Python Project

Estimation of obesity levels based on eating habits and physical condition

By Marc Lemaistre, Louis Martyr, Killian Lafaye

Why this subject ?

Mexico:

- 32th in the world
- 28,1% obese (16% for children)

France:

- 65th in the world
- 23,9% obese (8% for children)

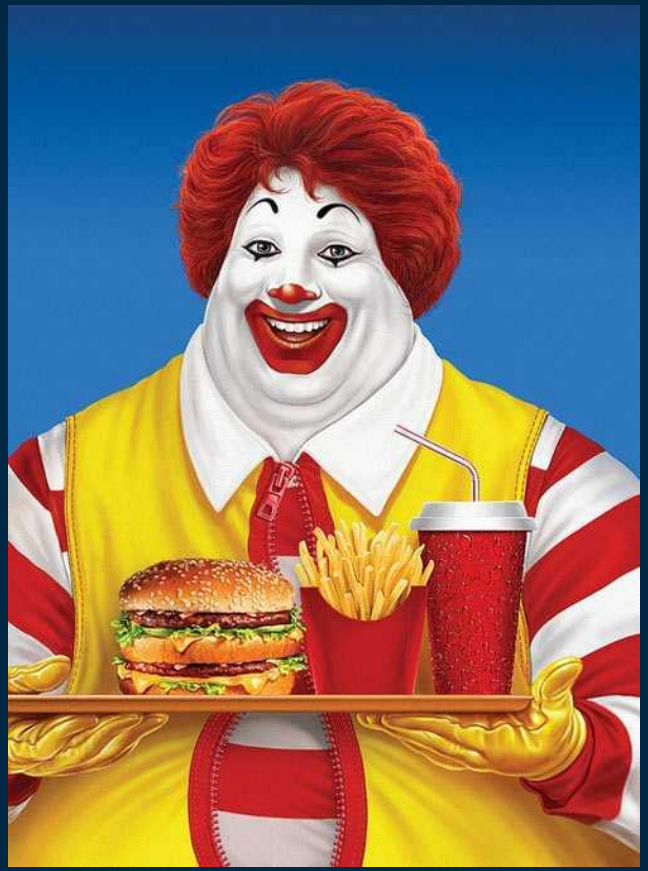
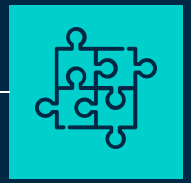


Table of Contents (1/2)



01 INTRODUCTION

02 PROBLEMATIC

03 DATA

Table of Contents (2/2)

Introduction

Dataset: Eating habits and physical conditions

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

Problematic

Can we estimate people’s obesity category from their eating habits and physical condition ?

Data Preprocessing

With the help of this website

Step 1

Understanding the features

Step 2

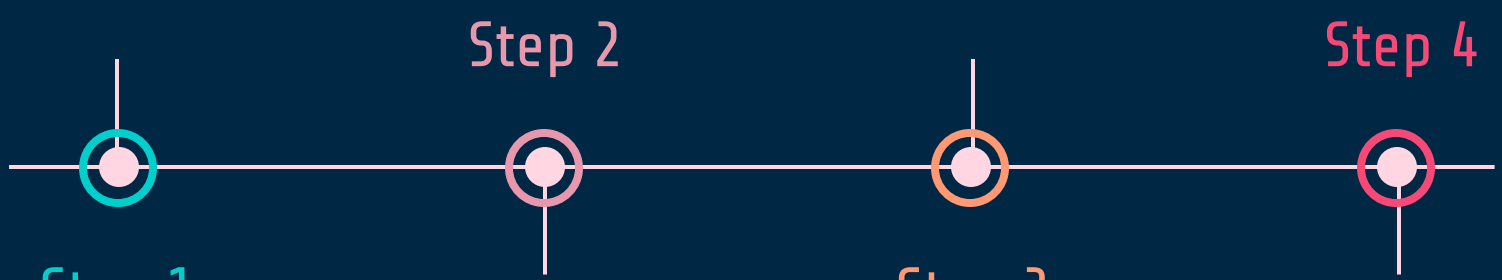
Renaming the features

Step 3

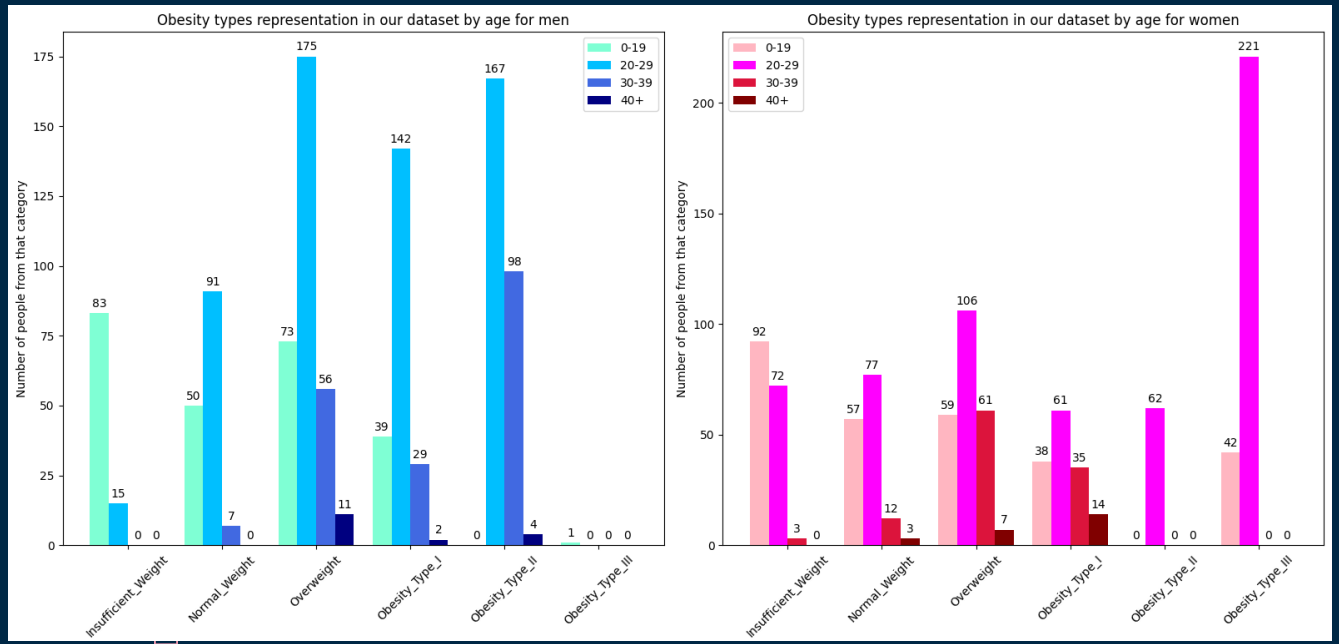
Cleaning the data

Step 4

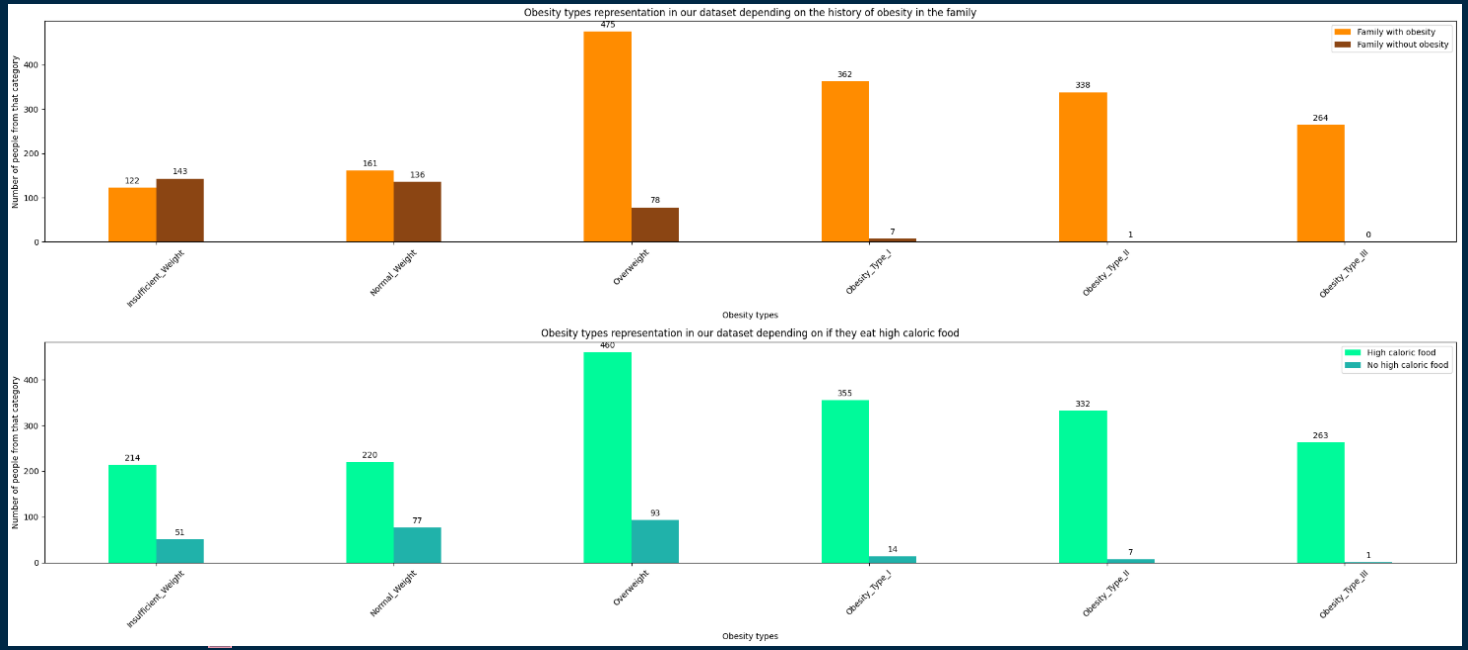
Selecting the data



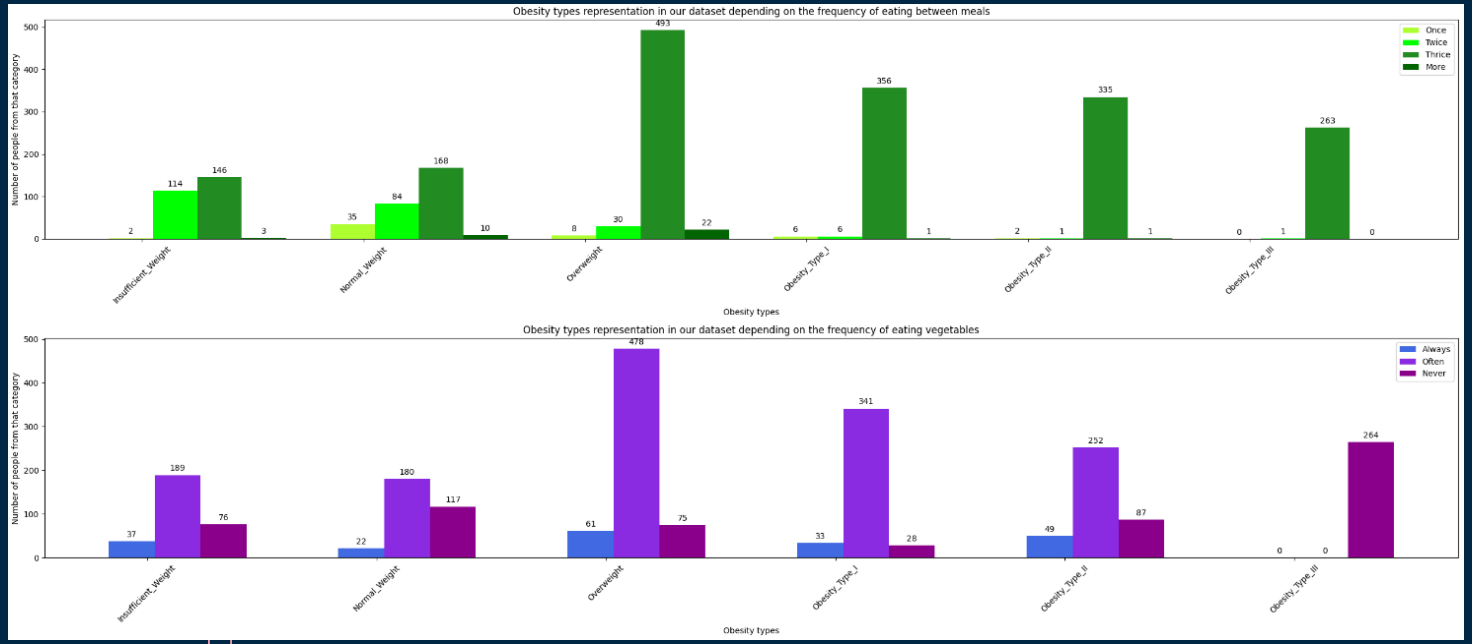
Data Visualisation (1/4)



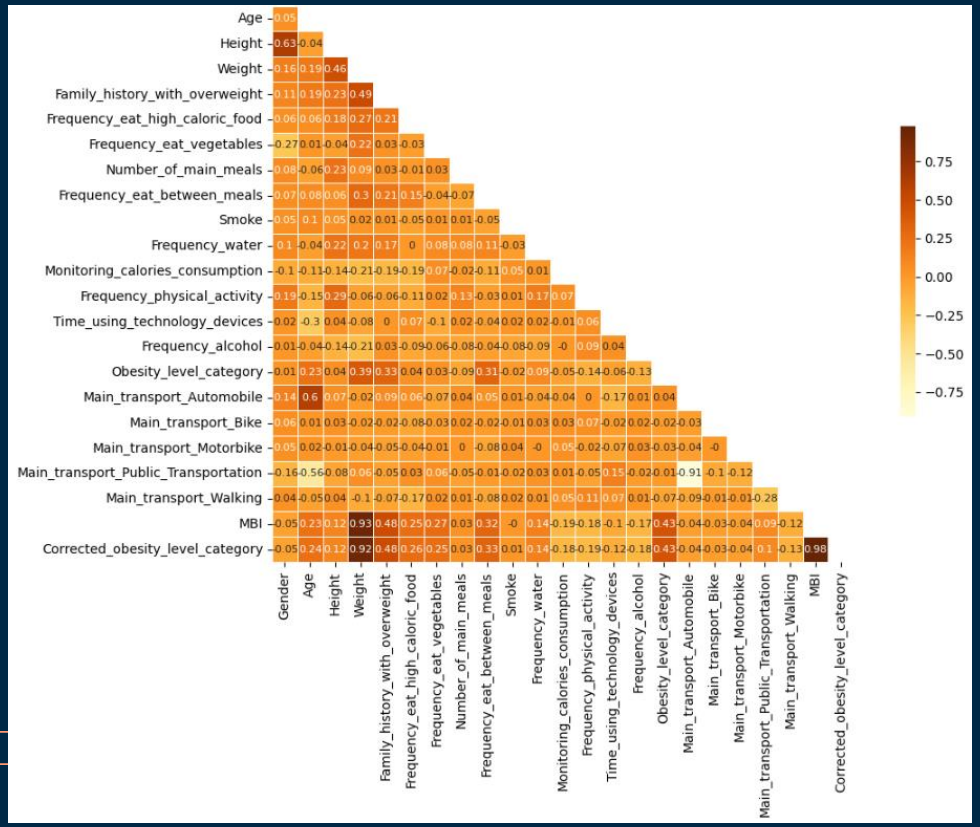
Data Visualisation (2/4)



Data Visualisation (3/4)



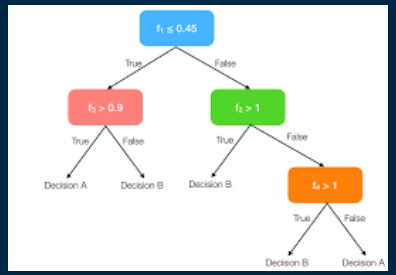
Data Visualisation (4/4)



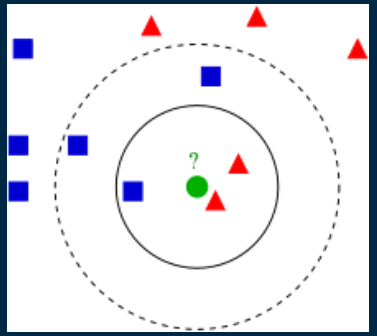
Prediction Model (1/5)

- Objective : Predict someone IBM category based on his eating habits and physical conditions
- Quick test on 4 different algorithms

Prediction Model (2/5)

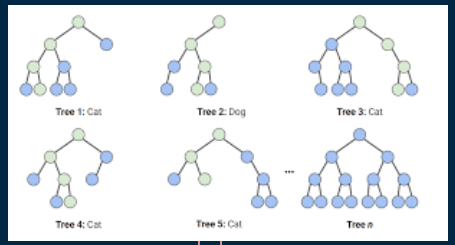


accuracy 0.71

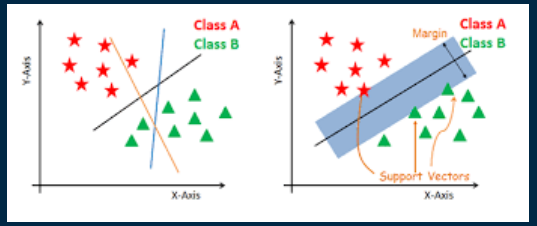


accuracy 0.74

We chose to go with the Random Forest Classifier



accuracy 0.86



accuracy 0.72

Prediction Model (3/5)

- Searching for the best parameters

```
Best Parameters: {'max_depth': 20, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 50}
Best Accuracy on Test Set: 0.8229665071770335
```

- Making sure our model generalize

```
Cross-validation scores: [0.80838323 0.80838323 0.82634731 0.8502994  0.76047904 0.82634731
 0.80838323 0.86826347 0.85628743 0.81927711]
Mean cross-validation score: 0.8232450761128345
```

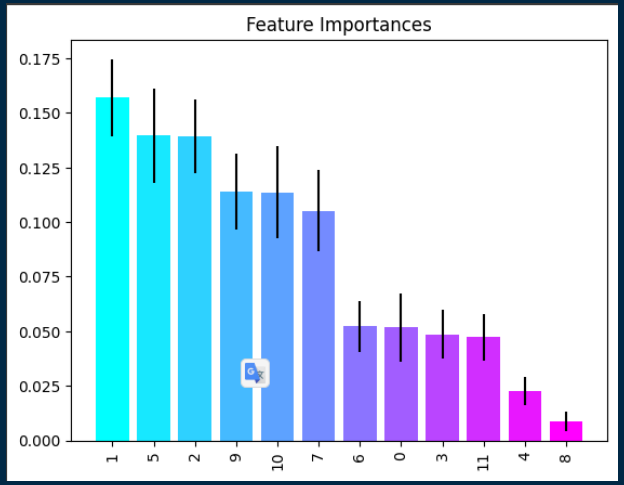
Prediction Model (4/5)

- Test on ourselves

```
[['Louis' 'Killian' 'Marc']
 ['20.679012345679013' '19.047004132231404' '21.469150482230148']
 ['Normal_Weight' 'Normal_Weight' 'Normal_Weight']]
```

Prediction Model (5/5)

- Observations about our features



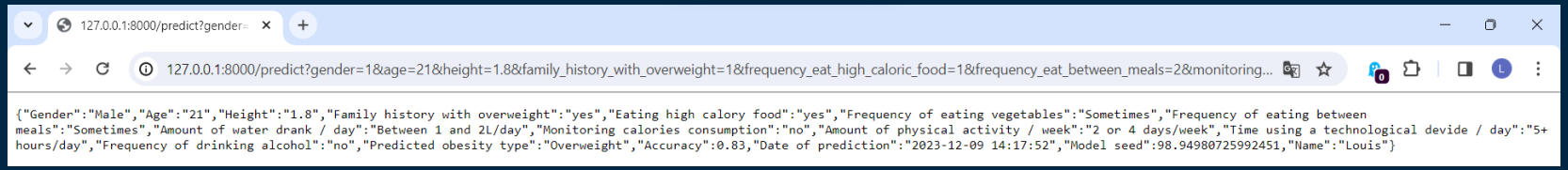
1. Age
5. Frequency eating vegetables
2. Height
9. Frequency physical activity
10. Time using technology devices
7. Frequency water
6. Frequency eat between meals
0. Gender
3. Family history with overweight
11. Frequency alcohol
4. Frequency eat high caloric food
8. Monitoring calories consumption

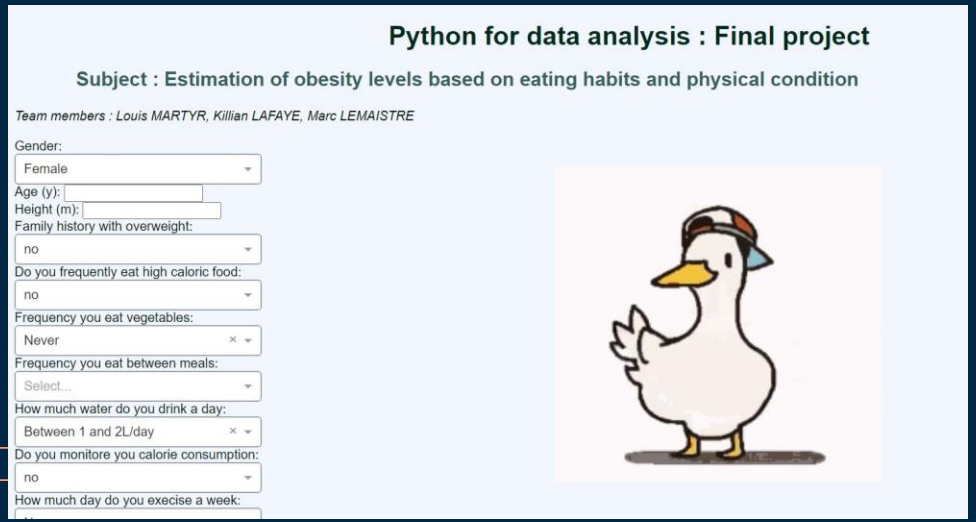
Conclusion

Can we estimate people's obesity category from their eating habits and physical condition ?

- Some features do have an impact
- Some obesity problem can be solved by modifying some eating habits

API & Bonus


127.0.0.1:8000/predict?gender=
127.0.0.1:8000/predict?gender=1&age=21&height=1.8&family_history_with_overweight=1&frequency_eat_high_caloric_food=1&frequency_eat_between_meals=2&monitoring...
{"Gender":"Male","Age":"21","Height":"1.8","Family history with overweight":"yes","Eating high calory food":"yes","Frequency of eating vegetables":"Sometimes","Frequency of eating between meals":"Sometimes","Amount of water drank / day":"Between 1 and 2L/day","Monitoring calories consumption":"no","Amount of physical activity / week":"2 or 4 days/week","Time using a technological devide / day":"5+ hours/day","Frequency of drinking alcohol":"no","Predicted obesity type":"Overweight","Accuracy":0.83,"Date of prediction":"2023-12-09 14:17:52","Model seed":98.94980725992451,"Name":"Louis"}


Python for data analysis : Final project
Subject : Estimation of obesity levels based on eating habits and physical condition
Team members : Louis MARTYR, Killian LAFAYE, Marc LEMAISTRE
Gender:
Female
Age (y):
Height (m):
Family history with overweight:
no
Do you frequently eat high caloric food:
no
Frequency you eat vegetables:
Never
Frequency you eat between meals:
Select...
How much water do you drink a day:
Between 1 and 2L/day
Do you monitore you calorie consumption:
no
How much day do you execise a week:

Area for improvements

- Data acquisition
- Dataset size

The survey

Pros

- Easy to ask
- Same poll for everyone
- Easy to numerize

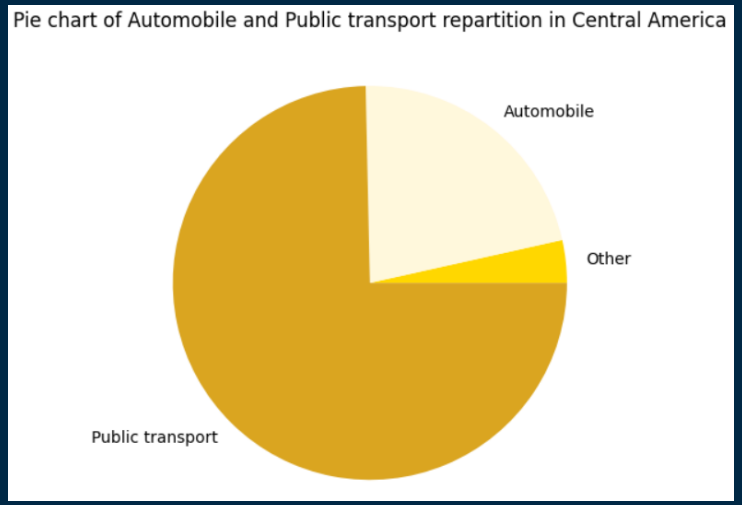
¿What is your gender?	• Female • Male
¿what is your age?	Numeric value
¿what is your height?	Numeric value in meters
¿what is your weight?	Numeric value in kilograms

Cons

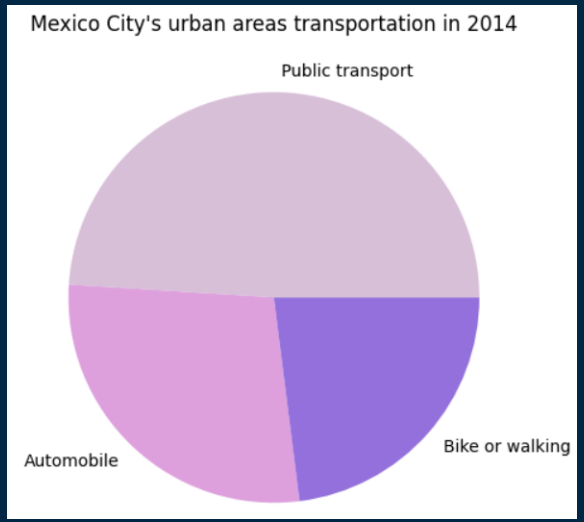
- Hard to answer
- Hard to understand
- Not representative

¿Which transportation do you usually use?	• Automobile • Motorbike • Bike

Not representative

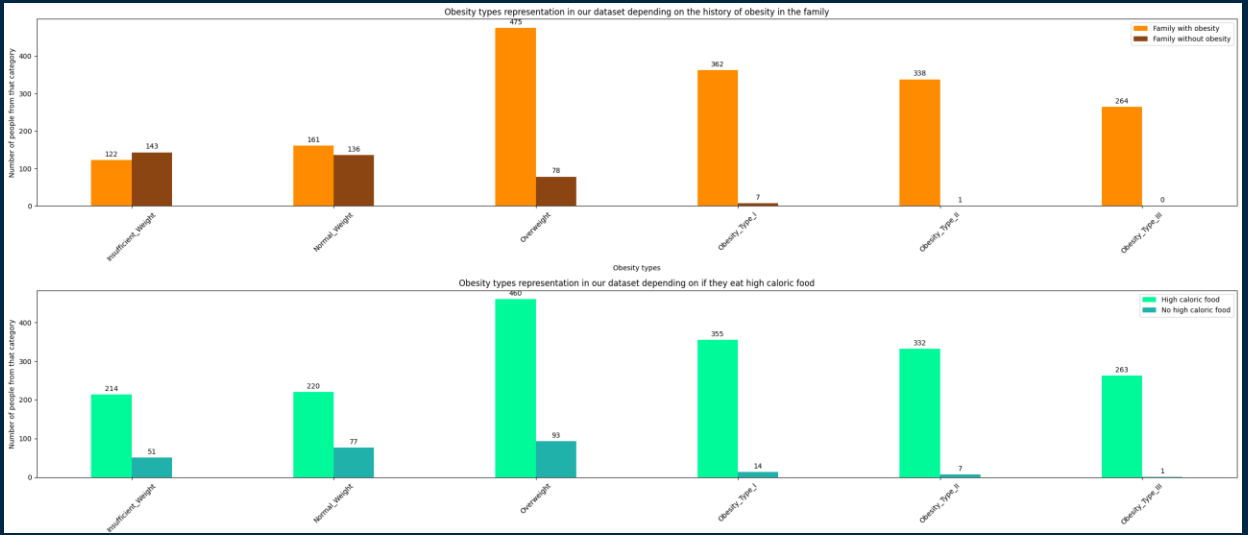


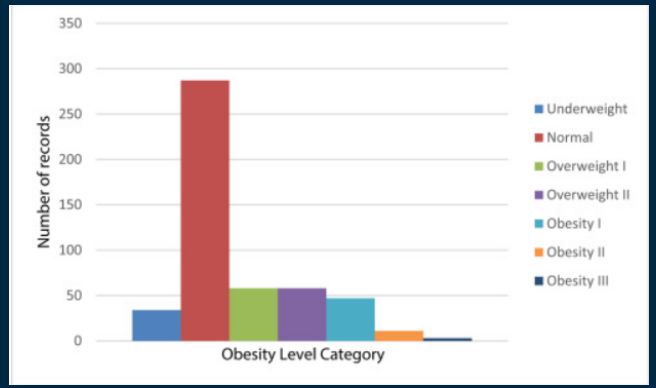
Based on our data



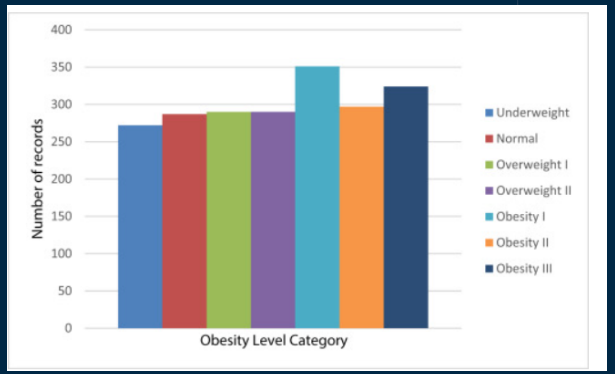
Reality

Oversampling incorrectly made





oversampling



Bibliography

- Our dataset: https://archive.ics.uci.edu/dataset/544/estimation+of+obesity+levels+based+on+eating+habits+and+physical+condition

- Obesity data: https://data.worldobesity.org/rankings/?age=c&sex=m

- Complementary information : https://doi.org/10.1016/j.dib.2019.104344