**Predictive Analytics with SAS**

**Marketing Insights for Wendy’s**



**GROUP 1**

**Yajaira Gonzalez - yxg140030**

**Vikalp Upadhyay - vxu180002**

**Ankit Raina - aar180001**

**Anindita Das - axd170034**

**Sweta Patra - sxp180081**

**Summary Highlights**

1. **Customer Behavior**
   1. We observe 5 key segments that drive revenue, however we should focus on the second to largest subsets that can increase total sales and drive higher revenue.
2. **When are customers likely to churn out?** 
   1. We see customer sets with larger families tend to churn out more, likely due to the unhealthiness of fast food.

**Overview**

* Wendy’s Dataset was our main source of data, including sales, transactions, frequency, HH income, HH size, etc by customer
* Clustering via Kmeans allowed us to segment the target base to be able to pull out key insights on customer behavior for detailed targeting and message tailoring
* Survival Analysis was done to further understand the likelihood that customers will churn or switch from Wendy’s.

**Recommendations for Wendys:**

**Target Promotions:**

* Drive a promotion for night owls that caters to their preference of sweet and fatty foods. Such as a buy one get one half off or adopt a crave box that is priced
* Example #2

**Targeted Segments:**

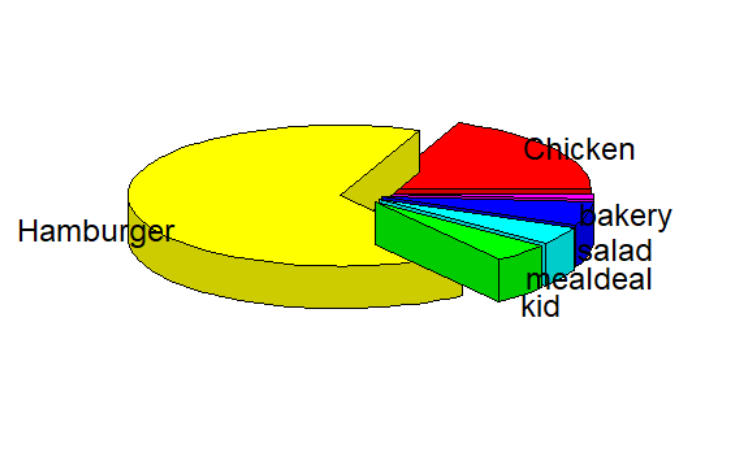
* Night Owls: Focus on this subset because these have a high revenue stream at night so a focus on night customers can increase sales dramatically.
* Lunch Rush: Focus on driving the lunch rush segment as there is opportunity for growth and have the largest transaction pattern.

**Retain Customers:**

* Strategy to retain customers: Wendy’s should focus on targeting smaller, individual families or single dwellings to increase and retain sales.
* Strategy to increase sales: Wendy’s has a relatively lower proportion of meal deal sales for dinner, therefore, it should give promotions on dinner meal deals as a prospective source of increased sales.

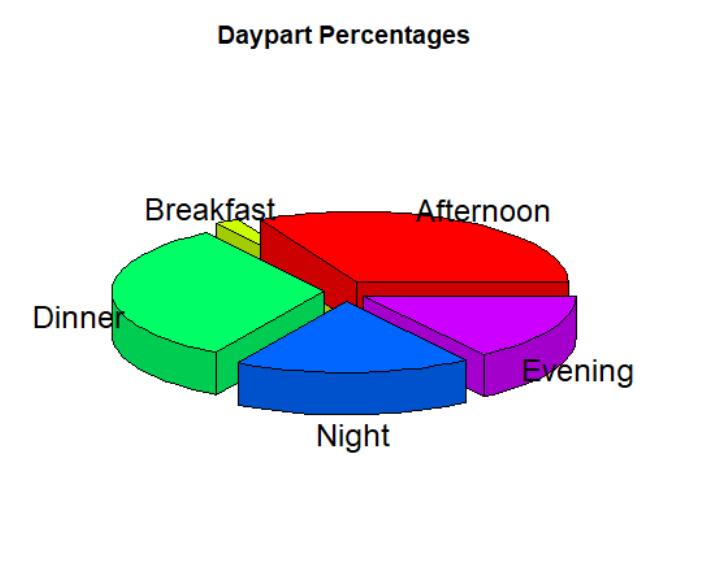
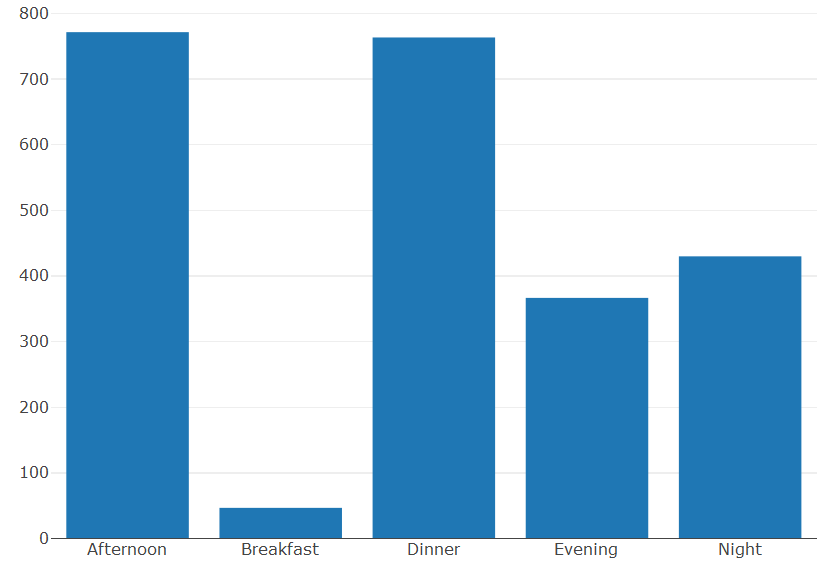
**Exploratory Data Analysis**

Top Products for Wendy’s (in terms of total units sold):

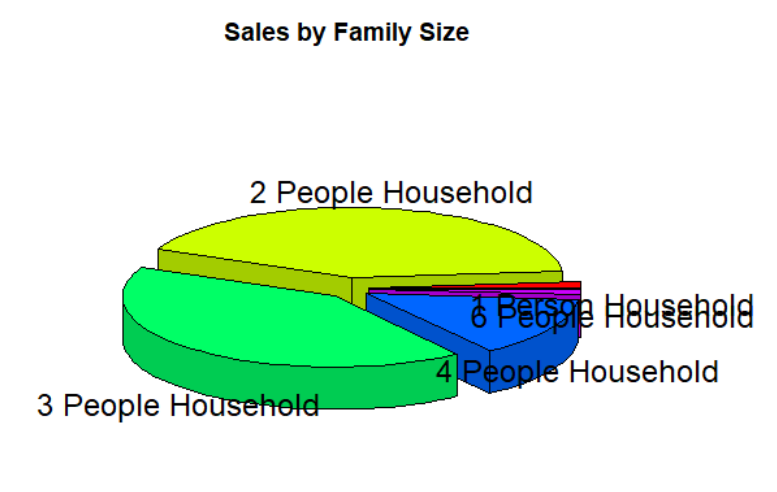


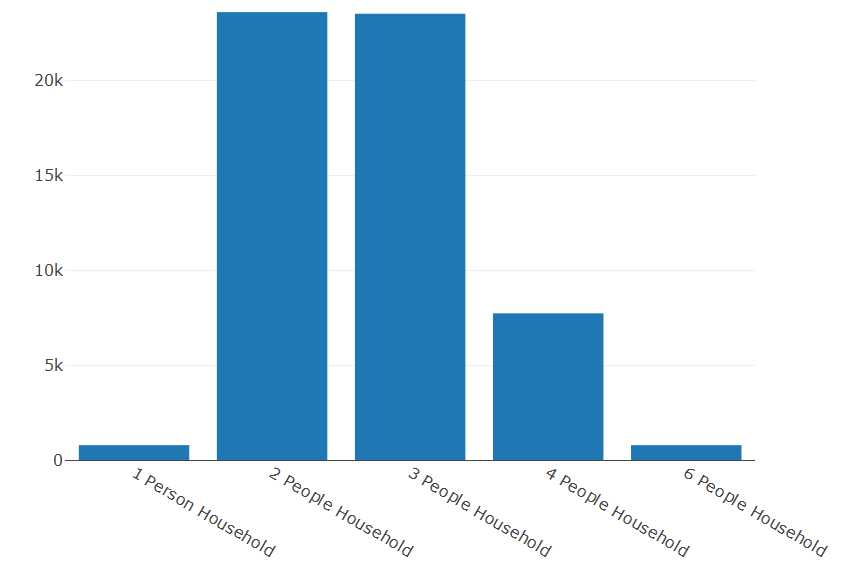
Hamburger and Chicken products have the highest amount sold. Kid product has the smallest amount sold.

Top Time Sets for Wendy’s (in terms of frequency and total amount sold):



Non-existent customer base for breakfast timeframe and smallest revenue stream.. Largest customer base for afternoon/lunch subset. Most valuable subset for the evening time set and onwards. Focusing on evening plus can drive 3x revenue stream than focus on earlier time sets.

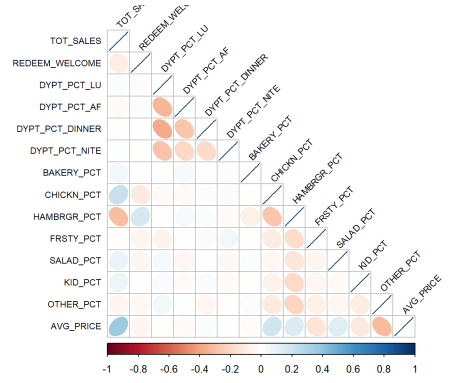
Top Household Sizes for Wendy’s (in terms of frequency and total amount sold):



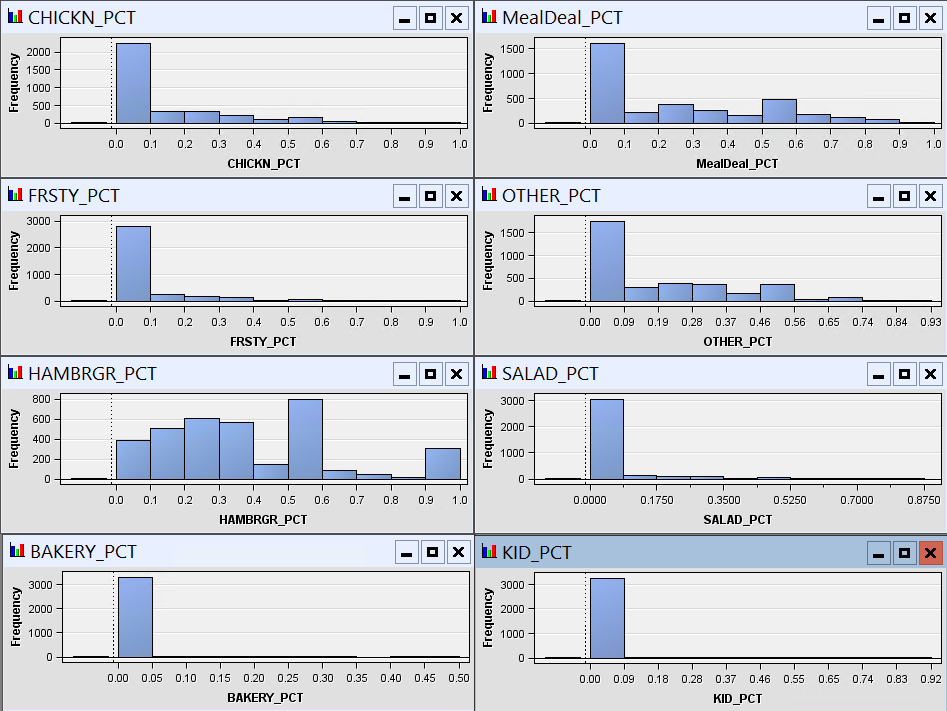
Smaller households make up a large % of the frequency seen in the dataset, Bigger households make up a smaller % of the dataset. When we look at revenue, we also see smaller subsets drive the majority of sales. A focus on small households would be a natural connection for a promotion. Or vice versa, a connect on bigger households would increase overall sales but they don’t seem like a good revenue stream.

**Analysing the Customer Behaviour of Wendy’s:**

To find what factors influence our customer base, we ran clustering against our dataset. Cluster analysis or clustering is an unsupervised method of grouping objects, such as customers in our case, in such a way that those in the same group (cluster) are more similar to each other than to those in other groups. By understanding the key customers who frequent Wendy’s then we can better profile and target them or expand small but valuable subsets.



*Variables for clustering:*

* AVG\_PRICE
* REDEEM\_WELCOME
* DYPT\_PCT\_LU
* DYPT\_PCT\_AF
* DYPT\_PCT\_DINNER
* DYPT\_PCT\_NITE
* CHICKN\_PCT
* HAMBRGR\_PCT
* FRSTY\_PCT
* SALAD\_PCT
* KID\_PCT
* OTHER\_PCT

We also looked at the dispersion of the variables to predict how the clusters would form. We see that hamburgers is dispersed throughout the dataset and the rest have key instances.

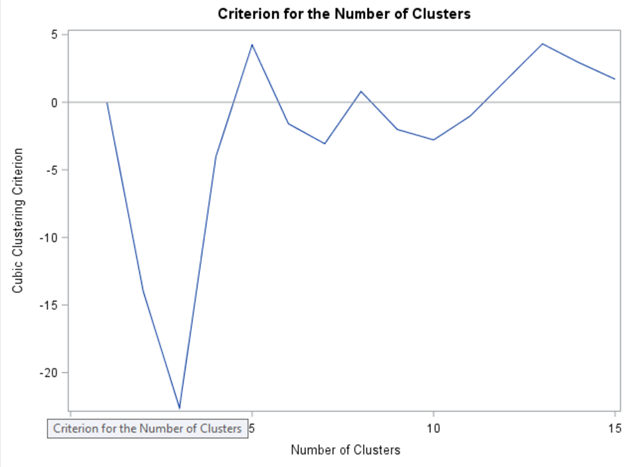
**Interpreting fit of the model**

1. Ran multiple clustering results of various variable combinations
2. Standardized Data
3. Focused on variables with smallest correlation

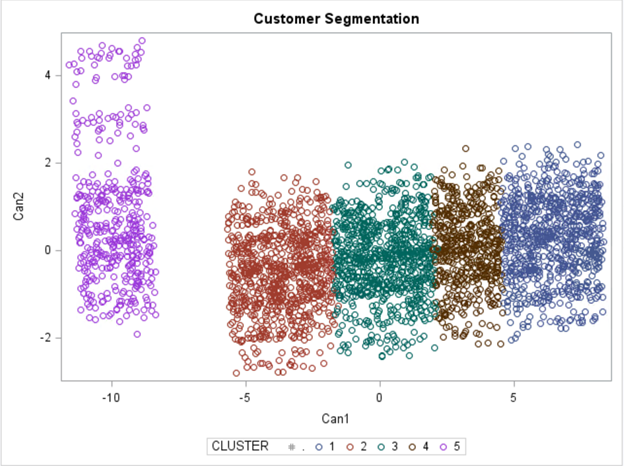
**Cluster Identification:**

We used Cubic Clustering Criterion (CCC) to ascertain the no. of clusters.

Based on the graph below, CCC is highest for the least cluster size = 5



We get the below 5 clusters on running the clustering algorithm using the above variables.



**Results:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Average Joe | Fast Food Junkie | Night Owls | Lunch Rush | Treat Ya Self |
| Size | 24% | 15.75% | 24% | 23.95% | 12.26%\*\* |
| Size\_HH | 1.93 | 1.93 | 1.92 | 2.11\*\*\* | 1.9 |
| HH\_Income | 68,259\*\* | 65,234 | 68258\*\* | 65,521 | 67,558 |
| Tenure | 57.86 | 81.24\*\*\* | 60.91 | 74.18\*\*\* | 60.5 |
| Num\_Txn | 1.779 | 7.897\*\* | 1.696 | 5.059\*\* | 1.655 |
| Tot\_sales | 7.811 | 46.69\*\* | 9.301 | 27.29\*\* | 7.32 |
| Tot\_Rewards | .172 | 1.828\*\* | .171 | 1.176\*\* | .178 |
| Lunch | .074 | .368\*\* | .031 | .434\*\* | .053 |
| Dinner | .022 | .16 | .031 | .208 | .645 |
| Evening | .032 | .286 | .037 | .208 | .221 |
| Night | .009 | .054 | .876 | .076 | .016 |
| Bakery | .462 | .307 | .5 | .235 | .428 |
| Chicken | 0 | .034 | 0 | .059 | 0 |
| Hamburger | .463 | .307 | .5 | .235 | .428 |
| Meal Deal | 1.234 | 7.345 | 1.646 | 5.471 | 1.5 |
| Frosty | 0 | 0 | 0 | .235 | 0 |
| Other\_PCT | .238 | .117 | .198 | .068 | .218 |
| Avg Price | 1.874 | 3.449 | 2.057 | 2.46 | 1.804 |
| Tot Units | 3.59 | 13.93 | 4.07 | 10.12 | 3.588 |
| Elasticity | .522 | .247 | .50 | .24 | .50 |

* Relatively wealthier people tend to eat less junk food.
* Bigger families purchase more frosties, which is intuitive as bigger family size suggests presence of kids.
* People who come during lunch tend to eat snacks instead of meals.

**Survival Analysis of Customers:**

**PHREG MODEL**

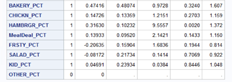
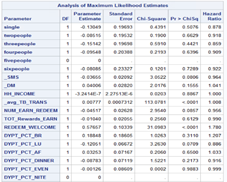
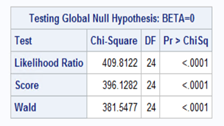
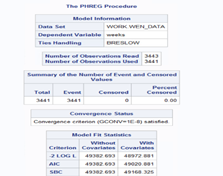
The phreg model performs the regression analysis of survival data based on the cox proportional hazards model. In this model, we are taking weeks as the dependent variable against several independent variables and our aim is to identify the time remaining for a customer to leave Wendy’s, so that we can take countermeasures to increase that time for as long as we can. A hazard function is always positive.

*Meaning of Coefficients:*

* \_SMS: 100(exp(beta) - 1) = -3.6%. The percent change in the hazard of a customer receiving sms, churning decreases by 3.6% compared to when they don’t receive sms
* \_avg\_TB\_TRANS: 100(exp(beta) - 1) = 0.8%. For every one unit increase in average time between transactions, the percentage change in the hazard of the customer churning increases by 0.8%
* REDEEM\_WELCOME: 100(exp(beta) - 1) = 78%. For a customer having redeem welcome gift card, the hazard of churning increases by 78%
* DYPT\_PCT\_LU: 100(exp(beta) - 1) = 11.4%. For a customer having daypart percent in lunch rather than night, the hazard of churning decreases by 11.4%
* HAMBURGER\_PCT: 100(exp(beta) - 1) = 37.2%. For a customer having a percentage contribution in hamburger rather than the other category, the hazard of churning increases by 37.2%

Interpreting fit of the model using -2logL, AIC and SC

Model with the smallest AIC, SBC and -2 Log L is considered the best model. From the aforementioned criterion, we can see that our model is good. Furthermore, the chi square test of the likelihood ratio is also significant. Hence, we can say that our model is better.



**SURVIVAL ANALYSIS OF CUSTOMERS:**

The Survival analysis is best model to determine when the customer will switch brand and from that we could work on devising different retention strategies. To determine after how many weeks, the customer will be going to switch or will not be using the product. The Survival Analysis tell the time when the event of interest will occur.

We have performed survival analysis on week, which we calculated by multiplying tenure (in months) by 4. We have the life test with Family size (<=2 and more than two) as a strata. From the result, we could confirm that the survival probability across the two groups is different. The chi-sq test confirms the same.

From the life test, we could see that the survival probability of the customer with small and large family sizes greatly varies. The figure showing the survival probability, where we could see a nice trend. The survival probability decreases with time, here we cannot comment on whether the customer switched or did not visit Wendy’s after 444 weeks.