

## **NUTRITION PHYSICAL ACTIVITY AND OBESITY**

### **1. Project Overview**

This project focuses on analyzing publicly available health indicators related to nutrition habits, physical activity levels, and obesity prevalence across multiple demographic and geographic dimensions.

Using Power BI, the dataset is transformed into an interactive dashboard that:

- Aggregates obesity, nutrition, and physical activity indicators
- Applies stratification-based analysis to compare demographic groups
- Uses advanced visuals such as KPI cards, trend lines, bar charts, maps, and decomposition trees
- Incorporates risk classification to highlight high-risk populations
- Enables drill-down, filtering, and interactive exploration of health data

The final outcome is a comprehensive analytical dashboard that helps stakeholders:

- Monitor public health trends
- Identify root causes of obesity
- Forecast future health risks
- Recommend targeted interventions for improving population health outcomes

### **Project Objective:**

The primary objective of this analysis is to leverage the Nutrition, Physical Activity, and Obesity dataset to identify trends, risk patterns, and contributing factors to obesity using Power BI visual analytics.

Specific Objectives:

- To analyze obesity trends over time across different states and regions
- To examine how nutrition and physical activity behaviors impact obesity levels
- To compare obesity rates across demographic stratifications such as gender, income, education, and age
- To identify high-risk populations and regions requiring immediate attention

- To provide actionable insights that support public health strategies, awareness campaigns, and policy planning

## 2. Data Sources

- **Source Description and Timeline:** Data government **and** 2011-2023.
- **Data Source link:** <https://catalog.data.gov/dataset/>
- **Domain:** Public Health / Healthcare Analytics.

## 3. Problem Statement

- To understand how nutrition habits, physical activity levels, and obesity rates vary across different regions and demographic groups.
- To identify key lifestyle factors contributing to obesity
- To Identify high-risk regions and populations
- Provide actionable insights for health awareness and policy planning.
- Provide insights to support public health strategies and awareness programs

## 4. Attribute (Column /Features) Details:

Attribute Name	Data Type	Description
Year Start	Integer	Starting year of data collection
Year End	Integer	Ending year of data collection
Location Abbr	Text	Abbreviation of the location (e.g., CA, NY)
Location Desc	Text	Full name of the location (state or region)
Class	Text	Category of health indicator (Nutrition, Physical Activity, Obesity)
Topic	Text	Broad topic under the class (e.g., Obesity / Weight Status)
Question	Text	Survey question asked to collect the data
Data value	Decimal	Actual measured value (percentage or rate)
Low confidence Limit	Decimal	Lower bound of confidence interval
High confidence Limit	Decimal	Upper bound of confidence interval
Sample Size	Integer	Number of survey respondents
Latitude	Decimal	geographic location information
Longitude	Decimal	geographic location information
Class ID	Text	Unique identifier for the class
Topic ID	Text	Unique identifier for the topic
Question ID	Text	Unique identifier for the survey question
Location ID	Integer	Unique identifier for the location
Stratification Category	Text	Category used for stratification (Gender, Age, Income, etc.)

Stratification	Text	Specific stratification value (Male, Female, Age 18–24, etc.)
Stratification Category ID	Text	type of demographic stratification used to segment the data
Stratification ID	Text	specific coded value within the selected stratification category.

## 5. Tools & Technologies

- **Excel:** Data cleaning, transformation, and Pivot Tables.
- **Power BI:** Data modelling, DAX calculations, visualization, and interactive dashboard creation.

---

## 6. Data Pre-Processing (Excel / Power Query)

### Tasks Performed:

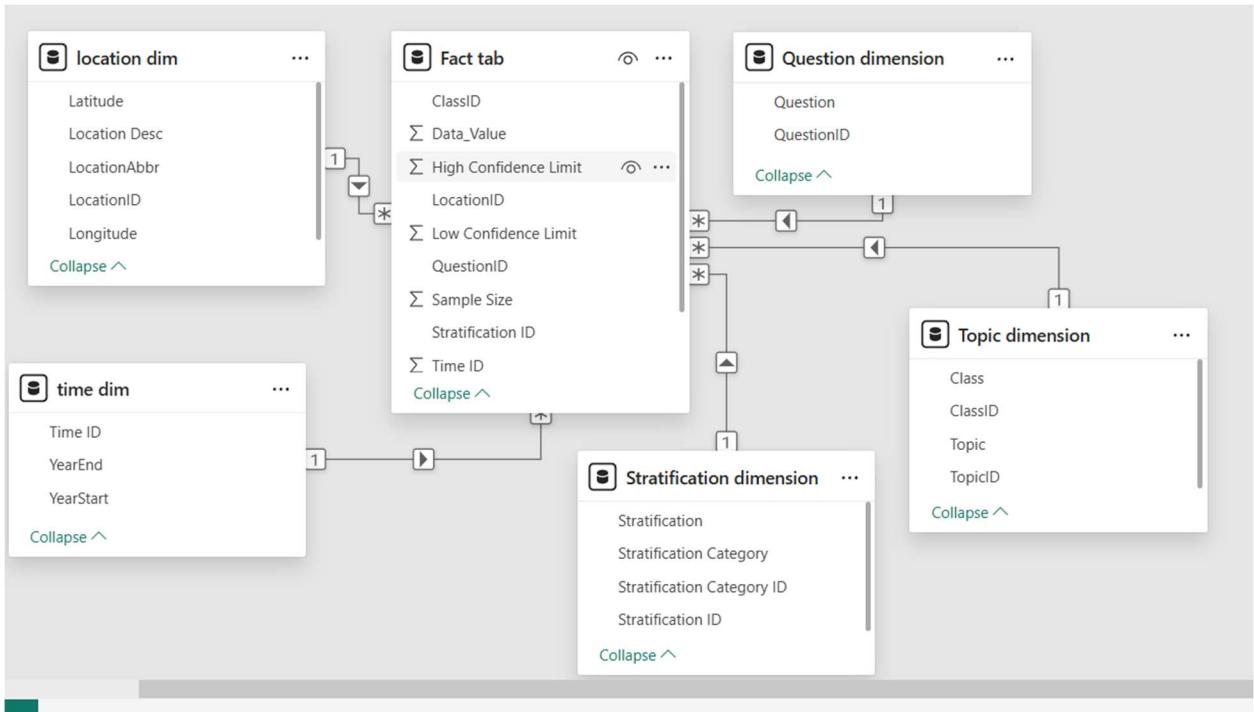
- **Data Cleaning :**
  - Removed duplicate records to avoid double counting.
  - Checked for missing or null values and handled them appropriately.
- **Data Transformation:**
  - Standardized **state names** to ensure accurate mapping in geographic visuals.
  - Cleaned and structured **Stratification Category** values (Age, Gender, Income, Education, Race/Ethnicity).
  - Created new calculated columns where required, such as:
    - ❖ **Risk Level (High / Medium / Low)**
    - ❖ **Risk Color codes** for conditional formatting.
- **Filtering & Sorting:**

Filtered the dataset to include only **relevant indicators** such as:

  - ❖ Obesity
  - ❖ Physical Activity
  - ❖ Nutrition-related indicators
  - Organized data to focus on relevant records.
  - Convert the data into Fact and Dimension Table
  - **Pivot Tables:** Generated Pivot Tables for data summarisation and initial insights.

## 7. Data Modelling and DAX (Power BI)

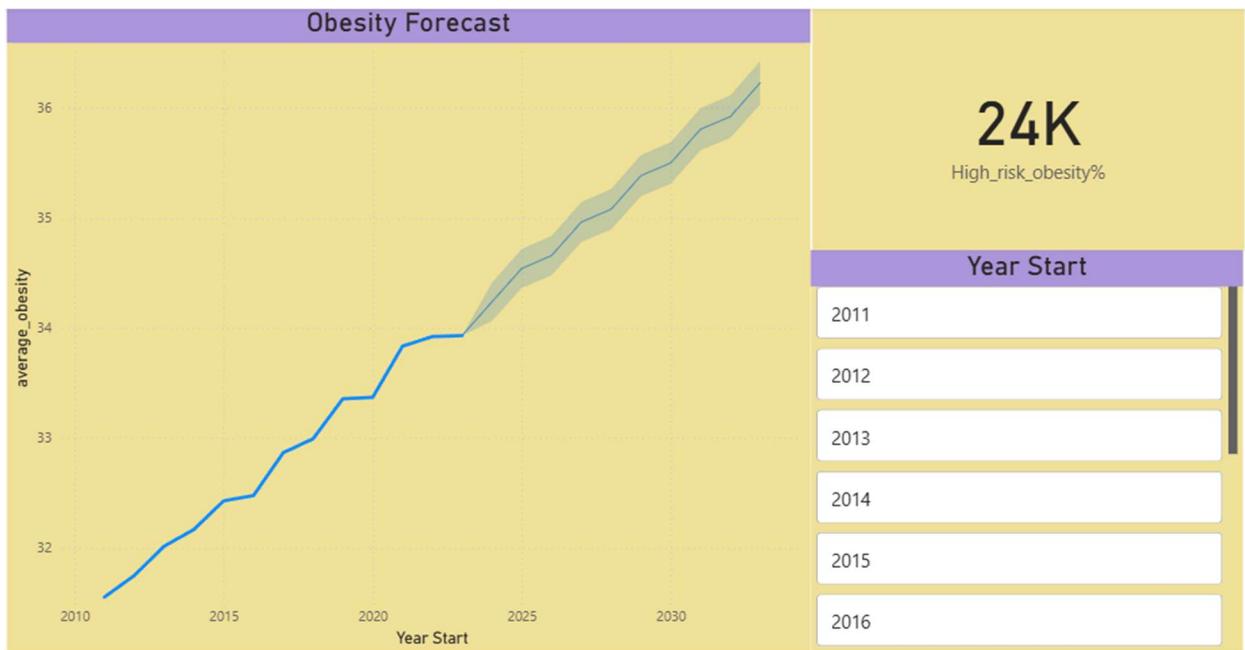
- **Cardinality:** One to many relationship



- **Calculated Columns & DAX Measures:**

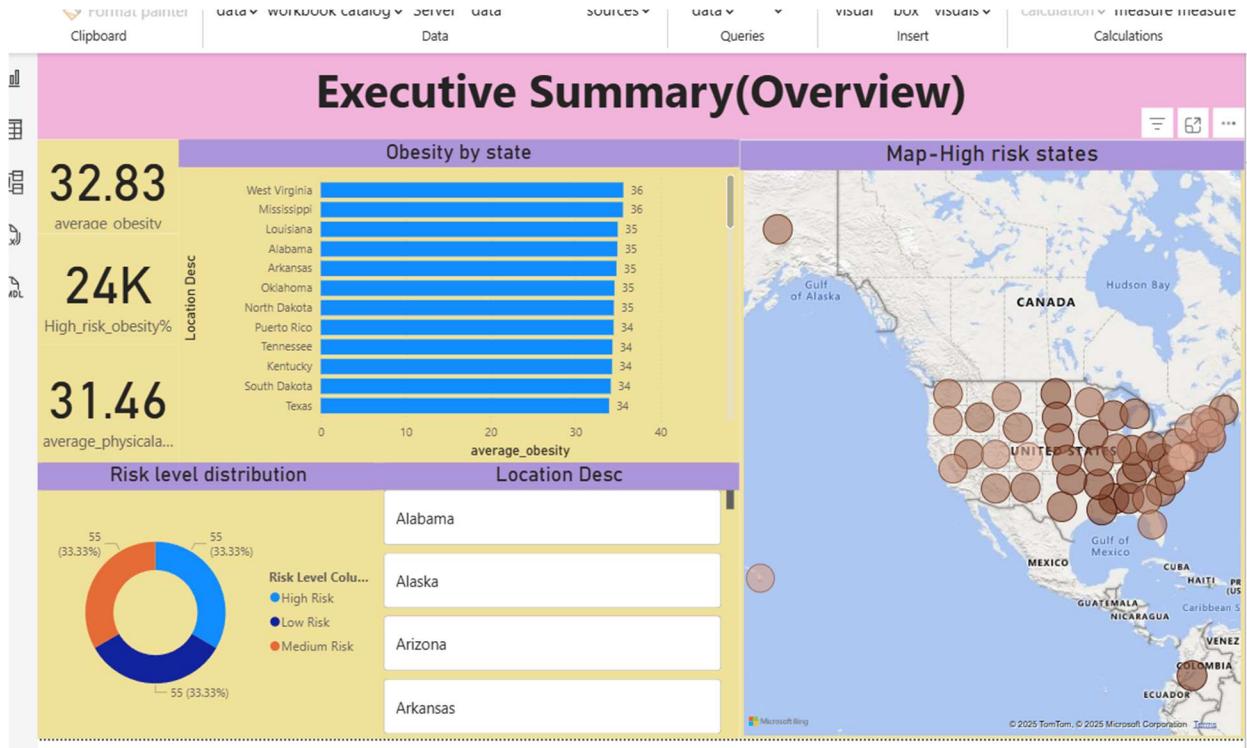
- average\_Nutrition = CALCULATE(AVERAGE('Fact tab'[Data\_Value]), FILTER('Topic dimension', CONTAINSSTRING('Topic dimension'[Topic], "Fruits and Vegetables"))))
- average\_obesity = CALCULATE(AVERAGE('Fact tab'[Data\_Value]), FILTER('Topic dimension', CONTAINSSTRING('Topic dimension'[Topic], "Obesity / Weight Status"))))
- average\_physicalactivity = CALCULATE(AVERAGE('Fact tab'[Data\_Value]), FILTER('Topic dimension', CONTAINSSTRING('Topic dimension'[Topic], "Physical Activity - Behavior"))))
- Female Avg Value = CALCULATE( AVERAGE('Fact tab'[Data\_Value]), 'Stratification dimension'[Stratification Category] = "Sex", 'Stratification dimension'[Stratification] = "Female") )
- High\_risk\_obesity% = CALCULATE(COUNTROWS('Fact tab'), 'Topic dimension'[Topic] = "Obesity / Weight Status", 'Fact tab'[Data\_Value] > 30)
- Risk Level = SWITCH(TRUE(), [Average\_datavalue] >= 30, "High Risk", [Average\_datavalue] >= 20, "Medium Risk", "Low Risk")
- Total\_obesity = CALCULATE(SUM('Fact tab'[Data\_Value]), FILTER('Topic dimension', CONTAINSSTRING('Topic dimension'[Topic], "Obesity / Weight Status"))))
- Total\_records = COUNTROWS('Fact tab')
- Confidence\_width = [Average\_HCL]-[Average\_LCL]
- 
- Avg Value by Gender = CALCULATE(AVERAGE('Fact tab'[Data\_Value]), 'Stratification dimension'[Stratification Category] = "Sex") )
- High Risk States = CALCULATE(DISTINCTCOUNT('location dim'[Location Desc]), 'Fact tab'[Risk Level Column] = "High Risk") )

## 8. Analysis and Visualizations (Power BI)



### KEY INSIGHTS FROM THE OBESITY FORECAST:

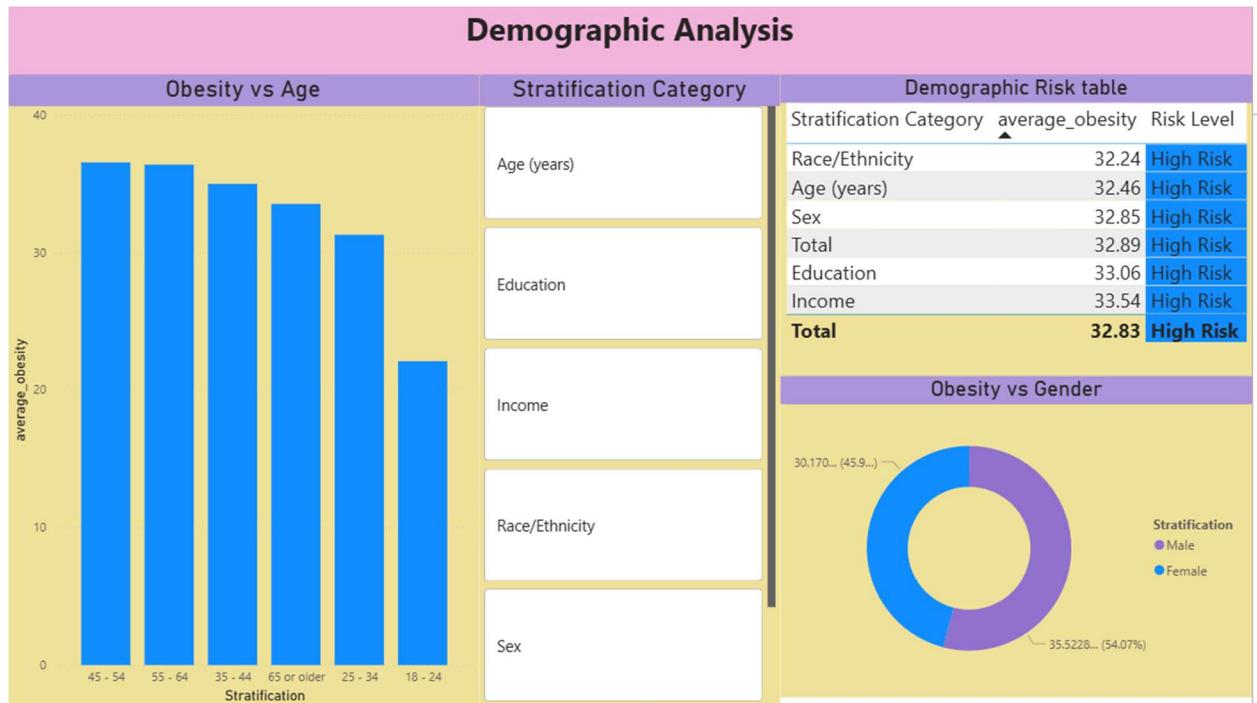
- From around **31.5% in 2010**, obesity steadily increases year by year.
- No significant decline or plateau is visible.
- This indicates a **persistent public health challenge** rather than short-term fluctuation. Obesity is **structural and long-term**, not seasonal.
- This Indicates
  - ❖ Lifestyle changes
  - ❖ Reduced physical activity
  - ❖ Possible long-term effects of social or economic changes
- Recent years show **accelerated obesity growth**, increasing urgency
- Even in the **best-case scenario**, obesity is expected to rise.
- The forecast line projects obesity reaching **~36–36.5%** in the coming years.
- The KPI value (**24K High\_risk\_obesity%**) highlights:
  - ❖ A large population segment already in the **high-risk category**
  - ❖ Rising averages mean **more people crossing into high-risk zones**.



### Key insights:

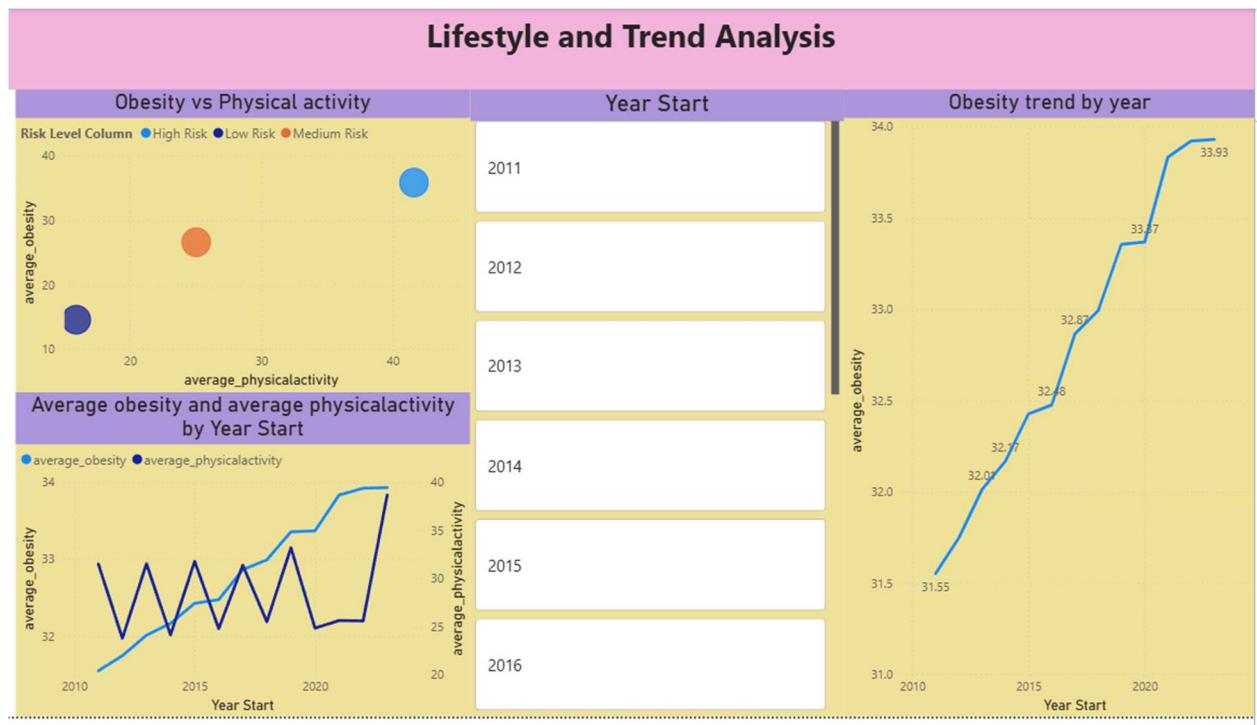
- The national average obesity rate is **32.83%**, indicating that **nearly one-third of the population** falls under obesity.
- This imbalance suggests a **strong inverse relationship** between activity and obesity.
- Insufficient physical activity is a key contributor to rising obesity rates.
- **West Virginia, Mississippi, Louisiana, Alabama, and Arkansas** rank among the **highest obesity rates**.
- Most top-ranking states exceed **34–36% obesity**.
- Regional socio-economic and lifestyle factors may play a major role.
- Strong concentration is visible in:
  - ❖ Southern states
  - ❖ Parts of the Midwest
- Preventive action is critical to stop **medium-risk states from escalating**.

- Low physical activity levels and an even distribution of risk categories indicate the need for immediate, targeted public health interventions to prevent further escalation



### Key Insights:

- The **highest obesity rates** are observed in:
  - **45–54 years**
  - **55–64 years**
- The **lowest obesity rate** is seen in the **18–24 years** age group.
- Younger adults show relatively lower obesity, presenting an opportunity for **early preventive interventions**.
- Obesity is **slightly more prevalent among males(54%)** than females(46%).
- Gender-specific behavioral and occupational factors may influence these differences.
- Socio-economic factors such as **income and education** significantly influence obesity risk.



#### Key Insights:

- **High Risk group** shows:
  - ❖ Higher obesity levels
  - ❖ Lower physical activity
- **Low Risk group** has:
  - ❖ Lower obesity
  - ❖ Higher physical activity
- **Medium Risk group** lies between high and low risk categories.
- There is a **clear inverse relationship** between physical activity and obesity.
- Obesity increases consistently over time despite fluctuations in physical activity.
- High-risk populations require **focused lifestyle interventions**.
- Preventive strategies must integrate **physical activity, nutrition education, and behavioral change**.

---

## 9. Insights :

### Key Findings:

- The **average obesity rate is 32.83%**, indicating obesity affects nearly one-third of the population.
  - Obesity prevalence shows a **steady upward trend over the years**, with no significant decline.
  - Certain states consistently appear as **high-obesity hotspots**, forming clear geographic clusters.
  - Middle-aged adults (**45–64 years**) experience the **highest obesity levels**.
  - **Lower physical activity** is strongly associated with **higher obesity rates**.
  - Obesity risk is **evenly distributed** across high, medium, and low categories, suggesting a **widespread issue rather than isolated cases**.
  - Socio-economic factors such as **income and education** show a noticeable influence on obesity prevalence.
- 
- **Analysis Insights:**

### Descriptive Analysis:

- The national average obesity rate stands at 32.83%.
- Several states exceed 35% obesity, placing them in the high-risk category.
- Obesity is higher among middle-aged adults and slightly higher among males.
- Physical activity levels remain moderate to low, compared to obesity prevalence.
- Obesity is widespread, persistent, and consistent across multiple demographics and regions.

### Diagnosis Analysis:

- States with lower physical activity levels show significantly higher obesity rates.
- Aging populations experience higher obesity due to reduced metabolism and lifestyle changes.
- Income and education disparities indicate limited access to healthy food and awareness.

- Behavioral and occupational differences contribute to gender-based obesity variation.
- Obesity is driven by a combination of lifestyle, demographic, and socio-economic factors, not a single cause.

### **Predictive Analysis:**

- Trend analysis indicates obesity will continue increasing if current patterns persist.
- Medium-risk groups are likely to shift into high-risk categories.
- Healthcare systems may face increased chronic disease burden related to obesity.
- Without effective intervention, obesity will become an even greater public health challenge in the future.

### **Prescriptive Analysis:**

- Implement targeted interventions in high-risk states and demographic groups.
- Promote physical activity programs in workplaces, schools, and communities.
- Launch nutrition education and awareness campaigns.
- Address socio-economic barriers through policy and healthcare access improvements.
- Use dashboards for continuous monitoring and evaluation.
- Data-driven, targeted, and preventive strategies can significantly reduce future obesity risk.

## **10. Conclusions :**

The integration of Excel and Power BI proved highly effective for conducting end-to-end data analysis, transforming raw public health data into meaningful, interactive, and actionable insights.

Through descriptive analysis, the study revealed that obesity prevalence is consistently high, with an average rate of 32.83%, affecting nearly one-third of the population. Geographic and demographic analysis identified specific states and population groups, particularly middle-aged adults and socio-economically disadvantaged groups, as high-risk segments.

Diagnostic analysis highlighted that low physical activity, age progression, and socio-economic factors are key contributors to rising obesity levels.

Using trend and predictive analysis, the data indicated a steady upward trend in obesity rates over time, suggesting that without timely and effective intervention, obesity will continue to increase and place greater pressure on healthcare systems.

Finally, prescriptive analysis provided actionable recommendations, including targeted interventions for high-risk states, promotion of physical activity, nutrition education, and policy-driven public health strategies. The interactive Power BI dashboards enable continuous monitoring and support data-driven decision-making for policymakers and health professionals.