Predictive Analytics

Group 8

Megha Syam Gutti
Ray Wong
Shambhavi Singh
Yueting Gao
Tolulupe Oduwole

Introduction:

Variables that explain the behavior of customers like their choice of food, discounts availed, and loyalty are used for the cluster analysis. Strategies that focus on improving the revenue, and customer insights are discussed in detail in this report.

Food category variables:

- fd_cat_bev
- fd_cat_brunc
- fd_cat_buffe
- fd_cat_combo
- fd_cat_dess
- fd_cat_drink
- fd_cat_kids
- fd_cat_other
- fd_cat_side

Discount category variables:

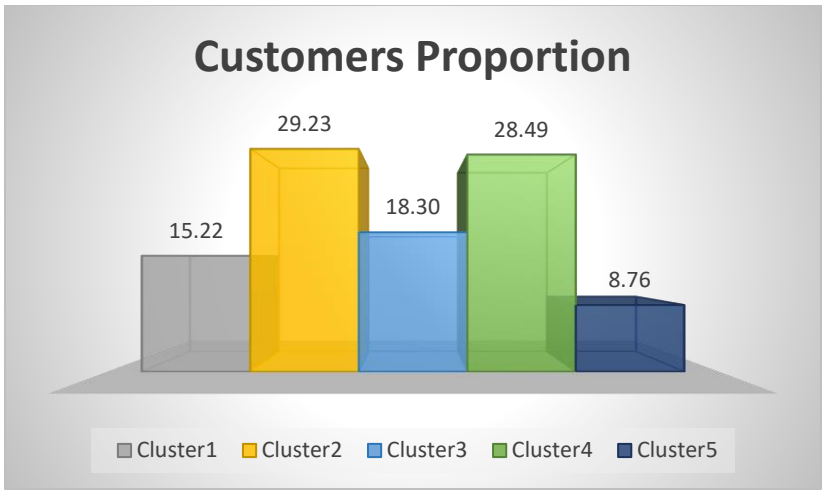
- disc_beverage
- disc_dessert
- disc_food
- disc_type_comp
- disc_chan_advo
- disc_chan_demo
- disc_chan_empl
- disc_chan_gmms
- disc_chan_gps
- disc_chan_local
- disc_chan_other
- disc_chan_value
- disc_pct_tot
- disc_pct_trans

Loyalty variables:

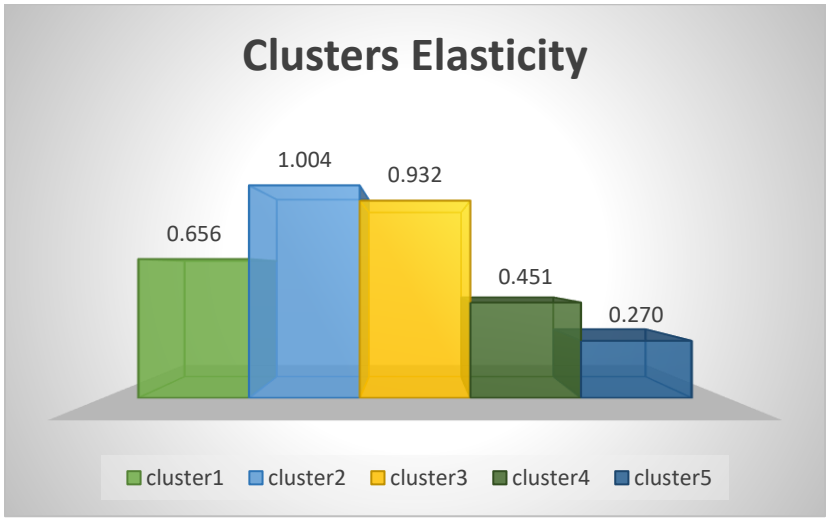
- tenure_day

This report includes,

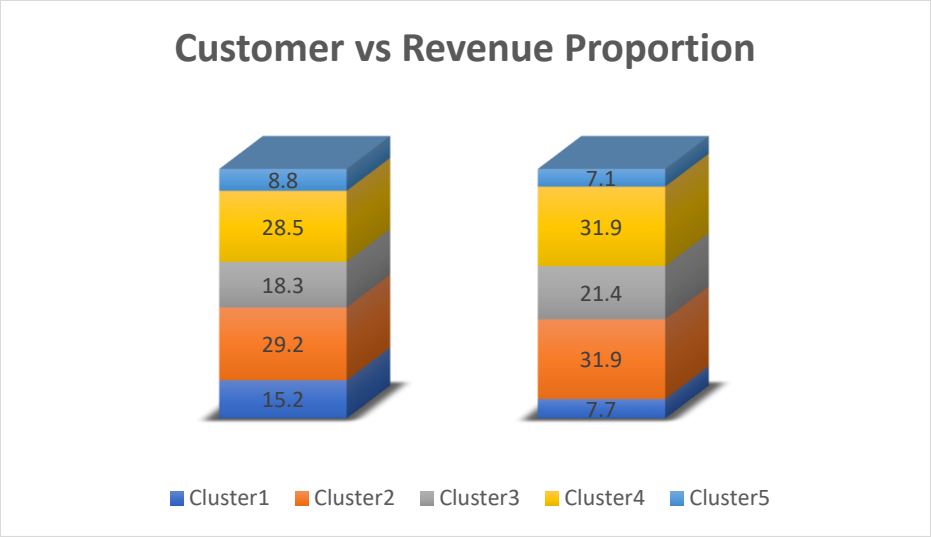
- Data analysis with data visualization.
- Cluster Analysis.
- Regression analysis to find price elasticity and the effect of email sent on the revenue.
- ANOVA to test a hypothesis.



- Cluster 2 has highest proportion of customers (3505), followed by Cluster 4 (3416), Cluster3 (2194), Cluster1 (1825) and Cluster 5 (1050).

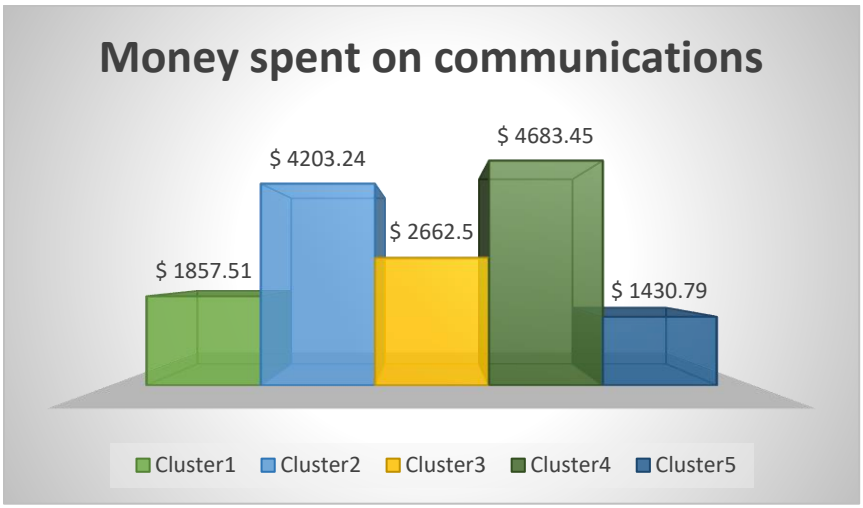


- Cluster 2 is price elastic.
- Clusters 1, 3, 4, 5 are not price elastic.



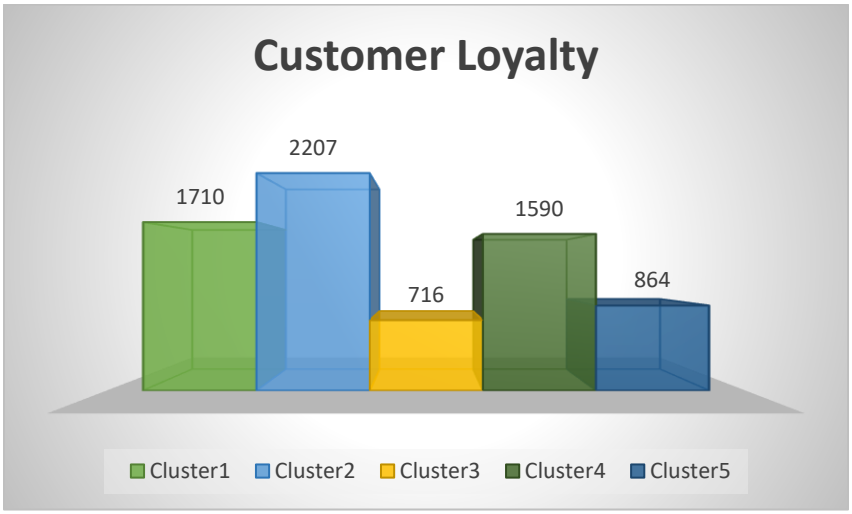
- Cluster 2 generates highest revenue with cluster 4 being very close to it, they each generate 40% of total revenue, around $620,000 each.
- Cluster 3 generates 21.4 % of total revenue, around $417,000.
- Cluster 1 and 5 generate 7.7 and 7.1 % of total revenue, around $150,000 and $140,000 each.

Relatively, Cluster 4 generates the highest and Cluster 1 generates the least revenue.

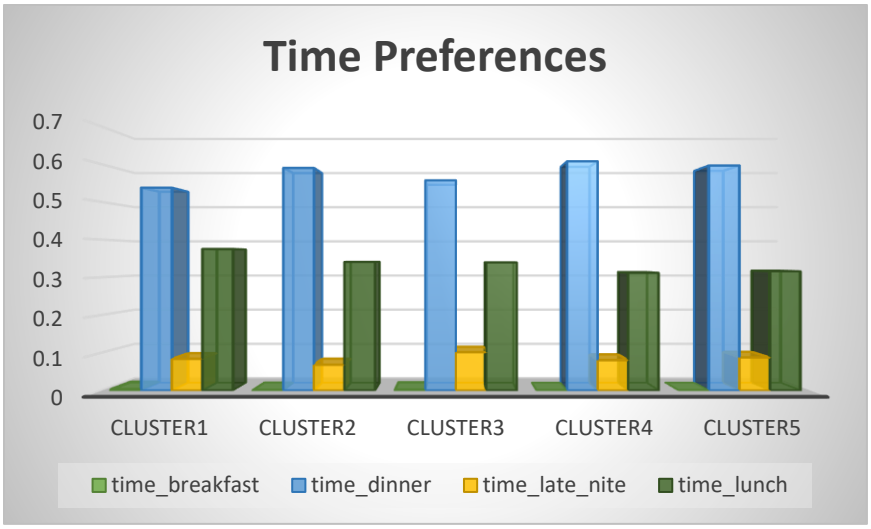


- Cluster 4 spent the most on email communications, it spent almost 0.8% of the total revenue it generated.
- Cluster 2 spent 0.7% of the total revenue it generated.
- Cluster 3 spent 0.6% of the total revenue it generated.
- Cluster 1 spent 1.2 % of the total revenue it generated.
- Cluster 5 spent 1 % of the total revenue it generated.

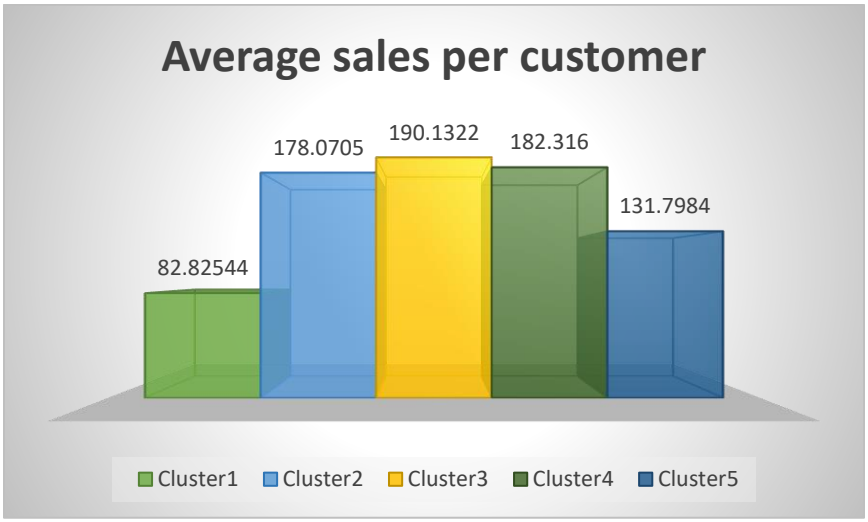
Cluster 1 used the most proportion of its revenue on email communications than the other clusters while Cluster 4 used the highest money.



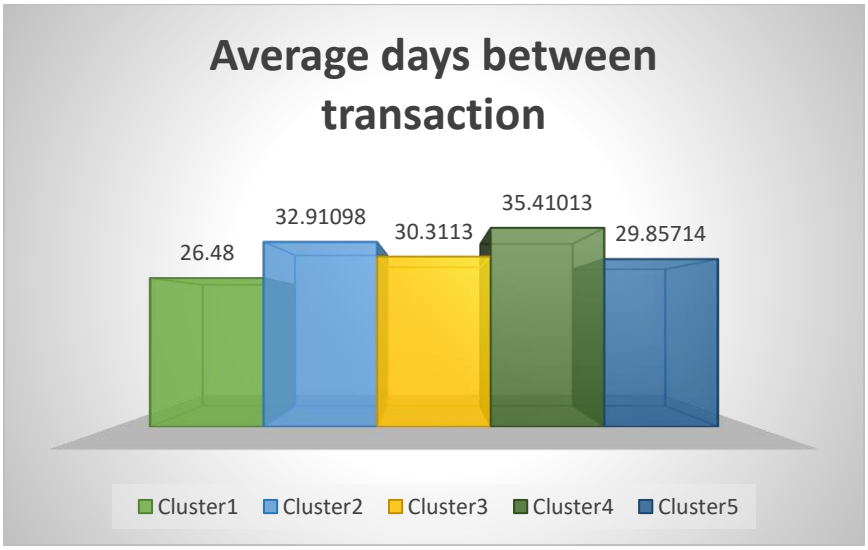
- Cluster 2 has the most loyal customers, they average number of days a customer is in the TGIF database is 2207 days.
- Cluster 1 has the 2nd most loyal customers, they average number of days a customer is in the TGIF database is 1710 days.
- Cluster 4 has the 3rd most loyal customers, they average number of days a customer is in the TGIF database is 1590 days.
- Cluster 5 has the 4th most loyal customers, they average number of days a customer is in the TGIF database is 864 days.
- Cluster 3 has the 5th most loyal customers, they average number of days a customer is in the TGIF database is 716 days.



- All the clusters have similar trend over the timing preferences, more than half of the customers in each cluster prefer going to TGIF for dinner, 30% of the customers prefer going to TGIF for lunch and 10% prefer late night.



- Cluster 3 customers spend more money compared to other clusters, followed by cluster 4, 2, 5, 1.



- On average, cluster 4 customers visit TGIF more frequently, followed by cluster 2, 3, 5, 1.

Cluster 1

Features:

Cluster 1 has 1825 customers, 15.2% of the total customers. It generates about 7.1 % of the total revenue, $140,000, each customer spends an average of $82.8. The average number of days a cluster 1 customer has been stayed in the TGIF database is 1701 days, making them the second most loyal customers. Around 55% of cluster 1 customers prefer going to TGIF for dinner, 38% of customers prefer going for lunch, and the rest for breakfast. 18% of the customers love the food category heavy entrée, 17% of the customers love appetizer and 14% love alcohol. These three food categories contribute to almost 50% of the total revenue of cluster 1.

Pricing Strategy:

Customers in cluster 1 are price inelastic, increasing the prices would increase the revenue. A proper pricing strategy aiming at the most liked food categories of cluster 1 should be incorporated. Trying the combinations of these preferred food categories using market basket analysis would benefit the company.

Marketing communications strategy:

$1857 was spent on email communications for cluster 1, which is about 1.2% of total revenue generated by cluster1, making it the top spender on email communications. An average of 34 emails are sent to each customer in cluster 1, spending almost $1.017 per customer. From the regression model total revenue = email sent; it is suggested to decrease the frequency of emails. Other modes of communications like SMS, DM, Internet, coupons, etc. should be explored. Marketing communications with better marketing strategy targeting the customer base behavior such as sending promotions on heavy entrée, appetizers, and alcohol or sending promotions on dinner followed by lunch should be incorporated.

Cluster2:

Features:

Cluster 2, the biggest cluster has 3505 customers, 29.2% of total customers, it generates close to 32% of total revenue, $620.000. On average each customer spends $178. Cluster 2 also has the most loyal customer base, the average number of days a customer in cluster 2 has been stayed in TGIF is 2207 days, this is the cluster of interest. Around 60% of cluster 2 customers prefer going to TGIF for dinner, 35% of customers prefer going for lunch, and the rest for breakfast. 20% of the customers love the food category heavy entrée, 17.5% of the customers love appetizer and 13.6% love alcohol. These three food categories contribute to more than 50% of the total revenue of cluster 2.

Pricing Strategy:

Customers in cluster 2 are price elastic, decreasing the price would increase the revenue. Giving special deals on the most preferred food categories such as heavy entrée, appetizers and alcohol or doing a market basket analysis and giving combo discounts on the most preferred combinations is suggested. Around 18% of the customers in cluster 2 have availed the discount_ticket category, more of these discount types should be introduced.

Marketing communications strategy:

$4203 were spent on email communications for cluster 2, about 0.7% of the total revenue generated by it. An average of 40 emails are sent to each customer, spending $1.2 on each customer. From the regression model total revenue = email sent; Like cluster 1, it is suggested to decrease the frequency of emails. Other modes of communications like SMS, DM, Internet, coupons, etc. should be explored. Marketing communications with better marketing strategy targeting the customer base behavior such as sending promotions on heavy entrée, appetizers, and alcohol or sending promotions on dinner followed by lunch should be incorporated.

Cluster3:

Features:

Cluster 3 has 2194 customers, 18.3% of the total customers. It generates about 21.4 % of the total revenue, $417,000, on average each customer spends $190. The average number of days a cluster 3 customer has been stayed in the TGIF database is 716 days, making them the least loyal customers. Around 59% of cluster 3 customers prefer going to TGIF for dinner, 35% of customers prefer going for lunch, and the rest for breakfast. 20% of the customers love the food category heavy entrée, 17% of the customers love appetizer and 15% love alcohol. These three food categories contribute more than 50% of the total revenue of cluster 3.

Pricing Strategy:

Customers in cluster 3 are price inelastic, increasing the prices would increase the revenue. A proper pricing strategy aiming at the most liked food categories of cluster 3 should be incorporated. Trying the combinations of these preferred food categories using market basket analysis would benefit the company.

Marketing communications strategy:

$2662 was spent on email communications for cluster 3, which is about 0.6% of total revenue generated by cluster3, making it relatively the least spender on email communications. An average of 40 emails are sent to each customer in cluster 3, spending almost $1.21 per customer. From the regression model total revenue = email sent; it is suggested to decrease the frequency of emails. Other modes of communications like SMS, DM, Internet, coupons, etc. should be explored. Marketing communications with better marketing strategy targeting the customer base behavior such as sending promotions on heavy entrée, appetizers, and alcohol or sending promotions on dinner followed by lunch should be incorporated.

Cluster4:

Features:

Cluster 4 has 3416 customers in it, about 28.5% of the total customers. It generates close to 32 % of the total revenue, about $620,000, on average each customer spends $182. The average number of days a cluster 4 customer has been stayed in the TGIF database is 1590 days, making them the 3rd most loyal customers. Around 61% of cluster 4 customers prefer going to TGIF for dinner, 31% of customers prefer going for lunch, and the rest for breakfast. 21.7% of the customers love the food category heavy entrée, 16.2% of the customers love appetizer and 14.7% love alcohol. These three food categories contribute more than 50% of the total revenue of cluster 4.

Pricing Strategy:

Customers in cluster 4 are price inelastic, increasing the prices would increase the revenue. A proper pricing strategy aiming at the most liked food categories of cluster 4 should be incorporated. Trying the combinations of these preferred food categories using market basket analysis would benefit the company.

Marketing communications strategy:

$4683 was spent on email communications for cluster 4, which is about 0.8% of total revenue generated by cluster4, making it highest spender on email communications in terms of money (not considering % of revenue spent). An average of 46 emails are sent to each customer in cluster 4, spending almost $1.37 per customer. From the regression model, total revenue = email sent; it is suggested to decrease the frequency of emails. Other modes of communications like SMS, DM, Internet, coupons, etc. should be explored. Marketing communications with better marketing strategy targeting the customer base behavior such as sending promotions on heavy entrée, appetizers, and alcohol or sending promotions on dinner followed by lunch should be incorporated.

Cluster5:

Features:

Cluster 5, the smallest cluster, has 1050 customers, 8.76% of the total customers. It generates close to 7.1% of the total revenue, about $140,000, on average each customer spends $138. The average number of days a cluster 5 customer has been stayed in the TGIF database is 864 days, making them the 4th most loyal customers. Around 60% of cluster 5 customers prefer going to TGIF for dinner, 31% of customers prefer going for lunch, and the rest for breakfast. 23% of the customers love the food category heavy entrée, 20% of the customers love appetizer and 11% love alcohol. These three food categories contribute more than 50% of the total revenue of cluster 5.

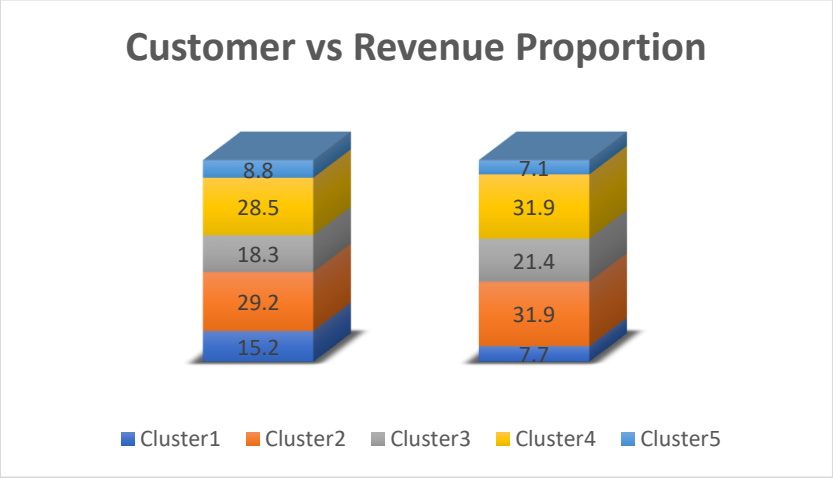
Pricing Strategy:

Customers in cluster 5 are price inelastic, increasing the prices would increase the revenue. A proper pricing strategy aiming at the most liked food categories of cluster 4 should be incorporated. Trying the combinations of these preferred food categories using market basket analysis would benefit the company.

Marketing communications strategy:

$1430 was spent on email communications for cluster 5, which is about 1.0% of total revenue generated by cluster5, making it 2nd highest spender on email communications (in terms of revenue generated). An average of 45 emails are sent to each customer in cluster 5, spending almost $1.36 per customer. From the regression model, total revenue = email sent; it is suggested to decrease the frequency of emails. Other modes of communications like SMS, DM, Internet, coupons, etc. should be explored. Marketing communications with better marketing strategy targeting the customer base behavior such as sending promotions on heavy entrée, appetizers, and alcohol or sending promotions on dinner followed by lunch should be incorporated.

Testing a hypothesis using ANOVA



From the graph above, can we conclude that cluster 2 and 4 are highest revenue generating clusters? Is there a difference in '29.2% proportion of customers generating 31.9% of revenue, 28.5% proportion of customers generating 31.9% of revenue, and 18.3 proportion of customers generating 21.4% of revenue'?

H_0: The average total revenue is same for all the clusters.

H_a: The average total revenue is not same for all the clusters.

The p-value of ANOVA in the model, net_sales_tot = cluster is significant. So, we reject the null hypothesis that the average total revenue is same for all the clusters and that there is at least one cluster with different average of Total revenue.

Means with the same letter are not significantly different.			
SNK Grouping	Mean	N	CLUSTER
A	190.132	2194	3
A			
A	182.316	3416	4
A			
A	178.071	3505	2
B	131.798	1050	5
C	82.825	1825	1

The table above shows that the clusters 3,4 and 5 are not significantly different but clusters 1 and 5 are different in terms of average total revenue.

- Clusters 2,3 and 4 are clusters of interest, they generate more than 85% of the revenue.
- Decrease the prices for cluster 2 customers.
- Increase the prices for the remaining customers.
- More than 50% customers in each cluster prefer having dinner, 30% having lunch. Conduct survey/feedback to know the customer expectations and if the customers are expecting new items in the menu and add them to the menu to increase the sales.
- Decrease the email frequency for all the clusters, send customized content communications targeting the food categories heavy entrée, appetizer and alcohol, and discounts availed.
- Explore other modes of communications such as Direct mail, coupons, SMS if they have not been tested already.
- Cluster 4 customers visit TGIF more frequently, implement more referral and reward programs.