SQL Project Report: Obesity Analysis Using SQL

1. Project Title:

Obesity Analysis Using SQL

2. Objective / Problem Statement:

The objective of this project is to analyze factors contributing to obesity using structured query language (SQL). The dataset includes various lifestyle and physiological factors. The goal is to extract insights regarding demographics, behavioral patterns, and health metrics related to obesity.

3. Tools & Technologies Used:

- Database: MySQL

- Tools: MySQL Workbench

- Languages: SQL

4. Dataset Description:

- Source: Uploaded dataset (ObesityDataSet_raw_and_data_sinthetic.csv)

- Records: 2,111 entries

- Attributes:

  - Gender, Age, Height, Weight

  - family_history_with_overweight

  - FAVC (Frequent consumption of high-caloric food)

  - FCVC (Frequency of vegetable consumption)

  - NCP (Number of main meals)

  - CAEC (Consumption of food between meals)

  - SMOKE, CH2O (Daily water intake), SCC (Calories monitoring)

- FAF (Physical activity frequency), TUE (Time using tech devices)

- CALC (Alcohol consumption), MTRANS (Mode of transport)

- NObeyesdad (Obesity level)

5. SQL Tasks and Analysis:

[OK] Data Preparation

```sql
CREATE DATABASE project;


CREATE TABLE project_obesity(

  Gender VARCHAR(30),

  Age INT,

  Height INT,

  Weight INT,

  family_history_with_overweight VARCHAR(30),

  FAVC VARCHAR(30),

  FCVC INT,

  NCP INT,

  CAEC VARCHAR(30),

  SMOKE VARCHAR(30),

  CH2O INT,

  SCC VARCHAR(30),

  FAF INT,

  TUE INT,

  CALC VARCHAR(30),

  MTRANS VARCHAR(50),
```

```
  NObeyesdad VARCHAR(50)

);


DESC project_obesity;

SELECT * FROM project_obesity;

```
```

6. Key Analysis Queries & Results:

1. Count by Obesity Category

SELECT NObeyesdad, COUNT(*) AS count

FROM project_obesity

GROUP BY NObeyesdad;

2. Average Weight by Gender

SELECT Gender, AVG(Weight) AS avg_weight

FROM project_obesity

GROUP BY Gender;

3. Obesity vs Physical Activity

SELECT NObeyesdad, AVG(FAF) AS avg_activity

FROM project_obesity

GROUP BY NObeyesdad;

4. Smoking Impact

SELECT SMOKE, AVG(Weight) AS avg_weight

FROM project_obesity

GROUP BY SMOKE;

5. Smoking vs Obesity Level

```
SELECT SMOKE, NObeyesdad, COUNT(*) AS count

FROM project_obesity

GROUP BY SMOKE, NObeyesdad

ORDER BY SMOKE;
```

6. Transportation Mode vs Obesity

```
SELECT MTRANS, NObeyesdad, COUNT(*) AS count

FROM project_obesity

GROUP BY MTRANS, NObeyesdad

ORDER BY MTRANS;
```

7. Obesity Level by Age Group

```
SELECT

  CASE

    WHEN Age < 18 THEN 'Teen'

    WHEN Age BETWEEN 18 AND 30 THEN 'Young Adult'

    WHEN Age BETWEEN 31 AND 45 THEN 'Adult'

    ELSE 'Senior'

  END AS age_group,

  NObeyesdad,

  COUNT(*) AS count

FROM project_obesity

GROUP BY age_group, NObeyesdad

ORDER BY age_group, count DESC;
```

8. Water Intake vs Obesity

```sql
SELECT CH2O AS daily_water_intake, NObeyesdad, COUNT(*) AS count

FROM project_obesity

GROUP BY CH2O, NObeyesdad

ORDER BY CH2O;
```

9. Caloric Food Consumption Impact

```sql
SELECT FAVC, NObeyesdad, COUNT(*) AS count

FROM project_obesity

GROUP BY FAVC, NObeyesdad

ORDER BY FAVC;
```

10. Physical Activity by Gender

```sql
SELECT Gender, AVG(FAF) AS avg_activity

FROM project_obesity

GROUP BY Gender;
```

11. BMI Calculation

```sql
SELECT Gender, Age, Weight, Height,

    ROUND(Weight / (Height * Height), 2) AS BMI, NObeyesdad

FROM project_obesity

ORDER BY BMI DESC

LIMIT 10;
```

7. Insights / Outcomes:

- People consuming high-calorie food (FAVC = Yes) are significantly more likely to fall into

overweight or obese categories.

- Average BMI and weight are higher in individuals with less physical activity (low FAF).

- Smoking does not have a strong direct correlation with obesity but has minor impact when cross-analyzed with NObeyesdad.

- Young adults showed the most diversity in obesity levels, highlighting lifestyle influence in this age group.

8. Visualizations (Optional Description):

Visualizations can be added using Power BI based on these SQL views:

- Pie chart for obesity categories

- Bar chart for average weight by gender

- Line chart for physical activity across age groups

- Heatmap for MTRANS vs Obesity

9. GitHub / Portfolio Link (If Any):

You can add your GitHub link here after uploading the SQL code and screenshots.