**Muestreo Estratificado en Python**

Yo lo entendí de esta forma:

1. Separa los datos por categorías (valores estratificados).
2. Define las categorías que quieres en tú muestro y su proporción.
3. Toma tantos datos como necesites de cada categoría para llegar a la proporción deseada.
4. Junta esas muestras de cada categoría para tu muestreo final.

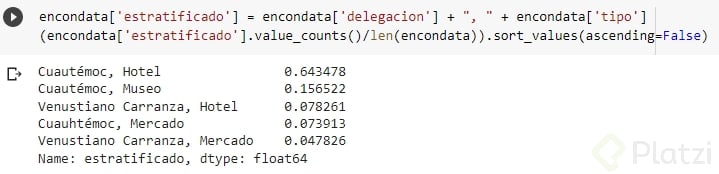
**Explicación del ejemplo**

Se crean dichas categorías a partir de la delegación y el tipo, esta información se almacena en la columna **estratificado**:

econdata['estratificado'] = econdata['delegacion'] + "," + econdata['tipo']

(econdata['estratificado'].value\_counts() / len(econdata)).sort\_values(ascending=**False**)

Con esto ya tenemos los datos por categorías (paso 1) y podemos ver su proporción real:



Para nuestro muestreo definimos las siguientes proporciones (paso 2):

* Cuautémoc, Hotel: 0.5,
* Cuautémoc, Museo: 0.2,
* Venustiano Carranza, Hotel: 0.1,
* Cuauhtémoc, Mercado: 0.1,
* Venustiano Carranza, Mercado: 0.1

Para crear dicha proporción a partir de los datos originales se utiliza la siguiente función que encuentra el número de muestras necesario de cada categoría (paso 3), hace un sampleo simple dentro de dicha categoria y luego junta esas muestras (paso 4):

**def** **data\_estratificada**(econdata, nombres\_columnas\_estrat, valores\_estrat, prop\_estrat, random\_state=**None**):

df\_estrat = pd.DataFrame(columns = econdata.columns)

pos = -1

**for** i **in** range(len(valores\_estrat)):

pos += 1

**if** pos == len(valores\_estrat) - 1:

ratio\_len = len(econdata) - len(df\_estrat)

**else**:

ratio\_len = int(len(econdata) \* prop\_estrat[i])

df\_filtrado = econdata[econdata[nombres\_columnas\_estrat] ==valores\_estrat[i]]

df\_temp = df\_filtrado.sample(replace=**True**, n=ratio\_len, random\_state=random\_state)

df\_estrat = pd.concat([df\_estrat, df\_temp])

**return** df\_estrat

Utiliza este código para crear el sampleo:

valores\_estrat = ['Cuautémoc,Hotel', 'Cuautémoc,Museo', 'Venustiano Carranza,Hotel', 'Cuauhtémoc,Mercado','Venustiano Carranza,Mercado']

prop\_estrat = [0.5, 0.2, 0.1, 0.1, 0.1]

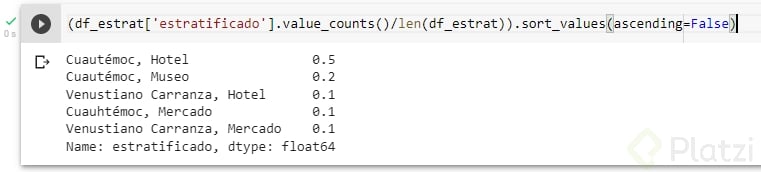
df\_estrat = data\_estratificada(econdata, 'estratificado', valores\_estrat, prop\_estrat, random\_state=42)

df\_estrat

Adicionalmente, puedes usar este código para ver la proporción final de tus datos:

(df\_estrat['estratificado'].value\_counts()/len(df\_estrat)).sort\_values(ascending=**False**)

Si todo salió bien, deberías ver algo similar a esto:



Con esto ya tienes una muestra en las proporciones que deseábamos en un principio.