An Analysis on the Factors Leading to Obesity

Submitted by:
Jose Baby (20-PST-020)

Submitted to:
Dr. Martin Luther William,
Head of the department,
Loyola College.

CONTENTS

- INTRODUCTION AND OBJECTIVE
- DATA DESCRIPTION
- DATA VISUALIZATION
- CONCLUSION

Introduction:

Obesity is a disorder involving excessive body fat that increases the risk of health problems. People who have obesity, compared to those with a normal or health weight, are at increased risk for many serious diseases and health conditions like High Blood Pressure (Hyper Tension), High LDL cholesterol (Dyslipidemia) and so on.

Objective:

The purpose of this study is to identify the various factors which causes Obesity in an individual and to further create a model to predict the various weight status resulting from the individual's lifestyle.

DATA DESCRIPTION

The data contains 14 attributes and 2111 records, the records are labelled with the class variable (Obesity Level) that allows the classification of the data using the values of Insufficient Weight, Normal Weight, Overweight and Obesity. The data contains information on respondents with ages between 14 and 61 with diverse eating habits and physical conditions.

1) **Gender -** Gender of the respondent (Male/Female)

2) **Age –** Age of the respondent (14-61)

3) **Family history with overweight –** If any of the respondents family members have obesity (Yes, No)

4) **FAVC –** If the respondent eats high caloric food frequently (Yes, No)

5) **FCVC –** If the respondent usually has vegetables along with their meals (Never, Sometimes, Always)

6) **NCP –** How many main meals does the respondent have (Three or less than three, More than three)

7) **CAEC –** The number of times the respondent eats food between meals (No, Sometimes, Frequently)

8) **SMOKE –** If the respondent has the habit of smoking (Yes, No)

9) **CH2O –** How much water does the respondent drink in a daily basis (Less than 2 litres, More than 2 litres)

10) **SCC –** If the respondent monitors the amount of calories they consume everyday (Yes, No)

11) **FAF –** Does the respondent indulge in physical activities (Yes ,No)

12) TUE – How much time does the respondent spend in using technological devices such as cell phone, video games, television, computer & Others (0-2hrs, 2-5 hours, more than 5 hours)

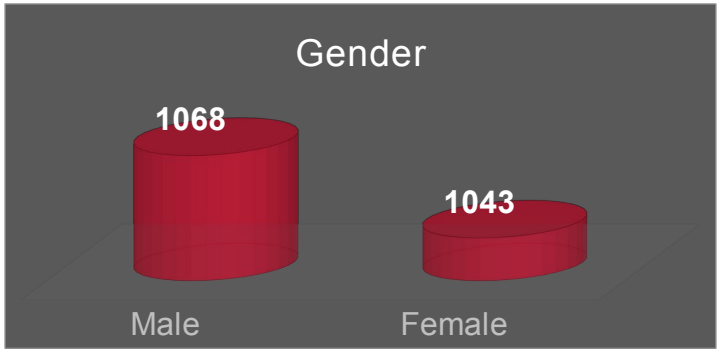
13) CALC – How often does the respondent consume alcohol (I do not drink, Sometimes, Frequently, Always)

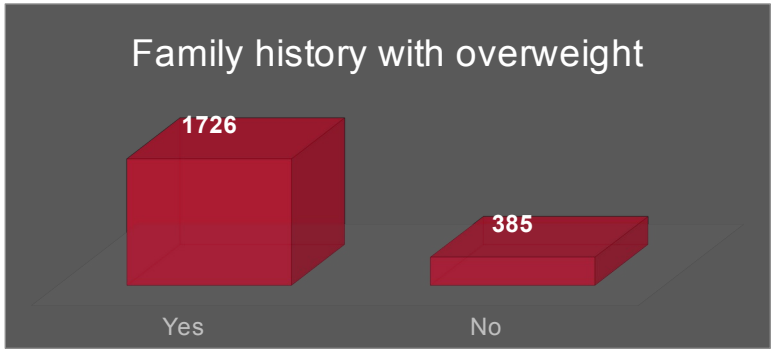
14) MTRANS – Which mode of transportation does the respondent usually use (Public Transportation, Private Transportation)

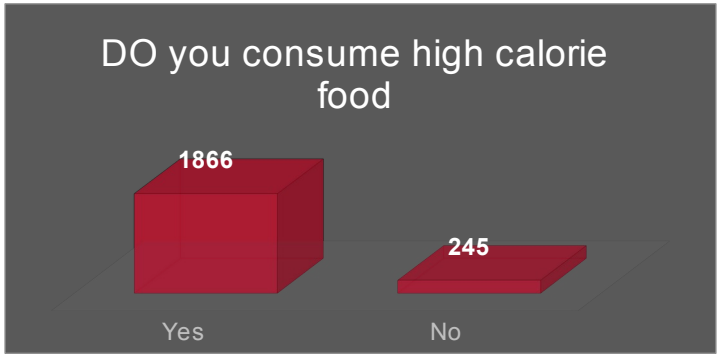
The dependent variable Obesity Level has 4 categories, namely

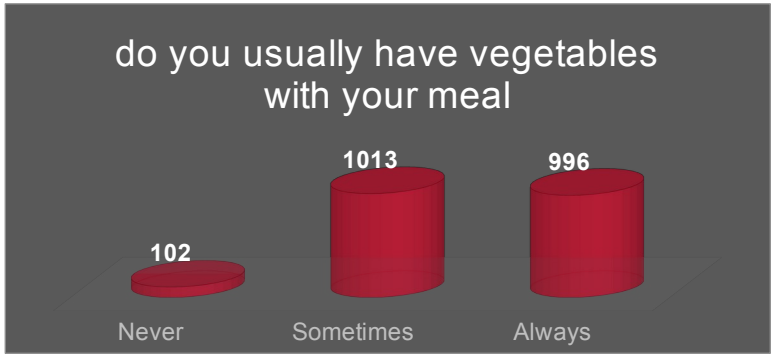
- Insufficient Weight
- Normal Weight
- Overweight
- Obese

VISUALIZATION

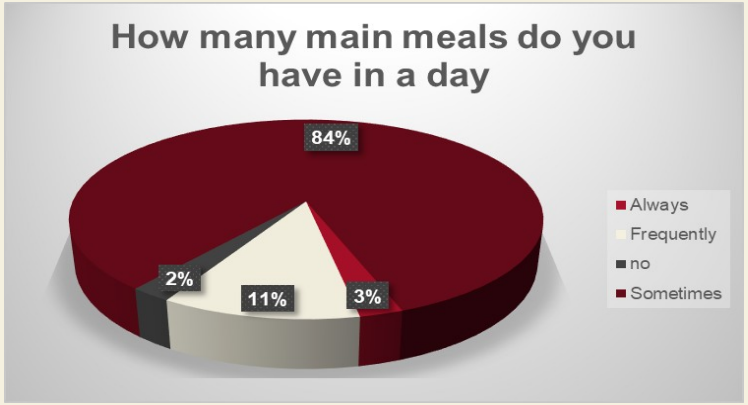


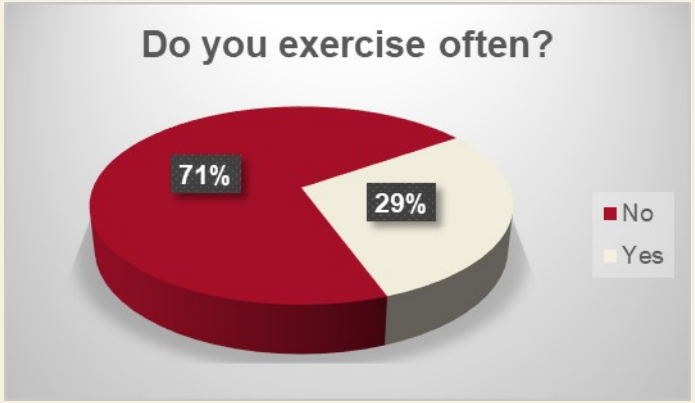


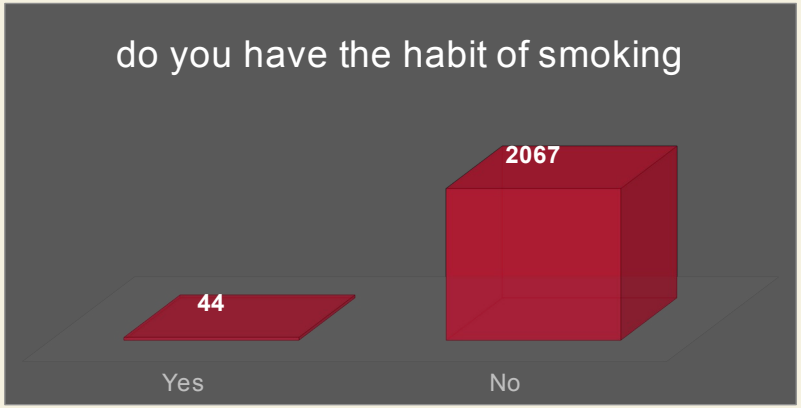


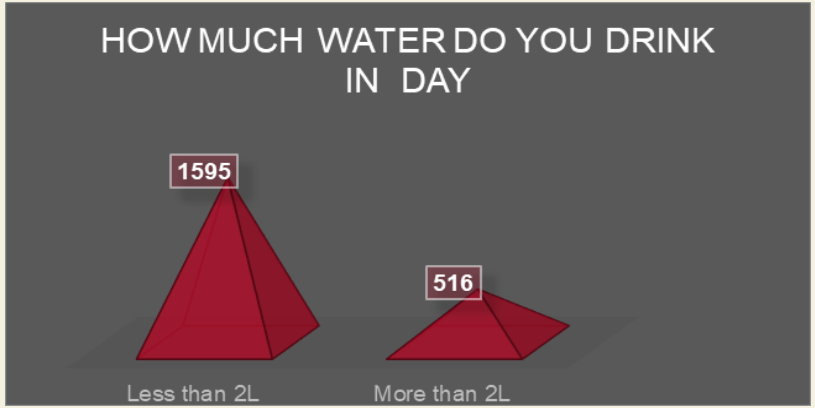


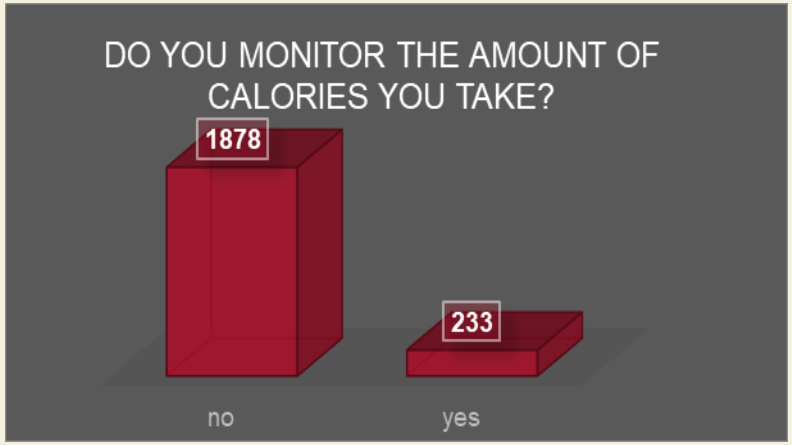

Do you have food between the meals
84%
2%
11%
3%
Always
Frequently
no
Sometimes

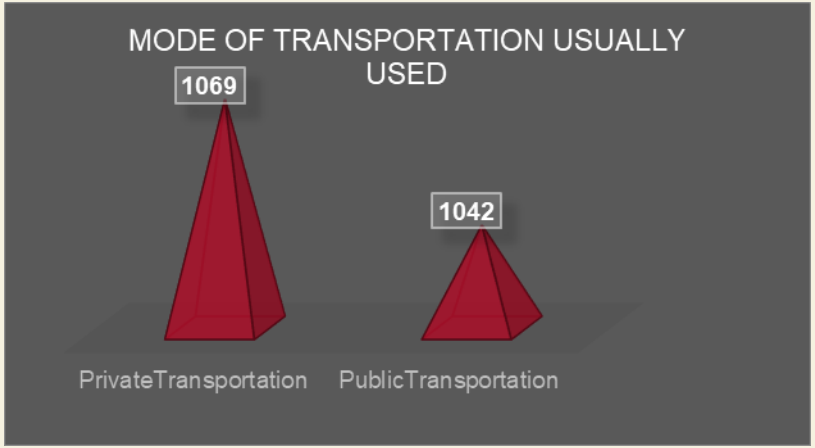

How many main meals do you have in a day
84%
2%
11%
3%
Always
Frequently
no
Sometimes

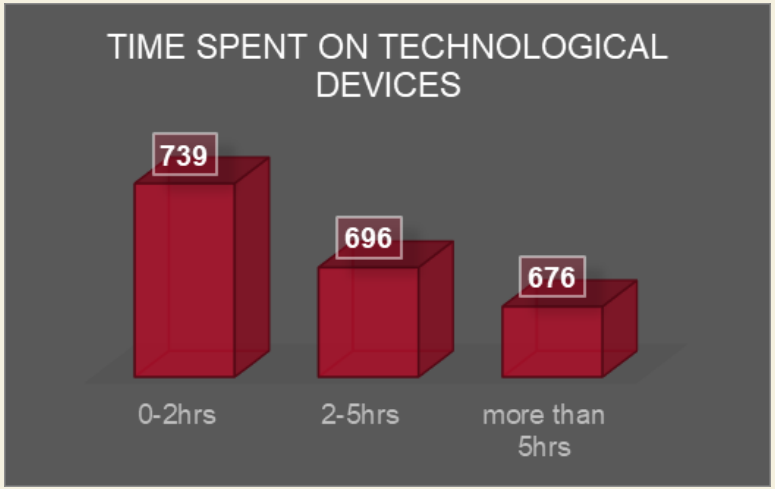

Do you exercise often?
71%
29%
No
Yes

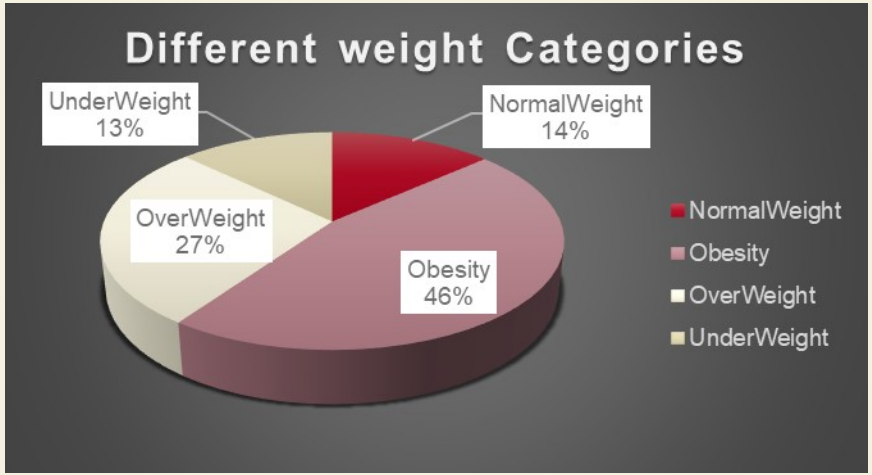

do you have the habit of smoking
2067
44
Yes
No


HOW MUCH WATER DO YOU DRINK IN DAY
1595
516
Less than 2L
More than 2L


DO YOU MONITOR THE AMOUNT OF CALORIES YOU TAKE?
1878
233
no
yes


MODE OF TRANSPORTATION USUALLY USED
1069
1042
PrivateTransportation
PublicTransportation


TIME SPENT ON TECHNOLOGICAL DEVICES
739
696
676
0-2hrs
2-5hrs
more than 5hrs


Different weight Categories
UnderWeight 13%
NormalWeight 14%
OverWeight 27%
Obesity 46%
NormalWeight
Obesity
OverWeight
UnderWeight

ANALYSIS

ODDS RATIO

1. Consumption of high calorie food x Weight Category

		Weight Category			
		Obese	Not Obese	ODDS	ODDS RATIO
Consumption of HCF	yes	1437	429	3.3496	3.7865
	no	115	130	0.8846	

Chi- Square of independence

```
        Pearson's Chi-squared test

data:  data$FAVC and data$Obesityl
X-squared = 100.59, df = 1, p-value < 2.2e-16
```

Interpretation

We have obtained the odds ratio $\theta^{\wedge} = 3.786$, which suggests that the odds favouring the risk of developing the condition of obesity when an individual consume high calorie rich food is 278% higher as compared to an individual who is not consuming.

Also, using the Chi-Square test of independence, we can observe that the p value is less than the level of significance 0.05, (i.e.) we reject the null hypothesis, and hence there is dependence.

2. Habit of monitoring the calorie intake vs Weight Category

		Weight Category			
		Obese	Not Obese	Odds	Odds Ratio
Do you monitor your calorie intake?	Yes	79	154	0.5129	0.141
	No	1473	405	3.637	

Chi- Square of independence

```
> chisq.test(table1)

        Pearson's Chi-squared test with Yates' continuity correction

data:  table1
X-squared = 38.123, df = 1, p-value = 6.641e-10
```

Interpretation

We have obtained the odds ratio $\hat{\theta} = 0.14105$, which suggests that the odds favouring the risk of developing the condition of obesity when an individual monitors the calories he/she consume is 86% lower as compared to an individual who is not.

Also, using the Chi-Square test of independence, we can observe that the p value is less than the level of significance 0.05, (i.e.) we reject the null hypothesis, and hence we can conclude that there is a dependence among the two variables.

3. Exercise vs Weight Category

		Weight Category			
		Obese	**Not Obese**	**ODDS**	**ODDS**
Do you Exercise?	**YES**	377	238	1.5840	0.4327
	NO	1175	321	3.6604	

Chi- Square of independence

```
> chisq.test(table2)

        Pearson's Chi-squared test with Yates' continuity correction

data:  table2
X-squared = 65.67, df = 1, p-value = 5.331e-16
```

Interpretation

We have obtained the odds ratio $\hat{\theta}$ = 0.43274, which suggests that the odds favouring the risk of developing the condition of obesity when an individual exercises regularly is 56.726 % lower as compared to an individual who does not exercises regularly.

Also, using the Chi-Square test of independence, we can observe that the p value is less than the level of significance 0.05, (i.e.) we reject the null hypothesis, and hence we can conclude that there is a dependence among the two variables.

4. Mode of Transportation vs Weight Category

		Weight Category			
		Obese	Not Obese	ODDS	ODDS RATIO
Type Of Transportation	PrivateTransportation	909	160	5.6812	3.5253
	PublicTransportation	643	399	1.6115	

Chi- Square of independence

```
        Pearson's Chi-squared test with Yates' continuity correction

data:  table1
X-squared = 13.135, df = 1, p-value = 0.0002899

>
```

Interpretation

We have obtained the odds ratio $\hat{\theta}$ = 3.5254, which suggests that the odds favouring the risk of developing the condition of obesity for an individual who generally travel by their private vehicle is 252.537% higher as compared to people who travel by public transportation facilities.

Also, using the Chi-Square test of independence, we can observe that the p value is less than the level of significance 0.05, (i.e.) we reject the null hypothesis, and hence we can conclude that there is a dependence among the two variables.

Marginal/Conditional Association:

1. Family History of Obesity x Weight Category / Gender

MARGINAL ODDS RATIO

	Weight Criterion			
		Obese	Not Obese	ODDS	ODDS RATIO
FAMILY HISTORY WITH OVERWEIGHT	yes	1445	281	5.142	13.36
	no	107	278	0.384	

PARTIAL TABLE

		Female			
		Obese	Not Obese	ODDS	ODDS RATIO
Family History	yes	691	120	5.7583	29.3978
	no	38	194	0.1958	
		Male			
		Obese	Not Obese	ODDS	ODDS RATIO
FAMILY HISTORY	Yes	754	161	4.683	5.7013
	No	69	84	0.8214	

INTERPRETATION

Marginal Odds Ratio

From the **marginal table**, we have obtained that $\Theta\hat{}_{XY} = 13.36$, which mean the odds ratio favouring a person to be obese is 1236% higher when the individual has a pre-existing history of condition of Obesity in his/her family.

Conditional Odds Ratio

Further from the **partial table** it can be observed that given the respondent is **Male**, the odds ratio for him to be obese, when he has a family history of unhealthy weight is:

$$\Theta\hat{}_{XY(1)} = 5.70$$

That is there is an increase of 470%, for a male to develop the condition of Obesity, due to his genes.

But, when the gender is female it has been observed that the odds ratio for her to be obese, when she has a family history of unhealthy weight is:

$$\Theta\hat{}_{XY(2)} = 29.397$$

(i.e.) there is an increase of 2840% , for a female to develop the condition of Obesity, due to her family history.

Thus, we can conclude that Obesity is inherent and it has more effect on Females as compared to Males.

2. Alcohol Consumption x Weight Category / Gender

MARGINAL TABLE					
		Weight Criterion			
		Obese	Not Obese	ODDS	ODDS RATIO
Alcohol consumption	Yes	1137	335	3.394	1.832
	NO	415	224	1.853	

PARTIAL TABLE					
		Female			
		Obese	Not Obese	ODDS	ODDS RATIO
Alcohol Consumption	Yes	552	187	2.952	2.118
	NO	177	127	1.394	
		Male			
		Obese	Not Obese	ODDS	ODDS RATIO
Alcohol Consumption	Yes	585	148	3.9527	1.611
	NO	238	97	2.4536	

INTERPRETATION

Marginal Odds Ratio

From the **marginal table**, we have obtained that $\Theta\hat{}_{XY} = 1.832$, which mean the odds ratio favouring a person to be obese is 83.2% higher when the individual has a habit of consuming Alcohol.

Conditional Odds Ratio

Further from the **partial table** it can be observed that given the respondent is **Male**, the odds ratio for him to be obese,:

$$\Theta\hat{}_{XY(1)} = 1.6109$$

(i.e.) there is an increase of approximately 62%, for a male to develop the condition of Obesity, due to his condition of drinking.

But, when the gender is female it has been observed that the odds ratio for her to be obese, when she has a drinking habit is:

$$\Theta\hat{}_{XY(2)} = 2.118$$

(i.e.) there is an increase of 118% , for a female to develop the condition of Obesity, due to the consumption of alcohol.

Thus, we can conclude that Obesity and Alcohol drinking habits are very much associated. Also it has a more severe impact on Females.

Multinomial (Ordinal) Logit Model:

Coefficients:	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept):1	0.938162	0.486854	1.927	0.05398	.
(Intercept):2	2.360452	0.488887	4.828	1.38E-06	***
(Intercept):3	4.272869	0.497594	8.587	< 2e-16	***
Age	-0.03907	0.009127	-4.28	1.86E-05	***
family_history_with_overweightyes	-1.86929	0.146428	-12.766	< 2e-16	***
FAVCyes	-0.86403	0.160648	-5.378	7.52E-08	***
FCVC2	0.159478	0.245893	0.649	0.51662	
FCVC3	-0.48976	0.246789	-1.985	0.0472	*
CAECFrequently	1.07438	0.349203	3.077	0.00209	**
CAECno	-1.14061	0.437358	-2.608	0.00911	**
CAECSometimes	-1.02525	0.319631	-3.208	0.00134	**
SCCyes	1.469683	0.170675	8.611	< 2e-16	***
FAF1	0.551258	0.110817	4.974	6.54E-07	***
TUE31	0.158426	0.125855	1.259	0.2081	
TUE32	-0.62034	0.132118	-4.695	2.66E-06	***
TypeOftransportPublicTransportation	0.299442	0.110065	2.721	0.00652	**

- A model has been build considering the **weight categories** as the dependent variables and 15 independent variables.
- The data was divided as ‘Train’(75%) and ‘Test’(25%)data.
- The model uses **cumulative probabilities** up to a threshold, thereby making the whole range of ordinal categories binary at that threshold. This model assumes that the coefficients that describe the relationship between all categories of response are the same. Because the relationship between all pairs of groups is the same, there is only one set of **intercept coefficients.**
- Among these 14 variables, eight were found to be significant

Model Obtained

1st Model:

log ($P_1(X)/P_{234}(X)$) = 0.938162-0.039069(Age)-1.869293(family_history_with_overweightyes) -0.864026 (FAVCyes)-0.489760(FCVC3) -1.140609 (CAECno) -1.025253(CAECSometimes) +1.07438 (CAECFrequently)+ 1.469683(SCCyes) +0.551258 (FAF1)-0.620343 (TUE32)+0.299442 (MTRANSPublic_Transport)

2nd Model:

log ($P_{12}(X)/P_{34}(X)$) =2.360452 -0.039069(Age)-1.869293(family_history_with_overweightyes) -0.864026 (FAVCyes)-0.489760(FCVC3) -1.140609 (CAECno) -1.025253(CAECSometimes) +1.07438 (CAECFrequently)+ 1.469683(SCCyes) +0.551258 (FAF1)-0.620343 (TUE32)+0.299442 (MTRANSPublic_Transport)

3rd Model:

log ($P_{123}(X)/P_4(X)$) = 4.272869 -0.039069(Age)-1.869293(family_history_with_overweightyes) -0.864026 (FAVCyes)-0.489760(FCVC3) -1.140609 (CAECno) -1.025253(CAECSometimes) +1.07438 (CAECFrequently)+ 1.469683(SCCyes) +0.551258 (FAF1)-0.620343 (TUE32)+0.299442 (MTRANSPublic_Transport)

INTERPRETATION:

- **Age:** It is seen from the coefficients that as the Age increases by 1 unit, the odds for a person to be lean against being obese decreases by a multiple of exp (-0.039069) = 0.96168 ,there is a decrease of 3.8% in the odds for the person to be lean.(i.e.) as the age increases for a person, they are more likely to gain weight.

- **family_history_with_overweightyes:** It is seen from the coefficients that as the family_history_with_overweight (yes) increases by 1 unit, the odds for a person to be lean against being obese decreases by a multiple of exp (-1.869293) = 0.154233 (i.e.) There is a approximately 84.57% decrease in the odds for the person to be lean. If the respondent's *family members have had condition of obesity*, then the respondent is also more likely to develop obesity.

- **FAVCyes:** It is seen from the coefficients that as the FAVCyes increases by 1 unit, the odds for a person to be lean against being obese decreases by a multiple of exp (-0.864026) = 0.4215 (i.e.) There is a 58% decrease in the odds for the person to be lean. As the *high caloric food consumption frequency* increases for a person they are more likely to become obese.

- **FCVC3:** It is seen from the coefficients that as the FCVC3 increases by 1 unit, the odds for a person to be lean against being obese decreases by a multiple of exp (-0.4898) = 0.6128 (i.e.) There is a 38.72% decrease in the odds for the person to be lean. If the *person consumes vegetables frequently along with their meals*, they are more likely to become obese.

- **CAECno:** It is seen from the coefficients that as the CAECno increases by 1 unit, the odds for a person to be lean against being obese decrease by a multiple of exp (-1.1406) = 0.3196243 (i.e.) There is a 68% decrease in the odds for the person to be lean. If the person does not have food between their meals, they are more likely to become obese.

- **CAECSometimes:** It is seen from the coefficients that as the CAECSometimes increases by 1 unit, the odds for a person to be lean against being obese decreases by a multiple of exp (-1.0252) = 0.3587 (i.e.) There is a 64.13% decrease in the odds for the person to be lean. If the person has food sometimes between their meals, they are more likely to become obese.

- **CAECFrequently:** It is seen from the coefficients that as the CAECFrequently increases by 1 unit, the odds for a person to be lean against being obese increases by a multiple of exp (-1.0743) = 2.92827(i.e.) There is a 192% increase in the odds for the person to be lean. If the person has food frequently between their meals, they are less likely to become obese.

- **SCCyes:** It is seen from the coefficients that as the SCCyes increases by 1 unit, the odds for a person to be lean against being obese increases by a multiple of exp (1.469683) = 4.34785 (i.e.) There is a 335% increase in the odds for the person to be lean. If the person monitors the amount of calories they consume, they are less likely to become obese.

- **FAF1:** It is seen from the coefficients that as the FAF1 increases by 1 unit, the odds for a person to be lean against being obese increases by a multiple of exp (0.551258) = 1.735435 (i.e.) There is a 74% increase in the odds for the person to be lean. If the person indulges in physical activities regularly, they are more likely to become lean.

- **TUE32:** It is seen from the coefficients that as the TUE32 increases by 1 unit, the odds for a person to be lean against being obese decreases by a multiple of exp (-0.620343) = 0.5377601 (i.e.) There is a 46% decrease in the odds for the person to be lean. If the person spends time in using technological devices such as cell phone, video games, television, computer & others for more than 5 hours, they are more likely to become obese.

- **MTRANSPublic_Transport:** It is seen from the coefficients that as the MTRANSPublic_Transport increases by 1 unit, the odds for a person to be lean against being obese increases by a multiple of exp (0.299442) = 1.39411 (i.e.) There is a 39% increase in the odds for the person to be lean. If the person usually uses public mode of transport, they are less likely to become obese.

ACCURACY

Train	Predicted			
Actual	UnderWeight	NormalWeight	OverWeight	Obesity
UnderWeight	84	21	65	34
NormalWeight	68	32	84	29
OverWeight	23	11	145	255
Obesity	10	1	44	678

Test	Predicted			
Actual	UnderWeight	NormalWeight	OverWeight	Obesity
UnderWeight	27	7	20	14
NormalWeight	33	4	17	17
OverWeight	9	0	47	88
Obesity	3	2	18	221

BIC	3195.602005
AIC	3109.718668
Accuracy(Train)	0.5928
Accuracy(Test)	0.5673

CONCLUSION

Obesity affects nearly every part of the body. If you're living with obesity you can treat or manage many of these risk factors with a combination of diet, exercise and lifestyle changes. Losing just 5 to 10 percent of your current weight can reduce your risk of developing those health issues.