

SRG

Data Analytics

REPORT

Predictive analytics with SAS

Spring 2020

Group 16

Savrin Darash

Karan Bhujbal

Jinx Mei

Rohanjit Singh Virdi

Prarthna Vasudevamurthy

# Introduction

SRG is a marketing consulting group that is active in designing advertising and marketing strategy, especially, food and grocery clients. Their services predominantly include; Brand design, Packaging design, Brand positioning, Culinary trends and mapping, Environmental designs Brand strategy, Consumer insight.

The dataset that we will be using in this project is result of a quantitative research study SRG conducted in 2015 to better understand health and wellness. Through the use of 30 questions about lifestyles, goals, and motivations, they scored a total of 1003 people. Respondents provided personal health and wellness information as well as behaviors associated with activity, shopping behaviors, and eating habits.

In this project we will be analyzing the data from a survey on health and wellness in order to provide Business insight for the company to advance their business and profits. In this project we also aim to find what influences Health & Wellness and how to predict someone’s state of health and wellness.

First we will define health and wellness and locate the pertinent answers to the Questions from the data survey and the corresponding variables in the dataset.

Definition of Health and Wellness

In 1946, the World Health Organization (WHO) defined health as “a state of complete *physical, mental, and social well-being*, and not merely the absence of disease or infirmity.”

Wellness

The Global Wellness Institute defines wellness as the *active pursuit of activities, choices and lifestyles that lead to a state of holistic health.*

In the scope of this project health will reflect the state of physical health and wellness will refer to mental state of health.

# Dataset

The dataset has 1003 rows and 1151 columns. Each column in the dataset corresponds to a Question or answer part from the survey. There are a lot of columns that are missing or completely blank. After data cleaning and transformation, we reduced the number of columns to 372 which only a part of that data was used in prediction (the relevant features(covariates)). Also after dropping outliers of the data we have reduced the number of the rows to 982.

The dataset data has been divided by these categories:

* Demographic Data: (D&S questions) Data such as age, race, weight, height, married/unmarried, living with children
* Wellness scoring: (Questions 1-6) Data such as state of life, mood, attitude toward current life, weight goals and frequency of exercising.
* Food choice and frequency of their consumption
* Stores
* Dinner choices
* Satisfaction with dinner
* Snack options
* Lifestyle

# Exploratory data analysis

1. Demographic Data

* *Marital Status*

A picture containing room, drawing

Description automatically generated

As we can see, most people are married (69.9%), only a small part of people are not married(31.1%).

* *Gender*

*A close up of a logo

Description automatically generated*

As we can see, the gender ratio is almost half and half.

* *Age*

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

As we can see, this survey covers all age from 18 to 65 and people’s age is pretty much equally distributed in this case which means every age group has enough number of people in it.

* *Education background*

*A close up of a logo

Description automatically generated*

As we can see that people with a high school degree is the most followed by people with Bachelor degree and people graduated from high school. These 3 parts together is the majority which takes more than 80% of the total population.

# *Household income*

# A screenshot of a cell phone Description automatically generated

As we can see that people’s household incomes are mostly fell on the “100-200K” group, followed by “15-25k” and 75-100k” groups. Interestingly, people’s household income are pretty much equally distributed compared to other variables in this case. Each income groups no matter it’s high income, median income or low income all have enough number of people in it, which is very different from what we expected.

* *Ethnicity*

*A screenshot of a cell phone

Description automatically generated*

As we can see that most people are white which is the majority, followed by black American, Hispanic, Asian and American Indian. The result is pretty consistent with the last US demographic census data.

* *BMI*

*A screenshot of a cell phone

Description automatically generated*

Body mass index which called BMI is a convenient [rule of thumb](https://en.wikipedia.org/wiki/Rule_of_thumb) used to broadly categorize a person as underweight, normal weight, overweight, or obese based on tissue mass ([muscle](https://en.wikipedia.org/wiki/Muscle), fat, and bone) and height. Commonly accepted BMI ranges which we used in this case are underweight (under 18.5 kg/m2), healthy (18.5 to 25), overweight (25 to 30), and obese (over 30).

From this picture we can see that obese people are the majority followed by healthy people, overweight people and underweight people. The result is telling us that obese is truly a common issue in the US nowadays which is worthy of notes and people should take some action to control their weights.

1. Wellness scoring

* *POINTS*

“Points” is the total point each people gets by answering question from 1 to 6. We used the scoring system from survey to make a judgement on people’s wellness. The scoring system is in below:

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generatedA picture containing device

Description automatically generated

From these 2 picture above we can see most people are sacrificing their health, followed by people who are aspiring to health, scrambling to health and neglected health. These 4 parts become the majority which takes more than 90% of total population.

1. Food choice and frequency of their consumption

* *Item that people consume a lot*

*Several times a week Never Several times a month Once a year or less*

*Every day Several times a week Once a month*

A picture containing drawing

Description automatically generated

From these bar charts we can see that most of the items that people consume a lot are healthy except for **Sugar** and **Packaged cheese**.

* *Item that people hardly consume*

A picture containing object, comb

Description automatically generated

It seems like there is no specific pattern between healthy/unhealthy and all those items. Some of them is because they are rare by itself, some of them is because people have other more popular alternatives like milk for soy milk.

1. Stores

* *Frequency of stores that people went to*

*Several times a week Never Several times a month Once a year or less*

*Every day Several times a week Once a month*

*A screenshot of a cell phone

Description automatically generated*

From these bar charts we can see that people went to pharmacy and convenient store more often than other type of stores. Few people went to a dollar store like *Dollar Tree* or a club store like *Costco.*

# Insights

1. **Most people don’t get enough sleep every night**

***Opinion on Most nights I don't get enough sleep***

A picture containing device

Description automatically generated

**25.5%** Strongly agree

A screenshot of a cell phone

Description automatically generated

**36.0%** Somewhat agree

Lack of sleep also called [Insomnia](https://www.webmd.com/sleep-disorders/insomnia-symptoms-and-causes) is a kind of sleep disorder in which you have trouble falling and/or staying asleep. The condition can be short-term (acute) or can last a long time (chronic). It may also come and go. Insomnia affects women more than men and older people more than younger ones. Young and middle-age African Americans also have a higher risk. Other risky factors include long-term illness, mental-health illness and working night shifts. From these 2 picture above we can see that more than 60% people believe they don’t get enough sleep every night which is telling us Insomnia is a common health issue that hurting people.

1. **Most people are obese or overweight**

Almost **half** of people are obese

*A screenshot of a cell phone

Description automatically generated*

Truly healthy(fit) people is less than **1/3**

Obesity has reached epidemic proportions in the United States. Over two-thirds of adults are overweight or obese, and one in three Americans is obese. The prevalence of obesity in children has increased markedly. Obesity has also been increasing rapidly throughout the world, and the incidence of obesity nearly doubled from 1991 to 1998. In 2015, nearly 40% of adults were obese in the U.S.

Body mass index which called BMI is a convenient [rule of thumb](https://en.wikipedia.org/wiki/Rule_of_thumb) used to broadly categorize a person as underweight, normal weight, overweight, or obese based on tissue mass ([muscle](https://en.wikipedia.org/wiki/Muscle), fat, and bone) and height. Commonly accepted BMI ranges which we used in this case are underweight (under 18.5 kg/m2), healthy (18.5 to 25), overweight (25 to 30), and obese (over 30).

From this picture we can see people who are truly healthy(fit) is less than 1/3 of the total population, almost half of people are obese. Which means obesity is a common and serious health issue today and people really should take some actions to counter this for the purpose of avoiding potentially diseases.

1. **A lot of people wrongly believe they are physically fit at their age.**

***Opinion on I think I am physically fit at my age***

***A picture containing device

Description automatically generated A screenshot of a cell phone

Description automatically generated***

From this 2 pictures we can see more than half of people believe they are physically fit at their age(Strongly Agree and Agree Some what), only 12.7% people strongly disagree they are physically fit. The interesting thing is from the BMI chart we are sure that most people are either overweight, obese or underweight, only less than 1/3 people are truly healthy. But in this case, it seems like people are not quite aware of the truth they are not fit at all.

**4. People do care about their health**

***Opinion on health is not my top priority***

*A close up of a logo

Description automatically generated*

**6/10** people think health is their **No.1** Priority

***Opinion on I do everything I should to take care of my health***

A picture containing device

Description automatically generated

**6/10** people are doing whatever they can do to keep health

***Opinion on I wish I had more willpower to make change on my health***

***A picture containing device

Description automatically generated***

**6/10** people wish they had more willpower to make change on health

**5. Health food is more helpful at lowering BMI**

**A close up of a map

Description automatically generated**

Correlation coefficient:

**-0.11**

**A close up of a piece of paper

Description automatically generated**

Correlation coefficient:

**-0.05**

We divided the food that people chose for question 7 into health food and unhealthy food. We did regression analysis between “healthy food”/ “unhealthy food” with “BMI” and we found out both healthy and unhealthy food has a negative relationship with BMI. But the healthy food has a stronger negative effect on people’s BMI compared to unhealthy food which means if people take more health or unhealthy food at the same time, if they choose to take healthy food their BMI would be lower compared to if they choose to take same amount unhealthy food.

**6. People don’t spend much time on dinner**

* *Time spent on preparation*

**A close up of a logo

Description automatically generated**

Almost **80%** people prepare their dinner less than **30 min**

* *Time spent on cooking*

**A picture containing device

Description automatically generated**

Almost **70%** people cook their dinner less than **30 min**

# Feature Engineering

# Machine learning and interpretations

# Relation, Question and feature importance

# Clustering and segmentation

# Conclusion and recommendation

# Appendix/Appendices