Prasanna Gopala Doddaballapura

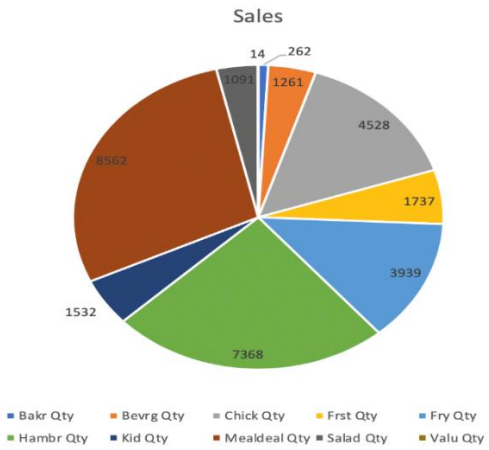
Objective:

Wendy's is an American international fast food restaurant chain. In this project we analyze Wendy's data to find meaningful insights which could help in making managerial decisions for the betterment of the business.

Descriptive Analysis: We have done some exploratory data analysis to figure out the major factors that we need to focus on to understand customer preferences.

Prasanna

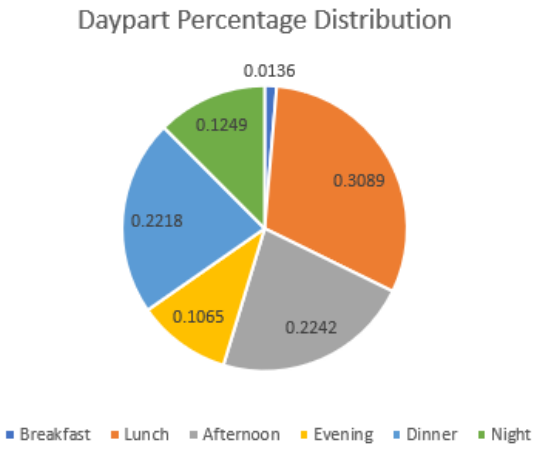
Predictive Analytics

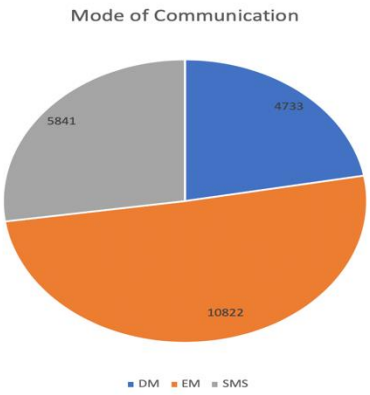


The adjacent plot shows us the most popular items at Wendy's. Clearly hamburgers sell the most followed by chicken sandwiches. The mealdeal attracts the most sales.

The value meals and fries at Wendy's are least popular among customers in terms of the quantity sold.

The daypart percentage distribution indicates that the greatest number of transactions are done during lunch and afternoon followed by breakfast. The least number of transactions are done for dinner.





As shown in the adjacent plot electronic mails are the most effective in attracting customers followed by direct mails and sms.

We can say that customers respond most to emails and these are the best way of promotions.

Regression Analysis: This will help us in identifying the impact of significant variables affecting total sales.

Prasanna **Predictive Analytics**

We applied the forward selection, backward selection and stepwise selection to arrive at the best features to analyze total sales. Log of total sales was taken as dependent variable and check for multicollinearity was done.

From the correlation matrix we saw that number of transactions and total units are highly correlated as they have a score of 0.82

Total rewards earned (TOT_Rewards_EARN) and number of times someone redeemed loyalty points are also highly correlated with a score of 0.87. So, we have dropped these features from the model.

Further no variable has a VIF greater than 10. The condition index is not greater than 100 for any variable. The eigenvalues are not very close to 0.

Best model after the above findings:

The REG Procedure
Model: MODEL1
Dependent Variable: log_TOT_SALES

Number of Observations Read	3443
Number of Observations Used	3441
Number of Observations with Missing Values	2

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	22	3972.75993	180.58000	2289.86	<.0001
Error	3418	269.54561	0.07886		
Corrected Total	3440	4242.30554			

Root MSE	0.28082	R-Square	0.9365
Dependent Mean	2.56359	Adj R-Sq	0.9361
Coeff Var	10.95420		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.38841	0.07160	19.39	<.0001
TOT_UNITS	1	0.02687	0.00106	25.44	<.0001
AVG_PRICE	1	0.42745	0.00436	98.09	<.0001
HAMBRGR_PCT	1	-1.04283	0.04069	-25.63	<.0001
SALAD_PCT	1	-0.56751	0.07061	-8.04	<.0001
OTHER_PCT	1	0.44736	0.04494	9.95	<.0001
MealDeal_PCT	1	0.28105	0.03465	8.11	<.0001
BAKERY_PCT	1	0.43329	0.13453	3.22	0.0013
FRSTY_PCT	1	-0.10043	0.04565	-2.20	0.0279
_avg_TB_TRANS	1	-0.00895	0.00022420	-39.90	<.0001
HAMBR_QTY	1	0.04820	0.00279	17.25	<.0001
SALD_QTY	1	0.01408	0.00537	2.62	0.0088
BEVRG_QTY	1	-0.01696	0.00506	-3.35	0.0008
FRY_QTY	1	0.00768	0.00329	2.33	0.0197
DYPT_PCT_LU	1	-0.07583	0.01290	-5.88	<.0001
DYPT_PCT_AF	1	-0.04284	0.01489	-2.88	0.0040
REDEEM_WELCOME	1	0.12052	0.02898	4.16	<.0001
TOT_Rewards_EARN	1	0.00648	0.00357	1.81	0.0697
_SMS	1	0.01770	0.00588	3.01	0.0026
_EM	1	0.06097	0.00586	10.41	<.0001
_DM	1	0.05605	0.00843	6.65	<.0001
inc_new	1	0.02007	0.00647	3.10	0.0019
DMA	1	-0.00004497	0.00001342	-3.35	0.0008

Model Interpretation: The adjusted R-square is 0.9365 which means that 93.65% of variance in log(TOT_SALES) is explained by the above model. All the variables are significant at a 95% confidence interval.

TOT_UNITS: If the total units purchased goes up by one unit the total sales goes up by 2.6% on average, all other variables remaining constant.

AVG_PRICE: If the average price paid by customer goes up by $1 the total sales goes up by 42.7% on an average, all other variables remaining constant.

Percentage of all the food items: Among different food items including hamburger, salad, mealdeal, bakery products, frosty and others we can say that the items that contribute the most towards increase in sales are classified under 'others'. This is followed by bakery products and mealdeal. Mealdeal indicates the presence of some deal and therefore attracts the customer and contributes in increasing the sale. The products which contribute the least are hamburgers, salad and frosty. Since hamburgers and salad are the main food items, they probably do not have a significant profit margin compared to the side items like bakery products and the ones classified as 'others'.

_avg_TB_TRANS: If the average time between transactions for any customer increases by 1 unit the total sales goes down by 0.8% on an average, all other variables remaining constant.

HAMBR_QTY: If the hamburger quantity increases by one unit the total sales go up by 4.8% on an average, all other variables remaining constant.

SALD_QTY: If the salad quantity increases by one unit the total sales go up by 1.4% on an average, all other variables remaining constant.

BEVRG_QTY: If the beverage quantity increases by one unit the total sales go down by 1.6% on an average, all other variables remaining constant.

FRY_QTY: If the fry quantity increases by one unit the total sales go up by 0.7% on an average, all other variables remaining constant.

DYPT_PCT_LU: The lunch menu during the lunch hours generate less revenues as compared to any other time of the day.

DYPT_PCT_AF: The afternoon hours generate greater revenue than the lunch menu but is lower as compared to other times of a day.

REDEEM_WELCOME: If a customer redeems a welcome gift the total sales go up by 12%.

TOT_REWARDS_EARN: If a customer earns a loyalty point the total sales go up by 0.6%.

Modes of contacts: Out of DM, EM and SMS emails are the most effective in getting customers and increasing total sales. Emails are followed by direct mails and sms.

Inc_new: Household income is scaled to get the new variable inc_new. If there is $10000 increase in the household income the total sales go up by 2%.

DMA: If the direct marketing area increases by 1 unit the total sales go up by 0.004%.

Insights: *Even though we saw in exploratory data analysis that lunch contributes the most when the daypart percentages are compared but the value associated with it is less as it does not boost the total sales as compared to other times of the day. This may be due to offers associated with lunch menus which attract many customers but does not generate much sales. Similar is the case of afternoon which amounts to second highest daypart percentage but does not generate proportionate sales.*

We should focus on these times of day where the total sales can be further boosted as the restaurant is frequented by many customers. One way of doing it is by making those items on menu available which have higher margins for profit.

Count Regression Modelling

As we want to maximize the number of transactions made by a customer at Wendy's we are building a statistical model to find the most significant factors affecting the number of transactions.

The FREQ Procedure

NUM_TXNS	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1580	45.92	1580	45.92
2	589	17.12	2169	63.03
3	336	9.76	2505	72.80
4	208	6.04	2713	78.84
5	148	4.30	2861	83.14
6	123	3.57	2984	86.72
7	90	2.62	3074	89.33
8	53	1.54	3127	90.87

We have summarized the Number of Transactions (Count data) using **Proc freq** command.

We see that Number of transactions vary from 1 to 63 and the frequency is rapidly decreasing. The rapidly decreasing characteristic proves that it is suitable for Count regression methods.

Here are the results for the two types of Count regressions that we have carried out the Negative Binomial model and the Poisson model. The negative binomial model is used with count data instead of the Poisson model if there is overdispersion in the data.

Unlike the Poisson model, the negative binomial model has a less restrictive property that the variance is not equal to the mean

Negative Binomial model

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	0.9671	0.1172	0.7374	1.1968	68.09	<.0001
TENURE	1	0.0081	0.0005	0.0071	0.0091	250.67	<.0001
AVG_PRICE	1	0.1385	0.0105	0.1179	0.1590	174.56	<.0001
NUM_EARN_REDEEM	1	0.2646	0.0095	0.2460	0.2832	779.77	<.0001
HH_INCOME	1	-0.1486	0.0147	-0.1774	-0.1199	102.66	<.0001
SIZE_HH	1	-0.0059	0.0121	-0.0297	0.0179	0.24	0.6247
Dispersion	1	0.2372	0.0114	0.2160	0.2606		

Poisson Model

Analysis Of Maximum Likelihood Parameter Estimates							
Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	1.6358	0.0849	1.4694	1.8023	371.14	<.0001
TENURE	1	0.0089	0.0004	0.0082	0.0097	569.91	<.0001
AVG_PRICE	1	0.1424	0.0070	0.1286	0.1562	409.72	<.0001
NUM_EARN_REDEEM	1	0.1445	0.0028	0.1390	0.1501	2625.79	<.0001
HH_INCOME	1	-0.2421	0.0106	-0.2630	-0.2213	518.98	<.0001
SIZE_HH	1	-0.0140	0.0082	-0.0301	0.0020	2.94	0.0864
Scale	0	1.0000	0.0000	1.0000	1.0000		

Here as the value for the Dispersion parameter is significantly different from zero, therefore we should use the Negative Binomial model for count regression methods.

Interpretation of significant variables

- **TENURE:** Individuals in the next month (Tenure is in months) are expected to have a 0.81% **increase** in number of transactions made by them
- **Average price:** If a customer's Average price spent goes up by 1 unit then it is expected to have a 14.24% increase in the number of transactions made by him.
- **Number of Times Redeemed loyalty earnings:** If a customer redeems his loyalty earnings one more time then this is expected to have a 14.45% increase in the number of transactions made by him

House Hold Income: This has a negative impact on the Number of transactions. If there is 10,000 dollars increase in the customer household income, then the number of transactions made by him is expected to **decrease** by 24.21%.