



Obesity Level Predictions

with eating habits and physical conditions



The Data

The original data:

“Dataset for estimation of obesity levels based on eating habits and physical condition in individuals from Colombia, Peru and Mexico” -Fabio Mendoza Palechor, Alexis de la Hoz Manotas

2111 rows of data with 17 features (23% online survey, 77% synthetic)

Additional Data: additional synthetic data was added to create 22845 total rows of data

Synthetic data created using the Weka tool with the SMOTE filter

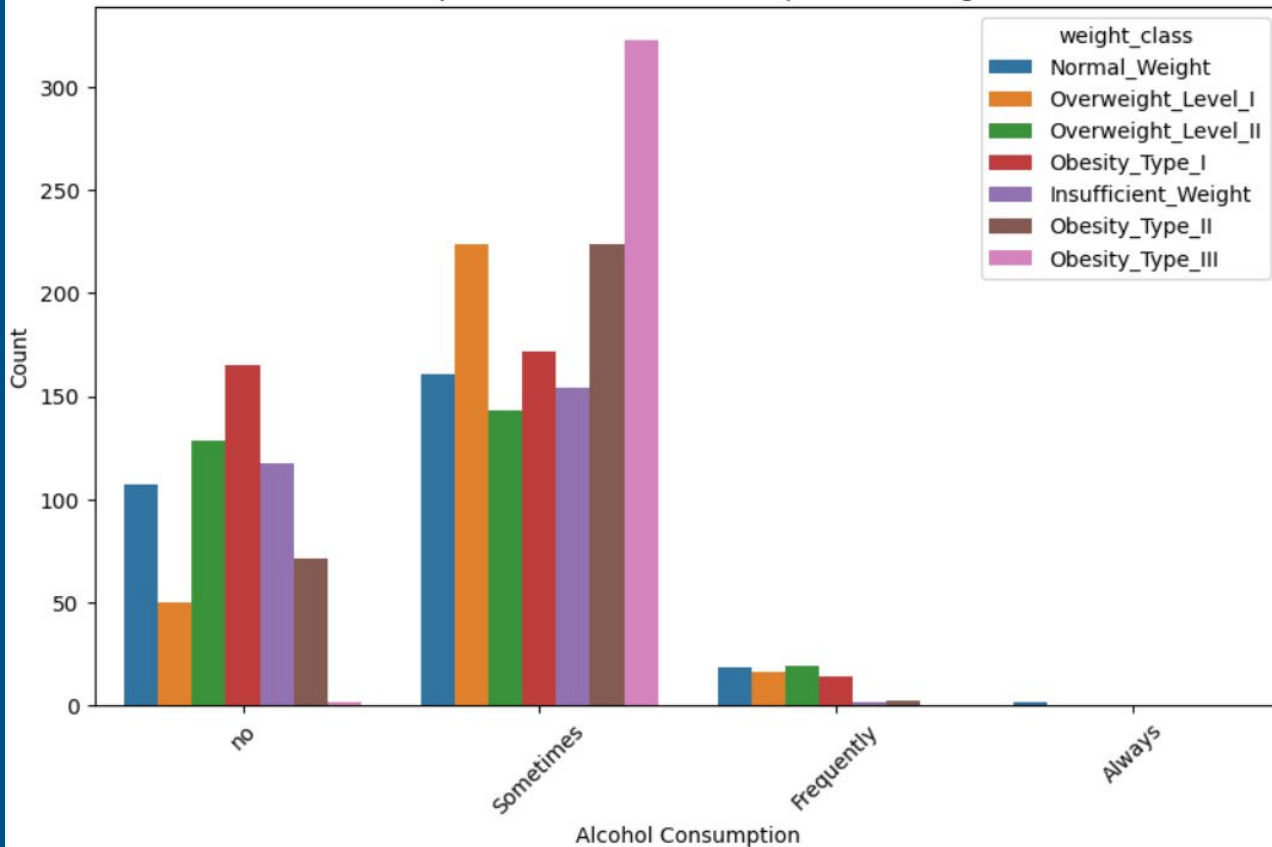
Analysis:

The original data has some flaws as seen here by looking at the original data (non synthetic)

Normal_Weight	287
Overweight_Level_I	58
Overweight_Level_II	58
Obesity_Type_I	47
Insufficient_Weight	34
Obesity_Type_II	11
Obesity_Type_III	3

Analysis:

Relationship between Alcohol Consumption and Weight Class

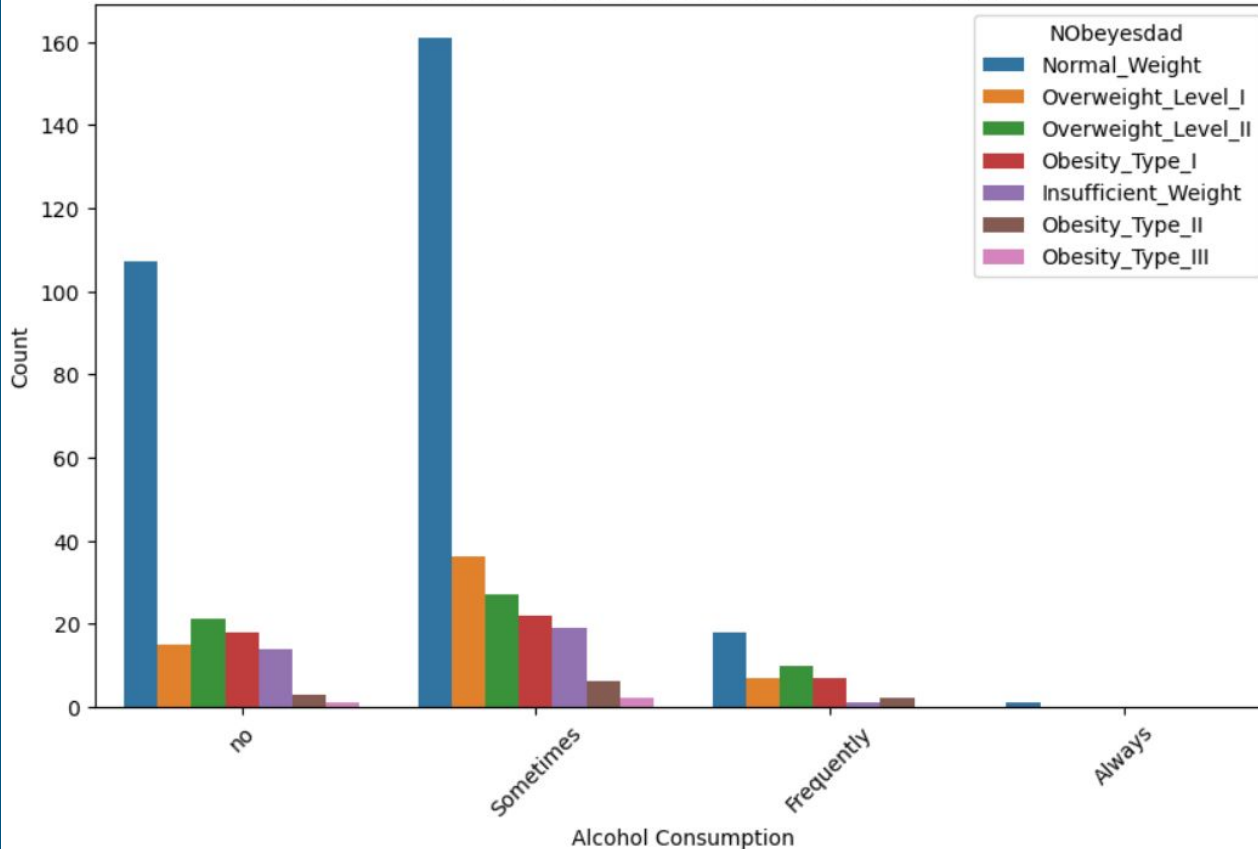


What happens when we build synthetic data on too small a sample:

-Obesity Type III exists almost exclusively in the sometimes column

Analysis:

Relationship between Alcohol Consumption and Weight Class



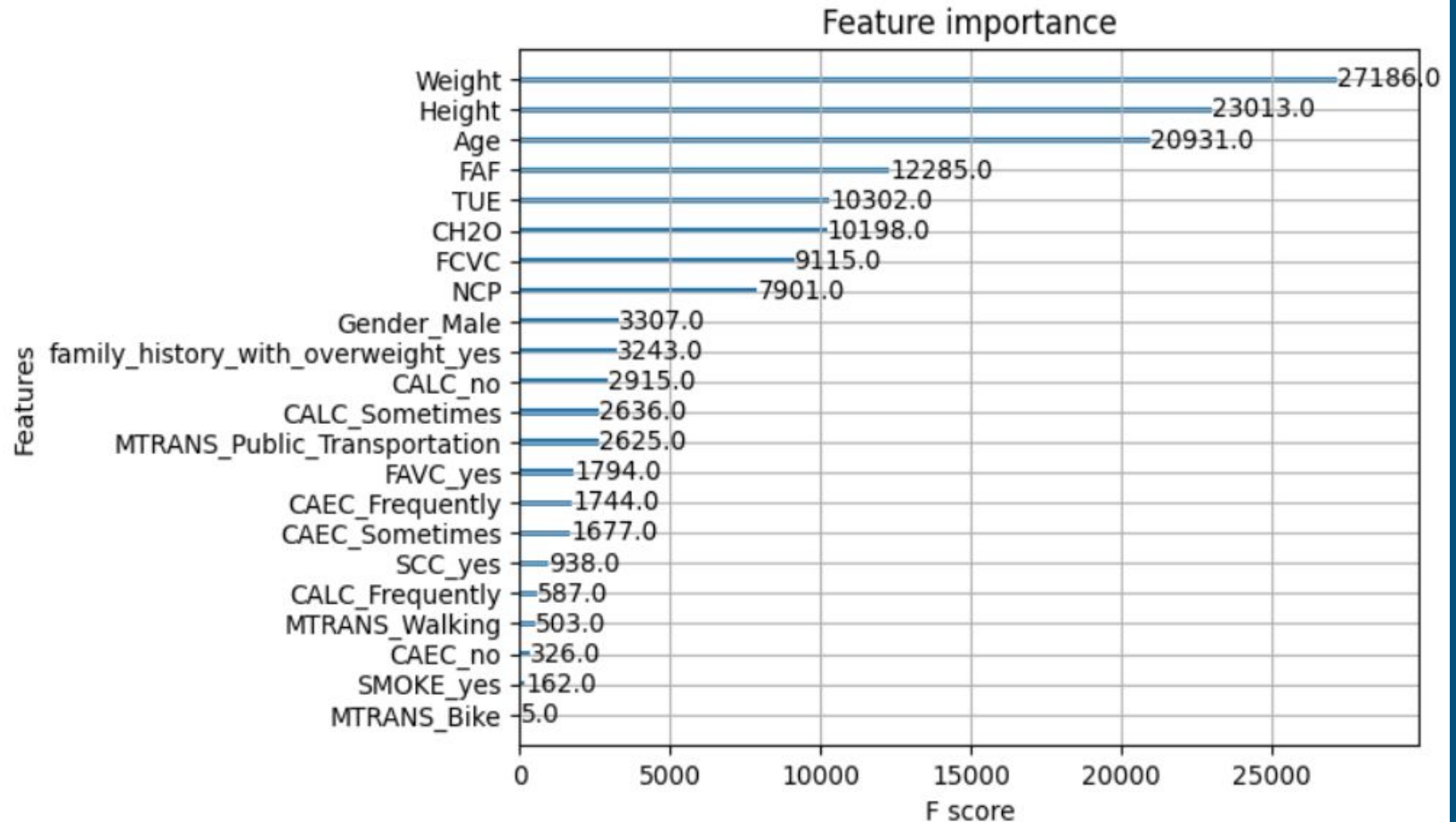
Why we need synthetic data:

- from this original we have the imbalance Normal_Weight class

- it suggests you should drink sometimes to have the best chance to fit the Normal_Weight class

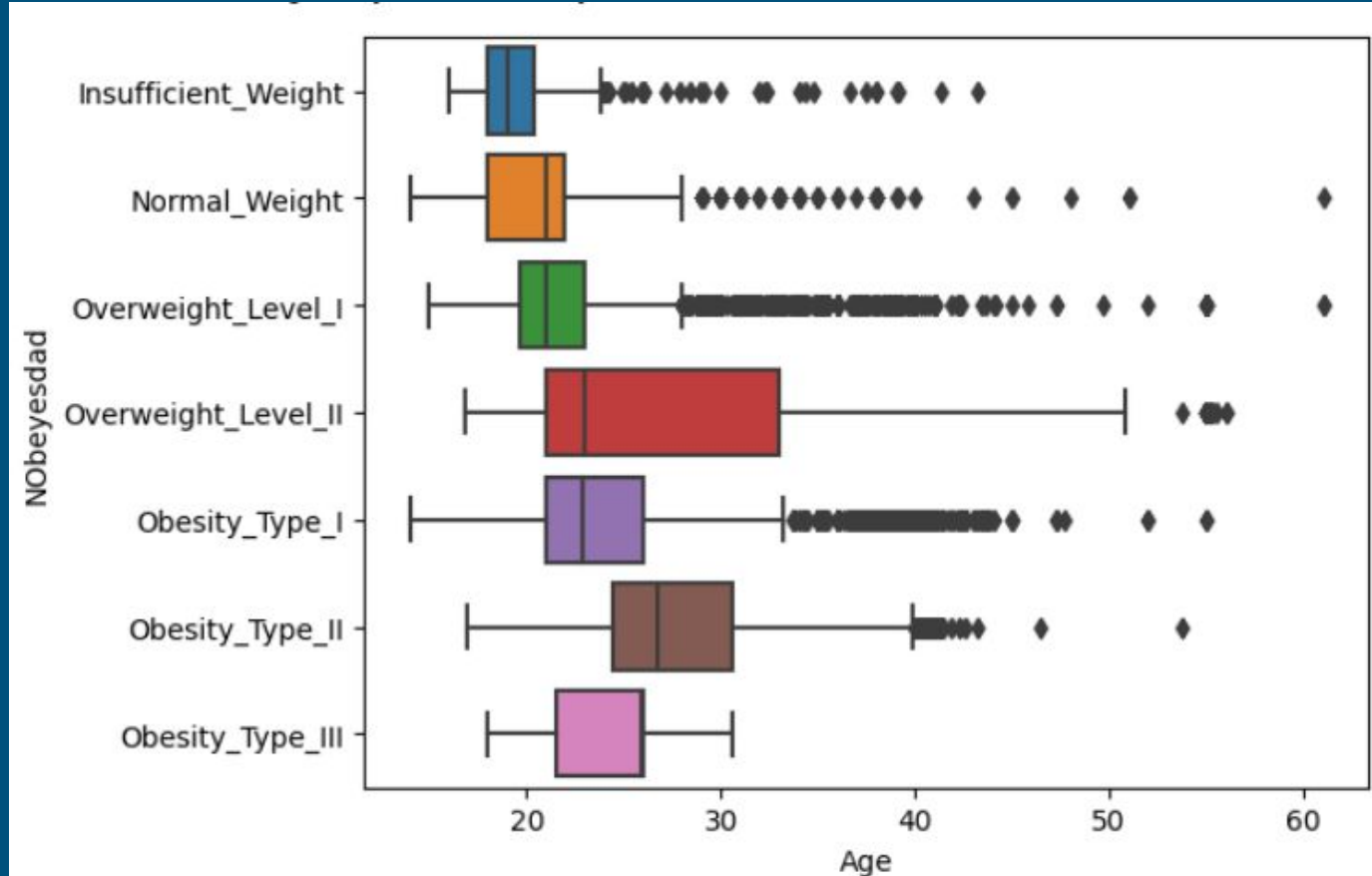
- We also see how the small sample size influenced the synthetic data for Obesity_Type_III

Analysis: The XGBClassifier top features



Analysis:

We see median age is generally higher as weight class increases



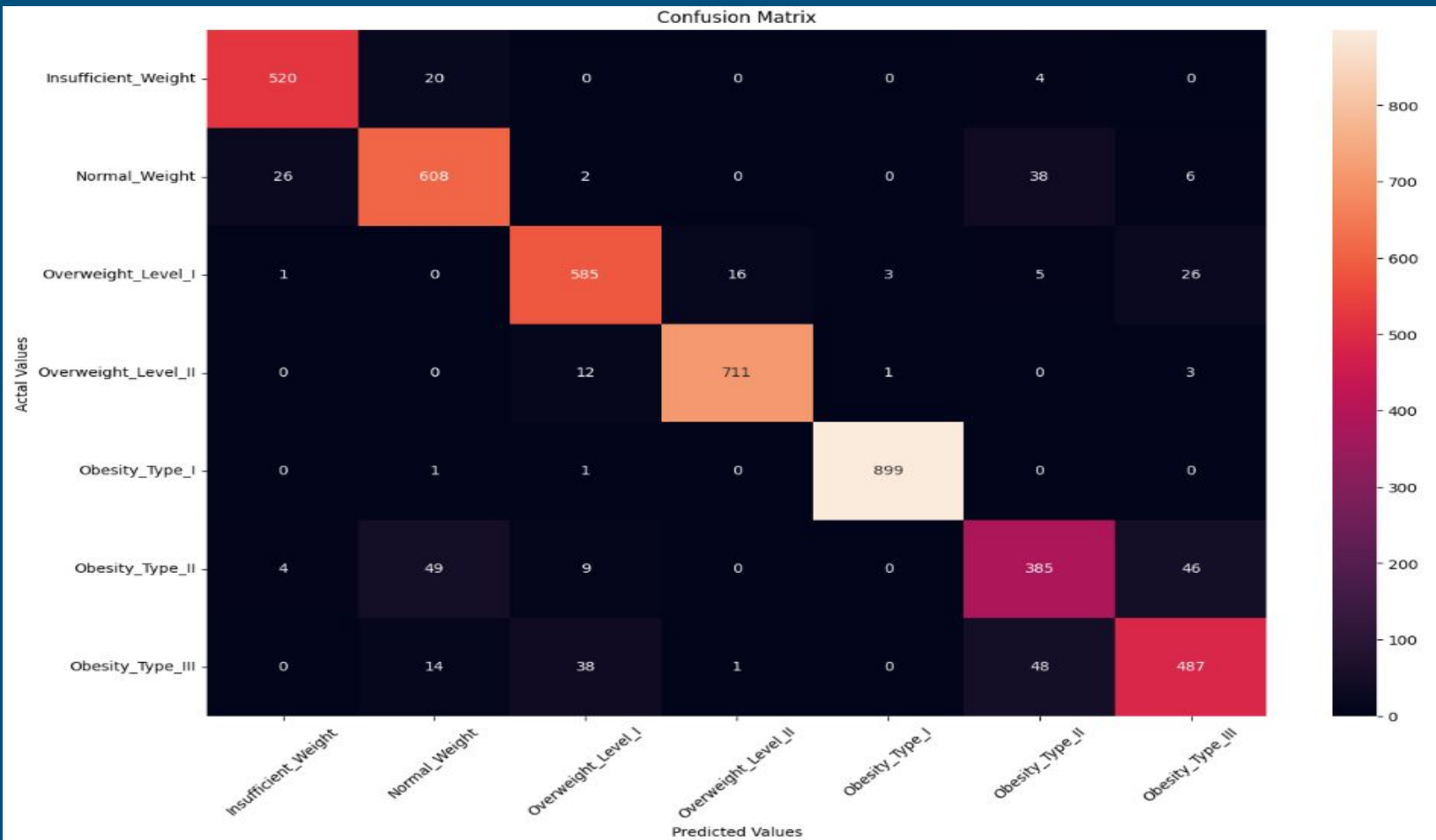
Analysis: Correlation and issues with synthetic data



Model Performance:

	XGBClassifier	LGBMClassifier	Logistic Regression
model	XGBClassifier	LGBMClassifier	Logistic Regression
train_score	0.95081	0.979481	0.85325
test_score	0.918144	0.91705	0.852265
f1_score	0.918488	0.917371	0.854451

Model Performance and Accuracy Across Classes



Streamlit app for sustainable options for improved weight management and overall health

Users can input their current conditions, and receive their weight class,
and make changes to see how it will impact their weight class

(In Progress) Snowflake Arctic LLM to offer recommendations
based on user input and preferences





Questions?

