```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import tensorflow
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Layer, Dense, Dropout
```

```
data = pd.read_csv('/content/weather.csv')
```

```
data.head()
```

	age	sex	on_thyroxine	query_on_thyroxine	on_antithyroid_meds	sick	pregnant	thyr
0	29	F	f	f	f	f	f	
1	29	F	f	f	f	f	f	
2	41	F	f	f	f	f	f	
3	36	F	f	f	f	f	f	
4	32	F	f	f	f	f	f	

5 rows × 31 columns

```
data.isnull().sum()
```

```
age                       0
sex                     307
on_thyroxine              0
query_on_thyroxine        0
on_antithyroid_meds       0
sick                      0
pregnant                  0
thyroid_surgery           0
I131_treatment            0
query_hypothyroid         0
query_hyperthyroid        0
lithium                   0
goitre                    0
tumor                     0
hypopituitary             0
psych                     0
TSH_measured              0
TSH                     842
T3_measured               0
T3                     2604
TT4_measured              0
TT4                     442
T4U_measured              0
T4U                     809
FTI_measured              0
FTI                     802
TBG_measured              0
TBG                    8823
referral_source           0
target                    0
patient_id                0
dtype: int64
```

```
data.dropna(subset=['target'],inplace=True)
```

```
data['target'].value_counts()
```

```
-      6771
K       436
G       359
I       346
F       233
R       196
A       147
L       115
M       111
N       110
```

```
S        85
GK       49
AK       46
J        30
B        21
MK       16
Q        14
O        14
C|I      12
KJ       11
GI       10
H|K       8
D         8
FK        6
C         6
P         5
MI        2
LJ        1
GKJ       1
OI        1
D|R       1
E         1
Name: target, dtype: int64
```

```
#Checking whether the age above 100
data[data.age>100]
```

	age	sex	on_thyroxine	query_on_thyroxine	on_antithyroid_meds	sick	pregnant
2976	455	F	f	f	f	f	f
5710	65511	M	f	f	f	f	f
6392	65512	M	f	f	f	f	f
8105	65526	F	f	f	f	f	f

4 rows × 31 columns

```
x=data.iloc[:,0:-1]
y=data.iloc[:,-1]
```

```
x
```

	age	sex	on_thyroxine	query_on_thyroxine	on_antithyroid_meds	sick	pregnant	t
0	29	F	f	f	f	f	f	
1	29	F	f	f	f	f	f	
2	41	F	f	f	f	f	f	
3	36	F	f	f	f	f	f	
4	32	F	f	f	f	f	f	
...	...	...	...	...	...	...	...	
9167	56	M	f	f	f	f	f	
9168	22	M	f	f	f	f	f	
9169	69	M	f	f	f	f	f	
9170	47	F	f	f	f	f	f	
9171	31	M	f	f	f	f	f	

9172 rows × 28 columns

```
x['sex'].unique()
```

```
array(['F', 'M', nan], dtype=object)
```

```
x['sex'].replace(np.nan, 'F', inplace=True)
```

```
x['sex'].value_counts()
```

```
F    6380
M    2792
Name: sex, dtype: int64
```

```
x.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9172 entries, 0 to 9171
Data columns (total 28 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   age                  9172 non-null   int64  
 1   sex                  9172 non-null   object 
 2   on_thyroxine         9172 non-null   object 
 3   query_on_thyroxine   9172 non-null   object 
 4   on_antithyroid_meds  9172 non-null   object 
 5   sick                 9172 non-null   object 
 6   pregnant             9172 non-null   object 
 7   thyroid_surgery      9172 non-null   object 
 8   I131_treatment       9172 non-null   object 
 9   query_hypothyroid    9172 non-null   object 
 10  query_hyperthyroid   9172 non-null   object 
 11  lithium              9172 non-null   object 
 12  goitre               9172 non-null   object 
 13  tumor                9172 non-null   object 
 14  hypopituitary        9172 non-null   object 
 15  psych                9172 non-null   object 
 16  TSH                  8330 non-null   float64
 17  T3                   6568 non-null   float64
 18  TT4_measured         9172 non-null   object 
 19  TT4                  8730 non-null   float64
 20  T4U_measured         9172 non-null   object 
 21  T4U                  8363 non-null   float64
 22  FTI_measured         9172 non-null   object 
 23  FTI                  8370 non-null   float64
 24  TBG_measured         9172 non-null   object 
 25  TBG                  349 non-null    float64
 26  referral_source      9172 non-null   object 
 27  target               9172 non-null   object 
dtypes: float64(6), int64(1), object(21)
memory usage: 2.0+ MB
```

```
from sklearn.preprocessing import OrdinalEncoder, LabelEncoder
ordinal_encoder=OrdinalEncoder(dtype='int64')
x.iloc[:, 1:16] = ordinal_encoder.fit_transform(x.iloc[:, 1:16])
#ordinal_encoder.fit_transform(x[['sex']])
```

```
x
```

	age	sex	on_thyroxine	query_on_thyroxine	on_antithyroid_meds	sick	pregnant	t
0	29	0	0	0	0	0	0	
1	29	0	0	0	0	0	0	
2	41	0	0	0	0	0	0	
3	36	0	0	0	0	0	0	
4	32	0	0	0	0	0	0	
...	...	...	...	...	...	...	...	
9167	56	1	0	0	0	0	0	
9168	22	1	0	0	0	0	0	
9169	69	1	0	0	0	0	0	
9170	47	0	0	0	0	0	0	
9171	31	1	0	0	0	0	0	

9172 rows × 28 columns

```
x.replace(np.nan, '0', inplace=True)
x
```

	age	sex	on_thyroxine	query_on_thyroxine	on_antithyroid_meds	sick	pregnant	t
0	29	0	0	0	0	0	0	
1	29	0	0	0	0	0	0	
2	41	0	0	0	0	0	0	
3	36	0	0	0	0	0	0	
4	32	0	0	0	0	0	0	
...	...	...	...	...	...	...	...	
9167	56	1	0	0	0	0	0	
9168	22	1	0	0	0	0	0	
9169	69	1	0	0	0	0	0	
9170	47	0	0	0	0	0	0	
9171	31	1	0	0	0	0	0	

9172 rows × 28 columns

```
label_encoder = LabelEncoder()
y_dt=label_encoder.fit_transform(y)


y=pd.DataFrame(y_dt, columns=['targest'])
y
```

	targest
0	0
1	1
2	2
3	3
4	4
...	...
9167	9167
9168	9168
9169	9169
9170	9170
9171	9171

9172 rows × 1 columns

```
from  sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.20,random_state=0)


from imblearn.over_sampling import SMOTE
y_train.value_counts()
```

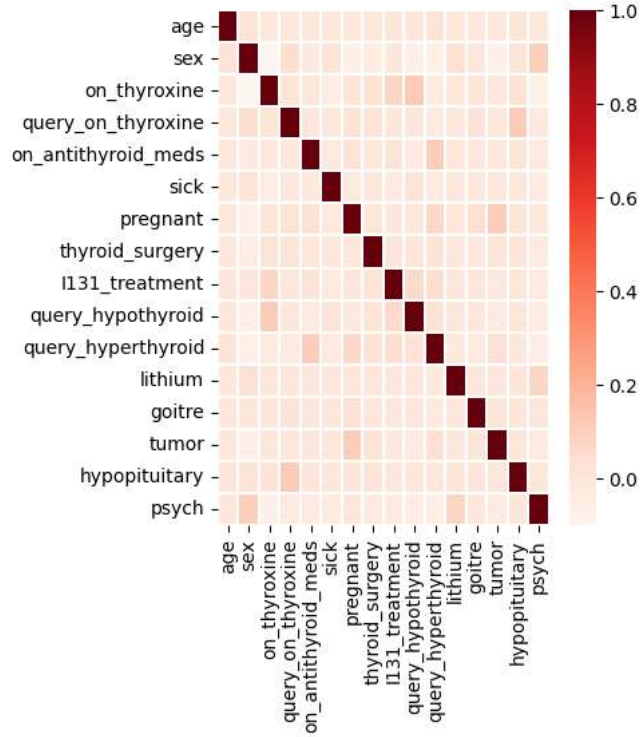
```
targest
0          1
6141       1
6138       1
6137       1
6135       1
          ..
3066       1
3064       1
3062       1
3061       1
9170       1
Length: 7337, dtype: int64
```

```
import seaborn as sns
```

+ Code + Text

```
#checking correlation using heatmap
f, ax = plt.subplots(figsize = (4,5))
sns.heatmap(corrmat, ax = ax, cmap="Reds", linewidths = 0.1)
```

```
<Axes: >
```



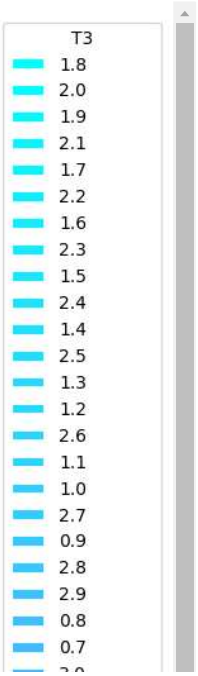

```
import os


from scipy import stats


from imblearn.ensemble import BalancedBaggingClassifier


os=SMOTE(random_state=0,k_neighbors=1)
x_bal,y_bal=os.fit_resample(x_train,y_train)
x_test_bal,y_test_bal=os.fit_resample(x_test,y_test)
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-39-5837100947c9> in <cell line: 2>()
      1 os=SMOTE(random_state=0,k_neighbors=1)
----> 2 x_bal,y_bal=os.fit_resample(x_train,y_train)
      3 x_test_bal,y_test_bal=os.fit_resample(x_test,y_test)

 ———————————————————————————— 7 frames ————————————————————————————
/usr/local/lib/python3.9/dist-packages/pandas/core/generic.py in __array__(self, dtype)
   2068
   2069     def __array__(self, dtype: npt.DTypeLike | None = None) -> np.ndarray:
-> 2070         return np.asarray(self._values, dtype=dtype)
   2071
   2072     def __array_wrap__(

ValueError: could not convert string to float: 't'
```

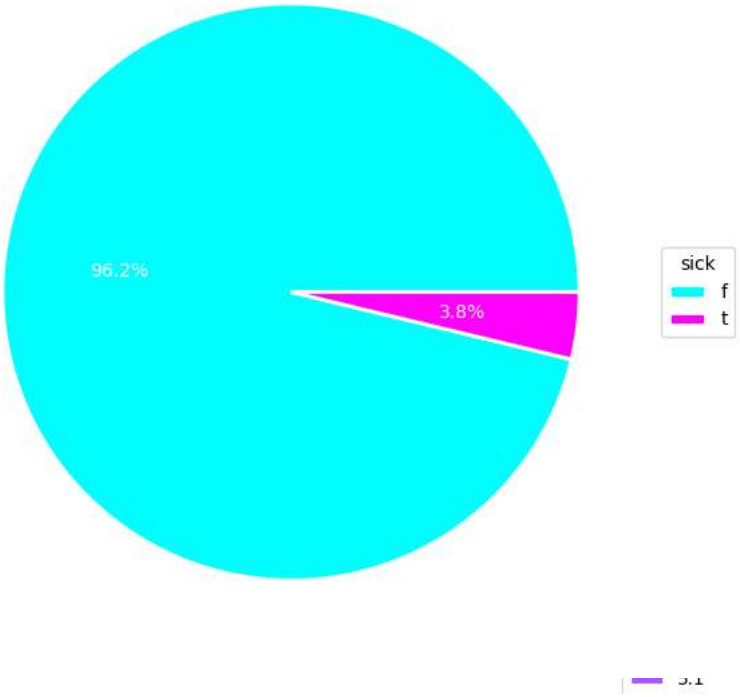
SEARCH STACK OVERFLOW

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib.cm as cm
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix


def pie_graph(col):
  vc = data[col].value_counts()
  colors = cm.cool(np.linspace(0, 1, len(vc)))
```

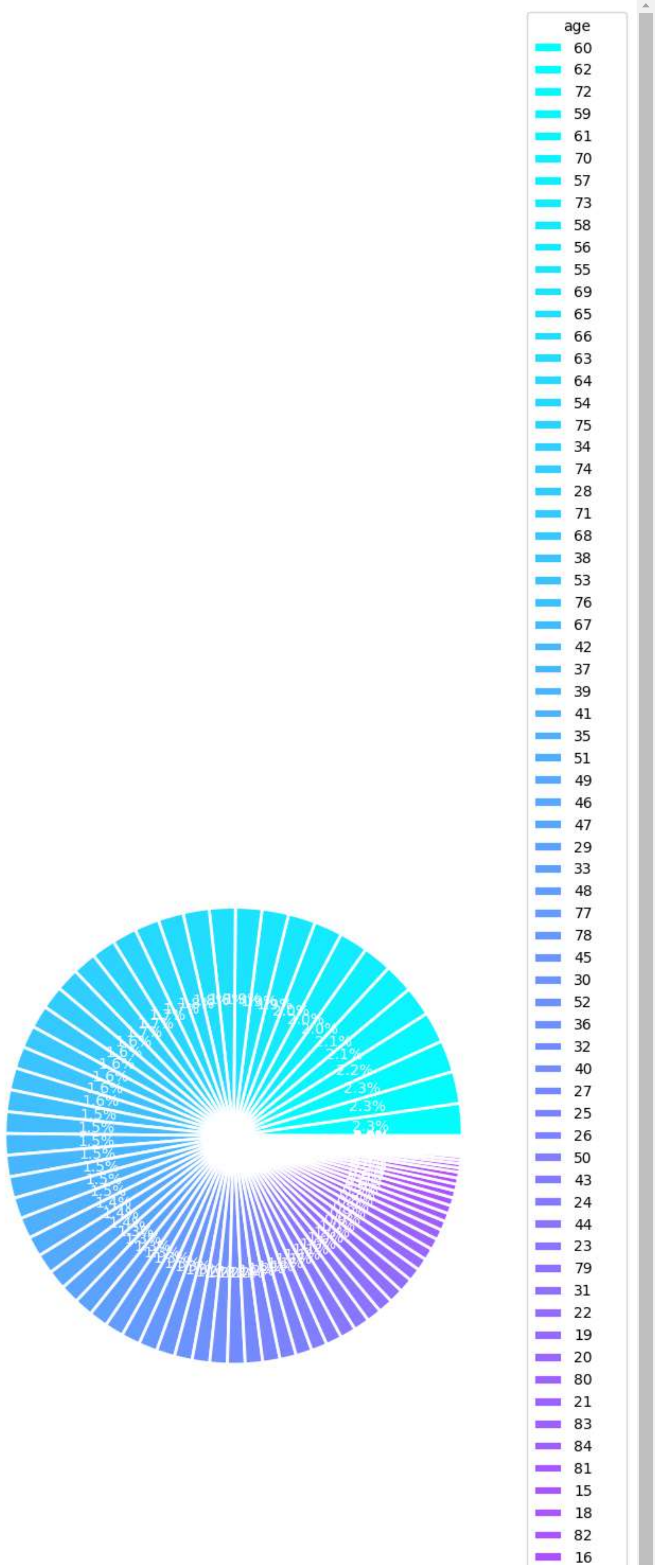
```
    fig, ax = plt.subplots(figsize = (8,7), subplot_kw = dict(aspect="equal"))
    wedges, texts, autotexts = ax.pie(vc, autopct = '%1.1f%%', textprops=dict(color="w", fontsize = 10),wedgeprops = {'linewidth': 2, 'edgecolo
    ax.legend(wedges, vc.index, title = col, loc = "center left", bbox_to_anchor = (1, 0, 0.5, 1))


pie_graph('T3')
```



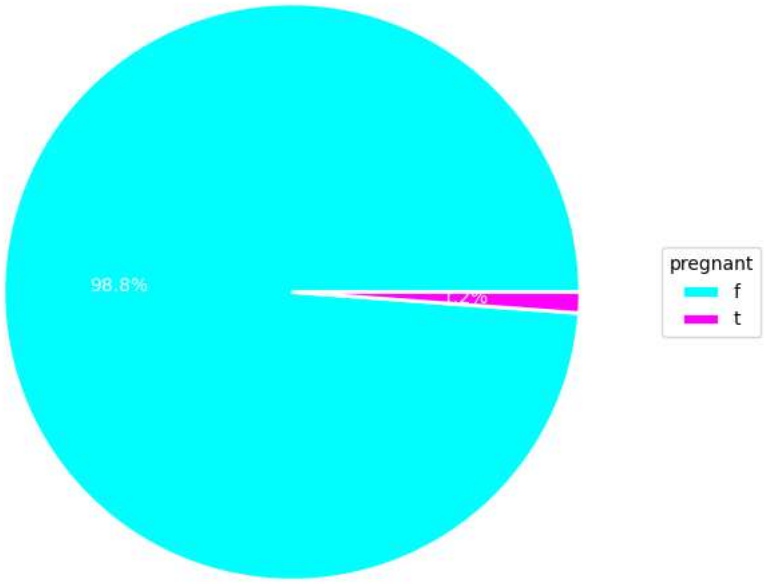

```
pie_graph('sick')
```




```
pie_graph('age')
```


age
60
62
72
59
61
70
57
73
58
56
55
69
65
66
63
64
54
75
34
74
28
71
68
38
53
76
67
42
37
39
41
35
51
49
46
47
29
33
48
77
78
45
30
52
36
32
40
27
25
26
50
43
24
44
23
79
31
22
19
20
80
21
83
84
81
15
18
82
16

```
pie_graph('pregnant')
```




```
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)


print(data.head())
```

```
   age sex on_thyroxine query_on_thyroxine on_antithyroid_meds sick pregnant  \
0   29   F            f                  f                   f    f        f
1   29   F            f                  f                   f    f        f
2   41   F            f                  f                   f    f        f
3   36   F            f                  f                   f    f        f
4   32   F            f                  f                   f    f        f

  thyroid_surgery I131_treatment query_hypothyroid query_hyperthyroid lithium  \
0               f              f                 t                   f       f
1               f              f                 f                   f       f
2               f              f                 f                   t       f
3               f              f                 f                   f       f
4               f              f                 f                   f       f

  goitre tumor hypopituitary psych TSH_measured  TSH T3_measured   T3  \
0      f     f             f     f            t  0.3           f  NaN
1      f     f             f     f            t  1.6           t  1.9
2      f     f             f     f            f  NaN           f  NaN
3      f     f             f     f            f  NaN           f  NaN
4      f     f             f     f            f  NaN           f  NaN

  TT4_measured    TT4 T4U_measured  T4U FTI_measured  FTI TBG_measured   TBG  \
0            f    NaN            f  NaN            f  NaN            f   NaN
1            t  128.0            f  NaN            f  NaN            f   NaN
2            f    NaN            f  NaN            f  NaN            t  11.0
3            f    NaN            f  NaN            f  NaN            t  26.0
4            f    NaN            f  NaN            f  NaN            t  36.0

  referral_source target  patient_id
0           other      -   840801013
1           other      -   840801014
2           other      -   840801042
3           other      -   840803046
4           other      S   840803047
```

```
plt.hist(data['sick'])
```

```
(array([8828.,    0.,    0.,    0.,    0.,    0.,    0.,    0.,    0.,
         344.]),
 array([0. , 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1. ]),
 <BarContainer object of 10 artists>)
```


```
#Splitting the data values as x and y
x=data.iloc[:,0:1]
y=data.iloc[:,-1]
```

x

	Age
0	23
1	47
2	47
3	28
4	61
...	...
195	56
196	16
197	52
198	23
199	40

200 rows × 1 columns

```
x['Sex'].unique()
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
/usr/local/lib/python3.9/dist-packages/pandas/core/indexes/base.py in get_loc(self,
key, method, tolerance)
   3628             try:
-> 3629                 return self._engine.get_loc(casted_key)
   3630             except KeyError as err:
```

```
x['Age']=x['Age'].astype('float')
x['BP']=x['BP'].astype('float')
```

```
pandas/_libs/hashtable_class_helper.pxi in
pandas._libs.hashtable.PyObjectHashTable.get_item()

KeyError: 'Sex'

The above exception was the direct cause of the following exception:

KeyError                                  Traceback (most recent call last)
/usr/local/lib/python3.9/dist-packages/pandas/core/indexes/base.py in get_loc(self,
key, method, tolerance)
   3629                 return self._engine.get_loc(casted_key)
   3630             except KeyError as err:
-> 3631                 raise KeyError(key) from err
   3632             except TypeError:
   3633                 # If we have a listlike key, _check_indexing_error will raise
```

✓ 2s completed at 9:15 AM