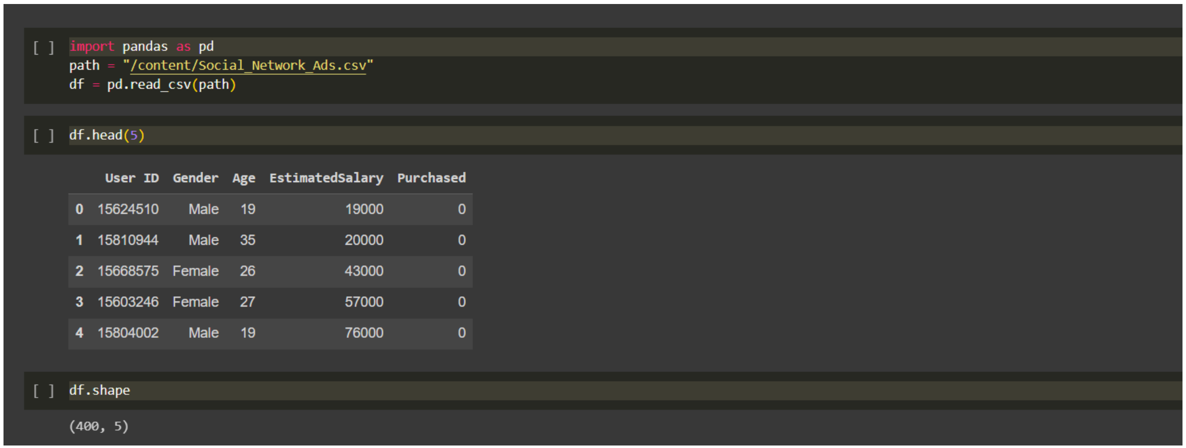
Brenda Estefania Castillo Fernandez

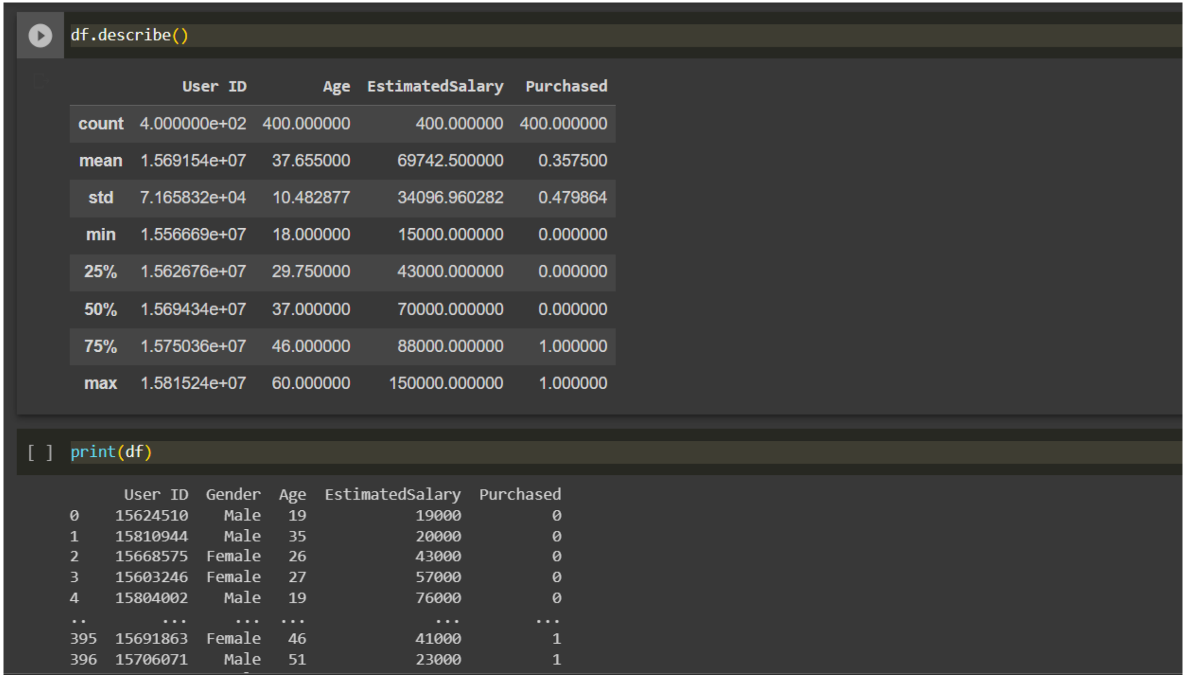


```
[ ] import pandas as pd
    path = "/content/Social_Network_Ads.csv"
    df = pd.read_csv(path)
```

```
[ ] df.head(5)
```

	User ID	Gender	Age	EstimatedSalary	Purchased
0	15624510	Male	19	19000	0
1	15810944	Male	35	20000	0
2	15668575	Female	26	43000	0
3	15603246	Female	27	57000	0
4	15804002	Male	19	76000	0

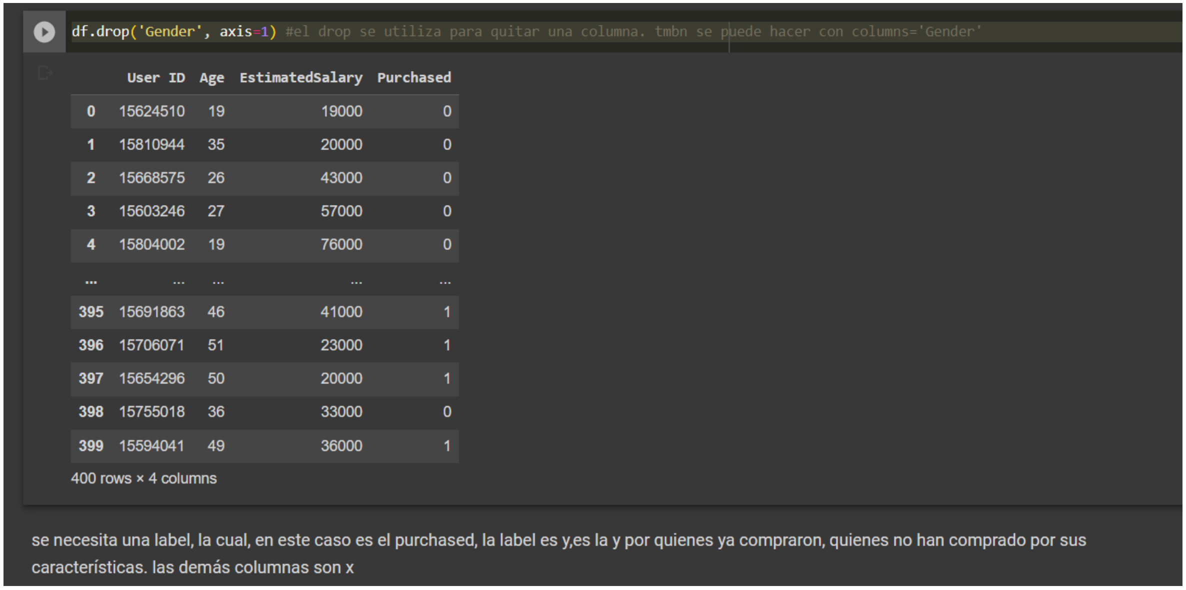
```
[ ] df.shape

(400, 5)
```



```
df.describe()
```

	User ID	Age	EstimatedSalary	Purchased
count	4.000000e+02	400.000000	400.000000	400.000000
mean	1.569154e+07	37.655000	69742.500000	0.357500
std	7.165832e+04	10.482877	34096.960282	0.479864
min	1.556669e+07	18.000000	15000.000000	0.000000
25%	1.562676e+07	29.750000	43000.000000	0.000000
50%	1.569434e+07	37.000000	70000.000000	0.000000
75%	1.575036e+07	46.000000	88000.000000	1.000000
max	1.581524e+07	60.000000	150000.000000	1.000000

```
[ ] print(df)

      User ID  Gender  Age  EstimatedSalary  Purchased
0    15624510    Male   19            19000          0
1    15810944    Male   35            20000          0
2    15668575  Female   26            43000          0
3    15603246  Female   27            57000          0
4    15804002    Male   19            76000          0
..        ...     ...  ...              ...        ...
395  15691863  Female   46            41000          1
396  15706071    Male   51            23000          1
```



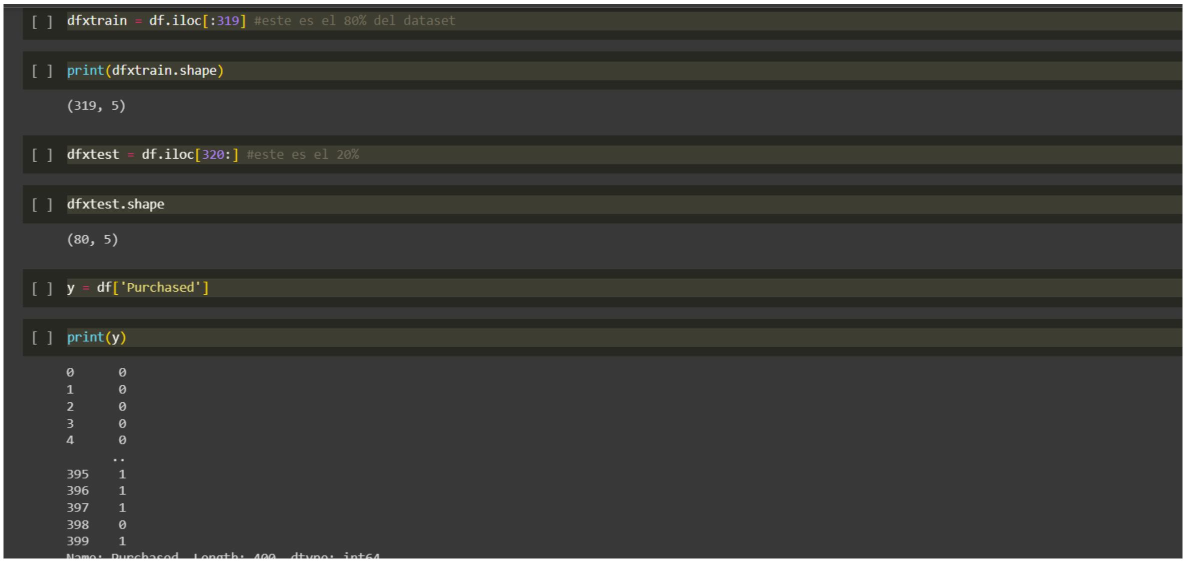
```
df.drop('Gender', axis=1) #el drop se utiliza para quitar una columna. tmbn se puede hacer con columns='Gender'
```

	User ID	Age	EstimatedSalary	Purchased
0	15624510	19	19000	0
1	15810944	35	20000	0
2	15668575	26	43000	0
3	15603246	27	57000	0
4	15804002	19	76000	0
...	...	...	...	...
395	15691863	46	41000	1
396	15706071	51	23000	1
397	15654296	50	20000	1
398	15755018	36	33000	0
399	15594041	49	36000	1

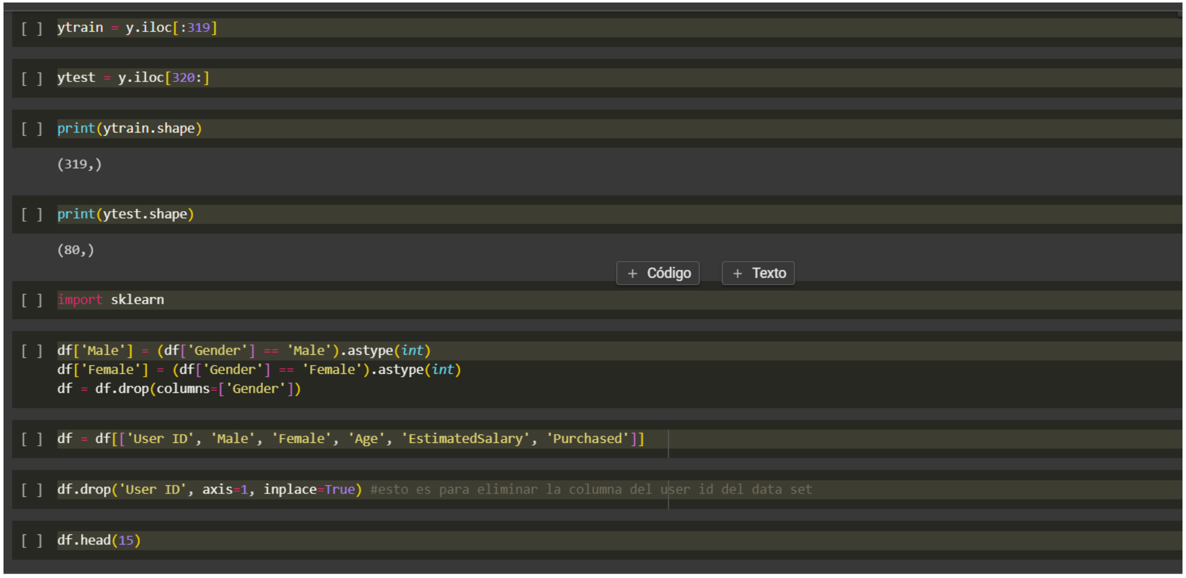
400 rows × 4 columns

se necesita una label, la cual, en este caso es el purchased, la label es y,es la y por quienes ya compraron, quienes no han comprado por sus características. las demás columnas son x

Brenda Estefania Castillo Fernandez

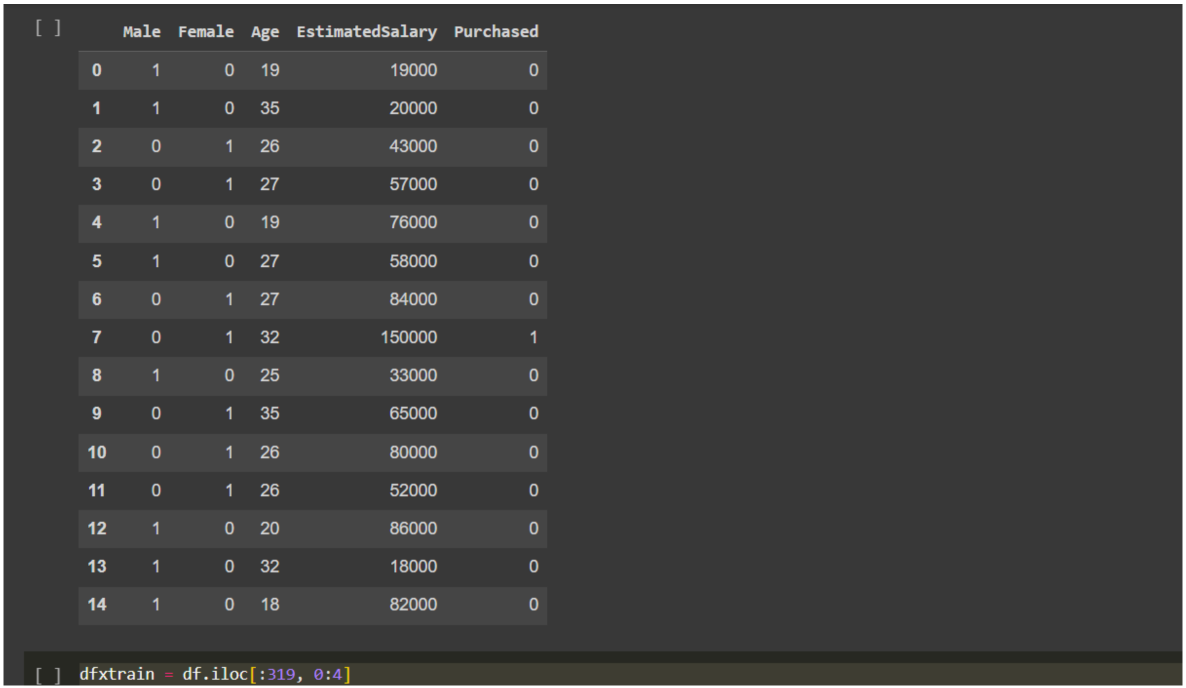


```
[ ] dfxtrain = df.iloc[:319] #este es el 80% del dataset

[ ] print(dfxtrain.shape)

    (319, 5)

[ ] dfxtest = df.iloc[320:] #este es el 20%

[ ] dfxtest.shape

    (80, 5)

[ ] y = df['Purchased']

[ ] print(y)

    0      0
    1      0
    2      0
    3      0
    4      0
          ..
    395    1
    396    1
    397    1
    398    0
    399    1
    Name: Purchased, Length: 400, dtype: int64
```



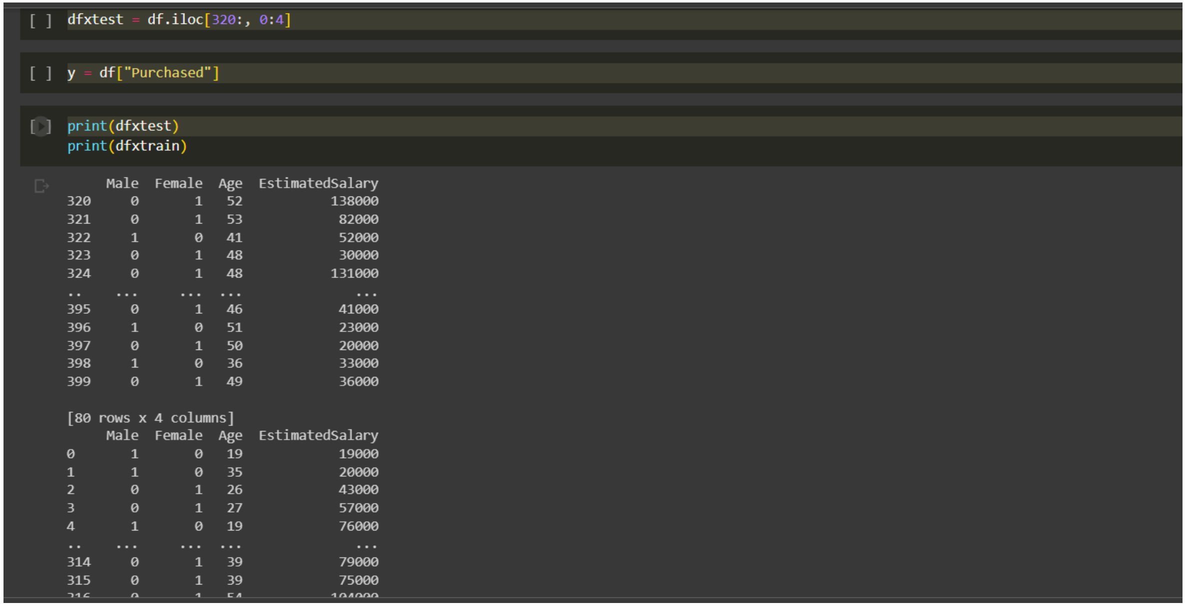
```
[ ] ytrain = y.iloc[:319]

[ ] ytest = y.iloc[320:]

[ ] print(ytrain.shape)

    (319,)

[ ] print(ytest.shape)

    (80,)

[ ] import sklearn

[ ] df['Male'] = (df['Gender'] == 'Male').astype(int)
    df['Female'] = (df['Gender'] == 'Female').astype(int)
    df = df.drop(columns=['Gender'])

[ ] df = df[['User ID', 'Male', 'Female', 'Age', 'EstimatedSalary', 'Purchased']]

[ ] df.drop('User ID', axis=1, inplace=True) #esto es para eliminar la columna del user id del data set

[ ] df.head(15)
```

Brenda Estefania Castillo Fernandez



	Male	Female	Age	EstimatedSalary	Purchased
0	1	0	19	19000	0
1	1	0	35	20000	0
2	0	1	26	43000	0
3	0	1	27	57000	0
4	1	0	19	76000	0
5	1	0	27	58000	0
6	0	1	27	84000	0
7	0	1	32	150000	1
8	1	0	25	33000	0
9	0	1	35	65000	0
10	0	1	26	80000	0
11	0	1	26	52000	0
12	1	0	20	86000	0
13	1	0	32	18000	0
14	1	0	18	82000	0

```
dfxtrain = df.iloc[:319, 0:4]
```



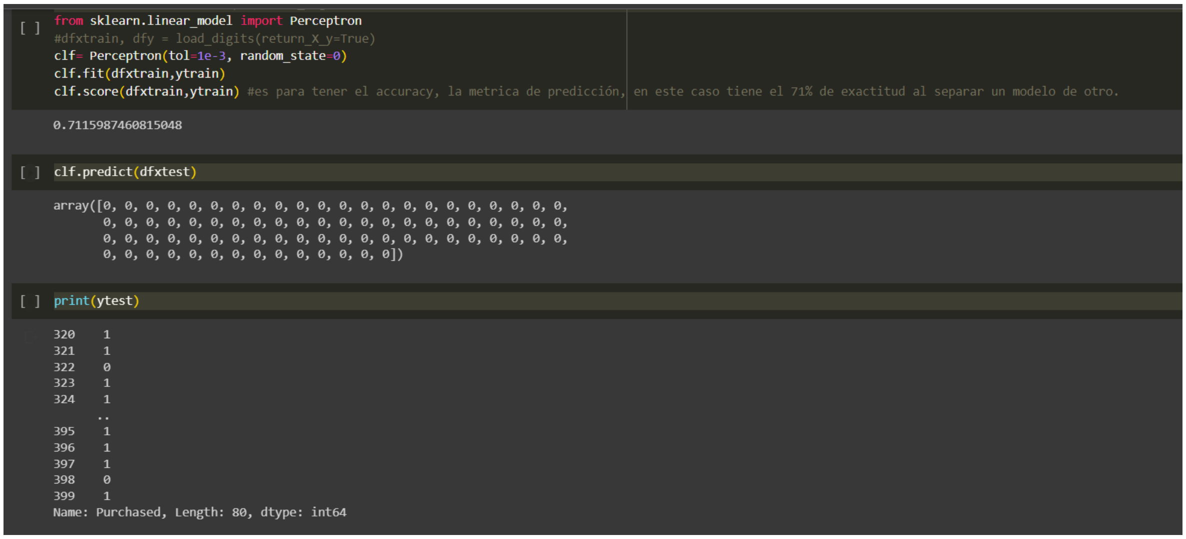
```
dfxtest = df.iloc[320:, 0:4]
```

```
y = df["Purchased"]
```

```
print(dfxtest)
print(dfxtrain)
```

```
     Male  Female  Age  EstimatedSalary
320     0       1   52           138000
321     0       1   53            82000
322     1       0   41            52000
323     0       1   48            30000
324     0       1   48           131000
..    ...     ...  ...              ...
395     0       1   46            41000
396     1       0   51            23000
397     0       1   50            20000
398     1       0   36            33000
399     0       1   49            36000

[80 rows x 4 columns]
     Male  Female  Age  EstimatedSalary
0       1       0   19            19000
1       1       0   35            20000
2       0       1   26            43000
3       0       1   27            57000
4       1       0   19            76000
..    ...     ...  ...              ...
314     0       1   39            79000
315     0       1   39            75000
316     0       1   54           104000
```

Brenda Estefania Castillo Fernandez



```
[ ] from sklearn.linear_model import Perceptron
    #dfxtrain, dfy = load_digits(return_X_y=True)
    clf= Perceptron(tol=1e-3, random_state=0)
    clf.fit(dfxtrain,ytrain)
    clf.score(dfxtrain,ytrain) #es para tener el accuracy, la metrica de predicción, en este caso tiene el 71% de exactitud al separar un modelo de otro.

    0.7115987460815048
```

```
[ ] clf.predict(dfxtest)

    array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
           0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
           0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
           0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0])
```

```
[ ] print(ytest)

    320    1
    321    1
    322    0
    323    1
    324    1
          ..
    395    1
    396    1
    397    1
    398    0
    399    1
    Name: Purchased, Length: 80, dtype: int64
```