import numpy as np

import pandas as pd

import seaborn as sb

import matplotlib.pyplot as plt

%matplotlib inline

from mpl\_toolkits.mplot3d import Axes3D

from matplotlib import cm

plt.rcParams['figure.figsize'] = (16, 9)

plt.style.use('ggplot')

from sklearn import linear\_model

from sklearn.metrics import mean\_squared\_error, r2\_score

#cargamos los datos de entrada

data = pd.read\_csv("./alex.csv")

#veamos cuantas dimensiones y registros contiene

data.shape

# Ahora veamos algunas estadísticas de nuestros datos

data.describe()

# Visualizamos rápidamente las caraterísticas de entrada

data.drop(['fecha','cantidad', 'cantidad'],1).hist()

plt.show()

filtered\_data **=** data[(data['cantidad'] **<=** 3500) **&** (data['edad'] **<=** 80000)]

colores**=**['orange','blue']

tamanios**=**[30,60]

f1 **=** filtered\_data['cantidad'].values

f2 **=** filtered\_data['edad'].values

*# Vamos a pintar en colores los puntos por debajo y por encima de la media de Cantidad de Palabras*

asignar**=**[]

**for** index, row **in** filtered\_data.iterrows():

**if**(row['cantidad']**>**1808):

asignar.append(colores[0])

**else**:

asignar.append(colores[1])

plt.scatter(f1, f2, c**=**asignar, s**=**tamanios[0])

plt.show()

# Asignamos nuestra variable de entrada X para entrenamiento y las etiquetas Y.

dataX =filtered\_data[["cantidad"]]

X\_train = np.array(dataX)

y\_train = filtered\_data['edad'].values

# Creamos el objeto de Regresión Linear

regr = linear\_model.LinearRegression()

# Entrenamos nuestro modelo

regr.fit(X\_train, y\_train)

# Hacemos las predicciones que en definitiva una línea (en este caso, al ser 2D)

y\_pred = regr.predict(X\_train)

# Veamos los coeficienetes obtenidos, En nuestro caso, serán la Tangente

print('Coefficients: \n', regr.coef\_)

# Este es el valor donde corta el eje Y (en X=0)

print('Independent term: \n', regr.intercept\_)

# Error Cuadrado Medio

print("Mean squared error: %.2f" % mean\_squared\_error(y\_train, y\_pred))

# Puntaje de Varianza. El mejor puntaje es un 1.0

#Vamos a comprobar:

# Quiero predecir cuántos "Shares" voy a obtener por un artículo con 2.000 palabras,

# según nuestro modelo, hacemos:

y\_Dosmil = regr.predict([[2000]])

print(int(y\_Dosmil))

#Vamos a comprobar:

# Quiero predecir cuántos "Shares" voy a obtener por un artículo con 2.000 palabras,

# según nuestro modelo, hacemos:

y\_Dosmil = regr.predict([[2000]])

print(int(y\_Dosmil))

from matplotlib import pyplot

def f1(x):

return (0.35347139\*x) + 34.45994729287628

x = range(-10, 15)

pyplot.plot(x, [f1(i) for i in x])

pyplot.axhline(0, color="black")

pyplot.xlim(-10, 10)

pyplot.savefig("output.png")

pyplot.show()