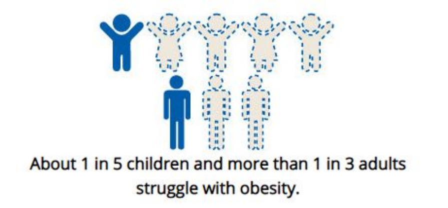
Analyzing Risk Factors Associated with Obesity/Overweight Using Machine Learning

DATA 606 - Capstone in Data Science

UMBC
Group Members:
Sandra Pinto, Shruthi Boban, Siyu Ma

Introduction

- ❖ Our project aims to analyze the relationship between obesity/overweight and different risk factors such as BMI, Race, Gender, physical activities, mental health, education level, and etc.
- ❖ Obesity constitutes a major public health concern in the U.S. and Globally
  - ➢ About 1 in 5 children and more than 1 in 3 adults struggle with obesity in the U.S. (CDC)
  - ➢ Adults with obesity have higher risk for developing Heart disease, Type 2 diabetes, and some types of cancer (CDC)
  - ➢ According to the The “World Health Organization” (WHO), 30% of global death will be caused by lifestyle diseases by 2030.
- ❖ There is a limited number of studies using machine learning to analyze obesity/overweight related datasets in the U.S.



Research Questions/Approach

Research Questions:

- ❖ Which variables are risk factors related to obesity/overweight?
- ❖ What are the correlations between different risk factors and BMI?
    - ➢ Is mental health an important factor that correlates with obesity/overweight?
- ❖ Which machine learning model can classify and conduct regression of the dataset?

Approach:

- ❖ Conduct EDA to find the relationship of different factors and produce visualizations
- ❖ Find the most accurate model for our dataset.
    - ➢ Classification and Regression models (e.g Random Forest, Support Vector Machines (SVM), Logistic Regression, and Decision Trees)

Literature/Industry Research Review

- The Technology and Health Departments of the University of Agder (Norway) identified potential risk factors associated with obesity/overweight using machine learning methods such as Support Vector Machines (SVM), Decision Trees, and Logistic regression models. (Chatterjee et al, 2021)
- The University of Bologna (Italy) used ML techniques to test for the predictive effects of emotional and affective variables over BMI values. (Delnevo et al, 2021)
- The Daffodil International University in Dhaka (Bangladesh) applied 9 prominent ML algorithms to predict the risk of obesity on the data collected from many varieties of people of different ages suffering from obesity and non-obesity. (Ferdowsy F. et al, 2021)

The Prediction of Body Mass Index from Negative Affectivity through Machine Learning: A Confirmatory Study - The University of Bologna (Italy)

Approach:

- Both classification and Regression models including K-Nearest neighbor, Classification and Regression Tree, support Vector Machine, Multi-Layer Perceptron, Ada boosting with decision tree, Gradient Boosting, Random Forest, LASSO (Least Absolute Shrinkage and Selection Operator), and Elastic Regression.

Findings:

- Using affect-related variables it is possible to predict the BMI with a good level of accuracy and the psychological variable that had most impact on the predictive capabilities of the algorithms is Depression
- The best performance was achieved by the LASSO and Elastic Net, with a MAE (mean absolute error) equal to 4.35 and the PCCs (Pearson correlation coefficient) respectively of 0.81 and 0.80, indicating a strong correlation between predictions and real value.

Limitations:

- Restricted number of subjects
- It did not employ newly collected data, thus making inferences limited
- Lack of other factors such as lifestyle habits

Dataset

Source:

CDC - National Center for Health Statistics. National Health and Nutrition Examination Survey March 2017 to 2020 Pre-pandemic

Data type (numerical & categorical):

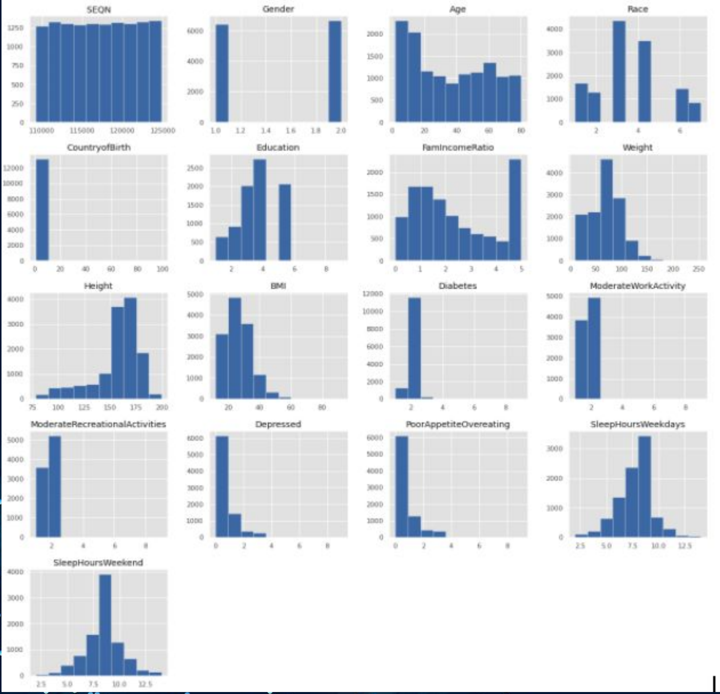
Demographics data, examination data, laboratory data, & questionnaire data including; Respondent sequence number, Gender, Race, Country of birth, Education level, Ratio of family income to poverty, Body measures(Weight, Height, BMI, BMI category), Diabetes status, Physical activity(Moderate work activity, recreational activity), Mental health(Depressed, Poor appetite or overeating), and Sleep disorders(Sleep hours on weekdays and weekends).

Dataset size:

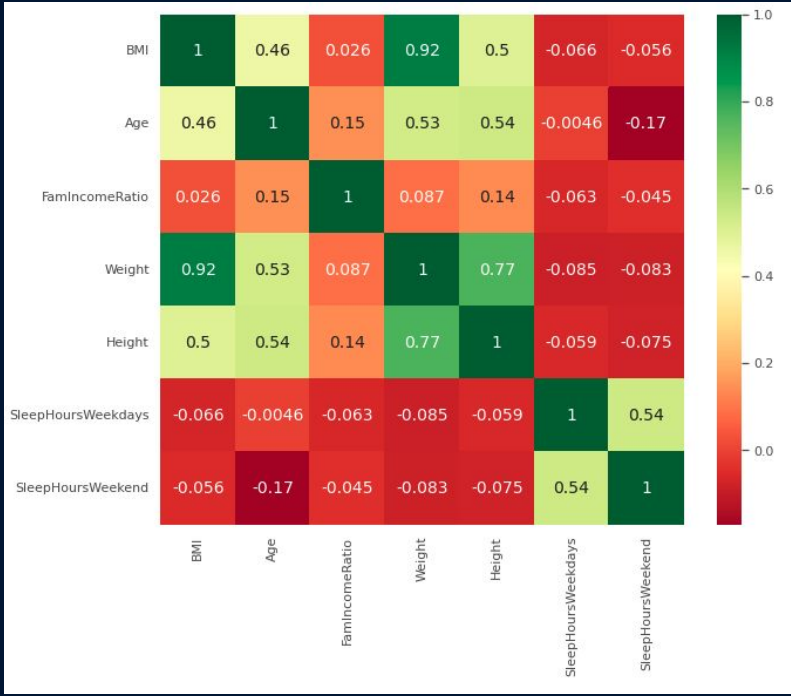
12.4MB - XPT. files / Zipped data file:1.4 MB

	SEQN	Gender	RIDAGEYR	Race	CountryofBirth	Education	FamIncomeRatio	Weight	BMI	BMICategory	Diabetes	ModerateWorkActivity	ModerateRecreationalActivities	Depressed	PoorAppetiteOvereating	SleepHoursWeekdays	SleepHoursWeekend
0	109263.0	1.0	2.0	6.0	1.0	4.0	4.66	65.42638	26.656847	2.0	2.0	2.0	2.0	5.397605e-79	5.397605e-79	7.64092	8.361768
1	109264.0	2.0	13.0	1.0	1.0	4.0	0.83	42.20000	17.600000	2.0	2.0	2.0	2.0	5.397605e-79	5.397605e-79	7.64092	8.361768
2	109265.0	1.0	2.0	3.0	1.0	4.0	3.06	12.00000	15.000000	2.0	2.0	2.0	2.0	5.397605e-79	5.397605e-79	7.64092	8.361768
3	109266.0	2.0	29.0	6.0	2.0	5.0	5.00	97.10000	37.800000	2.0	2.0	2.0	1.0	5.397605e-79	5.397605e-79	7.50000	8.000000
4	109267.0	2.0	21.0	2.0	2.0	4.0	5.00	65.42638	26.656847	2.0	2.0	2.0	2.0	5.397605e-79	5.397605e-79	8.00000	8.000000

Preliminary EDA & Visualizations



Basic histogram shows data distribution and frequency counts.



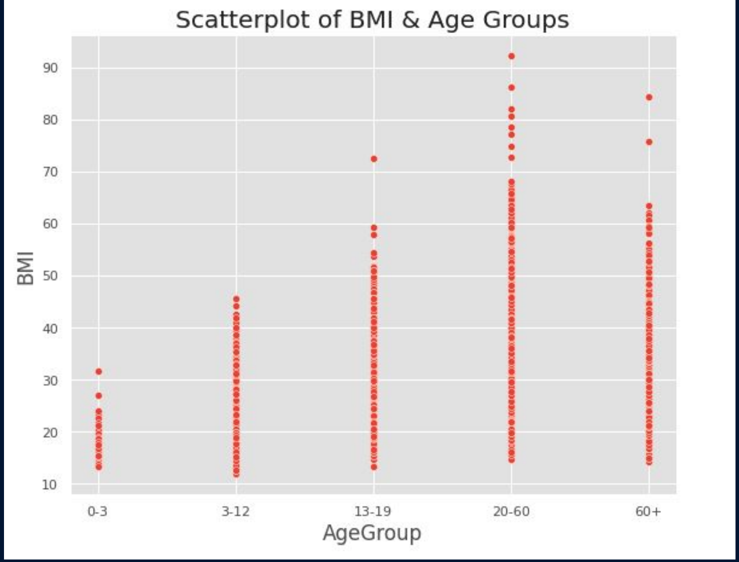
Headmap of numerical variables shows the correlation between different risk factors.

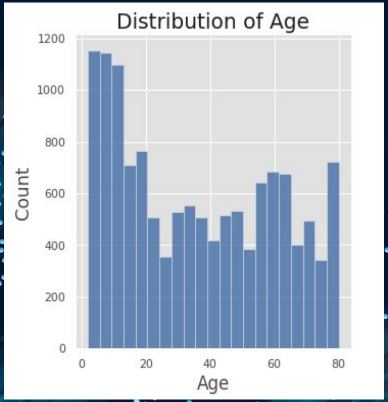
Preliminary EDA & Visualizations

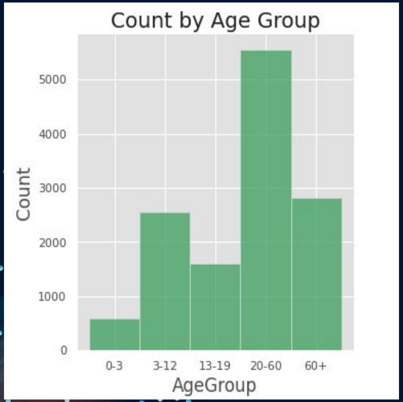
Dataset contains infants, children, teenagers, adults, and seniors.

- Added a new column and separated respondents to different age groups.
- Age groups: 0-3, 3-12, 13-19, 20-60, 60+

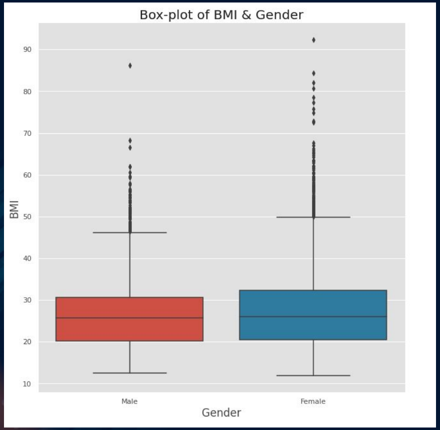
BMI	Weight Status
Below 18.5	Underweight
18.5 – 24.9	Healthy Weight
25.0 – 29.9	Overweight
30.0 and Above	Obesity

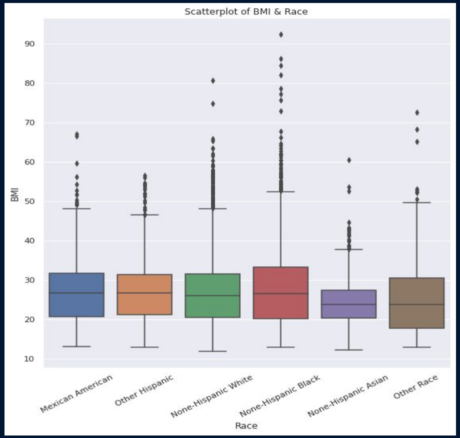


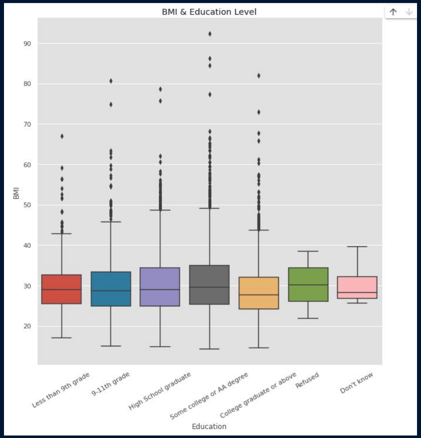


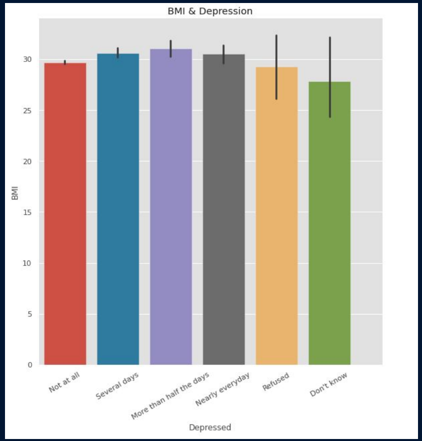


Preliminary EDA & Visualizations









1. Female respondents have slightly higher BMI than Male respondents.
2. None-Hispanic Asians have lower BMI compared to other race groups in our dataset.
3. People with college or above education level tend to have lower BMI.
4. Respondents with depression level as “More than half the days” have higher BMI.

References

Chatterjee A, Gerdes MW, Martinez SG. Identification of Risk Factors Associated with Obesity and Overweight—A Machine Learning Overview. Sensors. 2020; 20(9):2734. https://doi.org/10.3390/s20092734

Wilfley, D. E., Hayes, J. F., Balantekin, K. N., Van Buren, D. J., & Epstein, L. H. (2018). Behavioral interventions for obesity in children and adults: Evidence base, novel approaches, and translation into practice. American Psychologist, 73(8), 981–993. https://doi-org.proxy-bc.researchport.umd.edu/10.1037/amp0000293

CDC. (2021, March 1). Why It Matters. Centers for Disease Control and Prevention. https://www.cdc.gov/obesity/about-obesity/why-it-matters.html

Centers for Disease Control and Prevention. (2021, August 27). About adult BMI. Centers for Disease Control and Prevention. Retrieved September 29, 2021, from https://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html.

Ferdowsy, F.,Rahi, K. S. A., Jabiullah, Md. I., Habib, Md. T. (2021, August 5). A Machine Learning approach for obesity risk prediction. Current Research in Behavioral Science, 2, 2021. https://doi.org/10.1016/j.crbeha.2021.100053

Delnevo, G., Mancini, G., Roccetti, M., Salomoni, P., Trombini, E., & Andrei, F. (2021). The Prediction of Body Mass Index from Negative Affectivity through Machine Learning: A Confirmatory Study. Sensors, 21(7). https://doi.org/10.3390/s21072361

Pillai , R., Saravanan, S., & Shyam, D. G. K. (2020, December 8). The BMI and mental Illness NEXUS: A machine learning approach. IEEE Xplore. Retrieved September 29, 2021, from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9277446.