Projet Python for Data Analyse

A readme summarizing the task to be accomplished and your conclusions

We split the data set into 3 categories:

* The first one contains every attribute related with eating habits:
  + Frequent consumption of high caloric food (FAVC).
  + Frequency of consumption of vegetables (FCVC)
  + Number of main meals (NCP)
  + Consumption of food between meals (CAEC)
  + Consumption of water daily (CH20)
  + Consumption of alcohol (CALC)
* the second one contains every attribute related with physical condition:
  + Calorie’s consumption monitoring (SCC)
  + Physical activity frequency (FAF)
  + Time using technology devices (TUE)
  + Transportation used (MTRANS)
  + Smoke
* And finally, all the other variables:
  + Gender
  + Age
  + Height
  + Weight
  + Family\_history\_with\_overweight

All this data is classed into 7 categories in the variable **NObeyesdad**: Insufficient Weight, Normal Weight, Overweight level I, Overweight level II, Obesity Type I, Obesity type II and Obesity Type III

TASKS:

* Data cleaning (delete NaN values)
* Analysis of the Data with graphs
* Prepare the dataset for machine learning (put nonnumeric data into numbers/numeric data)

**Gender:** male = 0, female = 1

**Family\_history\_with\_overweight**: yes = 1, no = 0

**FAVC** (Frequent consumption of high caloric food): yes = 1, no = 0

**SCC** (Calorie’s consumption monitoring): no = 0, yes = 1

**SMOKE**: no = 0, yes = 1

**CAEC** (Consumption of food between meals): n0 = 0, Sometimes = 1, Frequently = 2, Always = 3

**CALC** (Consumption of alcohol): no = 0, Sometimes = 1, Frequently = 2, Always = 3

**MTRANS** (Transportation used): Automobile = 0, Motorbike = 1, Bike = 2, Public\_Transportation = 3, Walking = 4

**NObeyesdad** (target variable): Insufficient Weight = 0, Normal Weight = 1, Overweight level I = 2, Overweight level II = 3, Obesity Type I = 4, Obesity type II = 5, and Obesity Type III = 6.

“The data contains numerical data and continuous data, so it can be used for analysis based on algorithms of **classification**, **prediction**, **segmentation** and **association**.”

N ------------------------------------

ANALYSE DONNEES:

,,,,,,,,,,,,,,,,

|  |  |  |  |
| --- | --- | --- | --- |
| Nom | Type | Signification | Groupe |
| Gender | Qualitative | Gender | informatif |
| Age | Quantitative | Age | informatif |
| Height | Quantitative | Height | informatif |
| Weight | Quantitative | Weight | informatif |
| family\_history\_with\_overweight | Qualitative | family\_history\_with\_overweight | informatif |
| FAVC |  | Frequent consumption of high caloric food |  |
| FCVC |  | Frequency of consumption of vegetables |  |
| NCP |  | Number of main meals |  |
| CAEC |  | Consumption of food between meals |  |
| SMOKE |  |  |  |
| CH2O |  | Consumption of water daily |  |
| SCC |  | Calorie’s consumption monitoring |  |
| FAF |  | Physical activity frequency |  |
| TUE |  | Time using technology devices |  |
| CALC |  | Consumption of alcohol |  |
| MTRANS |  | Transportation used |  |
| NObeyesdad |  | target variable |  |

Ou as-tu trouvé les explications sur les variables?

Faire des groupes de variables?

Variables quantitative et qualitative.

Pour l’analyse :

I PRESENTATION

II ANALYSE

**III Classification of the type of Obesity.**

III CONCLUSION

**I PRESENTATION**

Presenter les données

Analyse des différentes populations d’obèses, explications des variables, (moyenne médiane écart type pour chaque population de type d’obèses)

* Combien de colonnes et combien de lignes. Faire des groupes pour les colonnes
* Peut etre pour chaque type de NObeyesdad

Distribution (moyenne, medianne +quartille + ecart type) des variables grises :

|  |
| --- |
| Gender |
| Age |
| Height |
| Weight |
| family\_history\_with\_overweight |

**Code couleur :**

**Romain**

**Leo**

**I PRESENTATION**

-présenter les données

-nettoyer les données

-encodage des données



**II ANALYSE**

Analyses des données, qu’elles variables ont le plus d’impacts ?

Var négative et var positive

* Pour chaque groupe laquelle a le plus de repercussion sur l’obésite
* Bonus: corrélation entre t’elle varaible

Pourcentage des gens qui fument qui sont en surpoids

Pourcentage des gens qui fument qui sont en surpoids

Pourcentage des gens qui fument qui sont en surpoids

Pourcentage des gens qui fument qui sont en surpoids

**III Classification of the type of Obesity.**

Partie modélisation

Matrice co-varaince

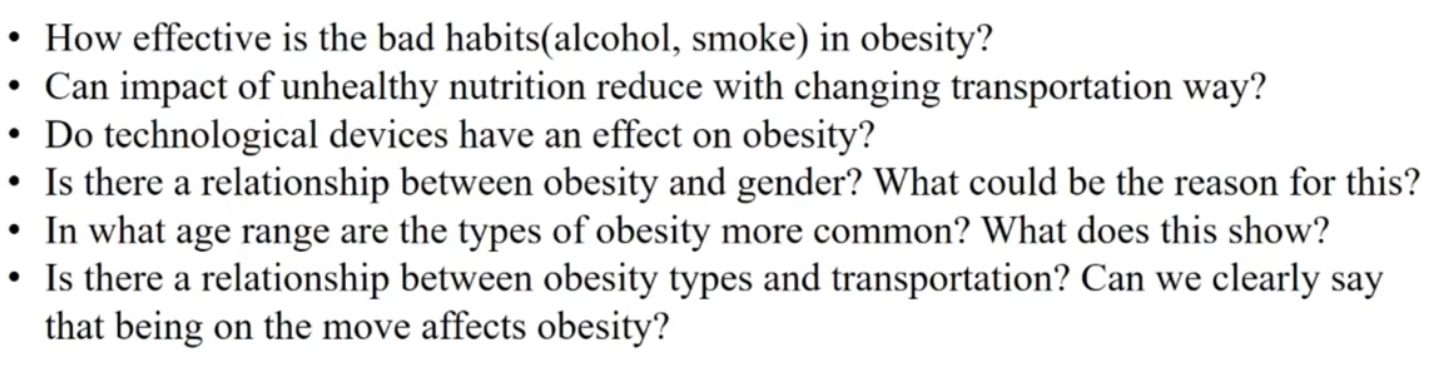
Prédire les meilleures paramètres

**III.2 FLASK**

**API FLASK**

**III CONCLUSION**

A voir^^



Mass body index = Weight/(height\*height)

After all calculation was made to obtain the mass body index for each individual, the results were compared with the data provided by WHO and the Mexican Normativity.

* Underweight Less than 18.5
* Normal 18.5 to 24.9
* Overweight 25.0 to 29.
* Obesity I 30.0 to 34.9
* Obesity II 35.0 to 39.9
* Obesity III Higher than 40

https://www.youtube.com/watch?v=FGz2mBCA1Jo

ID of the subject:

* + Gender
  + Age
  + Height
  + Weight
  + Family\_history\_with\_overweight

Harmful addictions:

* + Smoke
  + Consumption of alcohol (CALC)

Good eating habits:

* + Frequency of consumption of vegetables (FCVC)
  + Consumption of water daily (CH20)

Bad eating habits:

* + Frequent consumption of high caloric food (FAVC)
  + Consumption of food between meals (CAEC)

Quantification of food consomption :

* + Number of main meals (NCP)
  + Calorie’s consumption monitoring (SCC)

Lifestyle habits:

* + Physical activity frequency (FAF)
  + Time using technology devices (TUE)
  + Transportation used (MTRANS)

We decided to threat the subject of obesity.

The dataset we use to threat this subject is a dataset containing much information’s about a person like her eating habits, her physical conditions and other basic information like the gender, age, height, and weight. The dataset also gives use the level of obesity of this person.

Our goal was to determine the factors that can bring a person to be in a certain level of obesity.

First, we verified if the data had to be cleaned however the data was already ready to be used.

Then, we made an analysis of the different variables to see from what is formed the population we were going to analyze.

After this preliminary analysis we studied the different obesity type groups to see which factor could be an important obesity factor.

To confirm our analysis of the obesity factor we made a

Finally, we created a classification model of the types of obesity.