Qi-Wen Ng | DSI 14

# Customer Segmentation Analysis with STARBUCKS®

# Outline

# 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

The ‘average customer’ is a **concept of the past.**

Hitting the average does not mean hitting the majority, and this one-size-fits-all strategy **does not work** anymore.

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

## Customer Segmentation:

*process of dividing customers into groups based on common characteristics*

**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

Initial 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**

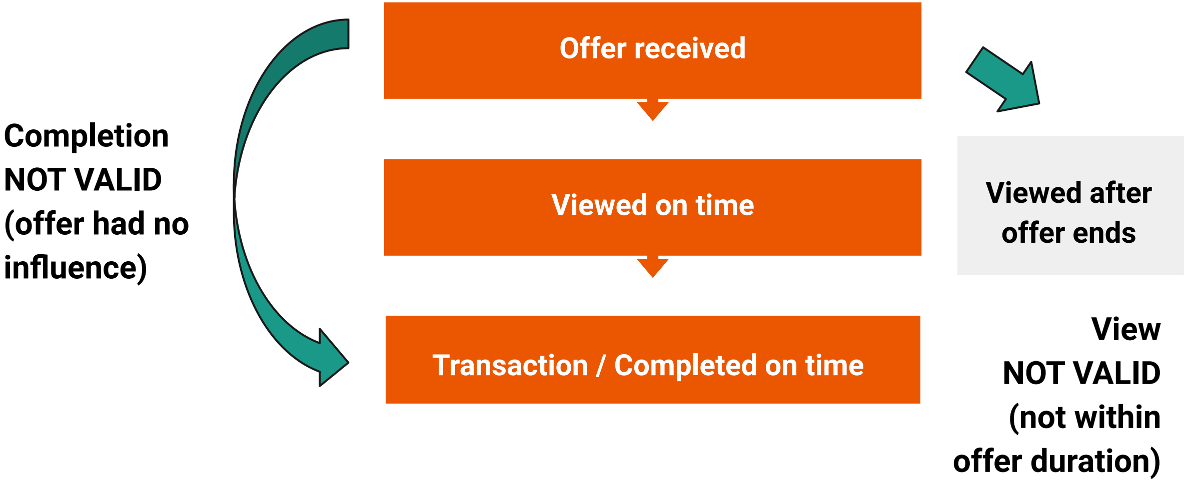
# Dataset

## Initial Dataset

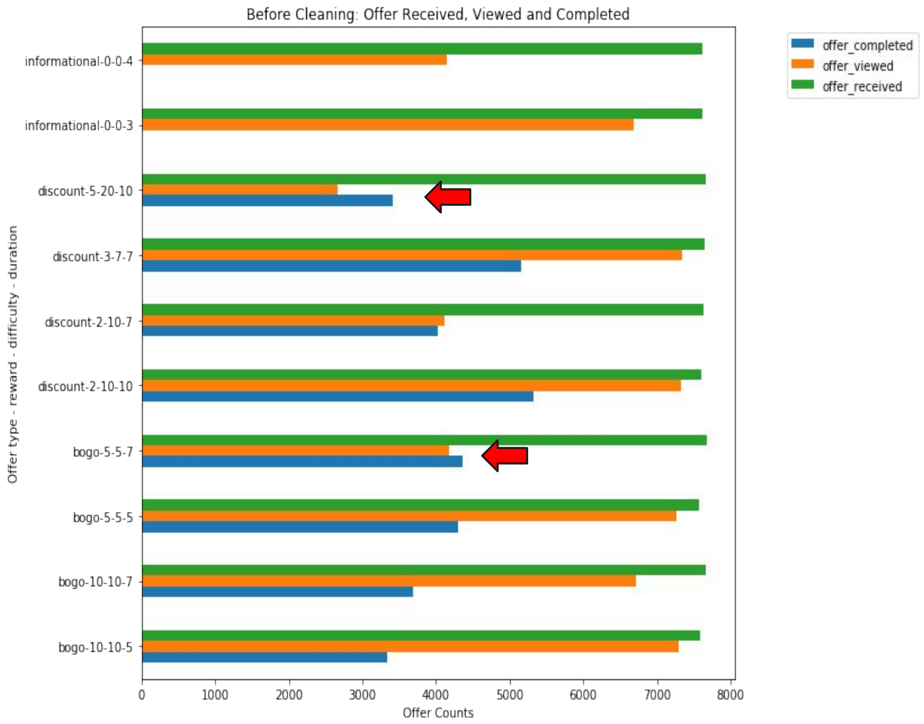
**promotion.json:**
10 promotional offers metadata: **BOGO, discount, informational**

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

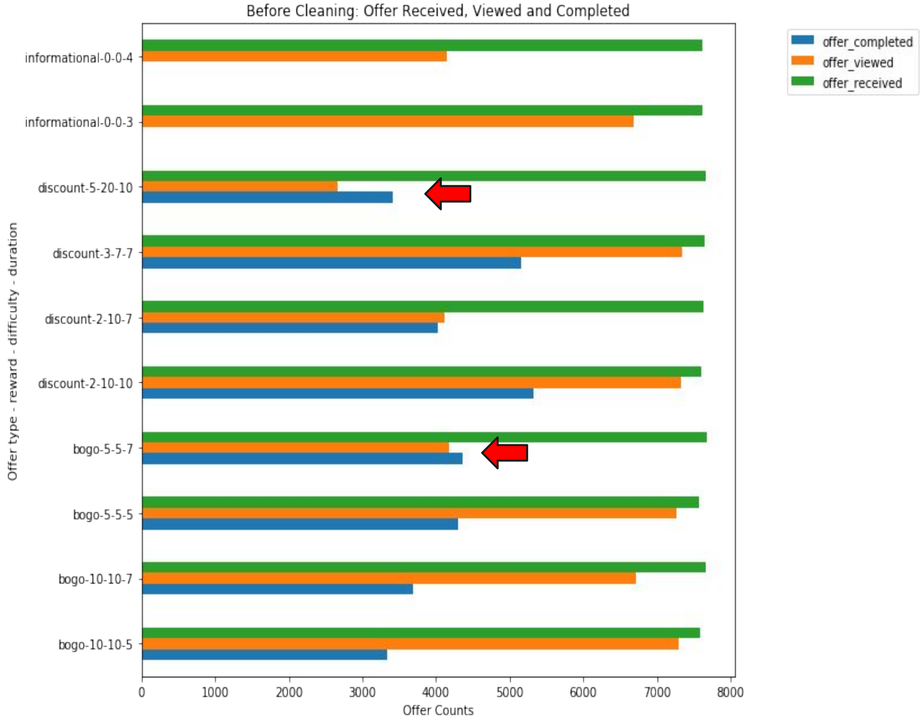
**event-log.json:**
306,534 customer timed activity logs: **transactions, offers received, viewed, completed**

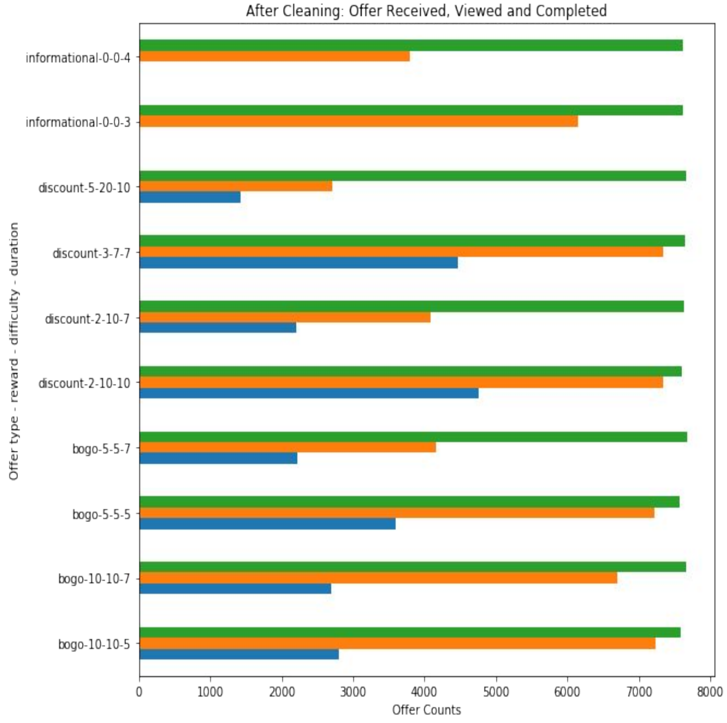


# Data Cleaning

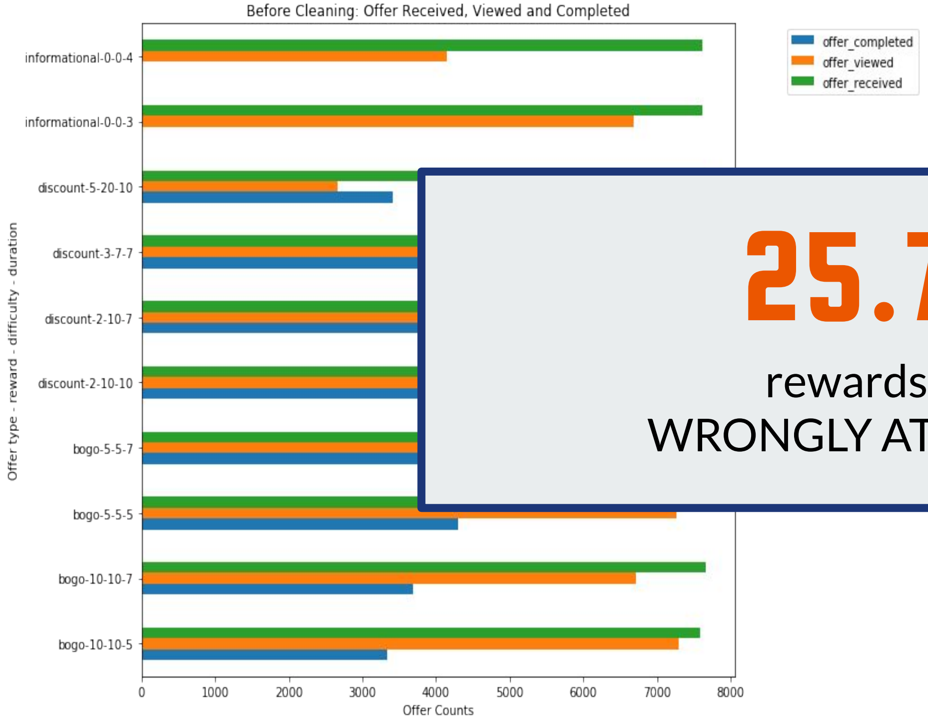


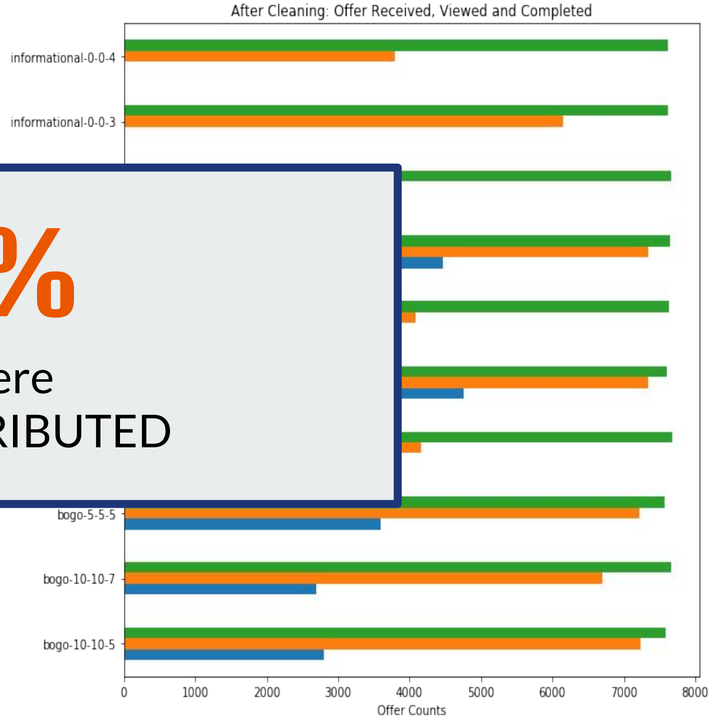
# Data Cleaning



# Data Cleaning




**25.7%**

rewards were
WRONGLY ATTRIBUTED

# Dataset

Initial 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**

- Valid views (viewed within offer duration)
- Valid completions (completed after viewing)
- 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

Initial 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**

- Valid views (viewed within offer duration)
- Valid completions (completed after viewing)
- Identifying repeat exposures to same offers by assigning unique offer ids

**Features engineered:**

- Days as member
- Viewed rates, convertion 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**

# K-Means Clustering

# 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

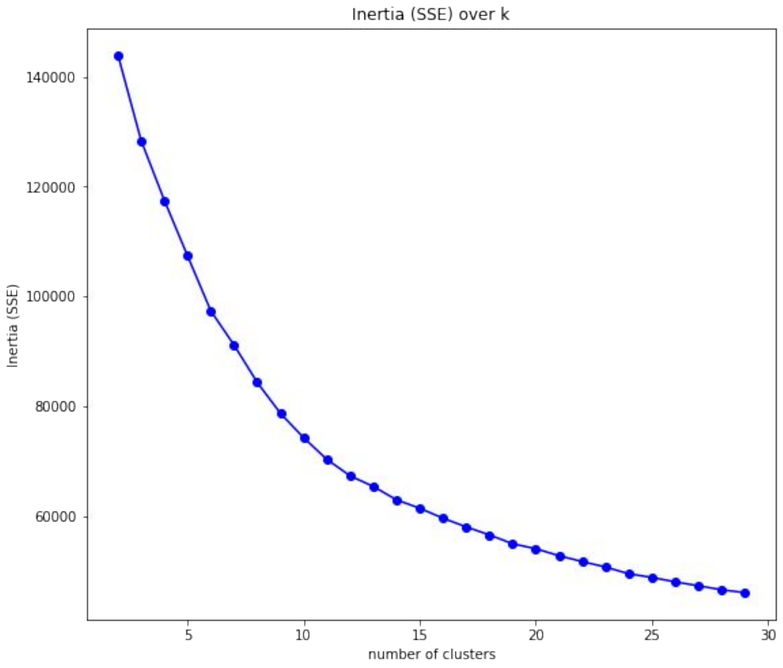
**Monetary / amount spent:**
offer_spent, actual_spent

**Recency / last visit:**
recency

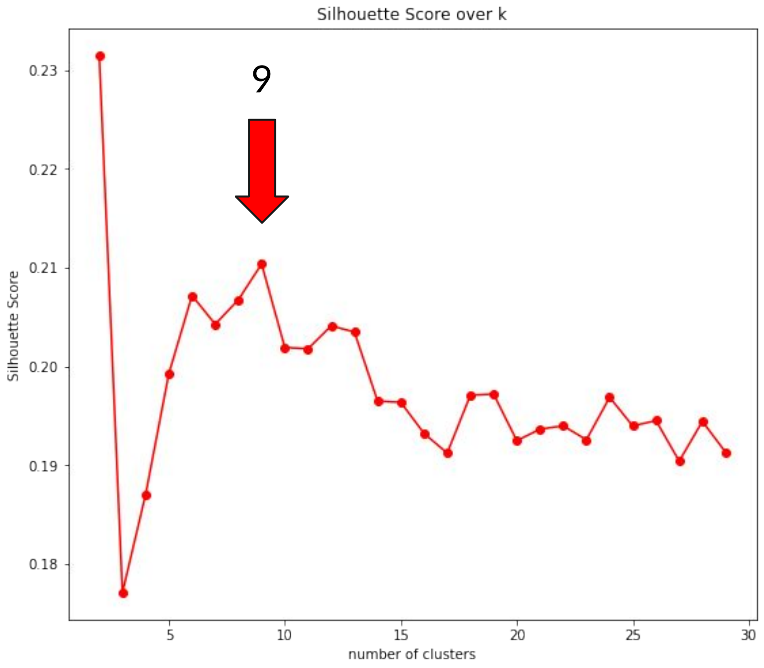
**StandardScaler**

# Clustering: Searching for optimal number of clusters

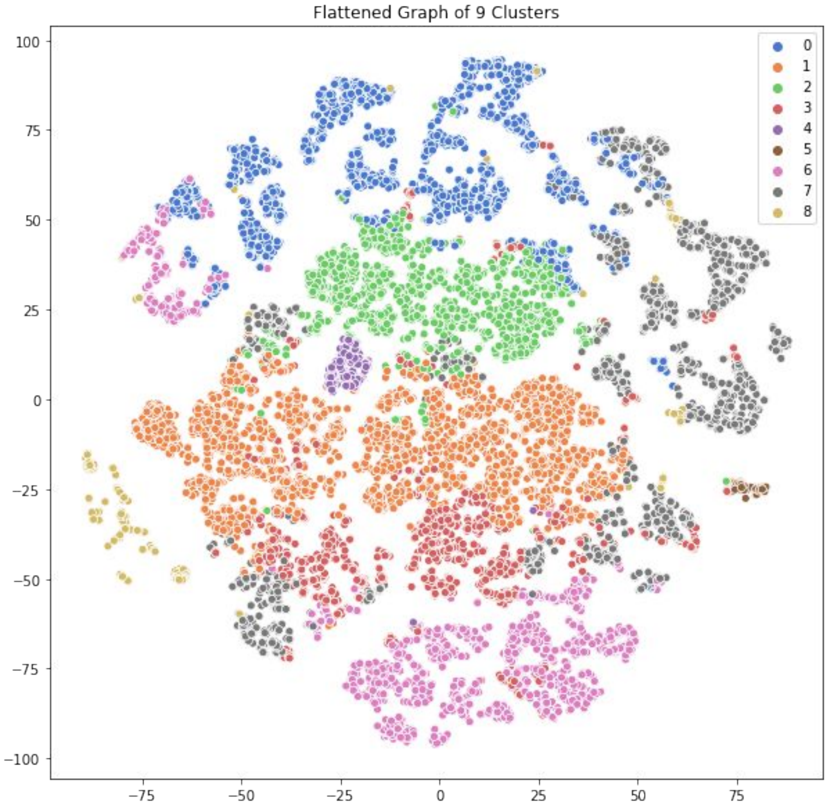
### Inertia (Intracluster distance)




### Silhouette (Intercluster distance)



# Clustering: Running K-Means & Visualizing Clusters



# Cluster Analysis: What offers to send to our customers?

View rates, conversion rates, offer-related transactions and spent, rewards

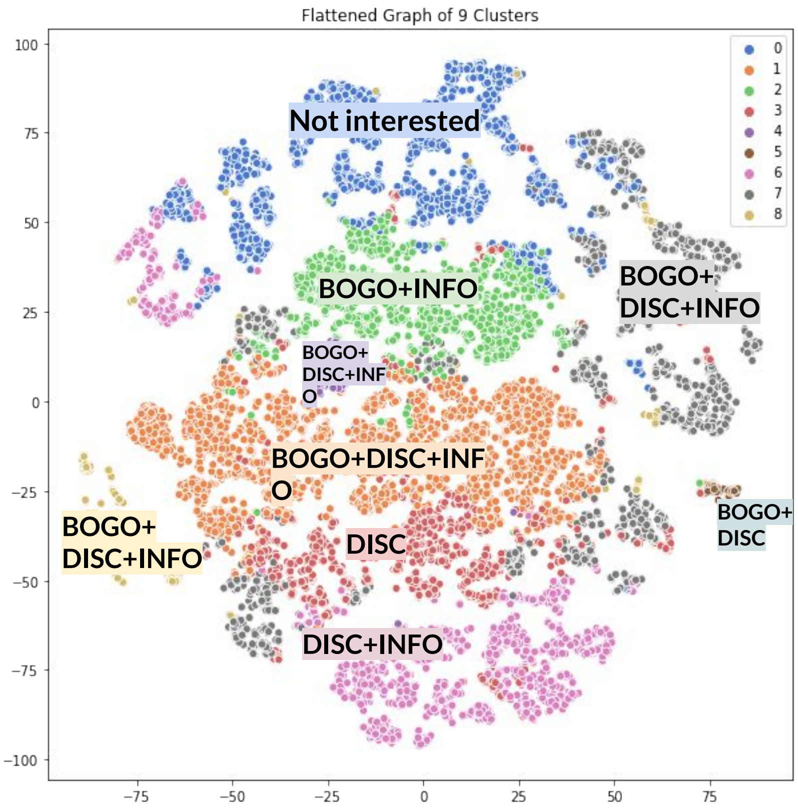
**BOGO & DISCOUNT**: Cluster 1, 4, 5, 7, 8

**BOGO only**: Cluster 2

**DISCOUNT only**: Cluster 3, 6

**INFORMATIONAL**: All clusters except 3 and 5

**NOT INTERESTED**: Cluster 0



# Cluster Analysis: Who are our most profitable customers?

Total spent

**Cluster 1, 4, 5**

- Common: **older age, high income, long-time members.**
- Other than the usual offers, it is important to make them **feel exclusive** and special such as tiered benefits.

**Cluster 2, 3, 6**

- Common: **long-time** members
- We want to continue **engaging** them and making sure of a positive experience.

**Cluster 0, 7, 8**

- Common: **newer** members.
- Reengage and drive subsequent purchases in these segments. **Retention** is important.

# Cluster Analysis: Which customers to focus for potential growth?

**Similar attributes to clusters in higher profitability levels**

**Cluster 0 & 7 (in low-profitability)**

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

**Cluster 2 (in mid-profitability)**

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

**Interested, frequent but no spending power yet**

**Cluster 3 (in mid-profitability)**

- 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**:
117k non-offer transactions = avg $11.46 spent per transaction
22k offer-related transactions =  avg $19.73 spent per transaction

- **Improve marketing focus**: Push offers only to segments that resonate better with specific offers, optimizing our costs.

- **Most and least profitable customers:** who to focus for retention, for maintaining the relationship / keeping them happy while increasing their purchases.

- **Pockets of potential growth**: Shifting customers from 4 segments into more profitable segments.

# Limitations

**- Missing demographics in 12.8% of our customers.**

**- Identifying misattributions:** confirming offer completions using separate redemption mechanisms (rather than auto) in the app or implement a tracking system.

**- Dataset across 30-day test period only:** unable to identify the spending patterns prior to this.

**- No information on how the data was collected**

# Next Steps

- **Explore other clustering algorithm:** hierarchical clustering

- **Identify the influence of channels:** tracking source of each activity

- **Explore influence of informational offers**

- **Explore the impact of offer attributes** (duration, rewards, difficulty)

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

# Thank you.