The Effects of Eating Habits on Likelihood of Being Overweight

Aaron Xiao

2020-05

Abstract

A health and wellness company has consulted our team to help determine what eating and health habits are associated with being overweight, and to determine high-risk groups to target for a health campaign. Using a 2014 Health Dataset collected by the U.S. Department of Agriculture's Economic Research Service, our team used two models, logistic regression, and random forest, to predict the likelihood of being overweight based on eating patterns. We found that those who spend more time eating, who do not drink soda and do not eat meat are significantly less likely to be overweight. The random forest model provided more accurate predictions than the logistic model. We used a clustering model to locate groups with similar eating patterns in our data set and noted two distinct groups, with group 1 at higher risk of being overweight compared to group 2. Our team also developed an app on the Shiny App platform so individuals can input their eating and health patterns to find out their likelihood of being or becoming overweight.

Introduction

Health companies are increasingly using a marketing strategy called "target marketing", which is a practice used to more effectively address the wants and needs of their customers (Elrod 2018). In an era where weight problems have become a serious concern for many individuals, health and wellness companies are trying to find out solutions to help their customers reach their health goals. Reaching the appropriate customers may be difficult as it is not always evident that an individual may be at risk of becoming overweight. Eating patterns and behaviours are major contributors to individuals' weight, and is subsequently the focus of our analysis. To predict the likelihood of being overweight based on their eating patterns, we used information such as the time spent in minutes eating a meal, if the individual consumes soft drinks (diet or regular), how often individuals take part on physical activities, or purchase take-out food, and who in the household does the primary grocery shopping and meal preparation.

Background

A health and wellness organization approached our analytics team and asked for our assistance in identifying groups at risk of being overweight and provide them with information on healthier eating and health habits. In doing so, our research question becomes: what types of eating and health patterns are associated with being overweight? Moreover, are there groups of individuals with similar eating and health habits that are at greater risk of being overweight? According to the Centers for Disease Control and Prevention, a Body Mass Index (BMI) higher than 25 points is an indication of being overweight (CDCP, 2020).

Once we determine what factors are associated with being overweight, and which groups that are more likely to be or become overweight, our business partner can target those groups with information on healthier eating habits to reduce their risk of being overweight. Our models will not only be useful to identify these groups and running predictions on the likelihood of being overweight. Our work will also be deployed on Shinny App which will allow users to identify their distinct group and risk of overweight based on the parameters he/she inputs into the app.

Drivers and relationship to input

Time spent while eating a meal may contribute to become overweight. Previous research points out that fast eaters see more weight gains, larger waistlines, and higher blood sugar levels than slow eaters (Laliberte, 2019). Eating too quickly may lead individuals to miss their satiety cues, resulting in overeating. Moreover, when individuals eat doing other activities such as working, they may unintentionally eat more and be unaware of the amount they consumed, also resulting in overeating.

Soft drink consumption has also been linked to higher rates of being overweight. The research conducted by Basu and colleagues (2013) concluded that soft drink consumption was significantly associated with obesity and overweight. More research is linked to relationship of physical activity and overweight. The study conducted by (Jakici et al., 2018) concluded that that physical activity of at least 150 min per week influenced body weight regulation and decreased the likelihood of being overweight. Take out and fast food can be high in sugar, salt, and saturated fats. As such, the frequency at which individuals purchase and consume take-out or fast food may also be associated with their likelihood of being overweight.

We also wanted to understand the relationship of grocery shopping and overweight. More specifically, we wanted to see if the individual that is being measured is doing the grocery shopping or someone else is doing it. Meat is likely to contribute to the weight of individuals and we wanted to identify if the amount of meat consumption within the last 7 days has an impact on the weight of people. Finally, we wanted to explore if there is a relationship of those individuals who prepare their own meals at home or somebody else does it for them.

Set of assumptions related to problem

We cannot assume that factors such as the amount of time spend on eating meals, soft drinks, and physical activity impact individuals equally. Some individuals have higher metabolism and may digests these foods easier than other people. We cannot account for premedical conditions that affect a person becoming overweight such as Hypothyroidism, Cushing's Syndrome, and Depression (Martin, 2019). Finally, this data set only measures the time spent doing physical activity and consumption of take-out food over the past 7 days. Our research does not consider if the individual has had unhealthy eating patterns for a very long time.

Define key metrics of success

Finding distinct groups in our data set and understanding which groups are more at risk of being or becoming overweight. Comparing our machine learning models (logistic regression and random forest) and pick the one with the highest accuracy score. It will also be important to note which IV's have a high correlation with our dependent variable.

Describe how the ML framework was applied

For this project, we made use of R programming language in order to find out the clusters in our data and predict the likelihood of being overweight. The following libraries were used to recode our data and build our ML model: cluster, dplyr, ggplot2, readr, Rtsne, reshape2, corrplot, rpart, rpart.plot, rattle, plyr, epiDisplay, randomForest, car.

Identify and prioritize means of acquisition

We used this data set to investigate the main drivers associated with individuals being overweight. We wanted to explore if we could find clusters of people based on their eating and health patterns. Health has and will always be important for humankind and we hope to make a difference by identifying what sort of behaviours lead to become overweight. This data set was retrieved from the website Kaggle and the survey was conducted by the US Bureau of Labor Statistics.

Data Preparation

To determine how eating pattern is associated with being overweight, we analyzed the 2014 Eating & Health Module of the American Time Use Survey collected by the U.S. Department of Agriculture's Economic Research Service. This survey was sponsored by the Bureau of Labor Statistics and was conducted by the U.S. Census Bureau. The entire dataset for this module contains three separate files: one is the primary respondent file which contains information such as respondents' BMI, eating habits, exercise frequencies, and household income; the second file is an activity file; and the third file contains bootstrapping weights for each respondent. Information captured in the activity file is also included in the respondent file, and bootstrapping practices fall outside of the purview of the current project. As such, only the main respondent file was used in the analysis.

The respondent file has 37 columns and 11,212 rows. Our target variable is whether or not the respondent is overweight. According to the U.S. Department of Health and Human Services (n.d.), a BMI of 25 or higher is considered to be overweight. As such, the BMI variable was categorized into "1" for those with BMI 25 and higher (i.e. overweight), and "0" for those with BMI lower than 25 (i.e. not overweight). There were 575 observations that had missing BMI. We excluded these cases given that this variable is our target variable and imputing for missing values may affect the results; moreover, the number of missing cases are small, so that even after excluding these cases our sample size is sufficiently large to complete the analysis.

Having identified the target variable, columns that were unrelated to our business question on how eating patterns are associated with the likelihood of being overweight were excluded. The dataset contains several columns that measure household income (relative to the poverty threshold). Although existing literature point to a negative relationship between being overweight and income in the US (i.e. the lower the income, the higher the likelihood of being overweight; see for example: Bentley, Ormerod, & Ruck, 2018), the focus of the business question is on eating patterns and income is largely outside of the control of whom the organization is targeting. Similarly, related variables such as the labour force status of the respondent, as well as their family members, whether they receive government benefits, and changes in household composition were excluded. Data on respondents' height and weight were excluded, given this is the information used to calculate BMI (and subsequently the target variable). Columns less relevant to the business question and that were highly imbalanced such as, whether respondents consumed raw milk or used meat thermometers when cooking were also excluded.

With the remaining data, some columns with overlapping information were combined for simplicity. For instance, respondents were asked whether they drink soda. Among those who said yes, they were then asked what type of soda. These two columns were combined so one column captures all activities: 0 – No Soda, 1 – Diet Soda, 2-Regular Soda. This combination was also done for those who reported engaging in physical activity and were asked in a subsequent question how many times they exercise per week. Once the columns were combined, the redundant columns were removed.

For the column on soda-drinking and meat-eating habits, there were large proportions of respondents who did not respond to the questions (32.7% and 27.9% respectively). Given how large these groups are, we did not impute the missing values to prevent creating unintended biases and instead kept them as separate groups, as they can potentially have a different profile from the remaining respondents who did respond to the question. Six of the remaining columns had small numbers of missing values. For continuous variables, such as time spent eating and exercise frequency, the missing values were imputed using the means. For categorical variables, such as whether the respondent is the primary grocery shopper or meal preparer of the household, missing values were imputed using the mode.

Once these transformations were made, the final dataset contains 9 columns and 10,647 observations (See Table 1 for summary). The columns retained included time individuals spend eating, how often they purchased take-out or fast food, who in the household does the grocery shopping and meal preparation, and whether respondent cooked meat. Although not directly associated with eating patterns, who the primary grocery shopper and meal preparer were included given that the person who usually does the grocery shopping and meal preparation are likely to have greater control over the food choices they make and also how the food is prepared, both of which may affect the types of food consumed, and subsequently weight-related outcomes. Variations in physical activity frequency must be included and adjusted for in the models, given its strong relationship with weight.

Table 1. Columns in Final Dataset

Column Name	Type of Variable	Coding	Description
overweight	Numeric	Dummy Coded (0,1)	Overweight = BMI 25 and higher
ertpreat	Numeric	Continuous	Total amount of time spent in primary eating and drinking (in minutes)
ertseat	Numeric	Continuous	Total amount of time spent in secondary eating (in minutes)
eudietsoda	Numeric	4 Categories: -1: No answer 0: No Soda 1: Diet Soda 2: Regular Soda	Whether or not respondent drinks soda, and if so, the type of soda
euexfreq	Numeric	Continuous	Number of times respondent engaged in physical activity outside of their job in the past 7 days
eufastfdfreq	Numeric	Continuous	Number of times respondent purchased prepared food from a deli, carry-out, delivery food, or fast food over the past 7 days
eugroshop	Numeric	3 Categories: 1: Primary Grocery Shopper 2: Not Primary 3: Shared Grocery Shopper	Who usually does the grocery shopping in respondent's household
eumeat	Numeric	3 Categories: -1: No answer 1: Ate Meat 2: Did Not Eat Meat	Whether respondent prepared any meals with meat, poultry, or seafood in the past 7 days
euprpmeal	Numeric	3 Categories: 1: Primary Meal Preparer 2: Not Primary 3: Split Meal Preparation	Who usually prepared meals in respondent's household

Basic Descriptive Results & Visualizations

Out of all the observations in our dataset, 64.4% are considered as overweight and 35.6% are non-overweight (see Figure 1 for distribution). For those who are overweight, the average total amount of time spent in primary eating and drinking is 65.8 minutes, with a median of 60.0 minutes. The maximum eating time is 508 minutes whereas the minimum time of eating is 0 minute. Comparing with the average eating time of those who are not overweight (68.2 minutes), the overweight observations spent 2.4 minutes more. (see Figure2).

igure 1. Overweight distribution

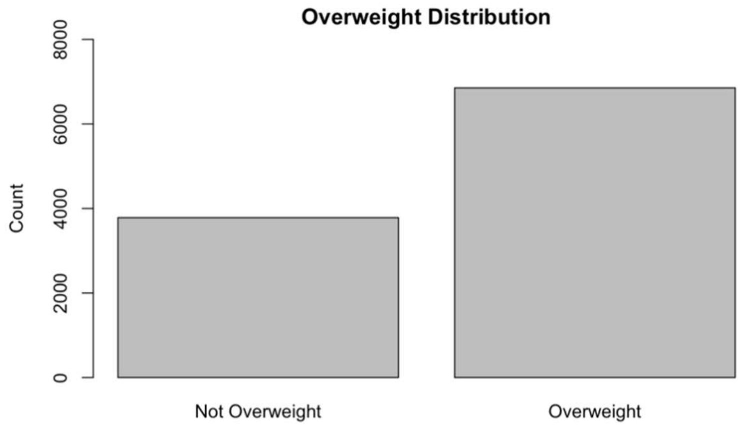
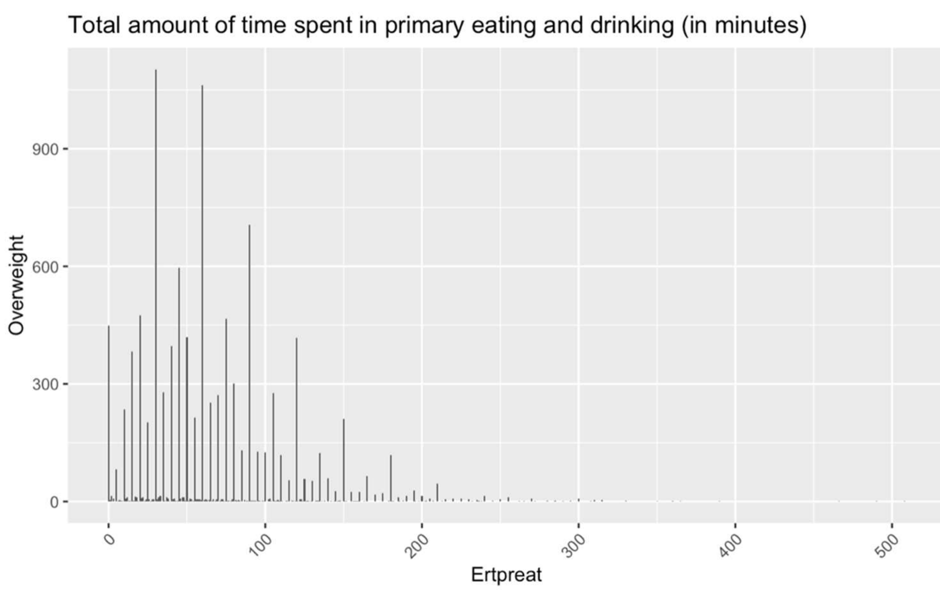
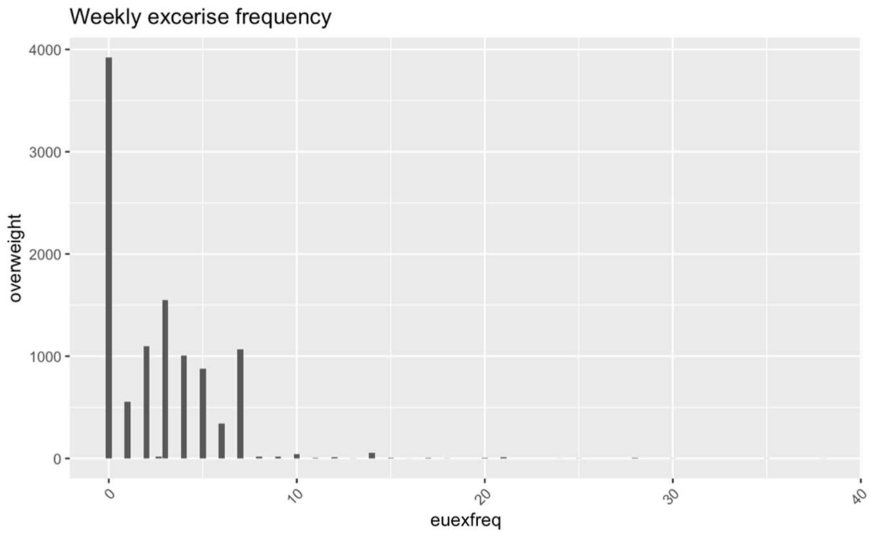


Figure2. Total amount of time spent in primary eating and drinking (in minutes)



The average weekly exercise frequency for those who are overweight is 2.6 times per week with a median of 2.0 times. The maximum times of exercise is 38 times per week, whereas the minimum is 0 time. For those who are not overweight, the average of times of weekly exercise is 3.1 which is 0.5 times more than those who are overweight. The majority of those who are overweight do not engage in any physical activity outside of, since the 0 time has the highest number of overweight counts (see Figure3).

Figure3. Weekly exercise frequency



We would also like to see if there exists any relationship between the overweight and other variables by showing this pair plot. There exists a negative relationship between overweight and weekly exercise frequency which suggest that the more people exercise, the least likely they become overweight and following by a positive relationship between the fast food consumption frequency and overweight that which indicates that the more people consume fast food, they are more likely to be overweight. Other than these two above, there doesn't exist a strong and clear relationship between overweight and each other variable.

Figure 4. Variables pair plot

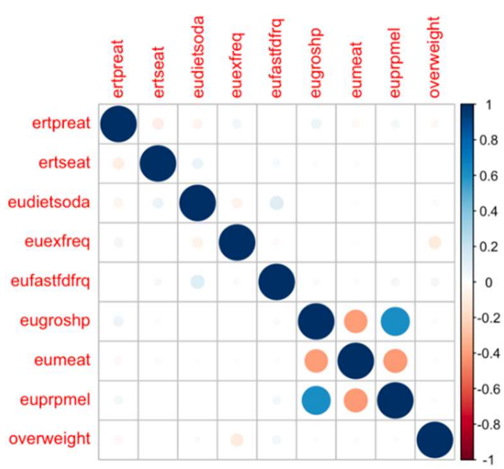


Table 2. Summary of Numeric Columns in Final Dataset

ertpreat	ertseat	eudietsoda	euexfreq	eufastfdfrq
Min. : 0.00	Min. : 0.00	-1: 3481	Min. : 0.000	Min. : 0.000
1st Qu.: 30.00	1st Qu.: 0.00	0: 4291	1st Qu.: 0.000	1st Qu.: 0.000
Median : 60.00	Median : 5.00	1: 1121	Median : 2.000	Median : 1.000
Mean : 65.85	Mean : 16.99	2: 1744	Mean : 2.646	Mean : 1.585
3rd Qu.: 90.00	3rd Qu.: 15.00		3rd Qu.: 4.000	3rd Qu.: 2.000
Max. :508.00	Max. :990.00		Max. :38.000	Max. :21.000
eugroshp	**eumeat**	**euprpmel**	**overweight**	
:6501	0: 2973	1: 6606	0: 3783	
2:2841	1: 6793	2: 2970	1: 6854	
3:1295	2: 871	3: 1061		

Modeling

Random Forest

After the data preparation, we decided to select the attributes listed below in our modeling section:

We split the data into 80% training data and 20% testing data and checked the distribution of the attribute “overweight”:

We created our random forest model by using the following code:

mod_rf <- randomForest(overweight ~ ., method = "class", data = train, na.action = na.fail,
nodesize = 1, importance = TRUE, ntree = 200)

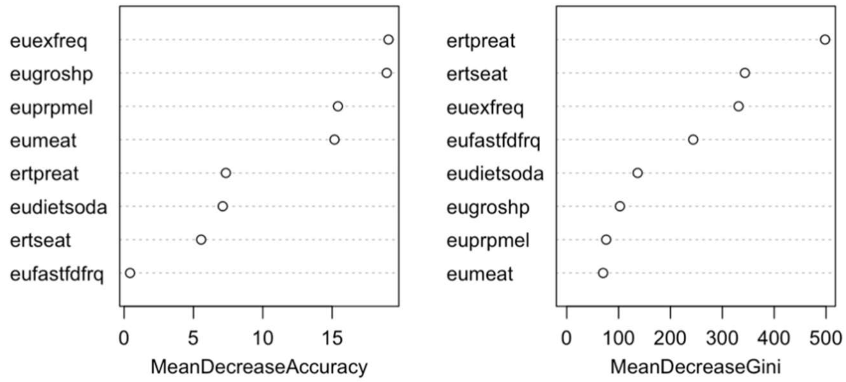
varImpPlot(mod_rf, main="")

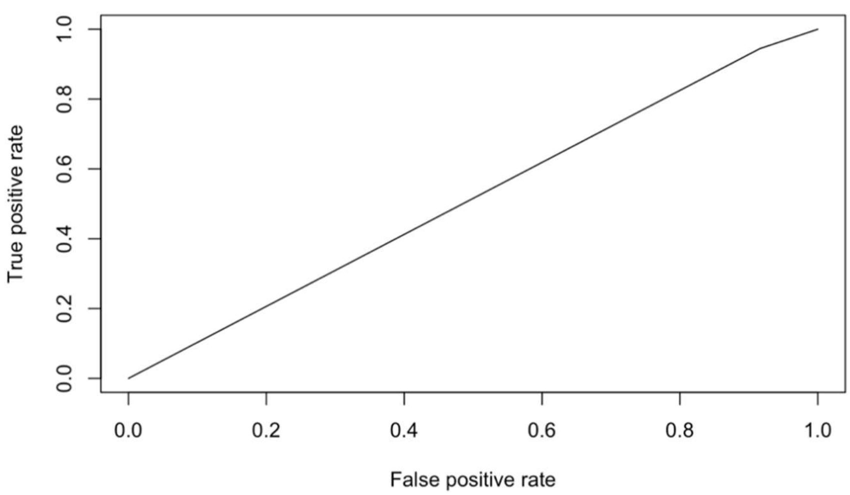
auc(test$overweight, as.numeric(Pred_rf))
pr_rf <- prediction(as.numeric(Pred_rf), as.numeric(test$overweight))
perf_rf <- performance(pr_rf, measure = "tpr", x.measure = "fpr")

plot(perf_rf)

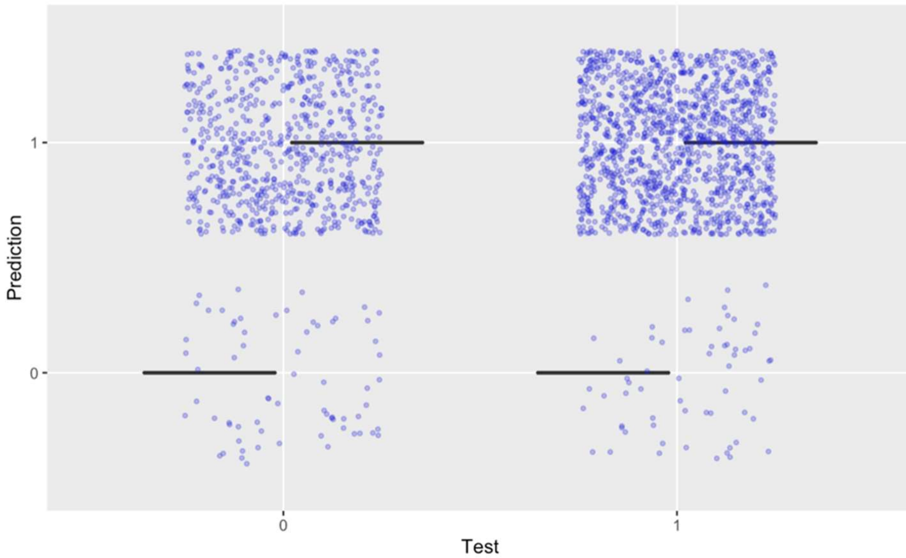
auc(as.numeric(test$overweight), as.numeric(Pred_rf))

Random forest uses bagging (picking a sample of observations rather than all of them) and random subspace method (picking a sample of features rather than all of them, in other words - attribute bagging) to grow a tree. In our case, the model only contains 9 independent variables. We leave the ntree as 200 and the nodesize as 1 which stands for the method of classification. Below is importance of the dataset attributes for the prediction of the “class” attribute. The important attributes are euexfreq, eugroshp, euprpmel, eumeat, ertpreat, eudietsoda and ertseat. The AUC that provided by the random forest model is 0.5264 with an accuracy of 62.3%





Based on the visualization of the confusion matrix for the random forest model (see below), random forest model has the highest accuracy in predicting true positive outcome and the followed by false positive outcome.



Logistic Regression Modeling

Similar to the random forest, we split the data into 80% training data and 20% testing data. Categorical variables (factors) were converted into dummy variables for the logistic regression models. Our first model included all variables included in Table 1; however, after running the full model, two sets of variables were not significantly associated with being overweight: how long respondents spend eating as a secondary activity (i.e. while doing something else), whether the respondent is the primary grocery shopper, and whether respondent was the primary meal preparer in the household. We removed the variables to make the model more parsimonious. The final model also has a better AIC score of 10860 than the full model.

Final Regression Model

For our final model, we used the following code to model the logistic regression, with our dependent variable "overweight" coded as 1 overweight and 0 as not overweight with the independent variables selected above. Results of the calculations and analysis are presented below.

model2 <- glm(overweight ~. -tucaseid -ertseat -notgrocshopr -bothgrocshopr -grocshopr -noprpmeal -splitprpmeal -prpmeal, family = binomial(link = 'logit'), data = train)

summary(model2)

```
Call:
glm(formula = overweight ~ . - tucaseid - ertseat - notgrocshopr -
   bothgrocshopr - grocshopr - noprpmeal - splitprpmeal - prpmeal,
   family = binomial(link = "logit"), data = train)

Deviance Residuals:
   Min      1Q   Median      3Q      Max
```

```
-1.9145  -1.3507   0.8359   0.9379   1.8230

Coefficients: (2 not defined because of singularities)
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.7514974  0.0567427  13.244  < 2e-16 ***
ertpreat    -0.0014829  0.0004738  -3.130 0.001750 **
euexfreq    -0.0732786  0.0079613  -9.204  < 2e-16 ***
eufastfdfrq  0.0429017  0.0112434   3.816 0.000136 ***
sodanoans    0.2269262  0.0538248   4.216 2.49e-05 ***
dietsoda     0.5773587  0.0851388   6.781 1.19e-11 ***
regsoda      0.0861669  0.0677112   1.273 0.203172
nosoda              NA         NA      NA       NA
meatzero    -0.0587436  0.0525391  -1.118 0.263528
nomeat      -0.4395313  0.0828442  -5.306 1.12e-07 ***
yesmeat             NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 11043  on 8509  degrees of freedom
Residual deviance: 10842  on 8501  degrees of freedom
AIC: 10860

Number of Fisher Scoring iterations: 4
```

Our logistic regression indicates that the more time one spends eating a meal (ertpreat) the less likely the person will be overweight. The more an individual does exercise (euexfreq) frequently over the past 7 days the less likely the individual will be overweight. Each additional time reduces the likelihood of being overweight by 0.07 (logged odds). We also found that the more frequently a person purchases prepared food from a deli, carry-out, delivery food, or fast food (eufastfdfrq) the more likely the person is to be overweight. Relative to those who do not drink soda, those who drink diet soda are significantly more likely to be overweight. Those who did not answer this question (sodanoans) were also significantly more likely to be overweight, reaffirming the earlier decision to keep this group separate, as they are likely different from the rest of the sample. Surprisingly, those who drink regular soda are not significantly different from those who do not drink variable.

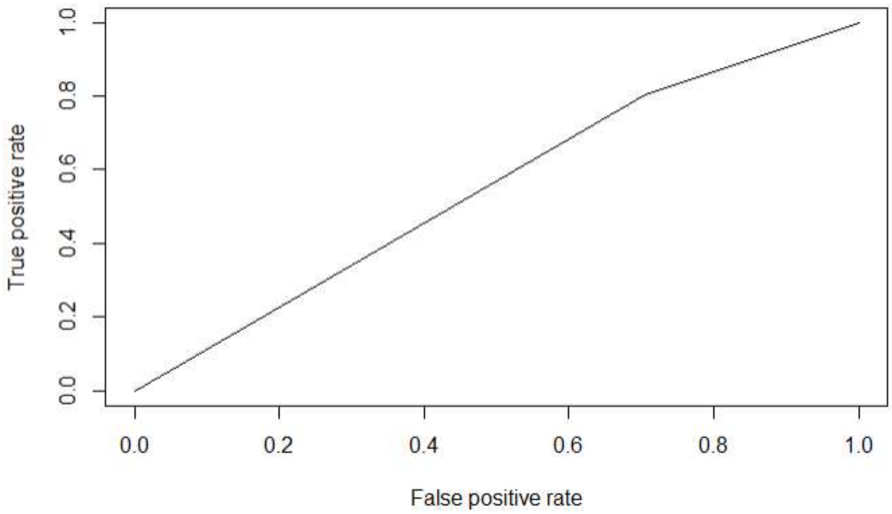
Compared to those who reported cooking with meat (yesmeat) in the past seven days, those who reported not cooking with meat (nomeat) were significantly less likely to be overweight. Those who did not respond to this question (meatzero) were not significantly different from those who reported cooking with meat.

Finally, we ran an F score that gave an accuracy output of 61.5% and an AUC score of 0.5497 while leaving the standard threshold of the Roc Plot to 0.5.

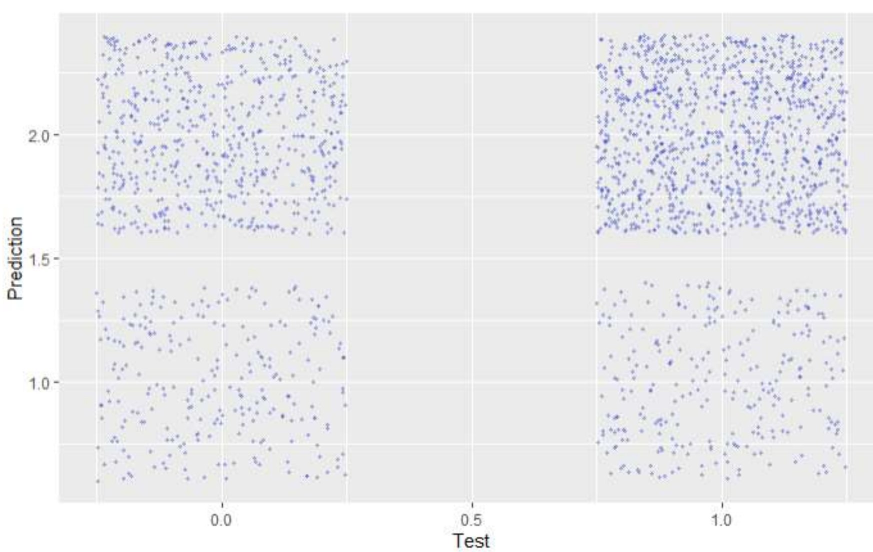
Visualizing the results of the logistic regression

pred_lm <- factor(pred_lm, levels = c("no","yes"), order=TRUE)
f <- table(test$overweight, pred_lm)

```
pred_lm
    no  yes
 0  239  569
 1  259 1060
```



Based on the results shown below, the logistic regression model does not perform as well as our random forest in predicting true positive outcome (accuracy rate of 61.5% compared to 62.3%). As well, our model predicted too many as overweight when they are not. Thus, random forest is a better model or predicting the likelihood of being overweight.



Clustering

To determine whether there are any natural groups in the data set based on eating and health behaviours, we will conduct a series of clustering. The final dataset used for clustering include all columns in Table 1. We excluded the dependent variables (overweight), however, given we want to look purely at the differences in eating and health behaviours, to later see if there are any significant differences among the groups in their likelihood of being overweight. Once the clusters are identified, we will conduct a chi-square test between he cluster groups and overweight to determine if there are any significant differences.

Hierarchical Clustering

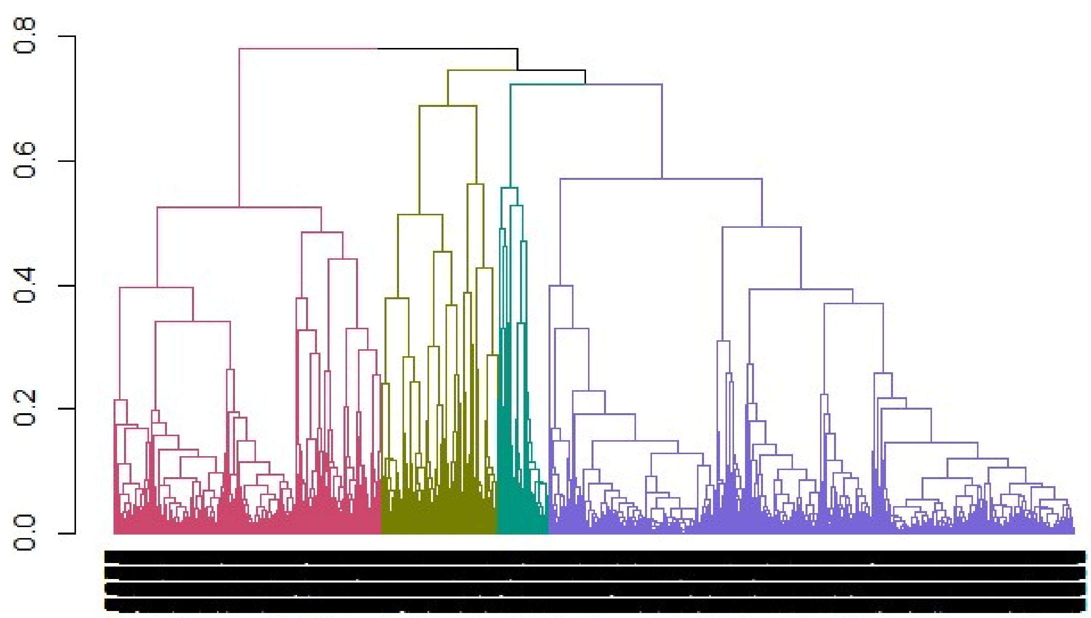
The data we prepared for clustering is a mixture of numerical and categorical variables. In order to calculate a dissimilarity matrix in this case, we utilized Gower distance.

gower.dist <- daisy(data_cluster[,3:10], metric = c("gower"))

We use gower distance to do the hierarchical clustering.

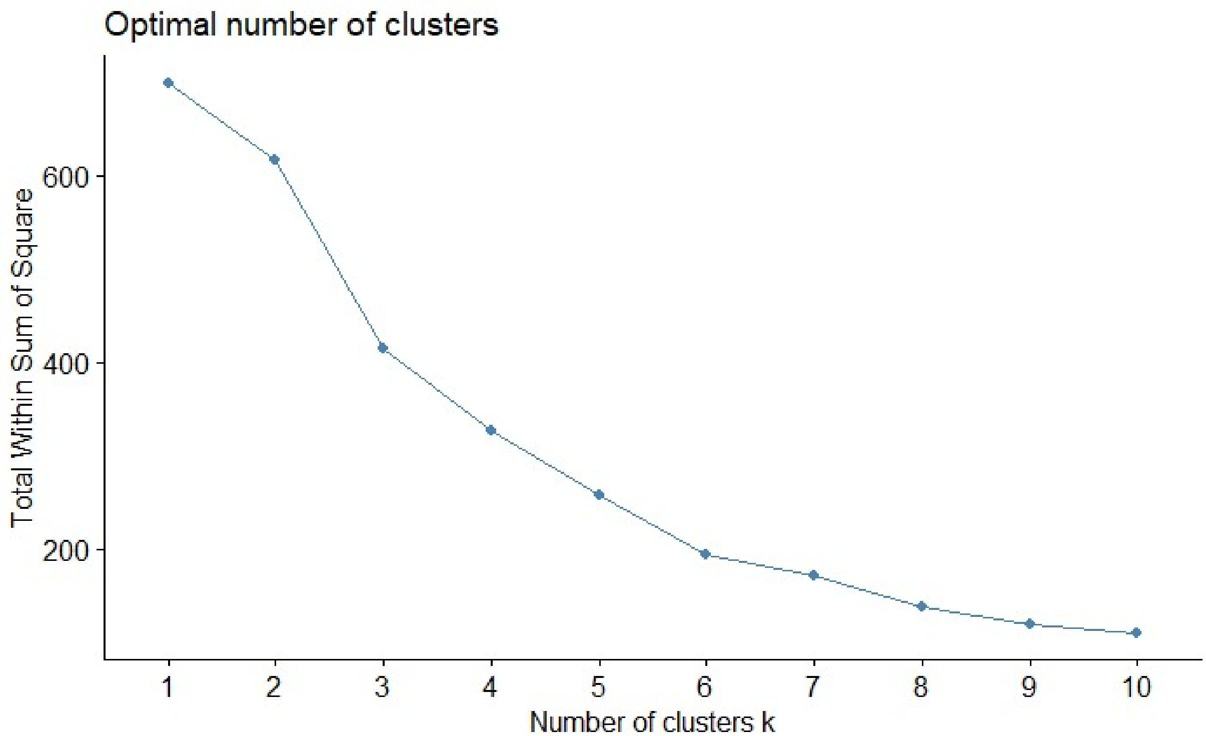
h_clust <- hclust(gower.dist, method = "complete")

After conducting hierarchical clustering, we are able to choose the number of groups we would like to have. Below is the plot of the clustering result when we target to 4 groups.

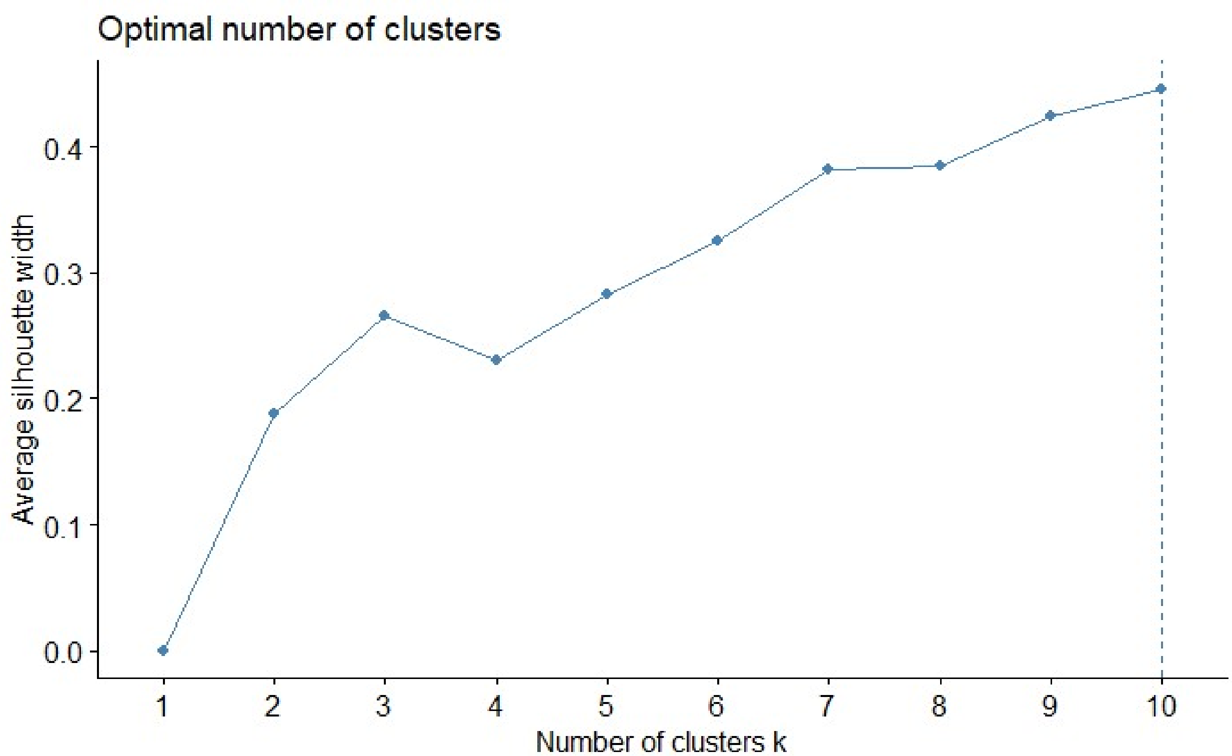


See how disproportional the size of clusters is. One of the reasons is that the dataset can be imbalanced, and some group of observations will outweigh all the rest in the analysis.

So, we've produced the "elbow" graph. It shows how the within sum of squares — as a measure of closeness of observations: the lower it is the closer the observations within the clusters are — changes for the different number of clusters. In the case of a graph below, we see a proper clustering number might be 2 to 6 observations.



Then we use silhouette assessment. We will choose the number that maximizes the silhouette coefficient because it means clusters that are distinctive (far) enough to be considered separate.



But in our case, it raises along with the increase of the cluster number. This imply that the more we break the dataset, the more distinctive the clusters become; however, it becomes inefficient and unhelpful to split until individual data points.

To verify whether the clustering make sense, we utilized another clustering method to compare the results.

K-Prototype clustering

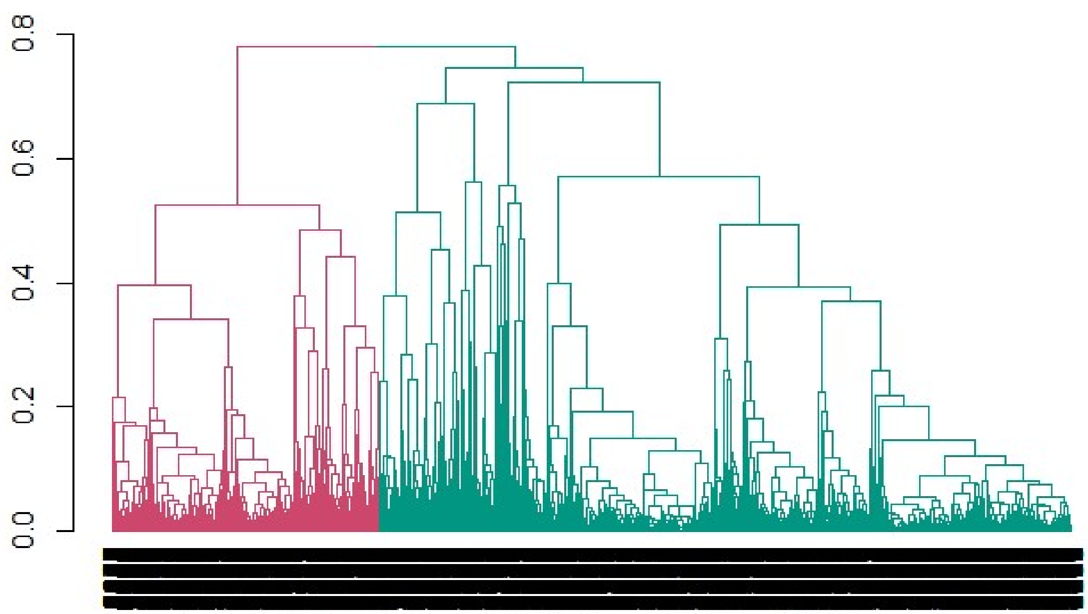
To compare the results with the hierarchical cluster, we utilized the K-Prototype clustering method, which is a Here comes the K-Prototype. It is a simple combination of K-Means and K-Modes in clustering mixed attributes.

kpres <- kproto(data_cluster[,3:10], 2, iter.max=1000, nstart=10, na.rm=TRUE, keep.data=TRUE, verbose=TRUE)

Just as K-means does, K-prototype is randomized in its starting centers. As such, different initial conditions will result with different clustering result. If, however, there are clearly distinguished groups, with multi nstart settings, we will able to get a convergent result.

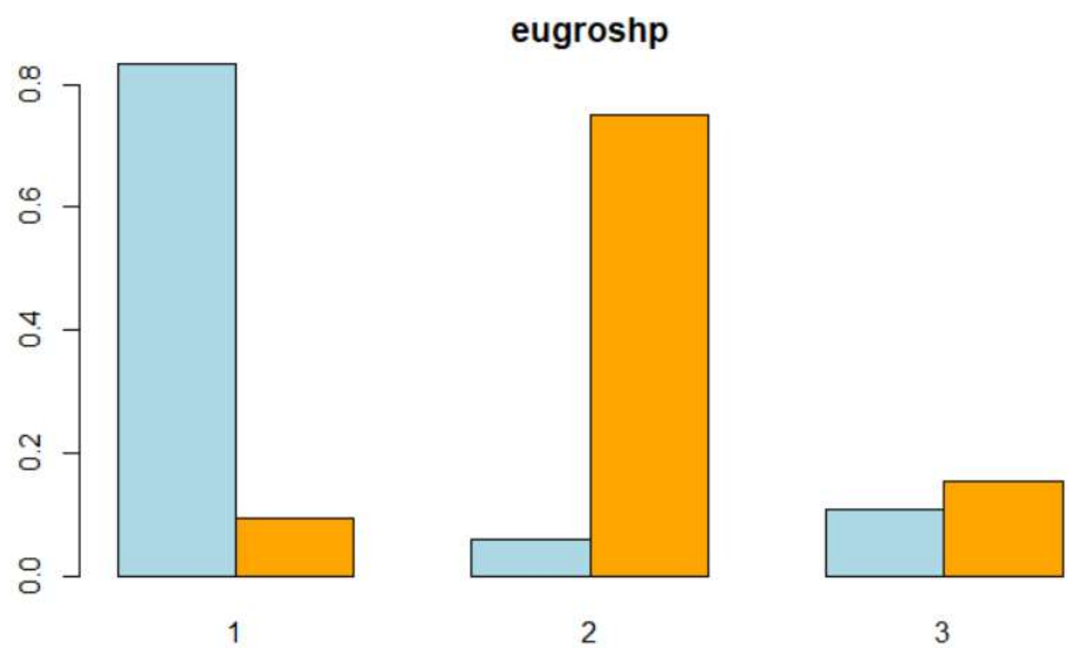
Clustering with 2 to 6 different groups resulted in a convergent result compared to the Hierarchical Clustering method if both of them cluster data to 2 groups. The correlation coefficient was 0.9103548, which can be considered very highly correlated.

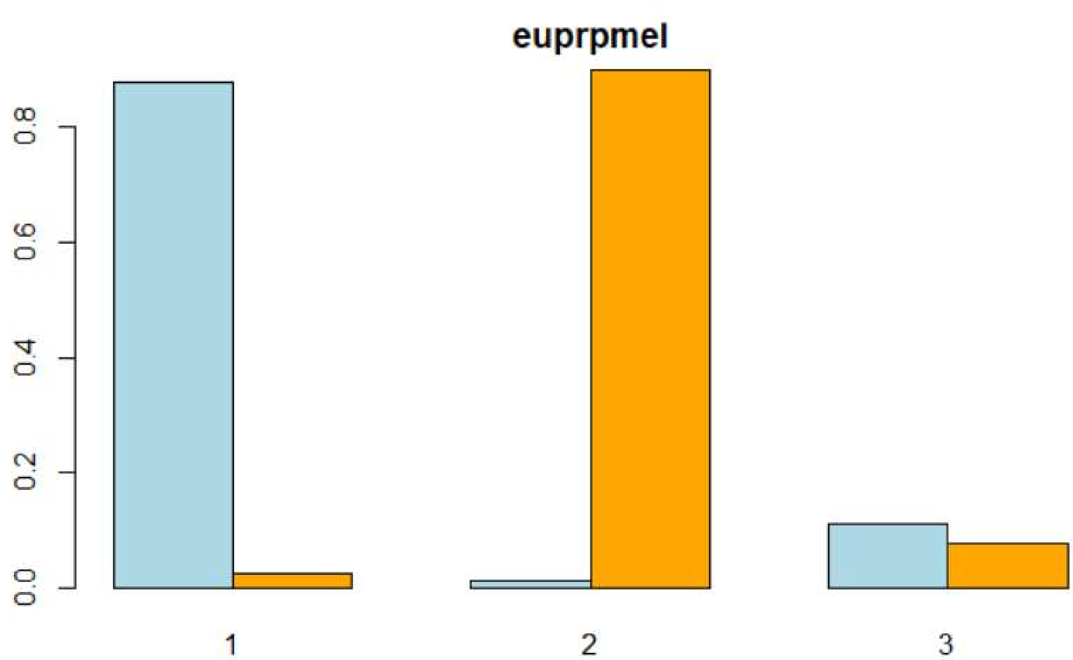
As a result of these findings, we decided to divide the cases in the dataset to 2 major groups. Below shows the separation in hierarchical clustering result.



Visualize the distribution of the variables with the clustering result

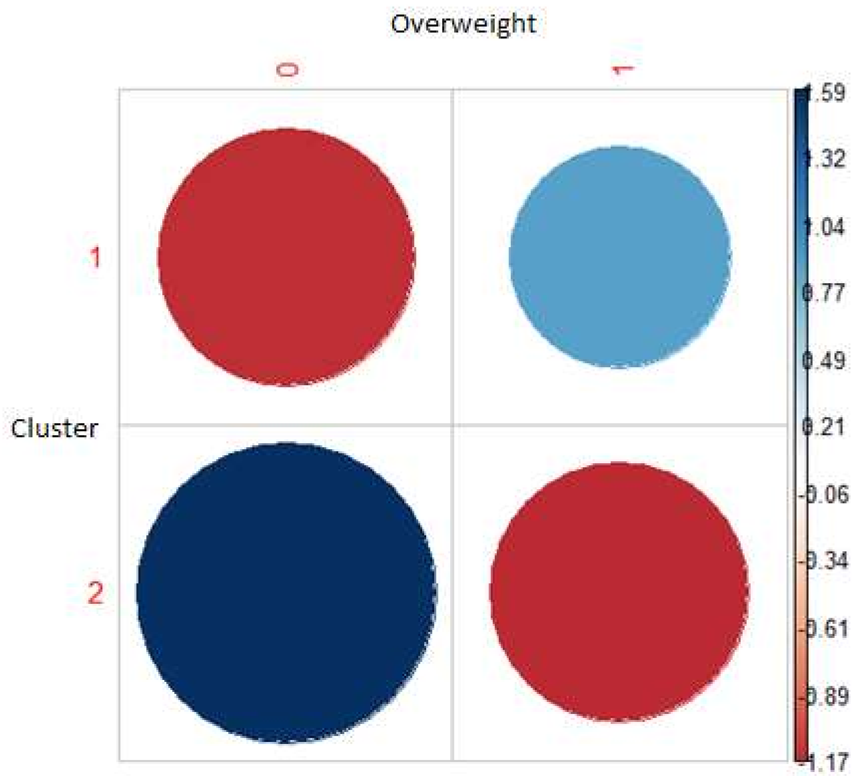
We plotted graphs to see the variables' status according to the clustering group. For example, from the two pictures below, we can see that group1 are those who less likely to do grocery shopping and prepare meal by themselves. Some variables do not present a clear split among the groups, however, such as the time individuals spend eating (both primary and secondary).





When comparing the likelihood of being overweight for the two groups, we found that group 1 was significantly more likely to be overweight than group 2 (X-squared = 5.954, df = 1, p-value =

0.01468). The graph below shows the visualization of the relationship between the clusters and being overweight.



Clustering Summary

Through exploring the dataset and using different methods to cluster, we split the dataset into two groups with different eating related behaviors. Some attributes show a clear difference between these two groups of people, while some attributes not. Most importantly, we found that Group 1 was significantly more likely to be overweight than Group 2.

Conclusion

Through exploring the eating and health module dataset, we developed an algorithm to predict how eating patterns are associated with being overweight. We predicted the likelihood of being overweight using both Random Forest and Logistic Regressions. The Random Forest model provided greater accuracy than the logistic regression mode (62.3% v.s. 61.5% respectively). More importantly, this model indicated some attributes have more impacts on the possibility of being overweight. For example, the more time one spends eating a meal the less likely the person will be overweight, and the more frequently one exercises, the less likely they are to be overweight. Finally, we adopted two types of clustering techniques (hierarchical and K-Prototype) to determine whether there are any natural groupings among individuals with different eating habits. Our key findings in this clustering model include people with less exercise and eat more fast food tend to be overweight, and even people with high frequency of exercise and rarely have fast food, they tend to be overweight as well. This may sound contradictory, but it provided some insight of being overweight.

With our findings, we can present to the health and wellness organization the types of eating patterns that are associated with being overweight. For instance, information they send can encourage people to slow down while they eat, consume less soda and take-out, and engage in physical activity more often. Our clustering analysis suggests that the information should be targeted to those who are in Group 1, as they are significantly more likely to be overweight. The next steps would be gathering more data and additional attributes to disclose the deeper causing factors for being overweight.

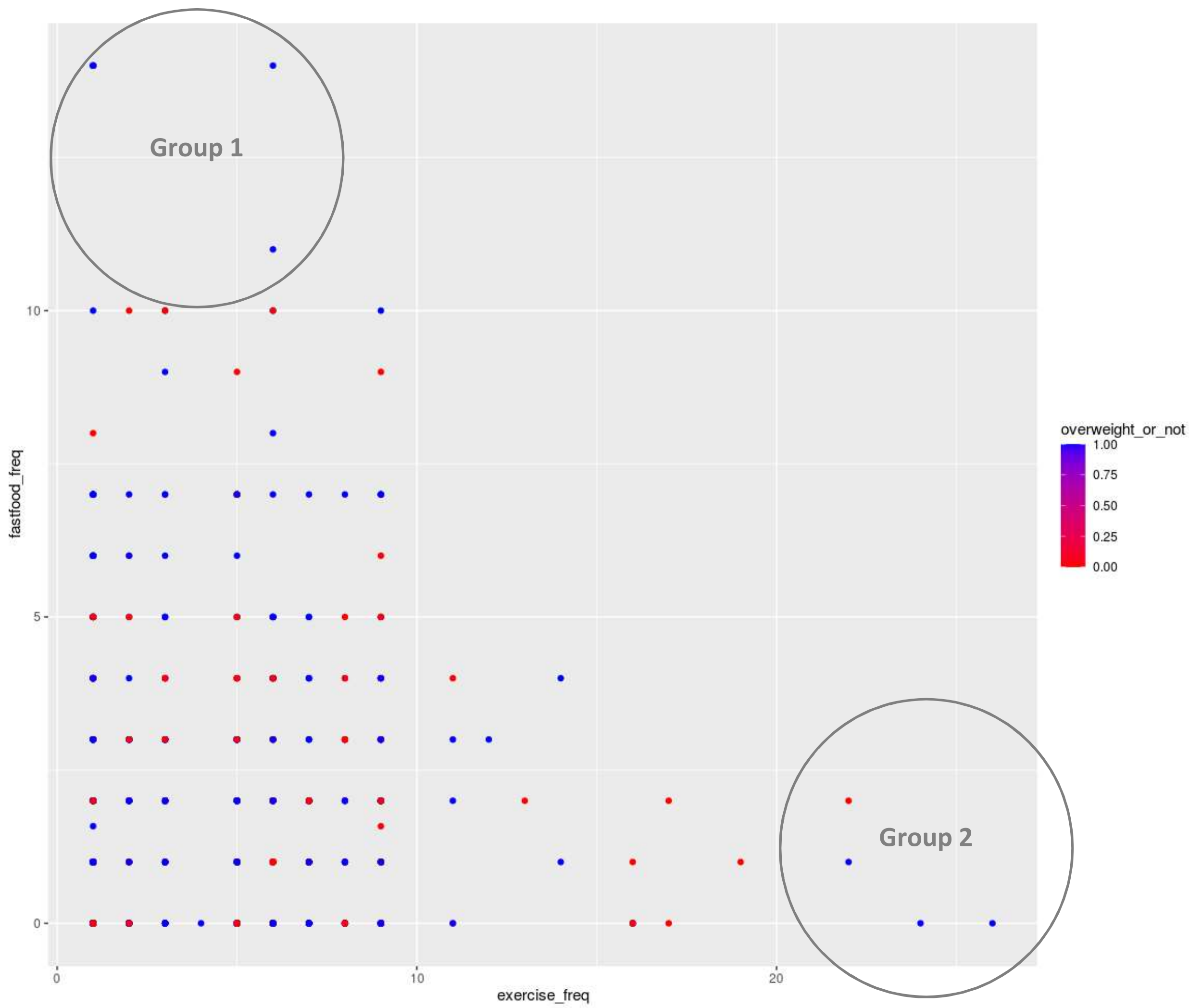
Shinyapp

In order to visualize the data more intuitively, we use shinyapp to deploy our result through the links:

Clustering: https://aaronrstudio.shinyapps.io/webapp/

Regression: https://aaronrstudio.shinyapps.io/CSDA1010_Regression_Model/

For clustering, we can have scatter plot with different variables separated by colour of a third variable. Sample picture shows below: people with less exercise and eat more fast food tend to overweight (1). But people with high exercise frequency and rarely have fast food also tend to overweight (2). One of the reason might be that these are those who passionate about fitness.



For regression, we can see the relationship between our dependent variable (over weight or not) and selected independent variable. And we can also see the significant coefficient of the variables if they are put in to the regression model. The scatter plot is for the selected variable and the target. However, you are not able to see the plot if you select more than one independent variable.

References

Basu, S., McKee, M., Galea, G., Stucker, D. (2013). Relationship of Soft Drink Consumption to Global Overweight, Obesity, and Diabetes: A Cross-National Analysis of 75 Countries. Am J Public Health. Retrieved on May 7,2020. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3828681/

Bently, A., P. Ormerod, D. Ruck (2018). Recent origin and evolution of obesity-income correlation across the United States. *Palgrave Communications,* 4.

Elrod, J., Fortenberry, J (2018). Target Marketing In The Health Services Industry: The Value of Journeying Off The Beaten Path. BMC Health Serv Res, 923. Retrieved on May 7,2020. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6293495/

Martin, L (2019). Medical Causes of Obesity. Retrieved on May 7,2020. Available at: https://www.webmd.com/diet/obesity/medical-reasons-obesity

Laliberte, M (2019). What Really Happens to Your Body When You East Fast. Retrieved on May 7,2020. Available at: https://www.thehealthy.com/weight-loss/eating-fast-weight-gain/

U.S. Department of Health & Human Services (n.d.). “Calculate your Body Mass Index”. Retrieved on May 5, 2020. Available at https://www.nhlbi.nih.gov/health/educational/lose_wt/BMI/bmicalc.htm