**GROUP – 5**
**MKT/BUAN 6340.001**

# PREDICTIVE ANALYTICS ON TGI FRIDAY'S DATASET

## OVERVIEW

The dataset consists of sales data from a leading American restaurant chain focusing on Casual Dining. As a sales manager of TGIF we want to analyze the different ways in which we can improve the overall sales. For this we have focused on customer segmentation and identified the further problem statements to boost the sales in each defined customer segment.

## PROBLEM STATEMENT

1. Identifying most valuable customers based on RFM model.
2. Identify and analyze segments on the basis of behavioral variables like food category, time and rest location. Further, profile segments using decision trees.
3. Perform Price Elasticity to suggest strategies for Pricing and Discount.

## 1) RFM Model

RFM model analyzes the customers based on their Recency, Frequency and Monetary value. The variables explaining these parameters are:

**Recency** – days_between_transaction – Recent customers will have less values for this field while infrequent customers will have high values. We divided the values in this field into 3 different buckets to make this a categorical variable.
**Frequency** – item_tot – It represents the total number of items purchased by the customer.
**Monetary** – net_sales_tot – It is total of the sales per customer. We further classified it into 3 different net sales range.

**The SAS System**

**The CORR Procedure**

| 3 Variables: | Recency Frequency Monetary |
|---|---|

Simple Statistics

| Variable | N | Mean | Std Dev | Sum | Minimum | Maximum |
|---|---|---|---|---|---|---|
| Recency | 11991 | 1.99817 | 0.82213 | 23960 | 1.00000 | 3.00000 |
| Frequency | 11991 | 1.98390 | 0.82587 | 23789 | 1.00000 | 3.00000 |
| Monetary | 11991 | 1.99992 | 0.81638 | 23981 | 1.00000 | 3.00000 |

Pearson Correlation Coefficients, N = 11991
Prob > |r| under H0: Rho=0

| | Recency | Frequency | Monetary |
|---|---|---|---|
| Recency | 1.00000 | -0.42334<br><.0001 | -0.42250<br><.0001 |
| Frequency | -0.42334<br><.0001 | 1.00000 | 0.89944<br><.0001 |
| Monetary | -0.42250<br><.0001 | 0.89944<br><.0001 | 1.00000 |

Table 1 of Frequency by Monetary
Controlling for Recency=Highrecency

| Frequency | Monetary | | | |
|---|---|---|---|---|
| Frequency | Highmonetar | Lowmonetary | Mediummonet | Total |
| Highfrequenc | 260 | 1 | 15 | 276 |
| Lowfrequency | 0 | 2882 | 261 | 3143 |
| Mediumfreque | 20 | 129 | 473 | 622 |
| Total | 280 | 3012 | 749 | 4041 |

Table 3 of Frequency by Monetary
Controlling for Recency=Mediumrecen

| Frequency | Monetary | | | |
|---|---|---|---|---|
| Frequency | Highmonetar | Lowmonetary | Mediummonet | Total |
| Highfrequenc | 2313 | 0 | 140 | 2453 |
| Lowfrequency | 0 | 272 | 86 | 358 |
| Mediumfreque | 111 | 89 | 876 | 1076 |
| Total | 2424 | 361 | 1102 | 3887 |

Table 2 of Frequency by Monetary
Controlling for Recency=Lowrecency

| Frequency | Monetary | | | |
|---|---|---|---|---|
| Frequency | Highmonetar | Lowmonetary | Mediummonet | Total |
| Highfrequenc | 1072 | 2 | 191 | 1265 |
| Lowfrequency | 1 | 476 | 209 | 686 |
| Mediumfreque | 218 | 145 | 1749 | 2112 |
| Total | 1291 | 623 | 2149 | 4063 |

### Interpretation

1. The correlation matrix shows high correlation between Frequency and Monetary which is as expected.
2. Majority of the frequent customers tend to buy less items per transaction. This accounts to less sales per transaction hence no data comes up for high bucket of the net total sales column for these customers.
3. Contrary to this, the customers who visits once in a month tend to buy more items which contributes to higher total sales.
4. We have decent number of some old customers who have not visited the restaurant from a long time. This type of customers tends to buy average number of items contributing to mid-range of total sales.

### Recommendations

1. Attract frequent customers to buy high price items by promoting an offer to get a discount on high priced items after a purchase of certain number of regular items. This will increase the item count as well as boost the sales of high-priced items. *'Buy 2 Hamburgers to get 50% off on TGIF Special Burger'.*
2. Target the once in a month visitor by promoting discount on monthly purchase to increase their recency which will boost the sales of highly priced items. *'Get 20% off on your 3rd visit of the month'*
3. Mail discount coupons to the old customers who have not visited since a long time. This will impart an impression of personal attention and preference.
4. Create RFM score to get most valued customer. As Frequency and Monetary are correlated they will weigh 25% each while Recency will contribute 50% for this score.

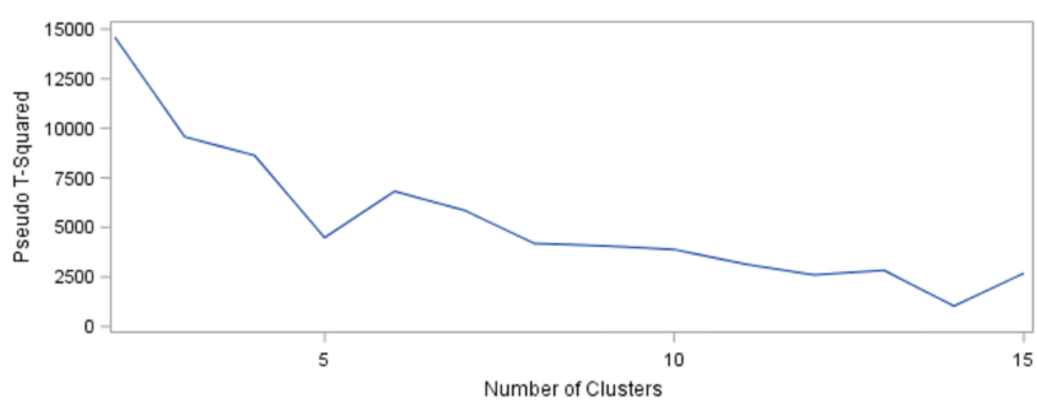
## 2) Behavioral Segmentation

We carried out segmentation on three main behavioral attributes of customers like food category, time of their visit and restaurant location preferred by them.

**fd_cat_X** – These are the types of food that customers ordered on their visit. It can be alcohol, appetizer, soup-salad, burger and so on.
**time_X** – This represents the time of day customer visited the restaurant like during breakfast hours, lunch, diner or late night.
**rest_loc_X** – This is the area of the restaurant preferred by customer for their dine-in like Patio, Bar, Cafe or if the opt for take-out option.

### Elbow Chart

We analyzed the elbow chart to get the optimal number of clusters for segmentation.



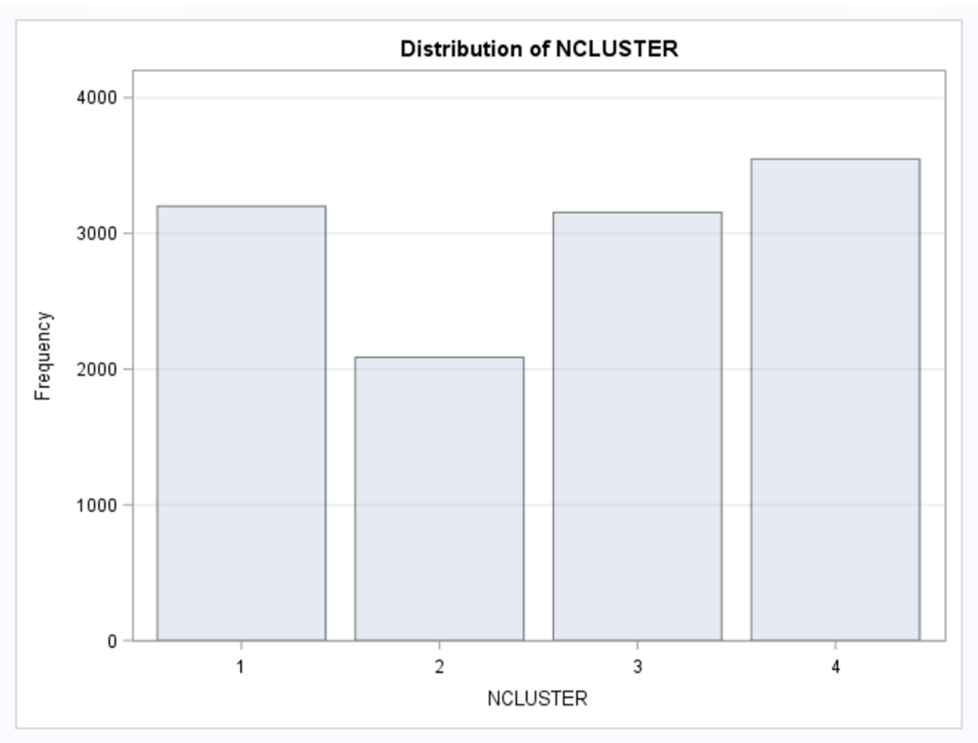
The elbow chart suggests 5 number of clusters to be optimal but after clustering it was observed that some clusters had more than 35% of data in it so we iteratively increase number of clusters. We got good data distribution for 7 clusters. But now some clusters had less than 15% of data in it which were merged with the nearest cluster to get the final number of clusters as 4.

| Criterion Based on Final Seeds = | 0.1073 |
|---|---|

| Cluster Summary | | | | | | |
|---|---|---|---|---|---|---|
| Cluster | Frequency | RMS Std Deviation | Maximum Distance from Seed to Observation | Radius Exceeded | Nearest Cluster | Distance Between Cluster Centroids |
| 1 | 3201 | 0.1115 | 1.0660 | | 6 | 0.9349 |
| 2 | 785 | 0.1267 | 1.6513 | | 5 | 1.0412 |
| 3 | 1274 | 0.0884 | 1.8176 | | 7 | 0.3249 |
| 4 | 438 | 0.1160 | 1.1857 | | 5 | 0.8921 |
| 5 | 1932 | 0.1165 | 1.5120 | | 4 | 0.8921 |
| 6 | 813 | 0.1204 | 1.6118 | | 3 | 0.8764 |
| 7 | 3548 | 0.0946 | 1.2642 | | 3 | 0.3249 |

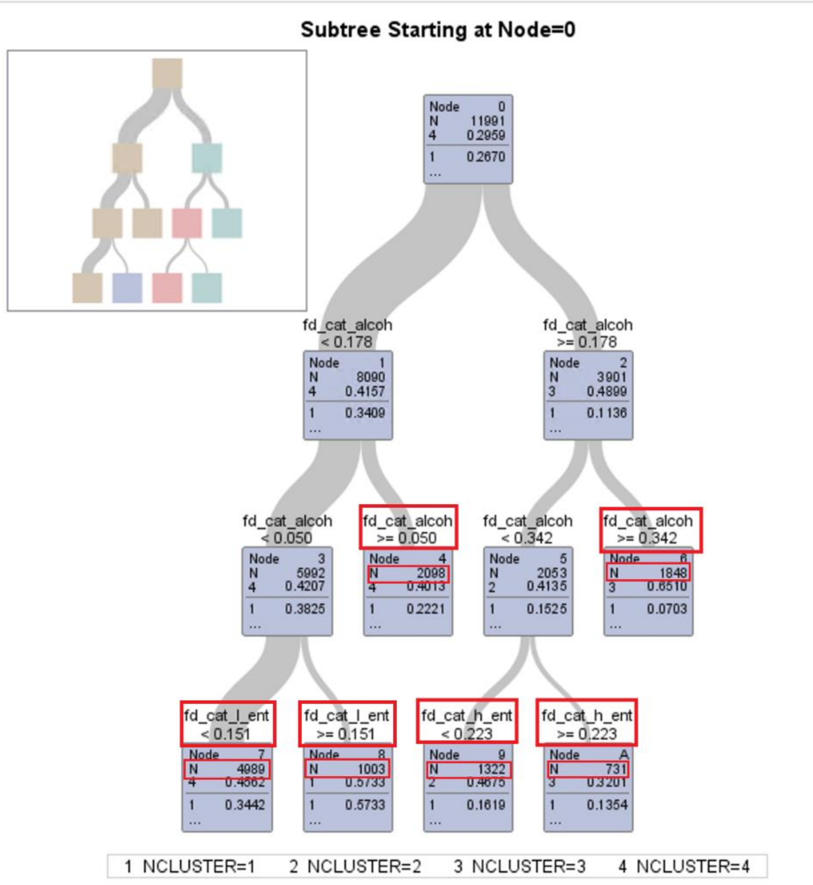
## Merging Cluster

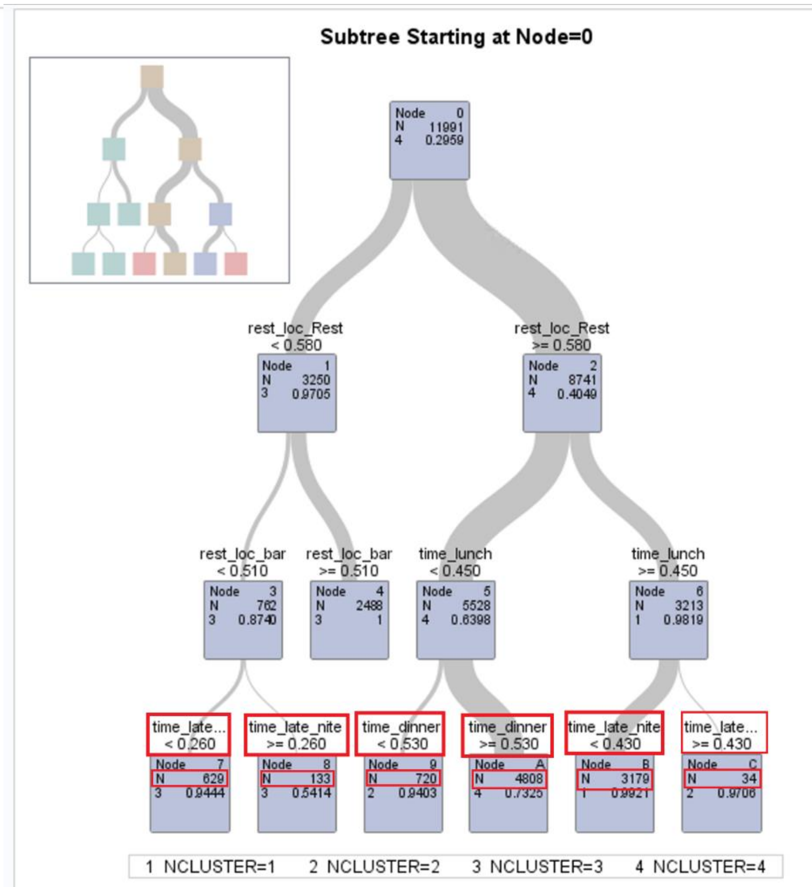
Cluster 1 has 26.7% of total customers
Cluster 2 has Cluster 3 and 6 merged into it which accounts to 17.4% of total customers.
Cluster 3 has Cluster 2, 4 and 5 merged into it which accounts to 26.31% of total customers.
Cluster 4 has 29.59% of total customers



## Profiling customer segment using Decision tree

Our early approach was to classify clusters using decision tree based on behavioral attributes mentioned above. We found that restaurant location and time dominates the decision splits in the tree, so we decided to build decision tree on restaurant location and time together while a separate decision tree for food category.




From above decision trees we profiled 4 segments as below:

### Cluster 1: Lunch-er

*Main Attributes:* Lunch time, Location rest, Light entree
These customers go to TGIF for lunch and they pick the Rest place. They prefer light entree and some proportion of alcohol.

### Cluster 2: Light Dinner with Alcohol

*Main Attributes:* Dinner, late night, Alcohol, Heavy entree
These customers go to TGIF at late night. However, they choose to have light dinner. They prefer heavy entrée with moderate proportion of alcohol.

### Cluster 3: Best in Bar

*Main Attributes:* Bar, Late night, Alcohol, Heavy entrée
These customers go to TGIF and prefer restaurant location as bar. They spend a lot of time at Bar with heavy entrees and high proportion of alcohol

## Cluster 4: Hungry peeps for Dinner

*Main Attributes:* Dinner time, Location rest, Alcohol, Light entree

These customers go to TGIF for dinner and prefer rest location to spend most of their time and order light entrée with less proportion of alcohol.

## Price Elasticity

Using Poisson regression, as quantity is count data, we modelled Quantity on Price to check price sensitivity of customers for each cluster. We used elasticity to recommend changes in discount and pricing strategies..

### Cluster 1

Analysis Of Maximum Likelihood Parameter Estimates

| Parameter | DF | Estimate | Standard Error | Wald 95% Confidence Limits | | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|---|---|
| Intercept | 1 | 1.3232 | 0.0787 | 1.1690 | 1.4774 | 283.02 | <.0001 |
| Price | 1 | -0.2836 | 0.0424 | -0.3667 | -0.2004 | 44.69 | <.0001 |
| days_between_trans | 1 | 0.0027 | 0.0003 | 0.0021 | 0.0034 | 69.04 | <.0001 |
| email_click_rate | 1 | 0.5414 | 0.2933 | -0.0334 | 1.1163 | 3.41 | 0.0649 |
| email_open_rate | 1 | 0.1082 | 0.0443 | 0.0213 | 0.1950 | 5.96 | 0.0146 |
| Scale | 0 | 1.0000 | 0.0000 | 1.0000 | 1.0000 | | |

### Cluster 2

Analysis Of Maximum Likelihood Parameter Estimates

| Parameter | DF | Estimate | Standard Error | Wald 95% Confidence Limits | | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|---|---|
| Intercept | 1 | 1.3669 | 0.1041 | 1.1630 | 1.5709 | 172.53 | <.0001 |
| Price | 1 | -0.2376 | 0.0544 | -0.3442 | -0.1310 | 19.08 | <.0001 |
| days_between_trans | 1 | 0.0024 | 0.0004 | 0.0016 | 0.0032 | 36.27 | <.0001 |
| email_click_rate | 1 | 0.6964 | 0.3372 | 0.0355 | 1.3573 | 4.26 | 0.0389 |
| email_open_rate | 1 | 0.0913 | 0.0553 | -0.0171 | 0.1997 | 2.73 | 0.0986 |
| Scale | 0 | 1.0000 | 0.0000 | 1.0000 | 1.0000 | | |

### Cluster 3

Analysis Of Maximum Likelihood Parameter Estimates

| Parameter | DF | Estimate | Standard Error | Wald 95% Confidence Limits | | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|---|---|
| Intercept | 1 | 1.4974 | 0.0647 | 1.3706 | 1.6242 | 535.54 | <.0001 |
| Price | 1 | -0.2800 | 0.0356 | -0.3498 | -0.2101 | 61.71 | <.0001 |
| days_between_trans | 1 | 0.0007 | 0.0003 | 0.0001 | 0.0014 | 4.57 | 0.0325 |
| email_click_rate | 1 | 0.5778 | 0.2257 | 0.1353 | 1.0202 | 6.55 | 0.0105 |
| email_open_rate | 1 | 0.1421 | 0.0430 | 0.0578 | 0.2265 | 10.91 | 0.0010 |
| Scale | 0 | 1.0000 | 0.0000 | 1.0000 | 1.0000 | | |

### Cluster 4

Analysis Of Maximum Likelihood Parameter Estimates

| Parameter | DF | Estimate | Standard Error | Wald 95% Confidence Limits | | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|---|---|
| Intercept | 1 | 1.0349 | 0.0571 | 0.9230 | 1.1468 | 328.51 | <.0001 |
| Price | 1 | -0.1205 | 0.0297 | -0.1787 | -0.0623 | 16.45 | <.0001 |
| days_between_trans | 1 | 0.0036 | 0.0003 | 0.0030 | 0.0042 | 143.45 | <.0001 |
| email_click_rate | 1 | 0.5021 | 0.2276 | 0.0560 | 0.9482 | 4.87 | 0.0274 |
| email_open_rate | 1 | 0.1090 | 0.0420 | 0.0266 | 0.1913 | 6.73 | 0.0095 |
| Scale | 0 | 1.0000 | 0.0000 | 1.0000 | 1.0000 | | |

Formula used for calculating elasticity: **Beta * (Average Price/Average Quantity)**

| Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|
| **-0.2011** | **-1.5692** | **-0.1702** | **-0.0843** |

## Interpretation:

For clusters with elasticity having absolute value greater than 1, quantity demanded is elastic so only 2nd cluster exhibits elasticity in demand.

## Recommendations:

1. For Clusters 1, 3 and 4, TGIF should increase Prices or offer less discounts as quantity demanded for Segment 1 is inelastic.
2. For Clusters 2, TGIF should decrease Prices or offer more discounts as quantity demanded for Segment 2 is elastic.

## Summary

| | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|
| % Revenue | 0.2084 | 0.1796 | 0.3565 | 0.2555 |
| % Customer | 0.267 | 0.174 | 0.2631 | 0.2959 |
| Number of Customers | 3210 | 2087 | 3155 | 3548 |
| Email Send | 40.62 | 42.32 | 40.77 | 41.63 |
| Days between transaction | 32.05 | 33.34 | 32.04 | 30.8 |
| Items total | 22 | 27.25 | 40.35 | 23.33 |
| Items total distinct | 11.43 | 14.81 | 15.97 | 12.15 |
| RFM Score | 1.94 | 2.02 | 2.04 | 1.99 |
| Time Dinner | 0.18 | 0.66 | 0.57 | 0.92 |
| Food Category Heavy Entrée | 0.2 | 0.17 | 0.16 | 0.27 |
| Email Open Rate | 0.23 | 0.22 | 0.23 | 0.23 |
| Food Category Appetizer | 0.15 | 0.19 | 0.2 | 0.16 |
| Food Category Steak | 0.08 | 0.1 | 0.07 | 0.13 |
| Food Category Burger | 0.12 | 0.08 | 0.08 | 0.1 |
| Food Category Beverage | 0.1 | 0.05 | 0.05 | 0.1 |
| Time Lunch | 0.81 | 0.09 | 0.31 | 0.07 |
| Food Category Other | 0.05 | 0.04 | 0.04 | 0.05 |
| Food Category Soup-Salad | 0.06 | 0.03 | 0.04 | 0.04 |
| Food Category Alcohol | 0.06 | 0.24 | 0.27 | 0.04 |
| Food Category Light Entree | 0.1 | 0.03 | 0.04 | 0.04 |
| Food Category Kids | 0.02 | 0.02 | 0.01 | 0.03 |
| Email click rate | 0.02 | 0.01 | 0.02 | 0.02 |
| Food Category Sides | 0.01 | 0.01 | 0.01 | 0.01 |
| Time Late Night | 0 | 0.15 | 0.05 | 0 |

## Conclusion

- Re-consider pricing and discount strategies for each segment according to price sensitivity.
- To target customers for email campaigns, sorting RFM score in descending order gets the most valuable customers for respective behavioral segments.


Report by-
**Group 5**

| NetID | Name |
|---|---|
| pxc173330 | Po-Chih Chen |
| ovc170000 | Oshin Chetule |
| fgw180000 | Fatima Wali |
| svy180000 | Sunny Yadav |