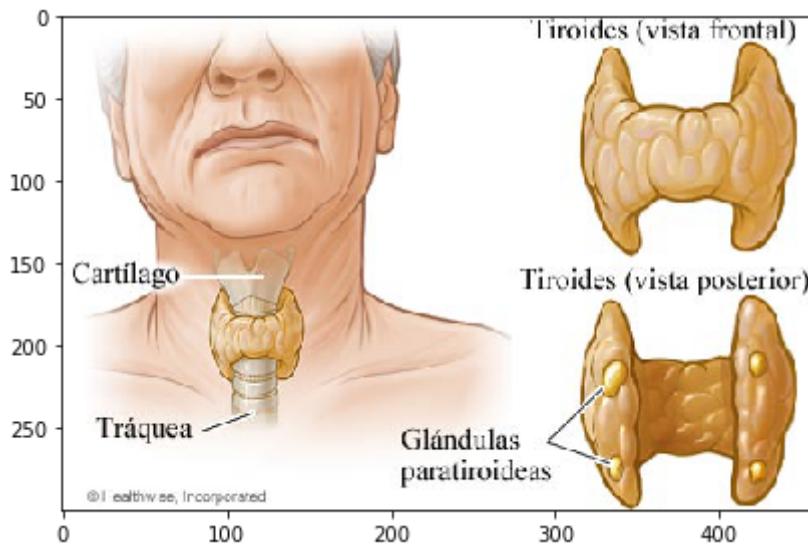# PROYECTO p1

Lectura de datos: https://docs.google.com/document/d/1zTVhqSv-
rKO5ki9c6h9dKqMSgrYWopX7RdarSlTL46I/edit?usp=sharing

In [358...
```python
from skimage import io
img_src='https://content.healthwise.net/resources/13.3/es-us/media/medical/hw/s_h99
image=io.imread(img_src)
io.imshow(image)
io.show()
```


In [359...
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
data = pd.read_csv("dis2.csv")
print(data)
```

```
       age sex on thyroxine query on thyroxine on antithyroid medication sick  \
0       41  F               f                  f                             f     f
1       23  F               f                  f                             f     f
2       46  M               f                  f                             f     f
3       70  F               t                  f                             f     f
4       70  F               f                  f                             f     f
...     ..  ..             ...                ...                           ...   ...
2795    70  M               f                  f                             f     f
2796    73  M               f                  t                             f     f
2797    75  M               f                  f                             f     f
2798    60  F               f                  f                             f     f
2799    81  F               f                  f                             f     f

      pregnant thyroid surgery I131 treatment query hypothyroid  ...  \
0            f               f              f                  f  ...
1            f               f              f                  f  ...
2            f               f              f                  f  ...
3            f               f              f                  f  ...
4            f               f              f                  f  ...
...        ...             ...            ...                ...  ...
2795         f               f              f                  f  ...
2796         f               f              f                  f  ...
2797         f               f              f                  f  ...
2798         f               f              f                  f  ...
2799         f               f              f                  f  ...

      TT4 measured  TT4 T4U measured   T4U FTI measured  FTI TBG measured TBG  \
0                t  125            t  1.14            t  109            f   ?
1                t  102            f     ?            f     ?           f   ?
2                t  109            t  0.91            t  120            f   ?
3                t  175            f     ?            f     ?           f   ?
4                t   61            t  0.87            t   70            f   ?
...            ...  ...          ...   ...          ...  ...          ...  ..
2795             t  155            t  1.05            t  148            f   ?
2796             t   63            t  0.88            t   72            f   ?
2797             t  147            t   0.8            t  183            f   ?
2798             t  100            t  0.83            t  121            f   ?
2799             t  114            t  0.99            t  115            f   ?

      referral source      class
0                SVHC  negative.
1               other  negative.
2               other  negative.
3               other  negative.
4                 SVI  negative.
...               ...        ...
2795              SVI  negative.
2796            other  negative.
2797            other  negative.
2798            other  negative.
2799              SVI  negative.

[2800 rows x 30 columns]
```

```
In [360…  data.head()
```

Out[360]:

	age	sex	on thyroxine	query on thyroxine	on antithyroid medication	sick	pregnant	thyroid surgery	I131 treatment	query hypothyroid
0	41	F	f	f	f	f	f	f	f	f
1	23	F	f	f	f	f	f	f	f	f
2	46	M	f	f	f	f	f	f	f	f
3	70	F	t	f	f	f	f	f	f	f
4	70	F	f	f	f	f	f	f	f	f

5 rows × 30 columns

In [361… `data.columns`

Out[361]:
```
Index(['age', 'sex', 'on thyroxine', 'query on thyroxine',
       'on antithyroid medication', 'sick', 'pregnant', 'thyroid surgery',
       'I131 treatment', 'query hypothyroid', 'query hyperthyroid', 'lithium',
       'goitre', 'tumor', 'hypopituitary', 'psych', 'TSH measured', 'TSH',
       'T3 measured', 'T3', 'TT4 measured', 'TT4', 'T4U measured', 'T4U',
       'FTI measured', 'FTI', 'TBG measured', 'TBG', 'referral source',
       'class'],
      dtype='object')
```

In [362… `data.tail()`

Out[362]:

	age	sex	on thyroxine	query on thyroxine	on antithyroid medication	sick	pregnant	thyroid surgery	I131 treatment	que hypothyr
2795	70	M	f	f	f	f	f	f	f	f
2796	73	M	f	t	f	f	f	f	f	f
2797	75	M	f	f	f	f	f	f	f	f
2798	60	F	f	f	f	f	f	f	f	f
2799	81	F	f	f	f	f	f	f	f	f

5 rows × 30 columns

In [363… `pd.unique(data['referral source'])`

Out[363]: `array(['SVHC', 'other', 'SVI', 'STMW', 'SVHD'], dtype=object)`

In [364… `data['class'].describe()`

Out[364]:
```
count        2800
unique          2
top      negative.
freq         2755
Name: class, dtype: object
```
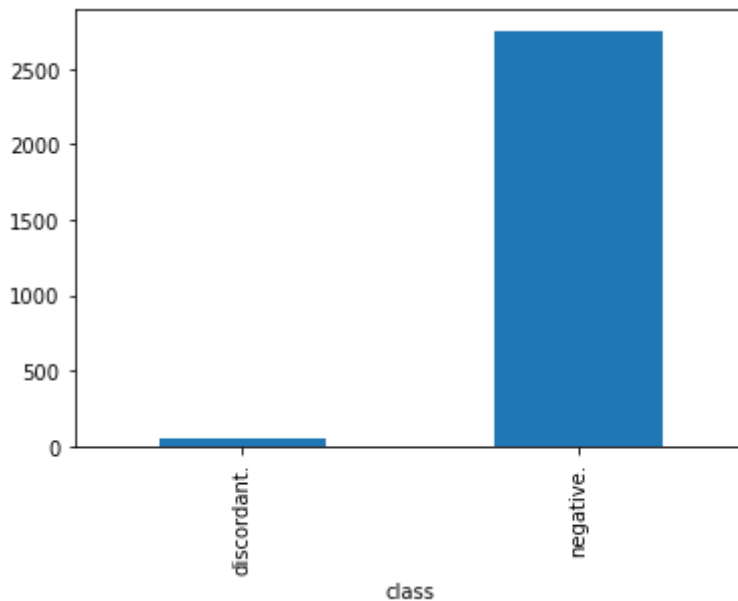
In [365…
```
data['age'].min()
data['age'].max()
data['age'].count()
```

Out[365]:   2800

In [366…
```python
class_counts = data.groupby('class')['referral source'].count()
print(class_counts)
```

```
class
discordant.        45
negative.        2755
Name: referral source, dtype: int64
```
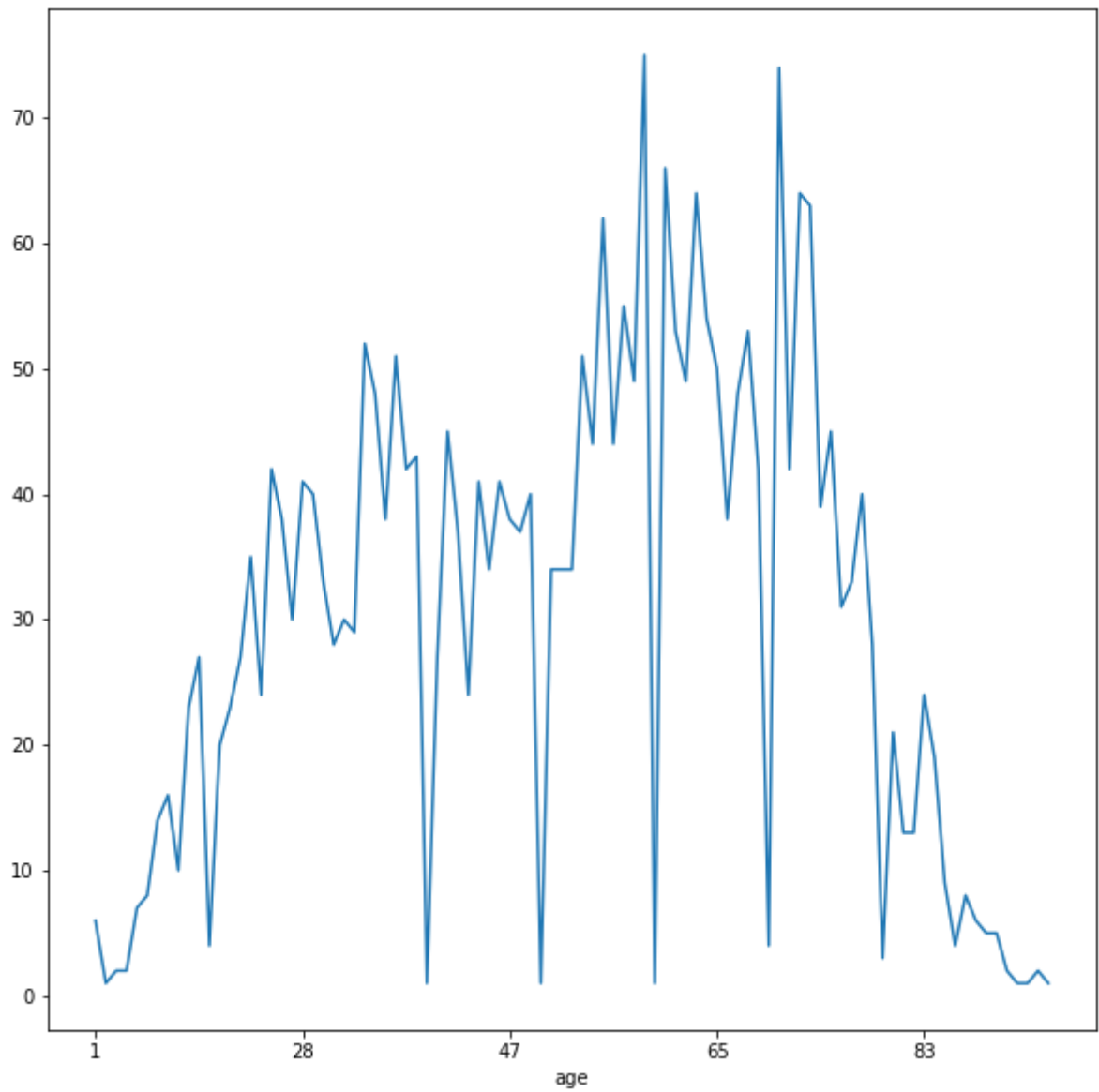
In [367…
```python
%matplotlib inline
# Creaemos una gráfica de barras
class_counts.plot(kind='bar');
```


In [368…
```python
grupos = data.groupby('age')
```

In [369…
```python
age_counts=data.groupby('age')['class'].count()
print(age_counts)
age_counts.plot(subplots=True,figsize=(10,10),sharey=False)
plt.show()
```

```
age
1      6
10     1
11     2
12     2
13     7
      ..
91     2
92     1
93     1
94     2
?      1
Name: class, Length: 93, dtype: int64
```

MAYORES a 90

```
In [370…  data_df = pd.read_csv("dis2.csv")
          data[data_df.age >= '90']
```

Out[370]:

	age	sex	on thyroxine	query on thyroxine	on antithyroid medication	sick	pregnant	thyroid surgery	I131 treatment	que hypothyre
440	90	F	f	t	f	f	f	f	f	f
556	90	F	f	f	f	f	f	f	f	f
1129	94	F	f	f	f	f	f	f	f	f
1403	90	M	f	f	f	f	f	f	f	f
1490	91	M	f	f	f	f	f	f	f	f
1558	90	F	f	f	f	f	f	f	f	f
1727	91	F	f	f	f	f	f	f	f	f
1985	?	F	t	f	f	f	f	f	f	t
2244	90	M	f	f	f	f	f	f	f	f
2418	93	F	f	f	f	t	f	f	f	f
2673	94	M	f	f	f	f	f	f	f	f
2760	92	F	t	f	f	f	f	f	f	f

12 rows × 30 columns

P2 clasificacion no supervisado

In [371…
```python
import csv
import numpy as np
from sklearn import preprocessing
with open('dis.csv') as f:
    datos = list(csv.reader(f, delimiter=","))

X_train= np.array(data)
```

In [372…
```python
try:
    scaler = preprocessing.StandardScaler().fit(X_train)
    print("datos procesados", scaler)
except:
    print ("datos con string")
```

datos con string

In [373…
```python
X, y = make_classification(random_state=42)
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=42)
pipe = make_pipeline(StandardScaler(), LogisticRegression())
pipe.fit(X_train, y_train)  # apply scaling on training data
Pipeline(steps=[('standardscaler', StandardScaler()),
                ('logisticregression', LogisticRegression())])

pipe.score(X_test, y_test)
```

Out[373]: 0.96

In [374…
```python
min_max_scaler = preprocessing.MinMaxScaler()
X_train_minmax = min_max_scaler.fit_transform(X_train)
print(X_train_minmax)
```

```
[[0.31429663 0.         0.62946    ... 0.48919562 0.5809746  0.54756063]
 [0.36809527 0.51664411 0.27233998 ... 0.77293684 0.49893783 0.51420651]
 [0.48077173 0.42402954 0.46950576 ... 0.684062    0.47072286 0.9332303 ]
 ...
 [0.69080869 0.60671411 0.39990917 ... 0.64183768 0.65992945 0.68982399]
 [0.62223193 0.65736885 0.77193918 ... 0.         0.45446132 0.68643192]
 [0.44135594 0.48457765 0.74302363 ... 0.36271249 0.56433746 0.87377042]]
```

In [375…
```python
#mapeo de datos
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=0)
quantile_transformer = preprocessing.QuantileTransformer(random_state=0)
X_train_trans = quantile_transformer.fit_transform(X_train)
X_test_trans = quantile_transformer.transform(X_test)
np.percentile(X_train[:,0], [0, 25, 50, 75, 100])
```

C:\Users\DSIE\anaconda3\lib\site-packages\sklearn\preprocessing_data.py:2590: Use
rWarning: n_quantiles (1000) is greater than the total number of samples (75). n_q
uantiles is set to n_samples.
  warnings.warn(

Out[375]:
```
array([-2.65096981, -0.52016716,  0.177701  ,  0.83301396,  2.40341559])
```

In [376…
```python
pt = preprocessing.PowerTransformer(method='box-cox', standardize=False)
X_lognormal = np.random.RandomState(616).lognormal(size=(3, 3))
X_lognormal
```
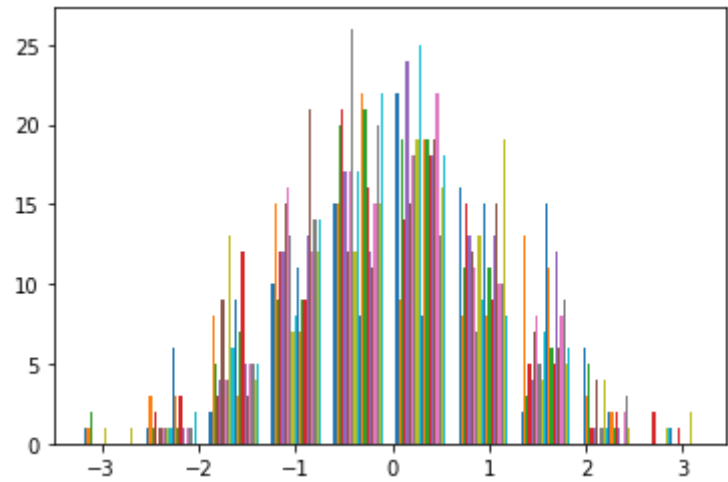
Out[376]:
```
array([[1.28331718, 1.18092228, 0.84160269],
       [0.94293279, 1.60960836, 0.3879099 ],
       [1.35235668, 0.21715673, 1.09977091]])
```

In [377…
```python
pt.fit_transform(X_lognormal)
```

Out[377]:
```
array([[ 0.49024349,  0.17881995, -0.1563781 ],
       [-0.05102892,  0.58863195, -0.57612414],
       [ 0.69420009, -0.84857822,  0.10051454]])
```

In [378…
```python
plt.hist(X_train)
```

Out[378]:
```
(array([[ 1.,  1.,  2., 10., 15., 22., 16.,  2.,  6.,  0.],
        [ 1.,  3.,  8., 15., 15.,  9.,  8., 13.,  3.,  0.],
        [ 2.,  1.,  5.,  9., 20., 19., 11.,  3.,  5.,  0.],
        [ 0.,  2.,  3., 12., 21., 14., 15.,  5.,  1.,  2.],
        [ 0.,  0.,  4., 12., 17., 24., 13.,  4.,  1.,  0.],
        [ 0.,  1.,  9., 15., 12., 15., 12.,  7.,  4.,  0.],
        [ 0.,  1.,  4., 16., 17., 18., 11.,  8.,  0.,  0.],
        [ 0.,  1.,  4., 13., 26., 18.,  7.,  5.,  1.,  0.],
        [ 1.,  1., 13.,  7., 12., 19., 13.,  4.,  4.,  1.],
        [ 0.,  1.,  6.,  8., 17., 25.,  9.,  7.,  1.,  1.],
        [ 0.,  6.,  9., 11.,  8.,  8., 15., 15.,  2.,  1.],
        [ 0.,  3.,  3.,  7., 22., 19.,  8., 11.,  2.,  0.],
        [ 0.,  1.,  7.,  9., 21., 19., 11.,  6.,  1.,  0.],
        [ 0.,  3., 12.,  9., 16., 18.,  9.,  5.,  2.,  1.],
        [ 0.,  1.,  5., 13., 12., 18., 13., 12.,  1.,  0.],
        [ 0.,  0.,  3., 21., 11., 19., 15.,  6.,  0.,  0.],
        [ 0.,  1.,  5., 12., 15., 22., 10.,  8.,  2.,  0.],
        [ 0.,  1.,  5., 14., 20., 13., 10.,  9.,  3.,  0.],
        [ 1.,  0.,  4., 12., 15., 16., 19.,  5.,  1.,  2.],
        [ 0.,  2.,  5., 14., 22., 18.,  8.,  6.,  0.,  0.]]),
 array([-3.24126734, -2.59782985, -1.95439236, -1.31095487, -0.66751738,
        -0.02407989,  0.6193576 ,  1.2627951 ,  1.90623259,  2.54967008,
         3.19310757]),
 <a list of 20 BarContainer objects>)
```

In [ ]:

In [ ]: