

## Python Food register analysis

## Statistics file

```

Created on Fri Oct 16 16:15:19 2020

@author: jlms3
"""
# José Luis Madrigal
# A01745419
import pandas as pd
from matplotlib import pyplot as plt
doc = pd.read_excel("A01745419_comidas.xlsx", sheet_name="Sheet1")
doc.drop(doc.columns[0:3], axis=1, inplace=True)
print(doc)

plt.style.use("fivethirtyeight")
plt.hist(doc["Calorias (kcal)"], color = "orange", edgecolor = "black", bins=15)
plt.title("Calorias (kcal)")
plt.show()

plt.style.use("fivethirtyeight")
plt.hist(doc["Carbohidratos (g)"], color = "blue", edgecolor = "black", bins=15)
plt.title("Carbohidratos (g)")
plt.show()

plt.style.use("fivethirtyeight")
plt.hist(doc["Lípidos (g)"], color = "red", edgecolor = "black", bins=15)
plt.title("Lípidos (g)")
plt.show()

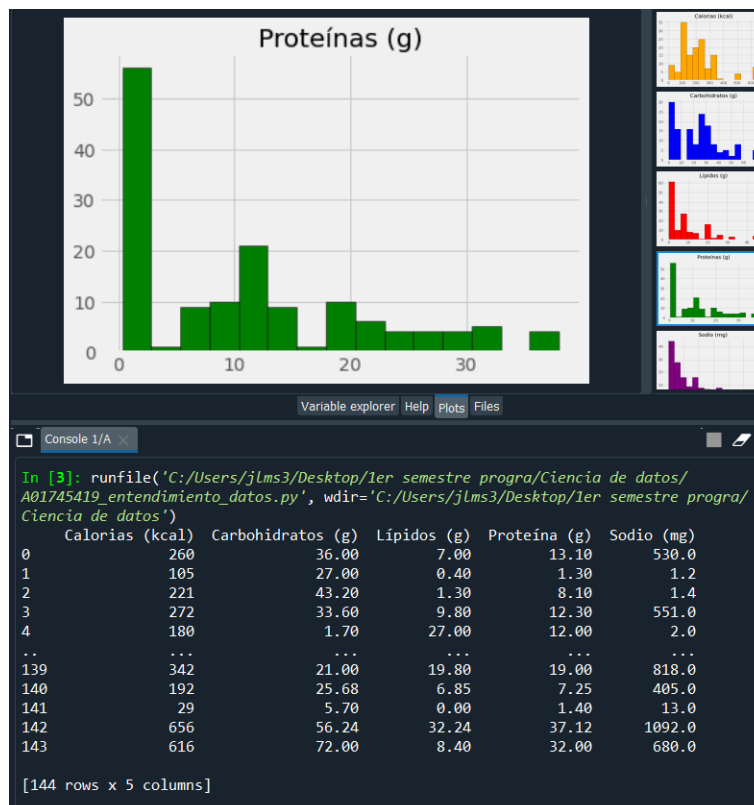
plt.style.use("fivethirtyeight")
plt.hist(doc["Proteína (g)"], color = "green", edgecolor = "black", bins=15)
plt.title("Proteínas (g)")
plt.show()

plt.style.use("fivethirtyeight")
plt.hist(doc["Sodio (mg)"], color = "purple", edgecolor = "black", bins=15)
plt.title("Sodio (mg)")
plt.show()

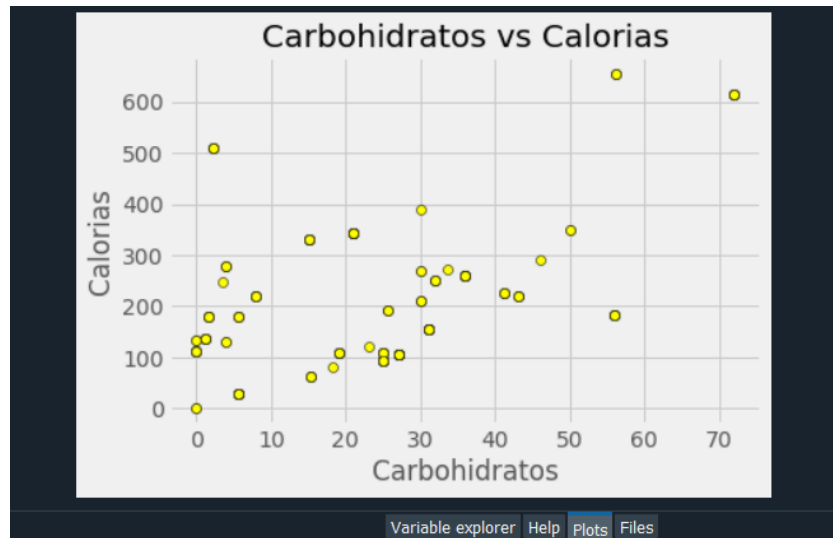
plt.style.use("fivethirtyeight")
plt.scatter(doc["Carbohidratos (g)"], doc["Calorias (kcal)"], color = "yellow", edgecolor = "black")
plt.title("Carbohidratos vs Calorias")

```

## Graphs of each biomolecule



### Graph one biomolecule vs calories



### Machine learning file

```
"""
Created on Tue Nov 10 09:39:21 2020

@author: jlms3
"""
import pandas as pd
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
import statsmodels as sm
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error

df = pd.read_excel("A01745419_comidas.xlsx")

#Variables X y y
features=["Carbohidratos (g)", "Lípidos (g)", "Proteína (g)", "Sodio (mg)"]

X = df[features]

features_sinsodio=["Carbohidratos (g)", "Lípidos (g)", "Proteína (g)"]
X_sinsodio =df[features_sinsodio]

# extraemos mpg como nuestra variable dependiente
y = df["Calorias (kcal)"]

# con machine learning

# separamos nuestra data dos partes para entrenar y para hacer el test.
X_train,X_test,y_train,y_test = train_test_split(X_sinsodio,y,test_size=.3,random_state=0)

model = LinearRegression() # iniciamos el modelo de LinearRegression
model.fit(X_train,y_train) # entrenamos el modelo con lo obtenido anteriormente

predictions = model.predict(X_test)

# mostramos el puntaje del entrenamiento
print('Puntaje entrenamiento: {}'.format(model.score(X_train,y_train)))
print('Puntaje Test: {}'.format(model.score(X_test,y_test)))
```

## Predictions

