customers to nurture or focus for potential growth?

Data Cleaning & Wrangling

Dataset

Raw Dataset

promotion.json:
10 promotional offers metadata:
BOGO, discount, informational

profile.json:
17,000 customer demographics:
age, income, gender, date as member

event-log.json:
306,534 timed customer activity logs: **transactions, offers received, viewed, completed**

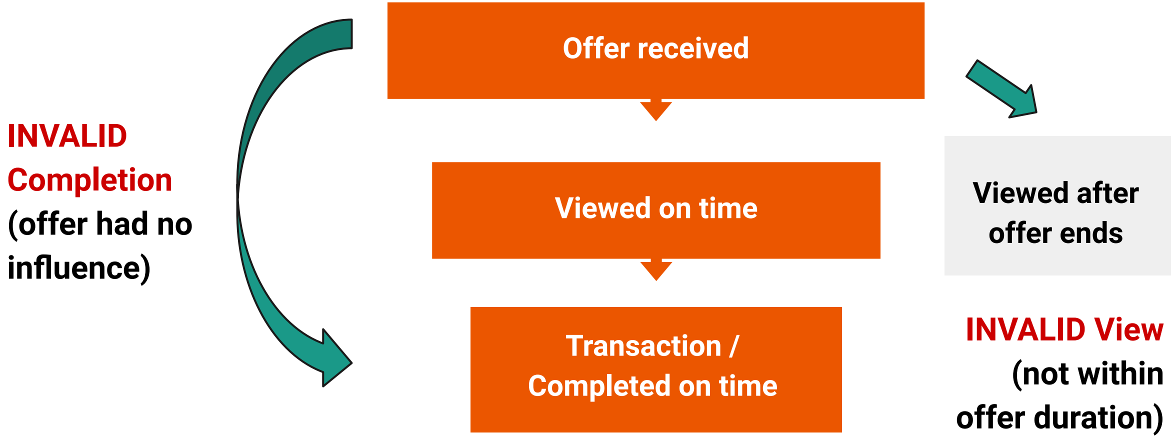
Dataset

Raw Dataset

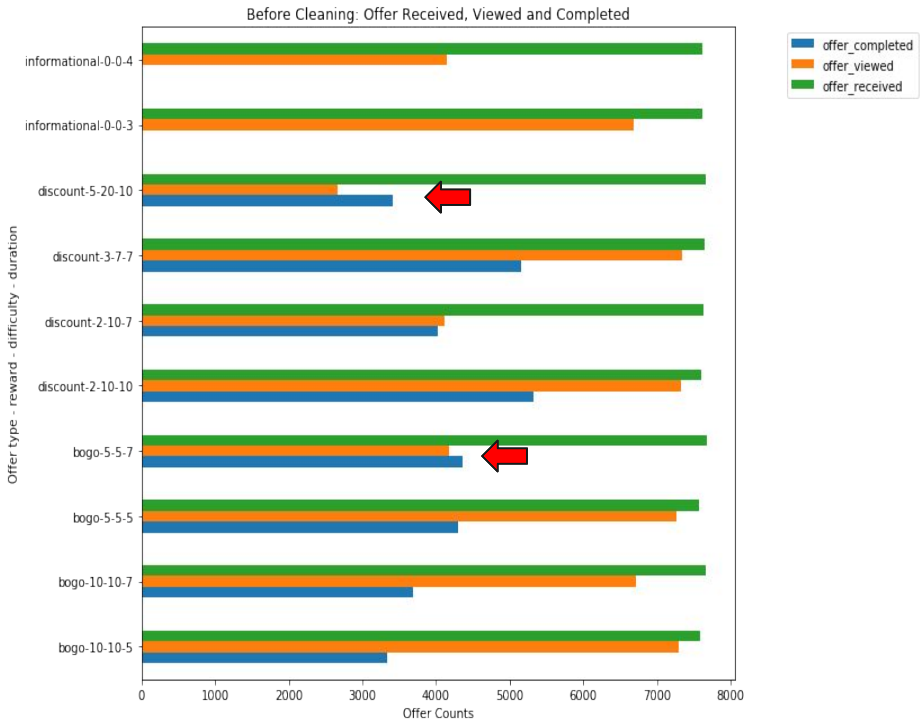
promotion.json:
10 promotional offers metadata: **BOGO, discount, informational**

profile.json:
17,000 customer demographics: **age, income, gender, date as member**

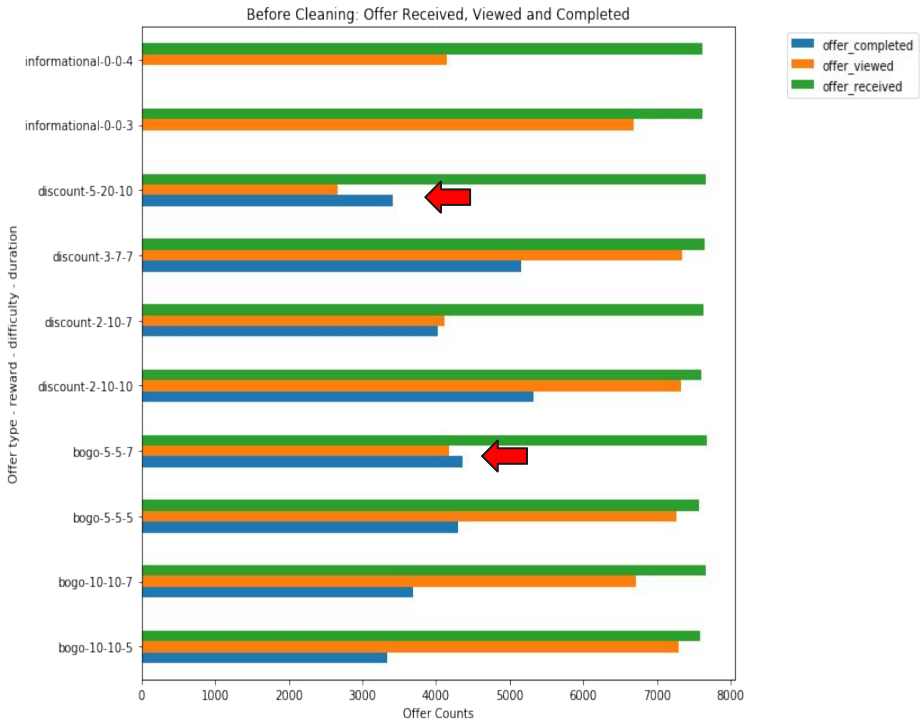
event-log.json:
306,534 customer timed activity logs: **transactions, offers received, viewed, completed**

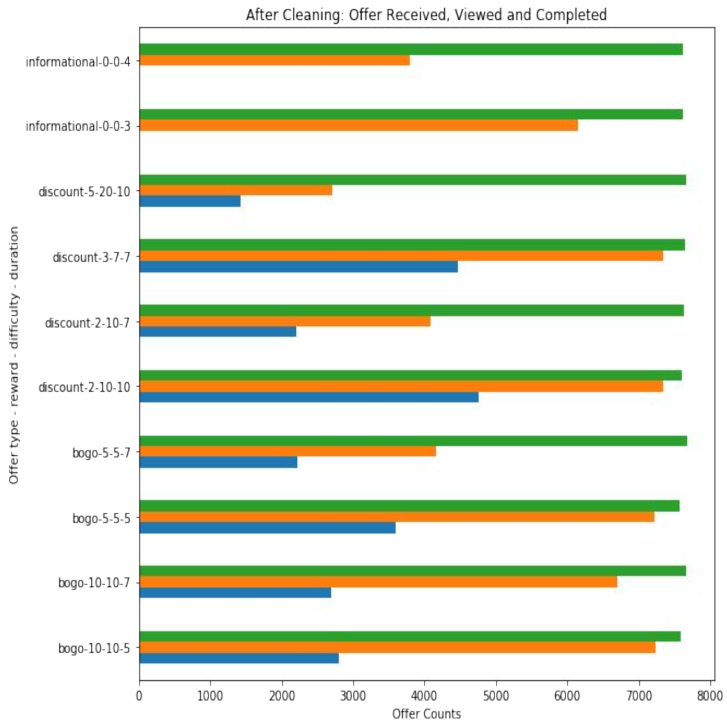


Data Cleaning

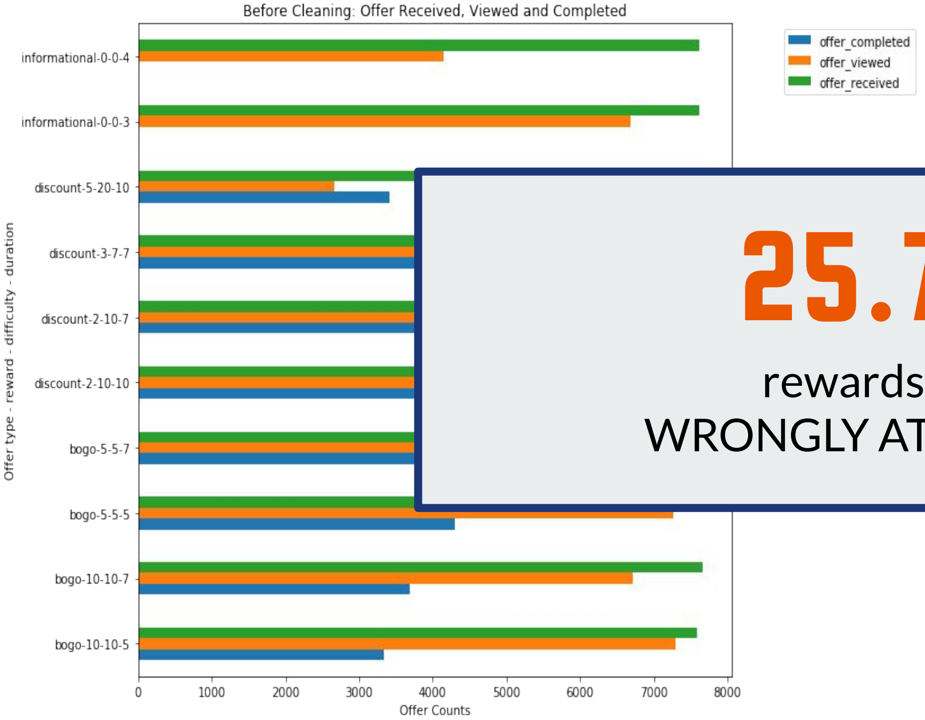


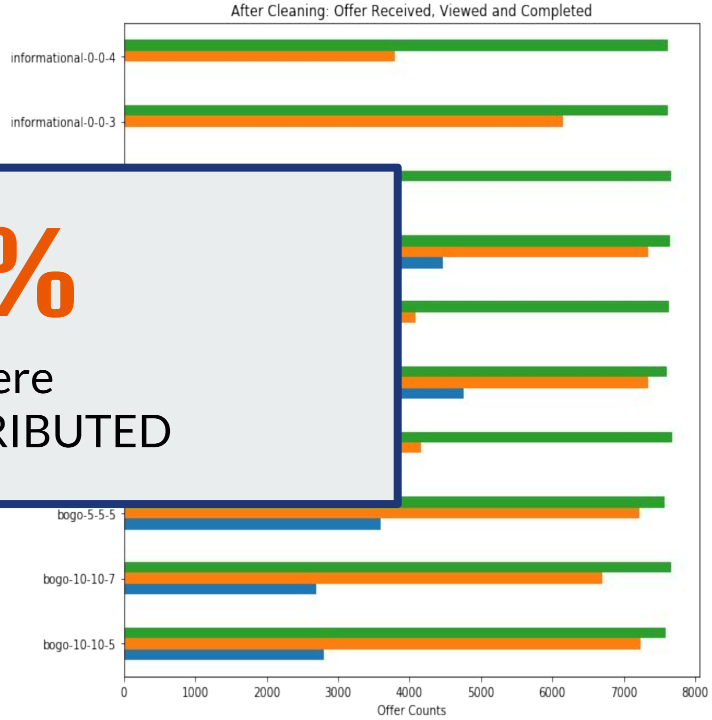
Data Cleaning





Data Cleaning





25.7%

rewards were
WRONGLY ATTRIBUTED

Dataset

Raw Dataset

Cleaning & Preprocessing

promotion.json:
10 promotional offers metadata:
BOGO, discount, informational

profile.json:
17,000 customer demographics:
age, income, gender, date as member

event-log.json:
306,534 customer timed activity logs: **transactions, offers received, viewed, completed**

- Misattributions: identify valid views, valid completions
- Repeat exposures to same offers by assigning unique offer ids

Features engineered:

- Days as member
- Viewed rates, conversion rates for each offer type
- Total spent and number of transactions (offer and non-offer related)
- Recency

Dataset

Raw Dataset	Cleaning & Preprocessing	Final Dataset

promotion.json:
10 promotional offers metadata: **BOGO, discount, informational**

profile.json:
17,000 customer demographics: **age, income, gender, date as member**

event-log.json:
306,534 customer timed activity logs: **transactions, offers received, viewed, completed**

- Misattributions: identify valid views, valid completions
- Repeat exposures to same offers by assigning unique offer ids

Features engineered:

- Days as member
- Viewed rates, conversion rates for each offer type
- Total spent and number of transactions (offer and non-offer related)
- Recency

Aggregate all into a final **customer-centric dataset**

Each unique customer: demographics, spendings, behaviours

Total 37 features

Clustering & Analysis

Clustering: Feature Selection & Scaling

View Rates:
bogo_viewed_rate, discount_viewed_rate, info_viewed_rate

Conversion Rates:
bogo_conversion_rate, discount_conversion_rate

Rewards:
total_rewarded

Frequency / number of transactions:
offer_num_transactions, actual_num_transactions

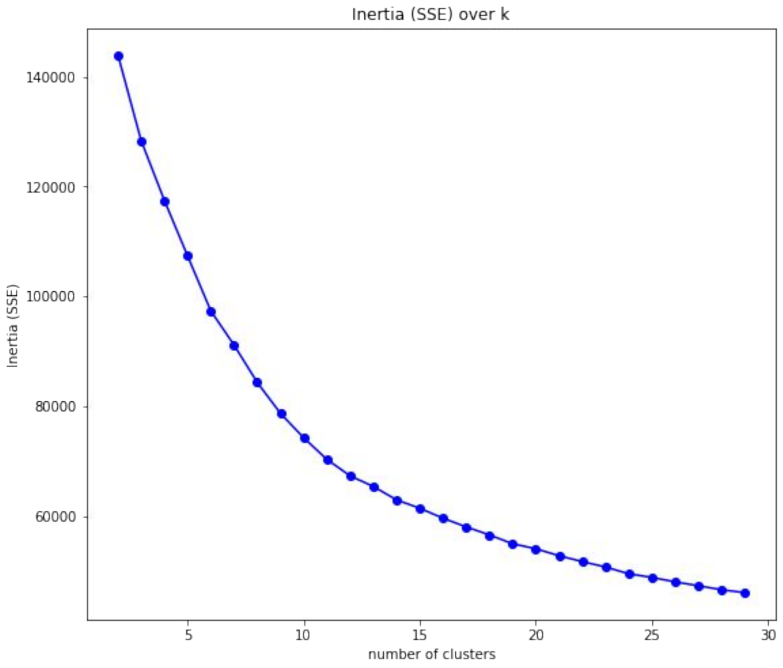
Monetary / amount spent:
offer_spent, actual_spent

Recency / last visit:
recency

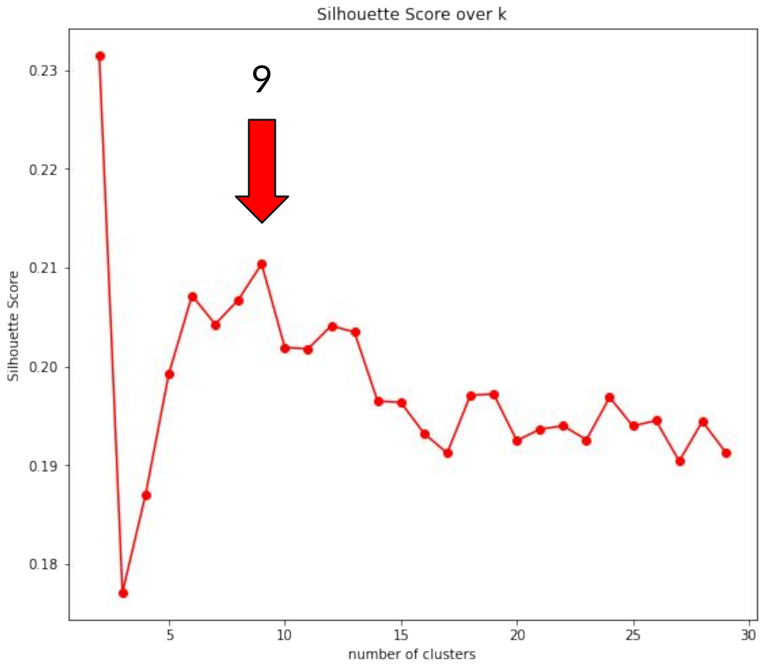
StandardScaler

Clustering: Searching for optimal number of clusters

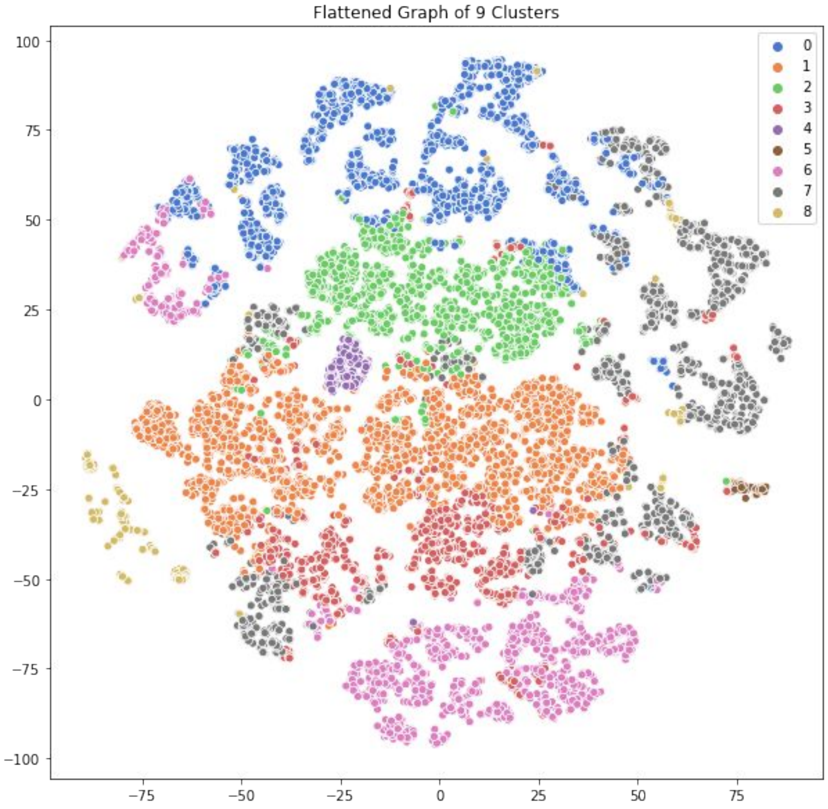
Inertia (Intracluster distance)



Silhouette (Intercluster distance)



Clustering: Running K-Means & Visualizing Clusters



Cluster Analysis: What offers to send to our customers?

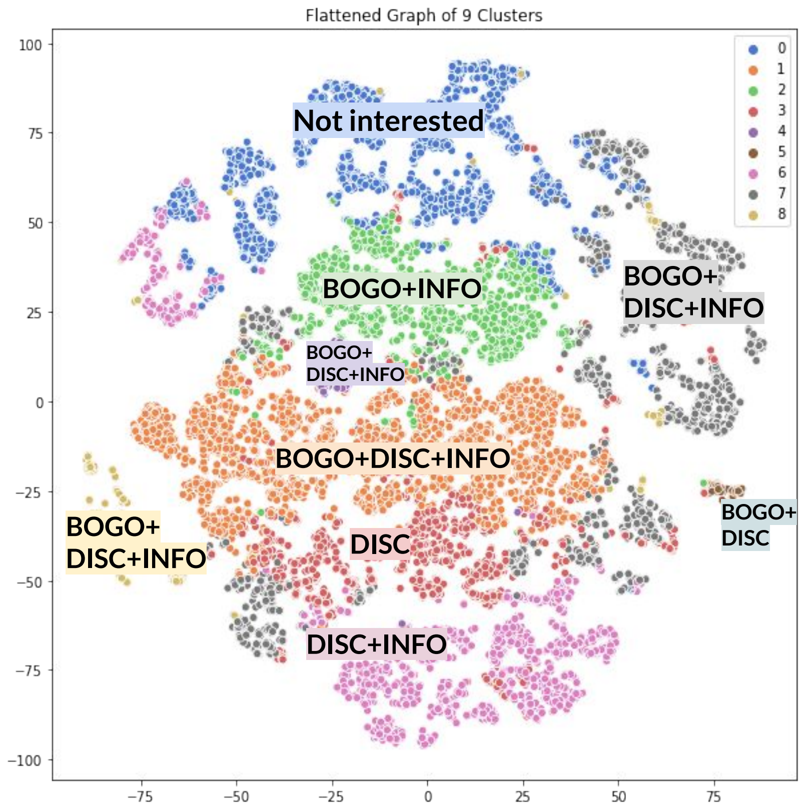
BOGO & DISCOUNT: Cluster 1, 4, 5, 7, 8

BOGO only: Cluster 2

DISCOUNT only: Cluster 3, 6

INFORMATIONAL: All clusters except 3 and 5

NOT INTERESTED (YET): Cluster 0



Cluster Analysis: Who are our most profitable customers?

Cluster 1, 4, 5

- Common: **older age, high income, long-time members.**
- Usual offers + make them **feel exclusive**

Cluster 2, 3, 6

- Common: **long-time** members
- Continue **engaging**, ensure positive experience.

Cluster 0, 7, 8

- Common: **newer** members.
- Reengage, drive subsequent purchases, **retention** is important.

Cluster Analysis: Which customers to focus for potential growth?

Similar attributes to clusters in higher profitability levels

Cluster 0 & 7

- Key difference: **very new members.**
- Already have the interest / desire in our offers, we want to develop and nurture them into the mid-profitability range with time.

Cluster 2

- Key difference: **not receiving as much offers**
- Send more offers to engage better

Interested, frequent but no spending power yet

Cluster 3

- Younger and lowest income.
- Assuming the **natural trend:** incomes increases as they gain work experience, this cluster would potentially develop into the high-profitability section.

Conclusion & Recommendations

Key Findings:

- Using offers to quickly generate additional revenue when needed

Non-offer transactions = ~$12 spent per transaction

Offer-related transactions = ~$20 spent per transaction

- Improve marketing focus

- Most and least profitable customers segments

- 4 pockets of potential growth

Limitations

- Missing demographics in 12.8% of our customers.

- Identifying misattributions

- Dataset across 30-day test period only

- No information on how the data was collected

Next Steps

- **Explore other clustering algorithm**

- **Identify the influence of channels**

- **Explore influence of informational offers**

- **Explore the impact of offer attributes**

- **Building prediction models:** predicting customer response to offers, customer lifetime value prediction, churn prediction, next purchase day

Thank you.

Qi Wen Ng
https://www.linkedin.com/in/qi-wen-ng/