Data-Driven Analytics for Obesity Management and Business Strategy

Author: Zhetan Zhang, Mingqi Yang

Team: Team ZY

1. Abstract

Obesity has become a serious global health issue and affects social economics to a certain degree. In this study, we will use the data that describe eating habits, physical activity, and lifestyle choices among individuals in Colombia, Peru, and Mexico to investigate the relationship between obesity and these factors. By analyzing the comprehensive dataset with these variables, we aim to provide some valuable insights into the prevalence of obesity in these countries and help to implement business action in the future. Our research performs a multifaceted approach, integrating various analytical methods to discern significant patterns and correlations within the data. We examine the impact of dietary choices, physical activity levels, and smoking habits on obesity prevalence. Furthermore, this study aims to go beyond identifying obesity-related factors to propose actionable strategies for obesity prevention or decreasing the obesity rate. We investigate effective interventions through analyzing the data set and come up with recommendations not only for obese people's lifestyle but also for companies(e.g. medical, transportation) to help obese people to change their lifestyle.

2. Project Description

2.1 Problem Definition

Obesity is a serious global health issue and affects social economics. According to World Obesity Atlas 2023, there are 14% men and 18% women in the world who have obesity for 2020 and this number will increase continuously in the future to 23% and 27% in 2035. Obesity not only affects individual health, but also affects the world economy. Our motivation is to contribute to addressing this escalating health crisis by conducting a comprehensive investigation into the factors contributing to obesity and proposing effective strategies for prevention.

Through our research, individuals may change their travel mode, eating habits, and lifestyle. Our findings can also guide some companies, like medical and transportation companies, to implement or improve some measures to help obese people. In addition, effective obesity prevention can reduce the burden on healthcare systems, resulting in cost savings and enhanced societal well-being. In addition, healthier individuals are likely to be more productive, benefiting both individuals and businesses.

This project is indeed a valuable chance for us to perform the whole data analysis cycle of how business decisions are made. The obesity dataset is a very comprehensive dataset that contains a wide range of data types, implying that we have the opportunity to enhance our skills in data preprocessing to adapt them for our specific needs. Besides, it will also be a great experience for us to learn how to optimize complex machine learning models by selecting appropriate features and parameters. With regard to these, the project itself will be beneficial for our future career development in data science.

2.2 Method

The dataset we are using for this project is called "Estimation of obesity levels based on eating habits and physical condition, " sourced from the UCI Machine Learning Repository. The dataset essentially serves as classification tasks containing 16 features and 2111 samples, with the goal of predicting obesity levels in different individuals. Among 16 features, 6 of them contain nominal data (e.g. Gender, Transportation Used), 2 of them contain ordinal data (e.g. Consumption of food between meals), 1 of them contains interval data, which is Age, and 7 of them contain ratio data. (e.g. Height, Weight) With such an abundant set of variables, we can generate a wide-ranging view about how various factors collectively contribute to obesity. For example, by analyzing the features related to individuals' eating habits and their corresponding obesity levels, we can offer guidance to nutrition companies on developing more effective slimming products that are not only appealing to obese population but also effective in reducing weight. Besides, examining physical condition features also provide insights on how to suggest infrastructure companies such as public transit services or medical rehabilitation centers on optimizing convenience or construct appropriate exercise regimens for obese populations. Therefore, the rich set of factors and substantial number of samples help us generate a clear graph of the future business actions. These could be designed for either pushing forward obesity treatment or improving life quality for people who experience obesity.

For the descriptive analytics part of our framework, we will employ statistical methods such as mean and median to understand the basic data distribution and relationships within the features to comprehend what is the past trends of the obesity level. For example, we may calculate the average obesity levels for people with different ages to understand which age group currently has the highest rate of obesity. For the predictive analytics part, we plan to use machine learning models such as SVM or random forest tree to predict future trends and outcomes. For instance, we may use logistic regression to estimate the likelihood of obesity occurrence across eating habits features and use F1 score, Recall and Precision to evaluate the accuracy of our prediction. For the prescriptive analytics part, we try to offer specific recommendations for companies based on the results of the predictive models and data summary. Specifically, we will optimize our models by selecting the best combination of exercise regimens and nutrition products based on the prediction of obesity levels with eating habits. Additionally, we'll determine how well our decision works with respect to the key factors that lead to obesity.

3. Bibliography

1. Estimation of obesity levels based on eating habits and physical condition . (2019). UCI Machine Learning Repository. https://doi.org/10.24432/C5H31Z.
2. Fabio Mendoza Palechor, Alexis de la Hoz Manotas, Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico, https://doi.org/10.1016/j.dib.2019.104344.(https://www.sciencedirect.com/science/article/pii/S2352340919306985)
3. World Obesity Federation, World Obesity Atlas 2023. https://data.worldobesity.org/publications/?cat=19