MINI REPORT: DATA-DRIVEN OBESITY ANALYSIS

Adina Adnan Mansoor , Muhammad Taqi , Muhammad Shayaan

1. INTRODUCTION & METHODOLOGY

This report analyzes behavioral determinants of obesity using machine learning. We compare Logistic Regression and a Neural Network classifier on the UCI Obesity dataset, perform lifestyle-based clustering using K-Means (k=7), and evaluate demographic fairness using Demographic Parity Difference (DPD) and Equal Opportunity Difference (EOD).

2. RESULTS

Model Performance

Metric	Logistic Regression	Neural Network
Accuracy	90.4%	96.2%
Macro F1	0.90	0.961
ROC-AUC(macro)	0.986	0.999

Interpretation:
The neural network outperforms logistic regression by ~6%, suggesting that obesity risk is influenced by non-linear interactions (e.g., diet × transport × activity) that linear models cannot easily capture. Key predictive features include Weight, Height, Age, FCVC (vegetable intake), NCP (meal frequency), and TUE (screen time).

Fairness Analysis

Model	**Gender - Demographic Parity**	**Gender - Equal Opportunity**	**Age - Demographic Parity**	**Age - Equal Opportunity**
Logistic Regression	0	0.09013	0	0.025131
Neural Network	0	0.126212	0	0.0088
Cluster Classifier	0	0	0	0

Interpretation: The neural network is the fairest model overall. It predicts obesity at equal rates across gender and age (DPD = 0.00) and shows very small differences in true-positive rates (EOD ≈ 0.03 gender, ≈0 age). Logistic regression is slightly less fair, with a larger gap in recall across groups (~0.09). The cluster classifier shows all zeros, but since it's not predicting obesity, these metrics don't reflect health fairness and should not guide policy decisions.

Clustering

Using K-Means (k=7, silhouette=0.1506), we identify behavioral archetypes:

Cluster	**Behavioral Pattern**	**Risk**
0	Physically active + high veg intake	Low risk
1	High screen time + moderate activity	Medium risk
2	High screen time + car commuters	High risk
3	High activity + high calorie intake	Medium risk
4	Low meals + low water intake (nutritional imbalance)	Medium risk
5	High veg intake (balanced diet)	Low risk
6	Sedentary + moderate veg intake	High risk

Interpretation : These clusters identify meaningful lifestyle segments, which is validated by a 93.6% accuracy Random Forest cluster-classification model. High-risk segments were dominated by car-dependent commuting and high screen-time behaviors, whereas lower-risk groups had higher physical activity and greater vegetable consumption. These clusters highlight how lifestyle patterns and habits are what may result in obesity.

3. KEY DRIVERS OF OBESITY

Driver	**Obesity Rate**	**Gap**	**Why it matters**
Sedentary lifestyle	54.2% vs 37.3% (active)	16.9pp	Metabolic dysfunction, largest
Auto development transport	45.2% vs 5.5% (walk)	39.7pp	Mobility is structural, not choice
High calorie diet (FAVC = yes)	51.7% vs 7.8%(no)	43.9pp	Food access + cost barriers

Age 40+	63.2% obesity	-	Accumulated inactivity, not age itself

These reflect structural conditions (urban design, food access), not personal failure.

4. POLICY RECOMMENDATIONS

- **Improve access to physical activity** through workplace programs + community initiatives (target sedentary populations).
- **Invest in walkable, transit-friendly infrastructure** to reduce car-dependency in high-risk areas.
- **Reform food pricing** by taxing ultra-processed foods and subsidizing fresh produce in low-income communities

5. SDG ALIGNMENT

SDG 3 (Good Health): Model accuracy (96.2%) supports equitable obesity screening; preventing sedentary obesity reduces NCD burden.

SDG 10 (Reduced Inequalities): Fair model (DPD=0.00) avoids demographic discrimination. Infrastructure investments in underserved areas narrow obesity disparities.

6. CONCLUSION

The neural network outperforms logistic regression (96.2% vs. 90.4%) and identifies non-linear behavioral drivers of obesity. Key risks include sedentary behaviour, automobile-dependent transport, and high-calorie diets, while fairness analysis shows no demographic bias, supporting ethical deployment. Effective policy must combine individual programs with structural change in mobility and food access.