Lifestyle Determinants of Obesity: A Detailed Analysis Using Synthetic Data

Drashti Miteshkumar Mehta(11686504)
Master's in Data Science
College of Information,University of North Texas
Denton, Texas
DrashtiMehta@my.unt.edu

***Abstract* — This paper presents an analysis on a comprehensive dataset related to obesity factors, which includes a variety of demographic, dietary, and lifestyle characteristics for 2111 people. Statistics and visuals are used to identify obesity predictors. This project examines obesity and factors like diet, exercise, and genetics to find ways to manage and prevent it. The findings of this study could help to inform targeted interventions and public health strategies.**

Keywords* — *Obesity, Data Analysis, Public Health, Lifestyle Factors, Dietary Habits, Physical Activity

I. Introduction (*Heading 1*)

Obesity is a global health concern with serious consequences for individual health and healthcare systems. Understanding the various causes of obesity, such as genetic, behavioural, and environmental factors, is critical for developing effective prevention and treatment strategies. This study examines obesity patterns and causes using a dataset of age, gender, diet, physical activity, and other factors.

The study's objectives are:

- Conduct a thorough exploratory analysis of the data, identifying key variables and distributions.
- To generate hypotheses regarding the links between lifestyle factors and obesity.
- Visually and statistically test these hypotheses to gain insights that can guide public health interventions.

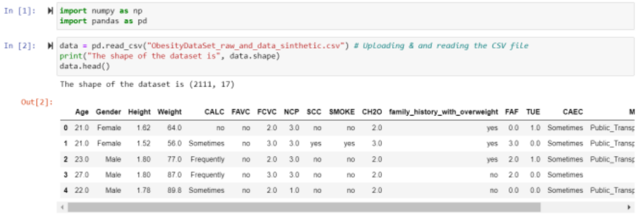
II. Data

A. Data Description

The dataset contains 2111 entries describing individual demographics, dietary habits, and lifestyle factors that may influence obesity. Attributes include age, gender, height, weight, meal frequency, physical activity, water intake, and familial health history. It categorizes each individual's obesity level based on their physical and lifestyle data.

Dataset -

https://www.kaggle.com/datasets/fatemehmehrparvar/obesity-levels



```
In [1]:  import numpy as np
         import pandas as pd

In [2]:  data = pd.read_csv("ObesityDataSet_raw_and_data_sinthetic.csv") # Uploading & and reading the CSV file
         print("The shape of the dataset is", data.shape)
         data.head()

         The shape of the dataset is (2111, 17)
Out[2]:
```

	Age	Gender	Height	Weight	CALC	FAVC	FCVC	NCP	SCC	SMOKE	CH2O	family_history_with_overweight	FAF	TUE	CAEC	M
0	21.0	Female	1.62	64.0	no	no	2.0	3.0	no	no	2.0	yes	0.0	1.0	Sometimes	Public_Transp
1	21.0	Female	1.52	56.0	Sometimes	no	3.0	3.0	yes	yes	3.0	yes	3.0	0.0	Sometimes	Public_Transp
2	23.0	Male	1.80	77.0	Frequently	no	2.0	3.0	no	no	2.0	yes	2.0	1.0	Sometimes	Public_Transp
3	27.0	Male	1.80	87.0	Frequently	no	3.0	3.0	no	no	2.0	no	2.0	0.0	Sometimes	
4	22.0	Male	1.78	89.8	Sometimes	no	2.0	1.0	no	no	2.0	no	0.0	0.0	Sometimes	Public_Transp

B. Attributes

1) Age: Age of the individual in years.

2) Gender: Gender of the individual (Male or Female).

3) Height: Height of the individual in meters.

4) Weight: Weight of the individual in kilograms.

5) CALC: Frequency of alcohol consumption, categorized as 'no', 'Sometimes', 'Frequently', and 'Always'.

6) FAVC: Binary indicator of frequent consumption of high caloric food ('yes' or 'no').

7) FCVC: Frequency of consumption of vegetables on a scale from 1 to 3, where 1 is low and 3 is high.

8) NCP: Number of main meals consumed daily, ranging from 1 to 4.

9) SCC: Binary indicator of self-monitoring of caloric intake ('yes' or 'no').

10) SMOKE: Smoking status of the individual ('yes' or 'no').

11) CH2O: Amount of water drunk daily, reported in liters.

12) family_history_with_overweight: Family history of overweight ('yes' or 'no').

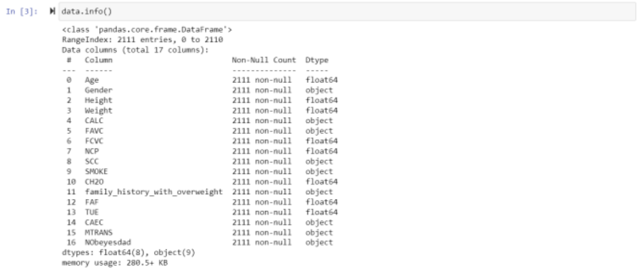
13) FAF: Frequency of physical activity per week, rated from 0 (none) to 3 (high).

14) TUE: Time spent using technology devices daily, rated from 0 (none) to 2 (high).

15) CAEC: Consumption of food between meals, categorized as 'no', 'Sometimes', 'Frequently', and 'Always'.

16) MTRANS: Mode of transportation used most frequently, including categories like 'Public_Transportation', 'Walking', 'Automobile', 'Motorbike', and 'Bike'.

17) NObeyesdad: Obesity classification of the individual, with categories ranging from 'Underweight' to 'Obesity_Type_III'.



```
In [3]:  data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2111 entries, 0 to 2110
Data columns (total 17 columns):
 #   Column                          Non-Null Count  Dtype  
---  ------                          --------------  -----  
 0   Age                             2111 non-null   float64
 1   Gender                          2111 non-null   object 
 2   Height                          2111 non-null   float64
 3   Weight                          2111 non-null   float64
 4   CALC                            2111 non-null   object 
 5   FAVC                            2111 non-null   object 
 6   FCVC                            2111 non-null   float64
 7   NCP                             2111 non-null   float64
 8   SCC                             2111 non-null   object 
 9   SMOKE                           2111 non-null   object 
 10  CH2O                            2111 non-null   float64
 11  family_history_with_overweight  2111 non-null   object 
 12  FAF                             2111 non-null   float64
 13  TUE                             2111 non-null   float64
 14  CAEC                            2111 non-null   object 
 15  MTRANS                          2111 non-null   object 
 16  NObeyesdad                      2111 non-null   object 
dtypes: float64(8), object(9)
memory usage: 280.5+ KB
```

C. *Tools Used*

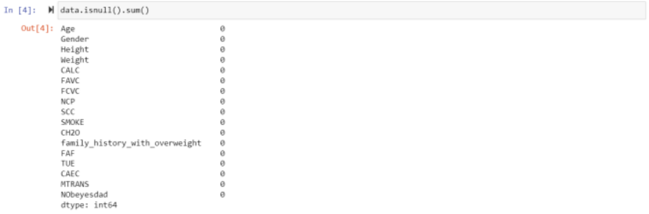
1) Jupyter Notebook

2) Tableau

III. Preprocessing of tHe Data

A. *Handling of Missing Values*

Missing data can introduce bias and affect the results of the analysis, so it is critical to address them either through imputation or removal. This dataset was checked for missing values using the isnull() method. There was no missing value found, so no imputation or removal was required.

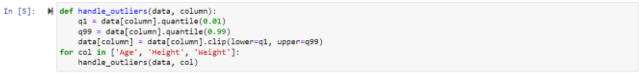


```
In [4]:  data.isnull().sum()

Out[4]:  Age                               0
         Gender                            0
         Height                            0
         Weight                            0
         CALC                              0
         FAVC                              0
         FCVC                              0
         NCP                               0
         SCC                               0
         SMOKE                             0
         CH2O                              0
         family_history_with_overweight    0
         FAF                               0
         TUE                               0
         CAEC                              0
         MTRANS                            0
         NObeyesdad                        0
         dtype: int64
```

B. *Handling the Outliers*

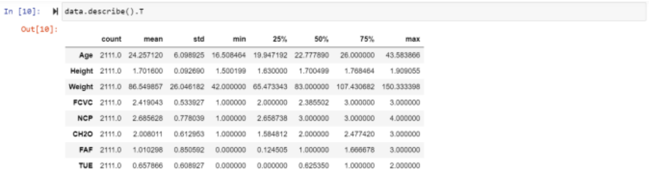
Outlier handling is intended to handle extreme values in the 'Age', 'Height', and 'Weight' columns. This function uses the 1st and 99th percentiles to trim outliers and keep data points within this range. By using this method, the analysis is protected from excessive impact of outlier values, which enables more accurate trend analysis and hypothesis testing within the dataset's expected bounds. Python uses Pandas' clip method to maintain dataset integrity for robust statistical inquiry.



```
In [5]:  def handle_outliers(data, column):
             q1 = data[column].quantile(0.01)
             q99 = data[column].quantile(0.99)
             data[column] = data[column].clip(lower=q1, upper=q99)
         for col in ['Age', 'Height', 'Weight']:
             handle_outliers(data, col)
```

C. *Descriptive Analysis*

Descriptive statistics tables show mean, standard deviation, and value distribution through minimum, median, and maximum readings. These insights guide initial data appraisal, normalization, and anomaly detection, laying the groundwork for advanced analytical tasks.



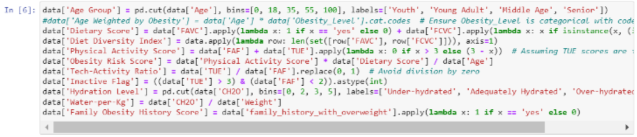
```
In [10]:  data.describe().T

Out[10]:
```

	count	mean	std	min	25%	50%	75%	max
Age	2111.0	24.257120	6.098925	16.508464	19.947192	22.777890	26.000000	43.583866
Height	2111.0	1.701600	0.092690	1.500199	1.630000	1.700499	1.768464	1.909055
Weight	2111.0	86.549857	26.046182	42.000000	65.473343	83.000000	107.430682	150.333398
FCVC	2111.0	2.419043	0.533927	1.000000	2.000000	2.385502	3.000000	3.000000
NCP	2111.0	2.685628	0.778039	1.000000	2.658738	3.000000	3.000000	4.000000
CH2O	2111.0	2.008011	0.612953	1.000000	1.584812	2.000000	2.477420	3.000000
FAF	2111.0	1.010298	0.850592	0.000000	0.124505	1.000000	1.666678	3.000000
TUE	2111.0	0.657866	0.608927	0.000000	0.000000	0.625350	1.000000	2.000000

D. *Feature engineering*

The feature engineering step of the project involves creating new features from existing data to better capture the patterns related to our hypotheses. We can enrich our dataset by adding relevant variables, which could provide deeper insights when analyzed. To examine how obesity levels vary across life stages, we've divided age into meaningful groups ('Youth', 'Young Adult', 'Middle Age', 'Senior'). To assess obesity, we've developed a 'Dietary Score' that quantifies an individual's diet and a 'Physical Activity Score' that considers activity frequency and technology use. To examine weight correlations, we added a 'Hydration Level' based on water intake and body weight. Finally, our 'Family Obesity History Score' quantifies family history’s impact.

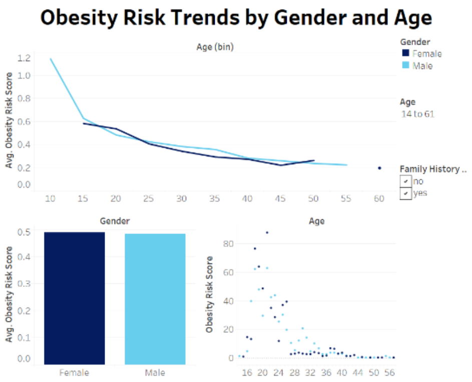


```
In [6]:  data['Age Group'] = pd.cut(data['Age'], bins=[0, 18, 35, 55, 100], labels=['Youth', 'Young Adult', 'Middle Age', 'Senior'])
         #data['Age Weighted by Obesity'] = data['Age'] * data['Obesity_Level'].cat.codes  # Ensure Obesity_Level is categorical with code
         data['Dietary Score'] = data['FAVC'].apply(lambda x: 1 if x == 'yes' else 0) + data['FCVC'].apply(lambda x: x if isinstance(x, (
         data['Diet Diversity Index'] = data.apply(lambda row: len(set([row['FAVC'], row['FCVC']])), axis=1)
         data['Physical Activity Score'] = data['FAF'] + data['TUE'].apply(lambda x: 0 if x > 3 else (3 - x))  # Assuming TUE scores are
         data['Obesity Risk Score'] = data['Physical Activity Score'] * data['Dietary Score'] / data['Age']
         data['Tech-Activity Ratio'] = data['TUE'] / data['FAF'].replace(0, 1)  # Avoid division by zero
         data['Inactive Flag'] = ((data['TUE'] > 3) & (data['FAF'] < 2)).astype(int)
         data['Hydration Level'] = pd.cut(data['CH2O'], bins=[0, 2, 3, 5], labels=['Under-hydrated', 'Adequately Hydrated', 'Over-hydratec
         data['Water-per-Kg'] = data['CH2O'] / data['Weight']
         data['Family Obesity History Score'] = data['family_history_with_overweight'].apply(lambda x: 1 if x == 'yes' else 0)
```

IV. Hypothesis and Vizulization

A. *Influence of Gender and Age on Obesity Risk*

I hypothesize that gender and age have a significant influence on obesity risk, with family history acting as a modifier. Specifically, I propose that obesity risk declines with age and varies by gender, with males potentially having a higher overall obesity risk. Furthermore, I believe that individuals with a family history of obesity have higher risk scores across all age groups than those without such a history. This dashboard aims to visually explore these relationships, providing insights into how demographic factors and family history influence obesity risk.



Visualization:

1. Line Graph

In this line graph, I show the average obesity risk scores across age groups, with separate lines for male and female trends. This graph clearly shows how obesity risk decreases with age for both genders, allowing for direct comparison to determine whether the trend is similar across genders or if one gender is consistently at a higher risk.

Development and Design Choices:

- **Format**: A line graph is used to depict trends across a wide age range.

- **Perceptual Tasks:** This graph allows for trend analysis across the age spectrum, as well as comparisons of male and female trends.

- **Design choices:** The use of blue for males and green for females creates an intuitive gender distinction, while the smooth lines allow for easy tracking of trends over time.

2. Bar Chart

The bar chart compares the average obesity risk scores for females and males. This provides a concise visual summary of the gender gap in obesity risk, encompassing the entire age range in a single comparative snapshot. This chart is useful for quickly identifying overall gender disparities in obesity risk.

Development and Design Choices:

- **Format:** A bar chart is useful for analyzing simple categorical data, in this case, female and male.

- **Perceptual Tasks:** The chart enables a simple evaluation of overall risk between genders.

- **Design Decisions:** Differentiating the bars with colors that correspond to those used in the line graph (blue and green) ensures visual consistency and facilitates quick identification.

3. Scatter Plot

This scatter plot shows individual data points of obesity risk scores plotted against age, with markers distinguishing whether people have a family history of obesity. This detailed breakdown aids in determining how family history affects obesity risk and whether this influence varies with age.

Development and Design Choices:

- Format: A scatter plot shows a wide distribution of data points across two variables, age, and obesity risk score.

- Perceptual Tasks: The plot identifies family history relationships and clustering.

- Design: Different markers (dots for no family history and stars for family history) help distinguish groups within the same age cohort, revealing obesity risk factors.

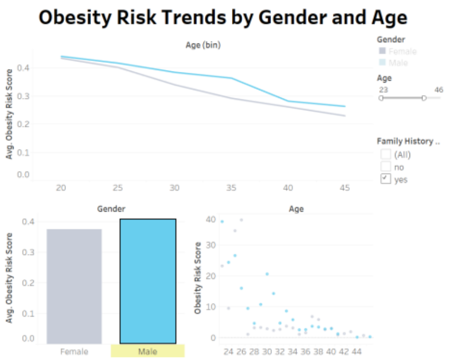
Interactive features:

- Gender and Age Filters

  A slider control for 'Age' allows users to refine the data presented in the line graph. This interactive tool enables a focused analysis on specific age groups, making it possible to observe how obesity risk scores vary within a narrower age range. The slider enhances the visualization's usability by providing a hands-on approach to exploring age-specific trends.

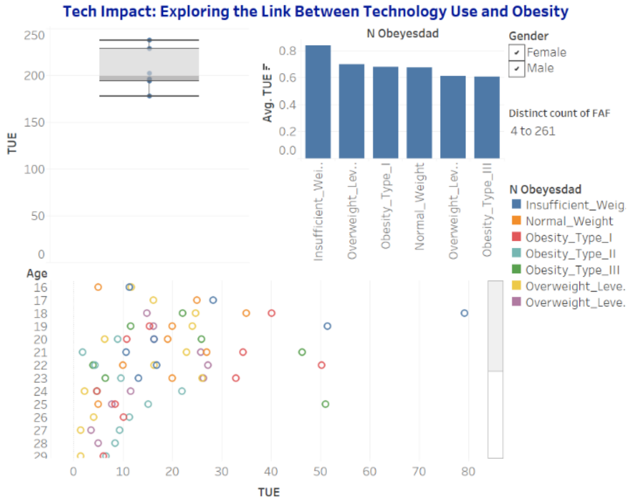
- Dropdown for Family History

  Checkboxes for 'Family History of Obesity' permit users to toggle the display of data for individuals with or without this background in the scatter plot. This interaction provides valuable insights into how genetic factors might affect obesity risks across different ages, enhancing the analytical depth of the visualization.



B. Tech Impact: Exploring the Link Between Technology Use and Obesity

I hypothesize that technology usage (measured in Technology Usage Equivalent, TUE) is associated with obesity across age groups and genders.



Visualizations:

1. **Box Plot**

In this box plot, I show the distribution of Technology Usage Equivalent (TUE) values among study participants. The plot shows the median, interquartile range, and potential outliers, which are important for understanding the overall distribution and central tendency of technology usage among participants. This visualization helps test my hypothesis by showing if technology use increases obesity.

Development and Design Choices:

- **Format:** I used a box plot to efficiently summarize the distribution of a continuous dataset and highlight outliers.

- **Perceptual Tasks:** The plot allows you to compare different quartiles and identify outliers in the data.

- **Design Choices:** Using a grey scale for the plot keeps the focus on the data distribution without being distracted by color.

2. **Bar Chart**

This bar chart shows the number of people who fall into various obesity categories, ranging from underweight to Obesity Type 3. Each bar is segmented by gender, allowing for a clear visual comparison of male and female distributions within each category. This chart clarifies gender differences in obesity prevalence, which is another layer of analysis in my research.

Development and Design Choices:

- **Format:** A bar chart was chosen because of its simple approach to comparing categorical data across groups.

- **Perceptual Tasks:** The chart allows for the comparison of quantities across obesity categories and between genders.

- **Design decisions:** Using different shades of blue allows for a more intuitive understanding of each category while maintaining a consistent overall aesthetic.

3. **Scatter Plot**

The scatter plot demonstrates the relationship between individual TUE scores and corresponding obesity categories across different age groups. Distinct colors and symbols are employed to represent each obesity category, facilitating rapid visual recognition and examination of patterns or trends in technology usage and obesity across different age groups. This plot is critical for observing how the relationship between TUE and obesity may change with age, adding depth to the insights gained from my hypothesis.

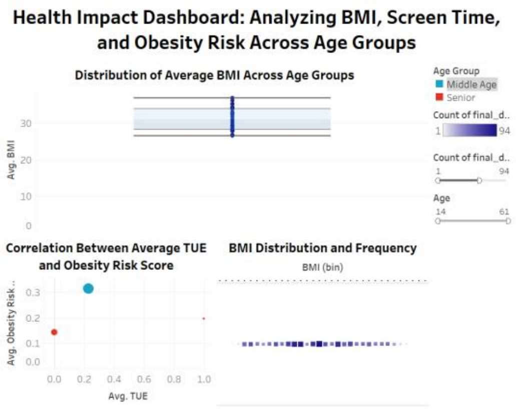
Development and Design Choices:

- **Format:** I chose a scatter plot because it is ideal for illustrating relationships between two continuous variables, in this case age and TUE.

- **Perceptual Tasks:** The plot allows for trend analysis and correlation identification across age groups and TUE values.

- **Design:** I used a color-coded scheme with obesity category markers for easy identification and comparison. Choosing a clear, orderly axis layout makes data easy to interpret.

Interactive Features


Health Impact Dashboard: Analyzing BMI, Screen Time, and Obesity Risk Across Age Groups
Distribution of Average BMI Across Age Groups
Avg. BMI
30
20
10
0
Age Group
Middle Age
Senior
Count of final_d..
1
94
Count of final_d..
1
94
Age
14
61
Correlation Between Average TUE and Obesity Risk Score
Avg. Obesity Risk ..
0.3
0.2
0.1
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Avg. TUE
BMI Distribution and Frequency
BMI (bin)

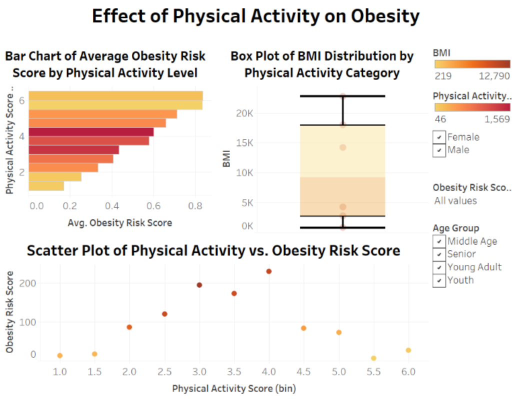
- Sliders for Age and Data Range Selection

  The incorporation of age and data range sliders gives users control over dataset granularity, allowing for a tailored view that focuses on specific age segments or changes dataset density. This interaction is critical for identifying trends and patterns that may vary significantly across age groups or data subsets.
- Dropdown for Age Group Selection

  The dashboard's data visualizations can be filtered by age group using a dropdown menu.

C. *Physical Activity's Role in Obesity Risk Reduction*

I hypothesize that there is a significant inverse relationship between physical activity levels and obesity risk scores across different demographics. Specifically, higher physical activity levels will correspond to lower obesity risk scores and BMI values among individuals. This dashboard explores the potential correlations between physical activity levels, BMI distributions, and obesity risk scores, examining how varying degrees of physical activity impact these metrics across different age and gender groups.


Effect of Physical Activity on Obesity
Bar Chart of Average Obesity Risk Score by Physical Activity Level
Physical Activity Score ..
6
4
2
0.0
0.2
0.4
0.6
0.8
Avg. Obesity Risk Score
Box Plot of BMI Distribution by Physical Activity Category
BMI
20K
15K
10K
5K
0K
BMI
219
12,790
Physical Activity..
46
1,569
☑ Female
☑ Male
Obesity Risk Sco..
All values
Age Group
☑ Middle Age
☑ Senior
☑ Young Adult
☑ Youth
Scatter Plot of Physical Activity vs. Obesity Risk Score
Obesity Risk Score
200
100
0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
Physical Activity Score (bin)

Visualization:

1. Bar Chart

The average obesity risk scores classed by different levels of physical activity are shown in this bar chart. Every bar stands for a distinct degree of activity, and color gradients show decreasing to increasing levels of activity. The chart effectively illustrates a possible inverse relationship between physical activity and obesity risk by showing how higher levels of physical activity are related with lower obesity risk scores.

Development and Design Choices:

- **Format:** A bar chart shows categorical and continuous variables—physical activity levels and average obesity risk scores—simply.

- **Perceptual Tasks:** The chart compares obesity risk across physical activity levels.

- **Design:** Gradients of color visualize higher activity with lower risk using intuitive color coding from red (higher risk) to yellow (lower risk).

2. Box Plot

The box plot represents the distribution of BMI in categories based on physical activity levels. It highlights the median, quartiles, and outliers within each category, providing information about BMI variability among people with different physical activity habits. This helps to determine whether more active people have a lower BMI.

Development and Design Choices:

- **Format:** A box plot is used to summarize and compare the distribution of a continuous variable (BMI) across multiple groups.

- **Perceptual Tasks:** The plot helps to identify distribution patterns and compare BMI ranges across physical activity levels.

- **Design Choices:** Using a consistent color scheme helps to differentiate between categories while maintaining visual cohesion with other dashboard elements.

3. Scatter Plot

This scatter plot illustrates the relationship between sphysical activity scores and obesity risk scores. Each data point represents an individual's level of physical activity and its correlation with their risk of obesity. This visualization is essential for detecting patterns and possible anomalies in the correlation between physical activity and obesity.

Development and Design Choices:

- **Format:** A scatter plot is ideal for investigating the relationships between two quantitative variables.

- **Perceptual Tasks:** It enables the detection of correlation patterns, trends, and outliers within data.

- **Design Options:** Color-coded dots could improve the plot by categorizing individuals based on additional variables such as age or gender, though it is currently using a uniform color.

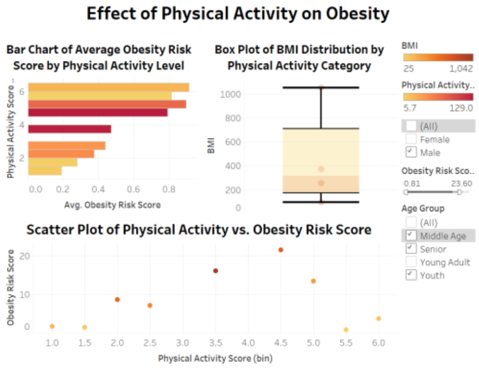
Interactive Features

- Sliders for Filtering Continuous Data

  BMI and Physical Activity Sliders - Interactive sliders for 'BMI' and 'Physical Activity Score' allow users to filter data along these dimensions to better understand the obesity risk-physical activity relationship. This feature helps isolate the effects of activity levels on BMI and obesity risk across demographic groups.

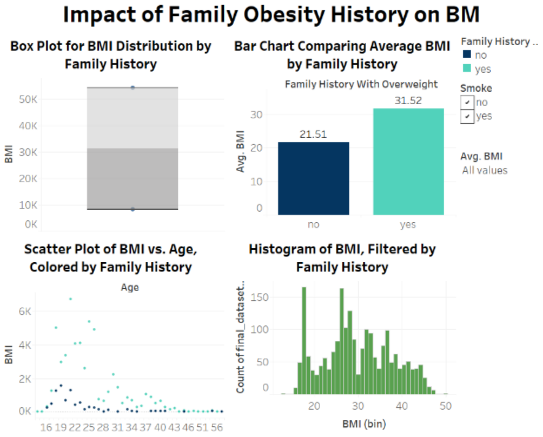
- Dropdown Selections for Demographic Filtering

  Interactive sliders for 'BMI' and 'Physical Activity Score' allow users to filter data along these dimensions to better understand the obesity risk-physical activity relationship. This feature helps isolate the effects of activity levels on BMI and obesity risk across demographic groups.


Effect of Physical Activity on Obesity
Bar Chart of Average Obesity Risk Score by Physical Activity Level
Physical Activity Score ..
6
4
2
0.0
0.2
0.4
0.6
0.8
Avg. Obesity Risk Score
Box Plot of BMI Distribution by Physical Activity Category
BMI
1000
800
600
400
200
0
BMI
25
1,042
Physical Activity..
5.7
129.0
☐ (All)
☐ Female
☑ Male
Obesity Risk Sco..
0.81
23.60
Age Group
☐ (All)
☑ Middle Age
☑ Senior
☐ Young Adult
☑ Youth
Scatter Plot of Physical Activity vs. Obesity Risk Score
Obesity Risk Score
20
10
0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
Physical Activity Score (bin)

D. Genetic Predisposition and Obesity

I hypothesize that a family history of obesity indicates a genetic predisposition that has a significant impact on BMI levels across various age groups. I anticipate that individuals who have a familial predisposition to obesity will exhibit elevated BMI values in comparison to those who do not possess such a genetic background. With this dashboard, my goal is to visually represent and examine the impact of genetic factors on BMI.


Impact of Family Obesity History on BM
Box Plot for BMI Distribution by Family History
BMI
50K
40K
30K
20K
10K
0K
Bar Chart Comparing Average BMI by Family History
Family History With Overweight
Avg. BMI
30
20
10
0
21.51
31.52
no
yes
Family History ..
no
yes
Smoke
☑ no
☑ yes
Avg. BMI
All values
Scatter Plot of BMI vs. Age, Colored by Family History
Age
BMI
6K
4K
2K
0K
16 19 22 25 28 31 34 37 40 43 46 51 56
Histogram of BMI, Filtered by Family History
Count of final_dataset..
150
100
50
0
20
30
40
50
BMI (bin)

Visualizations

1. Box Plot

   This box plot demonstrates the variation in BMI distribution between individuals who have a family history of obesity and those who do not. It presents important data points such as the median, quartiles, and outliers, providing an understanding of how a genetic inclination towards obesity is expressed in the general population. This visualization is crucial for determining whether individuals with a familial predisposition to obesity exhibit higher BMI values in comparison to those without such a predisposition.

   Development and Design Choices:

   - **Format:** A box plot efficiently summarizes and compares BMI distribution across two categorical groups (Family history of obesity and no family history).

   - **Perceptual Tasks:** The plot shows distribution patterns and BMI ranges between the two groups, helping explain genetic factors.

   - **Design:** A grayscale color scheme emphasizes distribution differences between people with and without family histories without distracting from the data.

2. Bar Chart

   This bar chart depicts the average BMI values for people classified by their family history of obesity. The comparison clearly demonstrates that individuals with a family history generally exhibit higher average BMI values, implying a potential genetic predisposition to obesity.

Development and Design Choices:

- **Format:** The bar chart format was chosen due to its clarity and effectiveness in demonstrating differences in average values between groups.

- **Perceptual Tasks:** It facilitates rapid comparison and facilitates clear comprehension of the average influence of family history on BMI.

- **Design Choices:** Employing different colors for each group enhances visual distinction and facilitates immediate identification of the group with a higher average BMI.

3. Scatter Plot

This scatter plot examines the correlation between BMI and age, with data points color-coded to indicate the presence or absence of a familial predisposition to obesity. It enables an examination of how BMI trends with age differ depending on genetic predisposition, revealing information about the long-term effects of genetics on obesity.

Development and Design Choices:

- **Format:** A scatter plot is ideal for displaying relationships and trends between two continuous variables, such as BMI and age.

- **Perceptual Tasks:** The plot aids in the analysis of trends and the detection of correlations, particularly in relation to age and genetic factors.

- **Design Choices:** The strategic utilization of color coding allows for clear distinction between individuals with and without a family history, facilitating the visualization and analysis of patterns.

4. Histogram

The histogram displays the frequency distribution of BMI values, categorized according to family history. This comprehensive analysis offers a meticulous examination of the distribution of BMI within each genetic category, emphasizing the frequency of specific BMI ranges among individuals with and without a family history of obesity.

Development and Design Choices:

- **Format**: Histograms show the number of data points within a range of values (bins).

- **Perceptual Tasks:** This visualization facilitates the analysis of distribution and comparison of frequencies in the context of genetic predisposition.

- **Design Options:** Different shades of green are used for each category, with varying intensities, to indicate the number of occurrences within each BMI range, emphasizing differences in distribution patterns.

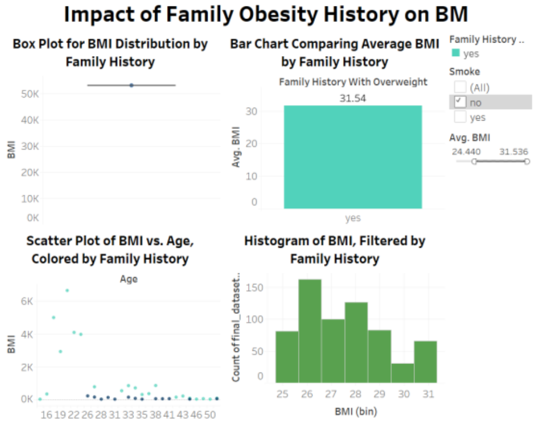
Interactive Features:

- **Dropdown and Checkbox Filters**

  Family History and Smoking Status Filters - Interactive dropdown and checkbox filters for 'Family History of Obesity' and 'Smoking Status' significantly enhance data navigation, enabling users to customize the data display according to specific criteria. These filters are crucial for segregating the dataset on the fly, allowing for immediate comparisons and focused analysis on how these factors influence BMI.

- Slider for Average BMI

  A slider for adjusting the BMI range offers a hands-on tool for users to explore BMI distributions within specific thresholds. This interactive feature aids in focusing the analysis on ranges of particular interest, such as higher or lower BMI values, enhancing the granularity and specificity of the insights derived.


Impact of Family Obesity History on BM
Box Plot for BMI Distribution by Family History
BMI
50K
40K
30K
20K
10K
0K
Bar Chart Comparing Average BMI by Family History
Family History With Overweight
31.54
Avg. BMI
30
20
10
0
yes
Family History ..
🟩 yes
Smoke
□ (All)
☑ no
□ yes
Avg. BMI
24.440
31.536
Scatter Plot of BMI vs. Age, Colored by Family History
Age
BMI
6K
4K
2K
0K
16 19 22 26 28 31 33 35 38 41 43 46 50
Histogram of BMI, Filtered by Family History
Count of final_dataset..
150
100
50
0
25
26
27
28
29
30
31
BMI (bin)

Conclusion

This study used a large dataset and sophisticated interactive visualizations to investigate various factors influencing obesity across different demographics. Through careful analysis and data science in a dynamic dashboard environment, we found significant relationships between obesity risk and physical activity, screen time, gender, age, and family history.

Our findings indicate that higher levels of physical activity are consistently correlated with lower obesity risk scores, highlighting the significance of active lifestyles in managing and potentially lowering obesity rates. Meanwhile, screen time analysis revealed that increased sedentary behavior, particularly among younger and middle-aged adults, is associated with a higher risk of obesity, emphasizing the impact of modern lifestyles on health.

The effect of family history on obesity was also significant, indicating that genetic predispositions play an important role in an individual's likelihood of becoming obese. This insight emphasizes the importance of targeted public health interventions that consider both genetic and lifestyle factors.

Gender-specific trends emerged from the data, with males and females exhibiting distinct patterns in how certain factors influence their obesity risk. Such distinctions emphasize the importance of gender-specific health strategies in addressing and mitigating obesity risks.

Interactive visualizations have been extremely useful in this study, providing a user-friendly platform for exploring complex datasets and discovering hidden patterns. These tools not only allowed for a more in-depth understanding of the data, but they also involved users in the analytical process, making the exploration both informative and accessible.

Finally, this project demonstrates the value of data visualization in public health research, providing clear insights that can inform future policies and interventions. As obesity continues to be a major global issue, utilizing such analytical tools will be critical in developing effective solutions tailored to diverse populations.

References

Yağın, F.H., Gülü, M., Gormez, Y., Castañeda-Babarro, A., Colak, C., Greco, G., Fischetti, F., & Cataldi, S. (2023). Estimation of Obesity Levels with a Trained Neural Network Approach optimized by the Bayesian Technique. *Applied Sciences –*

Estimation of Obesity Levels with a Trained Neural Network Approach optimized by the Bayesian Technique | Semantic Scholar

NOTE – (Used Quilbot for paraphrasing)