The George Washington University Data Analytics Bootcamp
Analysis of Obesity Risk Factors
Obesity Project – Group 2
Date: April 17, 2024

REPORT ON OBESITY RISK FACTORS:

Introduction: This report presents the findings of an analysis conducted by Group 2 as part of the Obesity Project. The aim was to identify primary risk factors associated with obesity, utilizing diverse datasets containing demographic, dietary, and lifestyle information.

Objective: The objective of this analysis was to explore the association between various factors, including dietary habits, smoking, alcohol consumption, physical activity levels, and obesity risk.

TEAM MEMBERS:

1. Kavita Lakhani
2. Monique Reid
3. Andrea Nimako
4. Lynda Sempele
5. Kevin Ngala

METHODS:

The methodology employed for this analysis encompassed several key steps, including data collection, preprocessing, exploratory data analysis (EDA), and statistical analysis.

Data Collection: Data were gathered from a web-based survey spanning individuals aged 14 to 61 from Mexico, Peru, and Colombia. The dataset consisted of 2111 records and 17 attributes, covering demographic information, dietary habits, physical condition, and independent variables.

1. **Data Preprocessing:** Prior to analysis, the dataset underwent meticulous preprocessing to ensure quality and consistency. This involved handling missing values, removing duplicates, and formatting data. Notably, no missing or duplicate data instances were found.
2. **Exploratory Data Analysis (EDA):** EDA involved a thorough examination to gain insights into distribution, relationships, and patterns. It included data cleaning, exploring variable distributions graphically and numerically, and conducting descriptive statistics.
3. **Statistical Analysis:** Following EDA, statistical analysis explored relationships between variables and potential obesity risk factors. This included correlation analysis, hypothesis testing, and regression modeling.
4. **Project Tasks and Requirements**: Integrated into the methodology, project tasks encompassed data collection, preprocessing, Pandas-based cleaning, Jupyter Notebooks documentation, Matplotlib-based visualization creation, and comprehensive write-up summarization. Each task was executed meticulously to maintain analysis integrity and rigor.

DATA ANALYSIS:

Following data preprocessing and exploratory analysis, the dataset underwent rigorous examination to discern correlations between variables and obesity risk.

1. **Correlation Analysis:** Assessment of relationships between factors like dietary habits, smoking, alcohol consumption, physical activity, and BMI.
2. **Hypothesis Testing:** Evaluation of statistical significance concerning the association between independent variables (e.g., dietary habits, smoking) and obesity.
3. **Regression Modeling:** Utilization of regression techniques to understand the predictive influence of various factors on obesity risk.
4. **Visualization:** Generation of informative visual aids, such as charts and graphs, to effectively communicate key findings and insights derived from the analysis.

This comprehensive data analysis approach unveiled significant associations, offering valuable insights for informing preventive strategies and health policies targeting obesity mitigation.

RESULTS:

1. **Gender Distribution:** The dataset exhibited a fairly equal distribution of gender, with 1,068 males and 1,043 females.
2. **Obesity Distribution:** The majority of individuals were categorized as underweight or normal weight, while 46.1% were classified as obese.
3. **Family History:** Respondents classified as overweight or obese were more likely to report a positive family history of overweight.
4. **Research Question 1: Dietary Habits and Obesity:** Poor dietary habits were more prevalent among overweight and obese individuals, highlighting the significant influence of diet on obesity risk.
5. **Research Question 2: Smoking, Alcohol, and Obesity:** A clear correlation was observed between alcohol consumption and obesity, with a significant portion of obese individuals reporting alcohol consumption.
6. **Research Question 3: Physical Activity and Obesity:** A weak negative correlation was found between physical activity levels and BMI, indicating lower physical activity associated with higher BMI.

CONCLUSION:

The analysis underscores the importance of dietary habits, alcohol consumption, and physical activity levels in influencing obesity risk. These findings are critical for informing targeted interventions and health policies aimed at addressing the obesity epidemic.

Recommendations:

1. Promoting healthier dietary habits and raising awareness of the impact of diet on obesity.
2. Implementing strategies to reduce alcohol consumption, particularly among populations at risk of obesity.
3. Encouraging regular physical activity to mitigate the risk of obesity and associated health complications.

DATASET SOURCES:

1. The dataset used in this analysis was sourced from Kaggle, provided by Fabio Mendoza Palechor and Alexis de la Hoz Manotas.

LIMITATIONS:

Given the reliance on self-reported data, potential biases may exist, and causal relationships cannot be determined definitively. Additionally, the dataset's focus on specific regions may limit the generalizability of findings to broader populations. For further details, please refer to the project documentation and materials in the repository.

ACKNOWLEDGMENTS:

We extend our gratitude to the contributors of the dataset used in this analysis for their valuable contribution to obesity research.